

**Note di statistica descrittiva
e
primi elementi di calcolo delle probabilità
Temi d'esame (testi e soluzioni)**

G. Boari, G. Cantaluppi

Dipartimento di Scienze statistiche
Università Cattolica del Sacro Cuore
Milano

UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Facoltà di Economia
Simulazione esame di Statistica - 4 dicembre 2002

1) La tabella seguente riporta la distribuzione dei pesi (in kg) dei pacchi spediti da un'azienda commerciale:

peso (kg)	n_i
5-10	3
10-20	7
20-30	7
30-50	6

1.1) Rappresentare graficamente la variabile statistica dei pesi e calcolarne la moda.

1.2) Calcolare il peso mediano.

2) Date le due seguenti distribuzioni di frequenza

COLORE CAPELLI	neri	rossi	biondi	castani
n_i	30	20	30	8

COLORE OCCHI	neri	marroni	verdi	celesti
n_i	20	51	10	7

confrontarne la variabilità.

3) Con i dati degli iscritti ad un certo ufficio di collocamento si è costruita la seguente tabella, riguardante i caratteri X ="occupazione" e A ="possesso della patente auto":

$X \setminus A$	si	no
si	480	35
no	70	1100

Valutare, con un opportuno indice di connessione, quanto il possesso della patente è determinante per ottenere un impiego.

4) Completare la seguente tabella di frequenze congiunte n_{ij} in modo che risulti, se possibile, $\chi^2_{\text{norm}} = 0$

	y_1	y_2	y_3
x_1	30	6	12
x_2	60		
x_3	10		

5) Date le seguenti coppie di osservazioni

x_i	20	35	35	28	42	35	10	14	20	50
y_i	10	19	23	16	21	14	17	12	19	36

rappresentare graficamente le coppie di punti.

Considerati, poi, i seguenti modelli

I) $Y^* = a + bX$ II) $Y^* = cX$ III) $Y^* = 20 + dX$

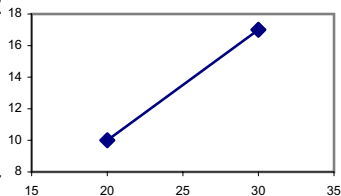
scegliere i due che meglio rispecchiano il precedente grafico.

Calcolarne, quindi, secondo il metodo dei minimi quadrati, i parametri e indicare quale si adatta meglio ai dati.

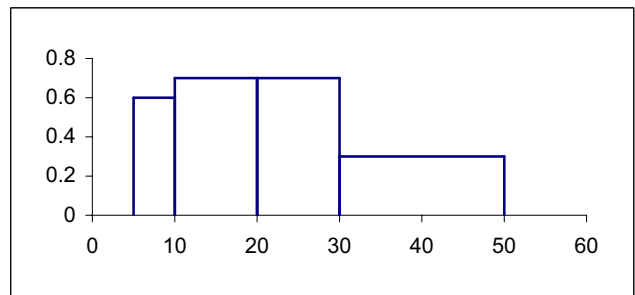
1	h(i-1)	h(i)	xi	ai
	5	10	7.5	5
	10	20	15	10
	20	30	25	10
	30	50	40	20

densita
0.6
0.7
0.7
0.3
P

mediana 22.85714



ni
3
7
7
6
23



2		ni	fi		ni	fi		
	capelli	30	0.340909	occhi	20	0.227273	Gcapelli	0.707645
		20	0.227273		51	0.579545	Gmax	0.75
		30	0.340909		10	0.113636	GNcapelli	0.943526 FN 0.7624
		8	0.090909		7	0.079545		
	totale	88	1	totale	88	1	Gocchi	0.593233
							Gmax	0.75
							Gocchi	0.790978 FN 0.5428

3	480	35	515	teo	168.1009	346.8991	cont	311.8991	-311.899
	70	1100	1170		381.8991	788.1009		-311.899	311.8991
	550	1135	1685						

chi2 1237.304
max 1685
chiN 0.734305 chi 0.8569

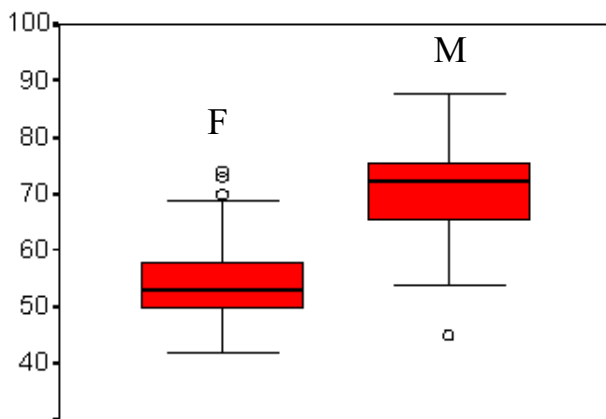
4		y1	y2	y3
	x1	30	6	12
	x2	60	12	24
	x3	10	2	4

5	20	35	35	28	42	35	10	14	20	50
	10	19	23	16	21	14	17	12	19	36

	xi	yi	xiyi	xi2	yi2	ystar		
	20	10	200	400	100	12.2375	Var(X)	146.69
	35	19	665	1225	361	21.41562	Var(Y)	47.61
	35	23	805	1225	529	21.41562	Cov	60.37
	28	16	448	784	256	17.1325		
	42	21	882	1764	441	25.69875	b	0.411548
	35	14	490	1225	196	21.41562	a	6.806258
	10	17	170	100	289	6.118749	rho2	0.521848
	14	12	168	196	144	8.566249	RMS	22.76484
	20	19	380	400	361	12.2375		
	50	36	1800	2500	1296	30.59375	c	0.611875
	289	187	6008	9819	3973			
	28.9	18.7	600.8	981.9	397.3	media	RMS	29.68554

UNIVERSITÀ CATTOLICA DEL SACRO CUORE
Facoltà di Economia
Prova di autovalutazione di Statistica - 14 novembre 2003

1) Con riferimento ai seguenti grafici a scatola (box plot) delle distribuzioni dei pesi di $n_1=74$ femmine e $n_2=80$ maschi (assumendo valori dei pesi distinti):



1.1) indicare la distanza inter-quartile approssimativa del peso

delle femmine:.....

1.2) indicare approssimativamente la mediana del peso

dei maschi:.....

1.3) indicare quante sono le femmine comprese nel box

.....

2) Data la seguente tabella:

X \ Y	1-3	3-6	6-8
x_1	4	2	6
x_2	3	1	n_{23}

dire, giustificando la risposta, per quali valori di n_{23} (anche non interi) sussiste:

2.1) indipendenza stocastica:

.....

2.2) dipendenza funzionale:

.....

2.3) disegnare il grafico della distribuzione marginale Y per il caso $n_{23}=6$.

3) Con riferimento ai seguenti dati:

y_i	7	8	9
n_i	n_1	4	3

3.1) definire n_1 in modo che $\mu = 7.5$:

.....

3.2) calcolare quindi $\mu^{(\infty)}$, Δ_∞

3.3) calcolare l'indice di Eterogeneità normalizzato:

3.4) costruire una tabella di minima variabilità;

4) Date le due seguenti distribuzioni delle frequenze relative condizionate:

$X y_1$	x_1	x_2	x_3
	0.30	0.20	0.50
$X y_2$	x_1	x_2	x_3
	0.10	0.60	0.30

sapendo che si riferiscono ad $n=110$ unità statistiche, delle quali 70 hanno valore $Y=y_1$, ricostruire la tabella a doppia entrata (X,Y) e calcolare un indice di connessione normalizzato:

.....

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 10.01.2008 (Tema 239)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

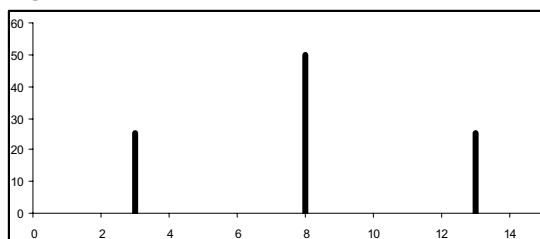
Studente _____ matricola _____

- 1) La seguente tabella riporta la distribuzione di frequenze congiunte riguardanti la durata media in ore degli spostamenti (Y) di 100 soggetti per raggiungere la località di vacanza (X).

$X \backslash Y$	3	8	13
montagna	10	31	10
mare	15	19	15

- 1.1) Si rappresenti graficamente la distribuzione della marginale Y .

y_j	$n_{\bullet j}$
3	25
8	50
13	25
	100



commenti Distribuzione simmetrica, carattere quantitativo, no presenza di classi, grafico a bastoncini

- 1.2) Si confrontino, con opportuni indici di posizione e di variabilità, le distribuzioni condizionate di $Y|x_1$ e $Y|x_2$.

indice di posizione utilizzato per il confronto: media aritmetica valore $Y|x_1$ 8 valore $Y|x_2$ 8

indice di variabilità utilizzato per il confronto: c.v. valore $Y|x_1$ 0.3914 valore $Y|x_2$ 0.4890

commenti Entrambe le distribuzioni delle frequenze condizionate sono simmetriche rispetto al valore 8 (media aritmetica); la distribuzione $Y|x_1$ presenta un livello di variabilità inferiore rispetto alla distribuzione $Y|x_2$ _____

- 1.3) Si calcoli un opportuno indice di connessione tra X e Y .

indice χ_N valore 0.22

commenti Le variabili X e Y presentano un grado di connessione molto basso. _____

- 2) Sapendo che per un aggregato macroeconomico il numero indice riferito al tempo 2006 con base 2005 è pari a ${}_{2005}I_{2006} = 0.95$ e che tra il 2006 e il 2007 l'aggregato ha subito una variazione percentuale del +5% si determini ${}_{2005}I_{2007}$, numero indice riferito all'istante temporale 2007 con base 2005.

formula: ${}_{2005}I_{2007} = {}_{2005}I_{2006} \cdot {}_{2006}I_{2007} = 0.95 \cdot 1.05$ valore: 0.9975

commenti L'aggregato macroeconomico ha subito una riduzione del -0.25% tra il 2005 e il 2007.

- 3) Completare la tabella, esplicitando anche i valori possibili per x_2 e y_2 , in modo tale che l'indice χ^2 normalizzato sia pari a zero:

$X \backslash Y$	1	$y_2 = \underline{\hspace{1cm}}$	3
5			
$x_2 = \underline{\hspace{1cm}}$	2		

$X \backslash Y$	1	$1 < y_2 < 3$	3
5	4	10	8
$x_2 > 5$	2	5	4

commenti $\chi^2_{\text{NORM}} = \chi^2 = 0$ se sussiste indipendenza stocastica, i.e. distribuzioni condizionate somiglianti (proporzionali).

Studente _____ matricola _____

4) Completare la tabella, esplicitando anche i valori possibili per x_2 e y_2 , in modo tale $\rho^2 = 0$ e $\eta_{X|Y}^2 = 1$:

$X \backslash Y$	1	$y_2 = \underline{\hspace{1cm}}$	3
5			
$x_2 = \underline{\hspace{1cm}}$	2		

$X \backslash Y$	1	$y_2 = 2$	3
5	0	5	0
$x_2 > 5$	2	0	2

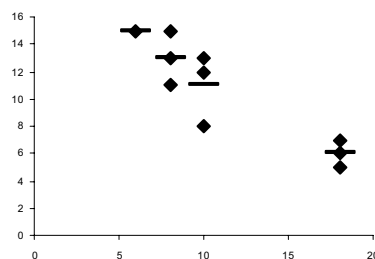
commenti $\rho^2 = 0$ e $\eta_{X|Y}^2 = 1 \rightarrow$ non correlazione e contemporaneamente dipendenza funzionale di X da Y .

Dal momento che $\eta_{Y|X}^2 = 0 \rightarrow \rho^2 = 0$, al fine di garantire la non correlazione, è sufficiente imporre $M(Y|x_1) = M(Y|x_2)$.

5) Su 10 filiali bancarie sono state rilevate le seguenti variabili: X = “numero dipendenti” e Y = “ore di straordinario pro-capite a settimana”.

X	18	10	8	6	18	8	10	8	10	18
Y	5	8	13	15	7	11	13	15	12	6

5.1) Si rappresentino graficamente le coppie di dati e la funzione di regressione per lo studio di $Y|X$.



5.2) Si stimino, attraverso il criterio dei minimi quadrati, i parametri dei modelli

$$\text{I) } Y^* = a + bX$$

$$\text{II) } Y^* = cX$$

quindi, attraverso un'opportuna misura di adattamento, si individui quale si adatta meglio ai dati.

modello I) $a = 18.521$ $b = -0.7036$ misura adattamento = 2.5293

modello II) $c = 0.704$ misura adattamento = 48.3576

espressione finale del modello migliore $Y^* = 18.521 - 0.7036X$

5.3) Si determinino, utilizzando il modello ritenuto migliore, le ore di straordinario pro-capite per una filiale con 20 dipendenti.

$$\text{formula } Y^*(20) = 18.521 - 0.7036 \cdot 20 = 4.4491$$

commenti Tra le variabili X e Y sussiste una relazione inversa, la covarianza risulta -14.1 ; il modello retta, che coincide con la funzione di regressione (dal momento che le medie condizionate sono allineate) è il modello migliore.

6) Uno studente deve superare un esame con 10 domande a risposta multipla, di uguale difficoltà, per le quali gli eventi “fornire risposta esatta” sono indipendenti ed hanno probabilità pari a p .

6.1) Sapendo che la probabilità di rispondere correttamente a tutte le 10 domande è 0.001, si ricavi il valore di p .

$$p = 0.5012$$

procedimento $p^{10} = 0.001 \rightarrow p = 0.001^{1/10} = 0.001^{0.1}$.

6.2) Se l'esame contenesse 100 domande, quale sarebbe la probabilità di rispondere correttamente a non più di 51 domande?

$$P(X \leq 51) = 0.5714$$

commenti $X \sim \text{Bin}(n=100, p=0.5012)$; $M(X) = 50.12$; $\text{Var}(X) = 25$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=50.12, \sigma^2=25)$ e

$$P(X \leq 51) = P\left(\frac{X - \mu}{\sigma} \leq \frac{51 - \mu}{\sigma}\right) = P\left(Z \leq \frac{51 - \mu}{\sigma}\right) = 0.5714.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 31.01.2008 (Tema 240)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

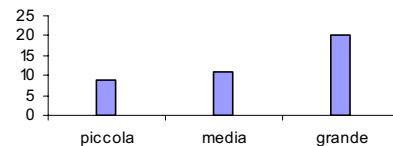
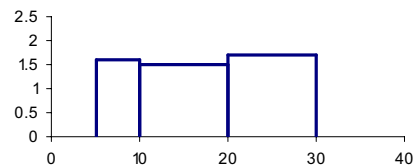
1) La tabella seguente contiene i dati dei pesi **W**, in kg, di 40 cani, distinti per taglia **T** (piccola, media, grande):

W \ T	piccola	media	grande
5-10	7	1	0
10-20	2	8	5
20-30	0	2	15

1.1) Rappresentare graficamente le distribuzioni delle marginali **W** e **T**.

W	$n_{i\cdot}$	$d_{i\cdot}$
5-10	8	1.6
10-20	15	1.5
20-30	17	1.7
	40	

T	$n_{\cdot j}$
piccola	9
media	11
grande	20
	40



commenti Dall'esame dell'istogramma della variabile **W** si rileva una distribuzione quasi uniforme dei pesi nelle 3 classi; osservando il grafico a bastoncini di **T** risulta più frequente la taglia "grande".

1.2) Calcolare opportuni indici di posizione e variabilità di **T**.

indice di posizione: MEDIANA valore "grande" indice di variabilità: Eterog. Frosini norm valore 0.747
 commenti il carattere **T** presenta un grado medio/alto di eterogeneità, la mediana coincide anche con la moda della distribuzione.

1.3) Confrontare le condizionate **W**|piccola e **W**|media con opportuni indici di posizione e variabilità.

indice di posizione utilizzato per il confronto: MEDIA valore **W**|piccola 9.167 valore **W**|media 16.136

indice di variabilità utilizzato per il confronto: c.v. valore **W**|piccola 0.34 valore **W**|media 0.291

commenti I cani di taglia media hanno un peso superiore a quelli di taglia piccola, la distribuzione di questi ultimi presenta maggiore variabilità.

1.4) Dire, senza effettuare calcoli e motivando teoricamente la risposta, se c'è indipendenza stocastica tra **W** e **T**.

commenti la presenza di zeri nella tabella esclude l'ipotesi di indipendenza stocastica.

1.5) Calcolare un indice di connessione tra **W** e **T**.

indice χ_N valore 0.6618.

commenti Tra le variabili **W** e **T** sussiste un livello di connessione medio/alto.

2) Completare la tabella (inserendo frequenze congiunte e marginali) in modo che $\chi^2=20$:

X \ Y	y_1	y_2	y_3	
x_1	15			
x_2				
				20

X \ Y	y_1	y_2	y_3	
x_1	15	n_{12}	0	$15 + n_{12}$
x_2	0	0	$20 - 15 - n_{12}$	$20 - 15 - n_{12}$
	15	n_{12}	$20 - 15 - n_{12}$	20

commenti Se $\chi^2=20$ allora $\chi_N^2 = 1$, quindi sussiste dipendenza funzionale (nel caso in esame di **X** da **Y**), ($0 < n_{12} < 5$)

Studente _____ matricola _____

- 3) Completare le tabelle (inserendo frequenze congiunte e marginali) in modo che ci sia dipendenza funzionale di Y in funzione di X :

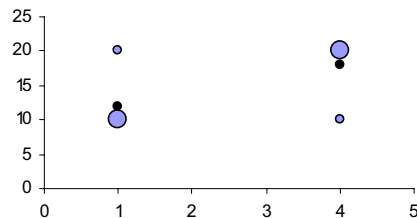
$X \backslash Y$	y_1	y_2	y_3	
x_1	15			
x_2				
				20

commenti Non è possibile completare la tabella in quanto le modalità della variabile Y sono in numero superiore a quelle della variabile X .

- 4) Con riferimento ai dati della seguente tabella:

$X \backslash Y$	10	20	
1	4	1	5
4	1	4	5
	5	5	10

- 4.1) Rappresentare, mediante un grafico *bubble diagram*, le coppie di dati e la funzione di regressione per lo studio di $Y|X$.



- 4.2) Si stimino, attraverso il criterio dei minimi quadrati, i parametri del modello $Y^* = a + bX$

$$a = 10 \quad b = 2$$

- 4.3) Calcolare ρ^2 . $\rho^2 = 0.36$

- 4.4) Calcolare i valori di $\eta_{Y|X}^2$ ed $\eta_{X|Y}^2$. $\eta_{Y|X}^2 = \eta_{X|Y}^2 = 0.36$

commenti Si osserva come sia studiando Y in funzione di X , che X in funzione di Y si hanno 2 distribuzioni condizionate, quindi 2 medie condizionate $\rightarrow$ modello retta e funzione di regressione coincidono e presentano lo stesso valore della misura di adattamento.

- 5) Un gioco a premi viene organizzato nel seguente modo: il concorrente lancia 2 dadi e se il prodotto dei numeri presenti sulle due facce è superiore a 10 vince un premio.

- 5.1) Calcolare la probabilità che su 3 lanci un concorrente vinca 1 premio.

$$P(\text{successo nella singola prova}) = \frac{17}{36} = 0.4722$$

$$P(\text{su 3 lanci un concorrente vinca 1 premio}) = 0.3946$$

procedimento $X = \text{numero vincite su 3 lanci}$ $X \sim \text{Bin}(3, 0.4722)$ $P(X = 1) = \binom{3}{1} p (1 - p)^2 = \binom{3}{1} 0.4722 \cdot (1 - 0.4722)^2$

- 5.2) Calcolare la probabilità che su 3 lanci vinca il premio solo al 3° tentativo.

$$\text{formula } (1 - p)(1 - p)p = (1 - 0.4722)(1 - 0.4722)0.4722 = 0.1315.$$

procedimento $P(\bar{V}_1) \cdot P(\bar{V}_2 | \bar{V}_1) \cdot P(V_3 | \bar{V}_1 \cap \bar{V}_2) = P(\bar{V}_1) \cdot P(\bar{V}_2) \cdot P(V_3)$.

- 5.3) Calcolare la probabilità che su 99 lanci vinca almeno 42 volte.

$$P(X \geq 42) = 0.8315$$

procedimento $X \sim \text{Bin}(n=99, p=0.4722)$; $M(X) = 46.75$; $\text{Var}(X) = 24.6736$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=46.75, \sigma^2=24.6736)$ e

$$P(X \geq 42) = P\left(\frac{X - \mu}{\sigma} \geq \frac{42 - \mu}{\sigma}\right) = P\left(Z \geq \frac{42 - \mu}{\sigma}\right) = P(Z \geq -0.96) = P(Z \leq 0.96) = 0.8315.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 14.02.2008 (Tema 241)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1) Con riferimento alle osservazioni della variabile mista (X,Y) effettuate su **150** unità statistiche, la seguente tabella riporta le distribuzioni condizionate relative della variabile $X|Y=alto$ e $X|Y=basso$.

$Y \backslash X$	0-15	15-25	25-35	35-55	
<i>alto</i>	0	0.30	0.40	0.30	1
<i>basso</i>	0.30	0.40	0.30	0	1

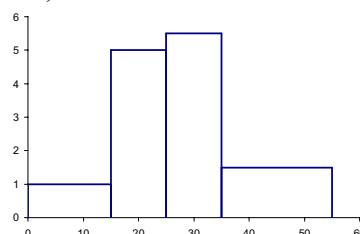
Sapendo che il numero di soggetti **alti** è **100**:

- 1.1) ricostruire la tabella doppia delle frequenze congiunte;

$Y \backslash X$	0-15	15-25	25-35	35-55	
<i>alto</i>	0	30	40	30	100
<i>basso</i>	15	20	15	0	50
	15	50	55	30	150

- 1.2) rappresentare graficamente la distribuzione della marginale X ;

X	$n_{\bullet j}$	$d_{\bullet j}$
0-15	15	1
15-25	50	5
25-35	55	5.5
35-55	30	1.5
	150	



- 1.3) indicare i valori possibili della media potenziata di ordine 15 e calcolare il valore della media armonica di X .

media armonica, formula: $\mu^{(-1)} = [M(X^{-1})]^{-1}$ valore 21.429

valori possibili per $\mu^{(15)}$: $21.429 = \mu^{(-1)} < \mu^{(15)} < 45 = x_{MAX}$

commenti in base alle proprietà delle medie potenziate

- 1.4) confrontare i livelli di variabilità delle due distribuzioni condizionate $X|Y=alto$ e $X|Y=basso$;

indice di variabilità utilizzato per il confronto: c.v. valore $X|Y=alto$ 0.3098 valore $X|Y=basso$ 0.4538

commenti la distribuzione $X|Y=basso$ presenta un livello di variabilità superiore alla distribuzione $X|Y=alto$

- 1.5) calcolare un opportuno indice di connessione tra le variabili X e Y .

indice χ_N valore 0.5592.

commenti Tra le variabili **W** e **T** sussiste un livello medio di connessione.

- 2) Sapendo che il valore di un aggregato macroeconomico nel 2006 è 80, che il numero indice riferito al tempo 2006 con base 2005 è pari a ${}_{2005}I_{2006} = 1.2$ e che tra il 2006 e il 2007 l'aggregato ha subito una variazione assoluta di -20 si determini ${}_{2005}I_{2007}$, numero indice riferito all'istante temporale 2007 con base 2005.

formula: ${}_{2005}I_{2007} = {}_{2005}I_{2006} \cdot {}_{2006}I_{2007} = 1.20 \cdot 0.75 = 0.90$

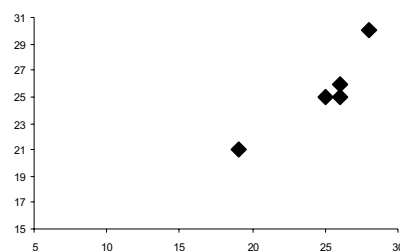
commenti: valore 2007: $80 - 20 = 60$, ${}_{2006}I_{2007} = \frac{60}{80} = 0.75$. Tra il 2005 e il 2007 l'aggregato ha subito un calo del 10%.

Studente _____ matricola _____

- 3) Per valutare l'efficacia delle lezioni in un corso universitario, della durata di 30 ore, sono disponibili i seguenti dati: numero di ore di lezione seguite (H) e voto d'esame (Y) di 5 studenti.

	Studente1	Studente2	Studente3	Studente4	Studente5
H	26	28	25	19	26
Y	25	30	25	21	26

- 3.1) Rappresentare graficamente le coppie di valori;



- 3.2) stimare attraverso il metodo dei minimi quadrati i parametri dei seguenti modelli

I) $Y^* = a + bH$

II) $Y^* = c + dH^2$;

$a = 3.991$

$b = 0.8632$

$c = 13.54$

$d = 0.019$

- 3.3) indicare quale modello si adatta meglio ai dati;

misura adattamento I) 0.8465

misura adattamento II) 0.8776

commenti il modello che si adatta meglio ai dati è $Y^* = 13.54 + 0.019 \cdot H^2$

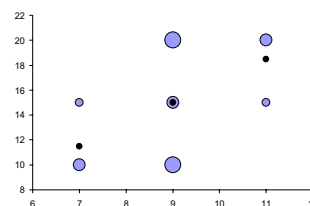
- 3.4) calcolare il voto teorico di uno studente che frequenta solo 10 ore.

$Y^*(10) = 13.54 + 0.019 \cdot 10^2 = 15.437 \approx 15$

- 4) Completare la seguente tabella in modo tale che $0 < \rho^2 = \eta_{Y|X}^2 < 1$:

$X \backslash Y$	10	15	20
7			0
9	12	7	12
11	0		

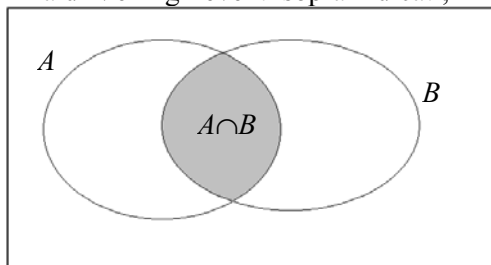
$X \backslash Y$	10	15	20
7	a	b	0
9	12	7	12
11	0	b	a



commenti medie condizionate allineate, no indipendenza in media, no dipendenza funzionale

- 5) In una lotteria si vince il premio a (evento A) con probabilità pari a 0.13 e il premio b (evento B) con probabilità pari a 0.15. Sapendo che la probabilità complessiva di vincere o uno o l'altro dei due premi è 0.20:

- 5.1) si rappresentino in un diagramma di Venn gli eventi sopra indicati;



- 5.2) si calcoli la probabilità di vincere entrambi i premi;

$p = 0.08$

procedimento $P(A \cup B) = P(A) + P(B) - P(A \cap B) \rightarrow P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.13 + 0.15 - 0.20$

- 5.3) si dica giustificando la risposta se gli eventi A e B sono indipendenti.

gli eventi A e B non sono indipendenti perché $0.08 = P(A \cap B) \neq P(A) \cdot P(B) = 0.13 \cdot 0.15 = 0.0195$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 05.06.2008 (Tema 242)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

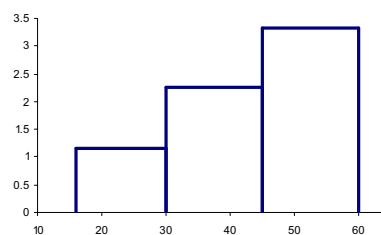
Studente _____ matricola _____

- 1) Con riferimento ad un'indagine conoscitiva effettuata su 100 lavoratori, la seguente tabella a doppia entrata riporta la distribuzione di frequenza per classi di età (X) e fasce di reddito (Y).

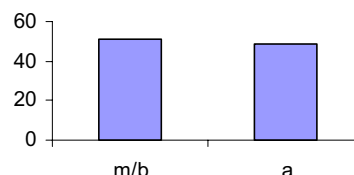
X	Y	16 - 30	30 - 45	45 - 60
$x_1 = \text{medio/bassa}$		11	15	25
$x_2 = \text{alta}$		5	19	25

- 1.1) Si dia una rappresentazione grafica delle distribuzioni marginali X e Y .

Y	$n_{\bullet j}$	$d_{\bullet j}$
16 - 30	16	1.14
30 - 45	34	2.27
45 - 60	50	3.33
	100	



X	$n_{i \bullet}$
$x_1 = \text{medio/bassa}$	51
$x_2 = \text{alta}$	49
	100



commenti: il carattere Y sembra presentare asimmetria negativa, il carattere X è quasi equidistribuito tra le modalità.

- 1.2) Si calcolino moda, mediana ed eterogeneità del carattere X .

moda(X), formula: $x_i; n_i = \max n_j$ _____ valore $x_1 = \text{medio/bassa}$

mediana(X), formula: _____ valore $x_1 = \text{medio/bassa}$

eterogeneità(X), formula: INDICE DI FROSINI NORM. _____ valore 0.98

commenti la distribuzione di X è quasi bimodale e presenta un grado di eterogeneità molto elevato

- 1.3) Si confrontino con opportuni indici di variabilità le distribuzioni di $Y|X=x_1$ e $Y|X=x_2$.

indice di variabilità utilizzato per il confronto: c.v. _____ valore $Y|X=x_1$ 0.2813 _____ valore $Y|X=x_2$ 0.2268

commenti la distribuzione $Y|X=x_1$ presenta maggiore variabilità rispetto alla distribuzione $Y|X=x_2$.

- 1.4) Si calcoli un opportuno indice di connessione e si costruisca la tabella a doppia entrata di indipendenza stocastica.

X	Y	16 - 30	30 - 45	45 - 60	
$x_1 = \text{medio/bassa}$		8.16	17.34	25.5	51
$x_2 = \text{alta}$		7.84	16.66	24.5	49
		16	34	50	100

indice χ_N _____ valore 0.1637

commenti Tra le variabili X e Y sussiste un livello di connessione molto basso.

2) È data la seguente distribuzione del carattere qualitativo S :

s_i	s_1	s_2	s_3
n_i	5	40	5

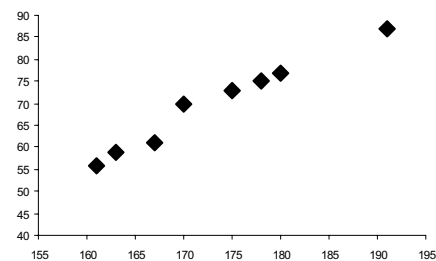
2.1) L'indice di eterogeneità di Gini risulta superiore quando il carattere S è sconnesso oppure se è ordinato?
 S ☐ sconnesso ☒ ordinato commenti l'indice di Gini è funzione solo delle frequenze e quindi il suo valore non dipende dalla tipologia del carattere oggetto di studio.

2.2) Si supponga ora S quantitativo con $s_1 = 10$ e $s_3 = 22$; si determini il valore di s_2 in modo che la media aritmetica di S sia eguale alla moda.
 $s_2 = 16$ commenti la distribuzione di S è simmetrica, quindi media aritmetica e moda coincidono

3) In un'indagine statistica è stato chiesto a 8 ragazzi di 18 anni, che svolgono regolare attività sportiva, di indicare rispettivamente A =altezza (in cm) e W =peso (in kg):

A	161	170	163	175	191	180	167	178
W	56	70	59	73	87	77	61	75

3.1) Si dia una rappresentazione grafica delle coppie (a_i, w_i) .



3.2) Si calcolino, secondo il principio dei minimi quadrati, i parametri dei seguenti modelli di regressione:

I) $W^* = \alpha + \beta A$

II) $W^* = \gamma A$

$\alpha = -110.23$

$\beta = 1.0396$

$\gamma = 0.4047$

3.3) Si individui, mediante un opportuno indice, il modello migliore, giustificando la risposta.

misura adattamento I) $RMS = 2.8488$

misura adattamento II) $RMS = 37.761$

commenti il modello che si adatta meglio ai dati è $W^* = -110.23 + 1.0396 A$ giustificazione: il modello retta completa è sicuramente migliore del modello retta con intercetta vincolata, come viene confermato dal confronto delle varianze residue

4) Siano A, B e C tre eventi caratterizzati da probabilità $P(A) = 0.2, P(B) = 0.5, P(C) = 0.8, P(A|B) = 0.2$ e $P(A|C) = 0$. Si dica, giustificando la risposta se:

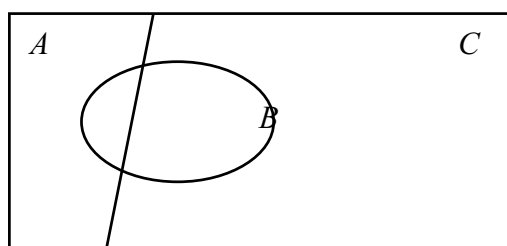
4.1) A e B sono indipendenti;

gli eventi A e B sono indipendenti ☒ SI ☐ NO perché $P(A) = P(A|B)$

4.2) A e C sono indipendenti.

gli eventi A e C sono indipendenti ☐ SI ☒ NO perché $P(A) \neq P(A|C)$

4.3) si rappresentino in un diagramma di Venn gli eventi A, B, C e si calcoli $P(A \cup B)$.



$P(A|C) = 0$

↓

$P(A \cap C) = 0 \leftrightarrow A \cap C = \emptyset$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.2 + 0.5 - 0.2 \cdot 0.5 = 0.6.$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 19.06.2008 (Tema 243)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Nel seguente prospetto sono riportati i risultati di una indagine statistica effettuata presso 10 soggetti, nella quale sono state rilevate le variabili X = “anzianità lavorativa” e Y = “n° giorni annui di formazione”.

soggetto	1	2	3	4	5	6	7	8	9	10
X	10	5	5	10	5	10	5	5	10	5
Y	5	7	6	7	9	8	10	11	9	13

- 1.1 Si sintetizzi la distribuzione della variabile Y attraverso mediana, media aritmetica e coefficiente di variazione.

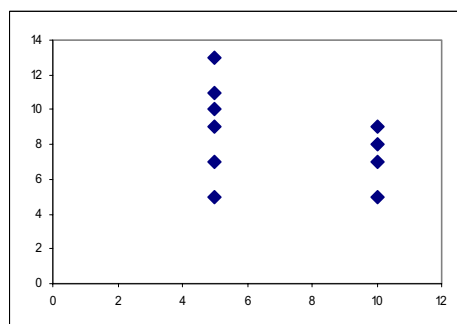
mediana(Y) 8.5

media aritmetica(Y) 8.5

coefficiente di variazione(Y) $\frac{\sigma}{\mu} = \frac{2.291}{8.5} = 0.2696$

osservazioni e commenti: per il calcolo della mediana occorre ordinare i dati; la serie Y risulta essere caratterizzata da un discreto grado di variabilità

- 1.2 Si rappresentino graficamente le coppie di valori (x_i, y_i) .



- 1.3 Si stimino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione

$$Y^* = a + bX \qquad Y^{**} = \alpha + \beta / X$$

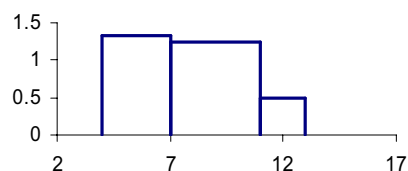
e se ne calcoli una opportuna misura di adattamento.

$$a = 11.4167 \qquad b = -0.4167 \qquad \alpha = 5.1667 \qquad \beta = 20.8333$$

$$\text{misura adattamento I) } \rho^2 = 0.1984 \qquad \text{misura adattamento II) } \rho^2 = 0.1984$$

- 1.4 Si raggruppino i valori di Y nelle 3 classi chiuse a destra $(4,7]$, $(7,11]$, $(11,13]$ e si dia una rappresentazione grafica della seriazione costruita con i dati riclassificati.

Y	n_i	a_i	d_i
4 → 7	4	3	1.33
7 → 11	5	4	1.25
11 → 13	1	2	0.5
	10		



- 1.5 Si costruisca la tabella a doppia entrata (con Y nelle 3 classi) e si valuti opportunamente il grado di connessione esistente tra X e Y .

$X \setminus Y$	4 → 7	7 → 11	11 → 13	
5	2	3	1	6
10	2	2	0	4
	4	5	1	10

indice χ valore 0.2887

commenti sussiste un basso livello di connessione tra le variabili X e Y .

2. Completare la seguente tabella, motivando le risposte, in modo tale che:

2.1 $\eta^2_{Y|X} = 0$ e $\rho^2 = 1$

$X \backslash Y$	10	15	20	
3				8
5				4
	3	6	3	12

commenti non è possibile in quanto $\rho^2 \leq \eta^2_{Y|X}$

2.2 $\rho^2 = 0$ e $\chi^2_{\text{NORM}} = 1$

$X \backslash Y$	10	15	20	
3				8
5				4
	3	6	3	12

commenti $\chi^2_{\text{NORM}} = 1 \leftrightarrow$ dipendenza funzionale;

con le frequenze marginali presenti nella tabella non è possibile realizzare la situazione di dipendenza funzionale

2.3 $\eta^2_{Y|X} = \chi^2_{\text{NORM}}$

$X \backslash Y$	10	15	20	
3	2	4	2	8
5	1	2	1	4
	3	6	3	12

commenti salvo situazioni particolari $\eta^2_{Y|X} = \chi^2_{\text{NORM}}$ solo nei casi estremi di indipendenza stocastica e di dipendenza funzionale; è, quindi, sufficiente costruire la tabella di indipendenza stocastica.

3. Da un'urna, che contiene 10 palline nere e 6 rosse e 14 blu, si estraggono 3 palline.

3.1 Si calcoli la probabilità che 2 delle palline estratte siano nere nell'ipotesi di estrazione senza reimmissione.

$$P = 0.2217$$

procedimento $P\{(N \cap N \cap \bar{N}) \cup (N \cap \bar{N} \cap N) \cup (\bar{N} \cap N \cap N)\} = P(N \cap N \cap \bar{N}) + P(N \cap \bar{N} \cap N) + P(\bar{N} \cap N \cap N)$;

$$P(N_1 \cap N_2 \cap \bar{N}_3) = P(N_1) \cdot P(N_2|N_1) \cdot P(\bar{N}_3|N_1 \cap N_2) = \frac{10}{30} \cdot \frac{9}{29} \cdot \frac{20}{28}$$

3.2 Si calcoli la probabilità che 2 delle palline estratte siano nere nell'ipotesi di estrazione con reimmissione.

$$P = 0.2222$$

procedimento $X \sim \text{Bin}(3, 0.3333)$ $P(X=2) = \binom{3}{2} p^2 (1-p) = \binom{3}{2} 0.3333^2 \cdot (1-0.3333)$

oppure $3 \cdot P(N_1 \cap N_2 \cap \bar{N}_3) = 3 \cdot P(N_1) \cdot P(N_2|N_1) \cdot P(\bar{N}_3|N_1 \cap N_2) = 3 \cdot P(N) \cdot P(N) \cdot P(\bar{N}) = 3 \cdot 0.3333 \cdot 0.3333 \cdot 0.6667$

3.3 Si calcoli la probabilità di ottenere 3 palline dello stesso colore nell'ipotesi di estrazione con reimmissione.

$$P = 0.1467$$

procedimento $P\{(N \cap N \cap N) \cup (R \cap R \cap R) \cup (B \cap B \cap B)\} = P(N \cap N \cap N) + P(R \cap R \cap R) + P(B \cap B \cap B) = \left(\frac{10}{30}\right)^3 + \left(\frac{6}{30}\right)^3 + \left(\frac{14}{30}\right)^3$

3.4 Effettuando 90 estrazioni con reimmissione si calcoli la probabilità di ottenere almeno 32 palline nere.

$$P(X \geq 32) = 0.3264$$

procedimento $X \sim \text{Bin}(n=90, p=0.3333)$; $M(X) = 30$; $\text{Var}(X) = 20$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=30, \sigma^2=20)$ e

$$P(X \geq 32) = P\left(\frac{X - \mu}{\sigma} \geq \frac{32 - \mu}{\sigma}\right) = P\left(Z \geq \frac{32 - \mu}{\sigma}\right) = P(Z \geq 0.45) = 1 - P(Z \leq 0.45) = 0.3264.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 03.07.2008 (Tema 244)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. In un'azienda viene rilevata la variabile Reddito annuo (X) espressa in migliaia di euro, separatamente per i maschi e per le femmine.

MASCHI	
x_i	d_i (densità)
0-20	2
20-40	1.5
40-60	0.4
60-100	0.6

FEMMINE	
x_i	f_i (freq. relative)
0-20	0.4
20-40	0.3
40-60	0.2
60-100	0.1

- 1.1 Confrontare i due gruppi con un opportuno indice di posizione e un opportuno indice di variabilità.

MASCHI

indice di posizione $M(X|\text{MASCHI}) = 35.490$

indice di variabilità c.v.($X|\text{MASCHI}$) = 0.770

FEMMINE

indice di posizione $M(X|\text{FEMMINE}) = 31$

indice di variabilità c.v.($X|\text{FEMMINE}$) = 0.713

commenti Il livello medio del reddito dei maschi è superiore a quello delle femmine, la distribuzione dei redditi dei maschi è leggermente più variabile di quella delle femmine.

- 1.2 Sapendo che le femmine sono 50, determinare la media di X .

$$M(X) = \frac{\mu_{\text{MASCHI}} n_{\text{MASCHI}} + \mu_{\text{FEMMINE}} n_{\text{FEMMINE}}}{n_{\text{MASCHI}} + n_{\text{FEMMINE}}} = \frac{35.490 \cdot 102 + 31 \cdot 50}{102 + 50}$$

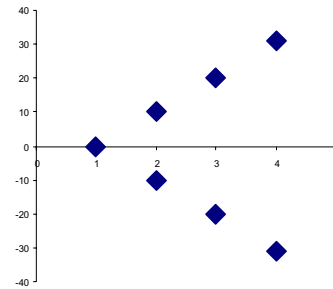
- 1.3 Calcolare la varianza totale di X con i dati dei due gruppi.

$$\text{Var}(X) = \sigma_{\text{Between}}^2 + \sigma_{\text{Within}}^2 = \frac{(\mu_X - \mu)^2 n_X + (\mu_Y - \mu)^2 n_Y}{n_X + n_Y} + \frac{\sigma_X^2 n_X + \sigma_Y^2 n_Y}{n_X + n_Y} = 4.451 + 661.681 = 666.131.$$

2. Date le seguenti coppie di valori:

(1;0); (2;-10); (3;-20); (4;-31); (2;10); (3;20); (4;31);

2.1 Rappresentare graficamente le coppie di valori (x_i, y_i) .



2.2 Si calcolino, secondo il principio dei minimi quadrati, i parametri dei seguenti modelli di regressione:

I) $Y = a + bX$

II) $Y = c \cdot X$

$a = 0$

$b = 0$

$c = 0$

2.3 Si individui, mediante un opportuno indice, il modello migliore, giustificando la risposta.

misura adattamento I) RMS = 417.429

misura adattamento II) RMS = 417.429

commenti Per il secondo modello non è possibile definire l'indice di adattamento

2.4 Calcolare $\eta_{Y|X}^2$.

Dal grafico si evince che sussiste indipendenza in media di Y rispetto a X quindi $\eta_{Y|X}^2 = 0$

2.5 Calcolare $\eta_{X|Y}^2$.

Dal grafico si evince che sussiste dipendenza funzionale di X rispetto a Y quindi $\eta_{X|Y}^2 = 1$

3. Date due variabili X e Y , sono note le seguenti informazioni:

$M(X) = M(Y) = 4$; $M(X^2) = M(Y^2) = M(XY) = 21$.

3.1 Determinare ρ^2 e χ^2 normalizzato.

commenti $\rho^2 = \frac{\text{Cov}(X,Y)^2}{\text{Var}(X) \cdot \text{Var}(Y)} = \frac{[M(X,Y) - M(X)M(Y)]^2}{[M(X^2) - M(X)^2][M(Y^2) - M(Y)^2]} = \frac{(21 - 4 \cdot 4)^2}{[21 - 4^2][21 - 4^2]} = 1$; tra le variabili X e Y

sussiste quindi dipendenza funzionale (lineare); quindi anche χ^2 normalizzato = 1.

4. Sia $X \sim N(\mu = 9; \sigma^2 = 25)$, determinare $P(8 < X < 11)$.

$$P(8 < X < 11) = P(X < 11) - P(X < 8) = P\left(\frac{X - \mu}{\sigma} < \frac{11 - \mu}{\sigma}\right) - P\left(\frac{X - \mu}{\sigma} < \frac{8 - \mu}{\sigma}\right) = P\left(Z < \frac{11 - \mu}{\sigma}\right) - P\left(Z < \frac{8 - \mu}{\sigma}\right) = \\ = P(Z < 0.4) - P(Z < -0.2) = 0.6554 - 0.4207 = 0.2347.$$

5. Ad un'uscita autostradale ci sono 3 caselli, il primo con pagamento con contanti, il secondo con pagamento con carta, il terzo con pagamento Telepass. Le probabilità di chiudere le 3 procedure di pagamento in meno di 25 secondi sono rispettivamente di 0.2, 0.5 e 0.8. Sappiamo che il 51% degli automobilisti paga in contanti, che il 20% con carta e il restante con Telepass.

5.1 Calcolare la probabilità di eseguire la procedura in meno di 25 secondi.

$$P = P((Co \cap Ok) \cup (Ca \cap Ok) \cup (Te \cap Ok)) = P(Co)P(Ok|Co) + P(Ca)P(Ok|Ca) + P(Te)P(Ok|Te) = 0.51 \cdot 0.2 + 0.20 \cdot 0.5 + 0.29 \cdot 0.8 = 0.434$$

5.2 Sapendo che un automobilista ha impiegato più di 25 secondi, calcolare la probabilità che abbia utilizzato la carta.

$$P = P(Ca|\overline{Ok}) = \frac{P(Ca \cap \overline{Ok})}{P(\overline{Ok})} = \frac{P(Ca)P(\overline{Ok}|Ca)}{P(\overline{Ok})} = \frac{0.2 \cdot 0.5}{1 - 0.434} = 0.1767.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 04.09.2008 (Tema 245)

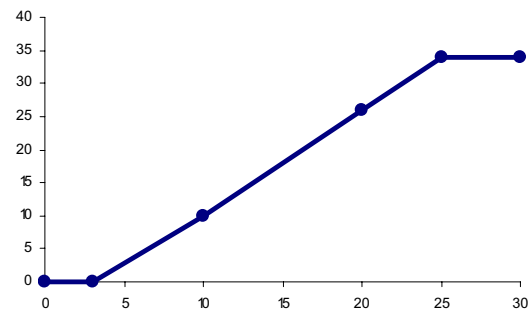
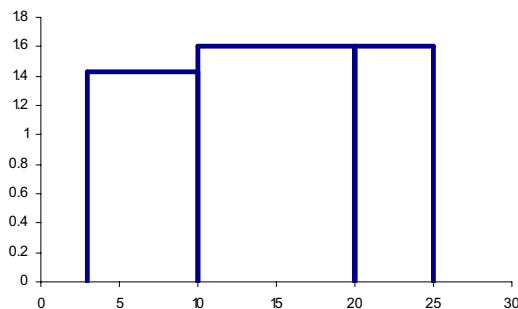
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la seguente seriazione X

x_i	n_i
3-10	10
10-20	16
20-25	8

1.1 Fornire una rappresentazione grafica della distribuzione X e della sua funzione di ripartizione.



1.2 Calcolare moda, media aritmetica, media armonica e varianza di X .

moda classe alla quale è associata la max densità di frequenza (10, 25], valore rapp. 17.5

media aritmetica $M(X) = 14.265$

media armonica $\mu^{(-1)} = 11.484$

varianza $\text{Var}(X) = 33.945$

1.3 Indicare l'intervallo dei possibili valori che può assumere la media quadratica di X .

$$M(X) = \mu^{(1)} < \mu^{(2)} < 22.5 = x_{\max}$$

2. Sono date la v.s. X , avente $M(X) = \text{Var}(X) = 4$, e la v.s. Y della quale è nota la relazione $Y = 3 - 3X$.

2.1 Calcolare media e varianza di Y .

$$M(Y) = 3 - 3 M(X) = 3 - 3 \cdot 4 = -9$$

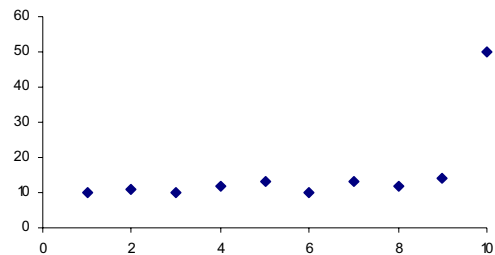
$$\text{Var}(Y) = 3^2 \cdot \text{Var}(X) = 3^2 \cdot 4 = 36$$

commenti si sono applicate le proprietà degli operatori media aritmetica e varianza

3. Sono date le seguenti coppie di valori (x_i, y_i) :

x_i	1	2	3	4	5	6	7	8	9	10
y_i	10	11	10	12	13	10	13	12	14	51

3.1 Rappresentare graficamente i dati.



3.2 Si determinino, secondo il criterio dei minimi quadrati, i parametri dei seguenti modelli:

$$\text{I) } Y^* = a + bX \quad \text{II) } Y^{**} = \alpha + \beta X^3.$$

$$a = 2.2667$$

$$b = 2.4242$$

$$\alpha = 7.1965$$

$$\beta = 0.0278$$

3.3 Si determini quello con migliore adattamento.

$$\text{misura adattamento I) } \rho^2 = 0.3438$$

$$\text{misura adattamento II) } \rho^2 = 0.5818$$

commenti Il secondo modello presenta un adattamento migliore rispetto ai dati

3.4 Calcolare $\eta_{Y|X}^2$. $\eta_{Y|X}^2 = 1$ sussiste, infatti, dipendenza funzionale di Y da X .

3.5 Indicare il valore della varianza residua del modello $Y^{***} = \text{costante}$.

$$\text{varianza residua} = \text{Var}(Y) = 141.04$$

4. Si faccia riferimento ad una *slot-machine* dotata di 3 finestrelle, all'interno delle quali si possono presentare, ad ogni lancio, i numeri da 0 a 9, in maniera casuale e indipendente.

4.1 Calcolare la probabilità che si presentino 3 numeri uguali.

$$P = 0.01$$

procedimento $10 \cdot P(x \cap x \cap x)$, con $x = 0, 1, 2, \dots, 9$ generico valore, $P(x \cap x \cap x) = P(x) \cdot P(x) \cdot P(x)$ per l'ipotesi di indipendenza, infine $P(x) = 0.1$.

4.2 Calcolare la probabilità che escano tutti pari.

$$P = 0.125$$

procedimento $P(\text{pari} \cap \text{pari} \cap \text{pari}) = P(\text{pari}) \cdot P(\text{pari}) \cdot P(\text{pari})$ per l'ipotesi di indipendenza, infine $P(\text{pari}) = 0.5$.

4.3 Eseguendo 101 lanci, calcolare la probabilità che almeno 20 presentino tutti numeri pari.

$$P(X \geq 20) = 0.0132$$

procedimento $X \sim \text{Bin}(n=101, p=0.125)$; $M(X) = 12.625$; $\text{Var}(X) = 11.0469$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=12.625, \sigma^2=11.0469)$ e

$$P(X \geq 20) = P\left(\frac{X - \mu}{\sigma} \geq \frac{20 - \mu}{\sigma}\right) = P\left(Z \geq \frac{20 - \mu}{\sigma}\right) = P(Z \geq 2.22) = 1 - P(Z \leq 2.22) = 1 - 0.9868 = 0.0132.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 18.09.2008 (Tema 246)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la seguente tabella, che riporta i risultati della rilevazione delle mutabili X = “abitudine al fumo” (con valori: $\bar{F}$ = fumatore non accanito; F = fumatore accanito) e Y = “efficacia di una terapia antifumo” (con valori: $\bar{E}$ = non efficace; E = efficace), espressi in termini di percentuali di colonna (distribuzioni condizionate).

	$X Y=\bar{E}$	$X Y=E$
$\bar{F}$	71%	45%
F	29%	55%
	100%	100%

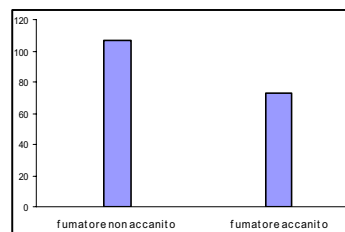
- 1.1 Sapendo che $n = 180$ e che $n_{\bullet 1} = 100$, si costruisca la tabella delle frequenze congiunte assolute.

$X \backslash Y$	$\bar{E}$	E	
$\bar{F}$	71	36	107
F	29	44	73
	100	80	180

- 1.2 Si costruisca la distribuzione di frequenze della marginale X e se ne fornisca una rappresentazione grafica.

Distribuzione della marginale X

X	$n_{i\bullet}$
$\bar{F}$	107
F	73
	180



- 1.3 Si calcolino opportuni indici di posizione e di variabilità per la distribuzione di X .

indice di posizione MODA

valore $\bar{F}$ = fumatore non accanito

indice di variabilità FROSINI normalizzato

valore 0.8111

- 1.4 Si valuti la connessione esistente tra le due variabili mediante l'impiego di un opportuno indice.

indice di connessione χ_N

valore 0.2631

- 1.5 Estratti a caso dai 180 fumatori 4 diversi nominativi, calcolare la probabilità che questi siano tutti fumatori accaniti.

$$P = 0.0257$$

$$\text{procedimento } P(F_1 \cap F_2 \cap F_3 \cap F_4) = P(F_1) \cdot P(F_2|F_1) \cdot P(F_3|F_1 \cap F_2) \cdot P(F_4|F_1 \cap F_2 \cap F_3) = \frac{73}{180} \cdot \frac{72}{179} \cdot \frac{71}{178} \cdot \frac{70}{177} = 0.0257.$$

2. Date due variabili X e Y , sia $Y^* = 3 - 0.5X$ la retta di regressione che spiega Y in funzione di X . Sapendo che $M(X) = 3$, $M(X^2) = M(Y^2) = 36$, si determini $\text{Cov}(X, Y)$.

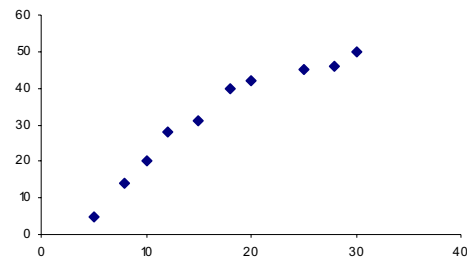
$$\text{Cov}(X, Y) = -13.5$$

$$\text{commenti } b = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} \rightarrow \text{Cov}(X, Y) = b \cdot \text{Var}(X) = b \cdot \{M(X^2) - [M(X)]^2\}.$$

3. Il seguente prospetto riporta i valori del tempo Y (in minuti) impiegato per lo svolgimento di 10 diversi test attitudinali e del livello di difficoltà X di tali test.

x_i	5	8	10	12	15	18	20	25	28	30
y_i	5	14	20	28	31	40	42	45	46	50

- 3.1 Si rappresentino graficamente le coppie (x_i, y_i) .



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b\sqrt{X} \quad \text{e} \quad II) Y^{**} = \alpha + \beta X.$$

$$a = -22.7123$$

$$b = 13.6869$$

$$\alpha = 3.3333$$

$$\beta = 1.6823$$

- 3.3 Si indichi quello che presenta il migliore adattamento.

$$\text{misura adattamento I)} \rho^2 = 0.9626$$

$$\text{misura adattamento II)} \rho^2 = 0.9131$$

commenti il modello migliore è $Y^* = -22.7123 + 13.6869\sqrt{X}$

- 3.4 In base al modello migliore, si determini il tempo previsto per svolgere un test di difficoltà pari a 40.

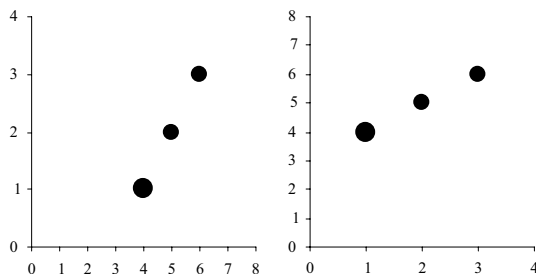
tempo previsto 63.85.

4. Si indichi, per ognuna delle seguenti tabelle a doppia entrata, il valore assunto dagli indici sotto indicati, motivando le risposte.

4.1

X \ Y	4	5	6
1	10	0	0
2	0	8	0
3	0	0	6

$$\eta_{X|Y}^2, \eta_{Y|X}^2, \rho^2, \chi_N^2$$



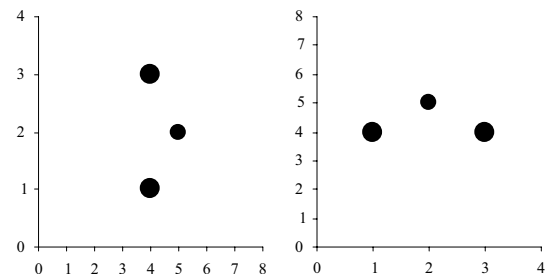
$$\eta_{X|Y}^2 = 1, \eta_{Y|X}^2 = 1, \rho^2 = 1, \chi_N^2 = 1 \quad (\chi^2 = 48)$$

dipendenza funzionale di X da Y e di Y da X, con medie condizionate allineate, quindi dipendenza funzionale lineare

4.2

X \ Y	4	6
1	10	0
2	0	8
3	10	0

$$\eta_{X|Y}^2, \eta_{Y|X}^2$$

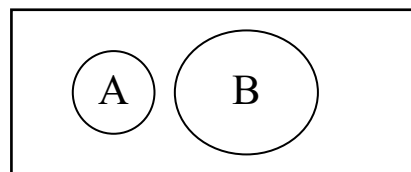


$$\eta_{X|Y}^2 = 0$$

$$\eta_{Y|X}^2 = 1$$

indipendenza in media di X da Y
dipendenza funzionale di Y da X

5. Siano A e B due eventi disgiunti, tali che $P(A) = 0.15$ e $P(B) = 0.4$. Indicando con $C = A \cup B$, calcolare:



- 5.1 $P(C)$;

$$P(C) = P(A) + P(B) - P(A \cap B) = P(A) + P(B) - P(\emptyset) = 0.15 + 0.4 - 0 = 0.55.$$

- 5.2 $P(A|C)$.

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(A)}{P(C)} = \frac{0.15}{0.55} = 0.2727.$$

- 5.3 Dire perché A e C non sono indipendenti.

$$P(A|C) \neq P(A).$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 09.01.2009 (Tema 247)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

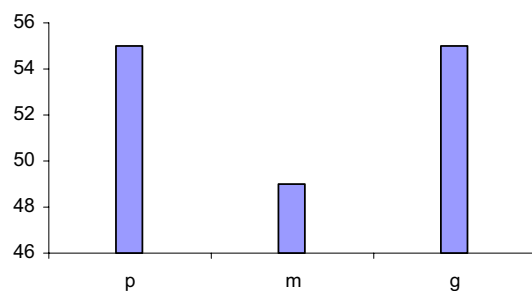
1. In un'indagine sulla produzione di lana in un allevamento italiano sono state rilevate su 127 pecore le seguenti variabili: X = taglia della pecora (p = piccola, m = media e g = grande) e Y = quantità di lana prodotta in un anno (espressa in hg). La tabella seguente riporta le frequenze congiunte:

$X \backslash Y$	5-10	10-20	20-25
p	29	12	14
m	11	27	11
g	14	12	29

- 1.1 Si rappresentino graficamente le distribuzioni delle marginali X e Y .

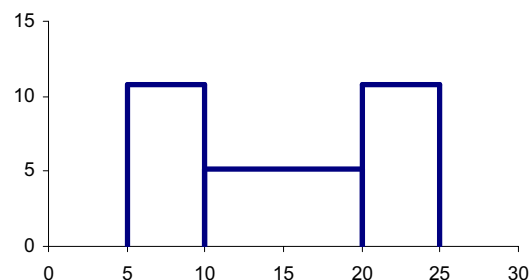
Distribuzione della marginale X

x_i	$n_{i\bullet}$
p	55
m	49
g	55



Distribuzione della marginale Y

$h_{i-1}-h_i$	$n_{\bullet j}$	$a_{\bullet j}$	$d_{\bullet j}$
5-10	54	5	10.8
10-20	51	10	5.1
20-25	54	5	10.8



commenti X carattere qualitativo: grafico a bastoncini; Y carattere quantitativo con classi: istogramma; la distribuzione di Y è simmetrica rispetto a 15; con riferimento a X sono maggiormente presenti le modalità p e g .

- 1.2 Si confrontino con opportuni indici di posizione e di variabilità le distribuzioni condizionate di $X|_{Y_1}$ e $X|_{Y_3}$, commentando opportunamente i risultati ottenuti.

indice di posizione	Moda / mediana	valore $X _{Y_1}$	p / p	valore $X _{Y_3}$	g / g
indice di variabilità	Indice di Frosini norm	valore $X _{Y_1}$	0.6906	valore $X _{Y_3}$	0.6906

commenti con riferimento alle pecore che producono poca lana la moda e la mediana sono taglia piccola, mentre per quanto attiene a quelle che producono molta lana la moda risulta taglia grande; il valore medio/alto degli indici di eterogeneità suggerisce che la moda non debba essere considerata quale valore rappresentativo.

- 1.3 Si determini, eventualmente senza effettuare calcoli, se il coefficiente di variazione di Y è maggiore, minore o uguale a quello di $W = Y+2$.

commenti $c.v.(W) = \frac{\sigma_W}{\mu_W}$; $\sigma_W = \sigma_Y$; $\mu_W = \mu_Y + 2$; quindi $\mu_W > \mu_Y$ e $c.v.(W) < c.v.(Y)$.

- 1.4 Si dica perché fra X e Y non c'è indipendenza stocastica e si calcoli un opportuno indice di connessione. X e Y non sono stocasticamente indipendenti perché le distribuzioni condizionate non sono somiglianti.

indice di connessione	χ_{NORM}	valore	0.3035 (livello connessione basso)
-----------------------	----------------------	--------	------------------------------------

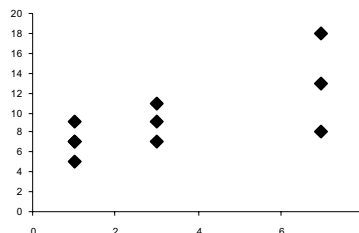
- 1.5 Si calcoli il valore della media geometrica della v.s. $Y|_{X=g}$.

media geometrica 15.5709

2. Da un'indagine sui consumi energetici familiari legati all'utilizzo di lavatrici di classe A sono state rilevate per 10 famiglie le seguenti variabili: X = numero settimanale di lavaggi e Y = consumo energetico settimanale (in kw) per il solo utilizzo della lavatrice. I risultati sono riportati nella seguente tabella:

X	1	3	1	3	1	7	3	7	7	1
Y	5	7	7	9	9	8	11	13	18	7

2.1 Si rappresentino graficamente le coppie di dati.



2.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

I) $Y^* = a + bX$ e II) $Y^* = cX$.

$a = 6$

$b = 1$

$c = 2.1461$

2.3 Si indichi quello che presenta il migliore adattamento.

misura adattamento I) RMS = 6.6

misura adattamento II) RMS = 19.2202

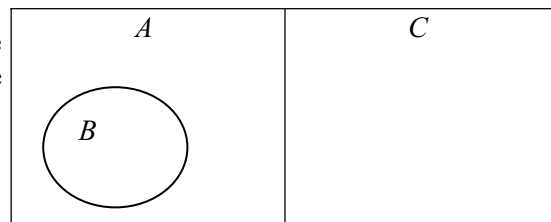
commenti il modello retta presenta un migliore adattamento rispetto al modello vincolato

2.4 Si dica, senza effettuare calcoli e motivando la risposta, che valore assume il rapporto di correlazione $\eta^2_{Y|X}$.

$\eta^2_{Y|X} = \rho^2 = 0.486$ commenti le medie condizionate sono allineate, quindi la funzione di regressione coincide con il modello retta completa.

3. Siano A , B e C tre eventi caratterizzati da: $P(A)=0.5$, $P(B)=0.22$, $P(C)=0.5$, $P(A|B)=1$, $P(A|C)=0$.

3.1 Si rappresentino in un diagramma di Venn gli eventi A , B , C e si indichi se (e perché) A , B e C costituiscono una partizione dello spazio campionario Ω



$P(A|B)=1 \rightarrow B \subset A$, $P(A|C)=0 \rightarrow A \cap C = \emptyset$

Gli insiemi A e C costituiscono una partizione di Ω , in quanto $A \cup C = \Omega$ e $A \cap C = \emptyset$

3.2 Si calcoli la probabilità di ottenere 2 successi in 10 prove indipendenti essendo 0.22 la probabilità di successo nella singola prova.

$P(X=2) = 0.2984$

procedimento $X \sim \text{Bin}(n=10, p=0.22)$; $P(X=2) = \binom{10}{2} \cdot 0.22^2 \cdot 0.78^8$.

3.3 Si calcoli la probabilità di ottenere almeno 21 successi in 100 prove indipendenti essendo 0.22 la probabilità di successo nella singola prova.

$P(X \geq 21) = 0.5948$

procedimento $X \sim \text{Bin}(n=100, p=0.22)$; $M(X) = 22$; $\text{Var}(X) = 17.16$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=22, \sigma^2=17.16)$ e

$P(X \geq 21) = P\left(\frac{X - \mu}{\sigma} \geq \frac{21 - \mu}{\sigma}\right) = P\left(Z \geq \frac{21 - 22}{4.1425}\right) = P(Z \geq -0.24) = P(Z \leq 0.24) = 0.5948$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 29.01.2009 (Tema 248)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

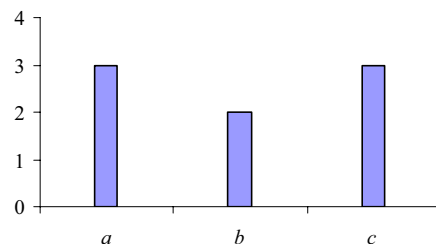
1. Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a un carattere qualitativo X e a un carattere quantitativo Y :

X	a	b	b	c	a	a	c	c
Y	30	20	24	50	40	50	40	30

- 1.1 Si ricostruisca la mutabile statistica X e se ne dia una rappresentazione grafica.

mutabile X

x_i	n_i
a	3
b	2
c	3
	8



commenti: la distribuzione della mutabile X è bimodale

- 1.2 Si riassumano con opportuni indici di posizione e di variabilità le distribuzioni X e Y , commentando opportunamente i risultati ottenuti.

X : indice di posizione Moda valore a, c indice di variabilità Frosini valore 0.8751

Y : indice di posizione $M(Y)$ valore 35.5 indice di variabilità c.v. valore 0.2978

commenti la mutabile X è bimodale; nel caso in esame le due modalità che presentano la massima frequenza non possono essere considerate rappresentative di X .

- 1.3 Si indichi l'intervallo dei possibili valori che può assumere la media armonica di Y .

Ricordando le proprietà delle medie potenziate: $20 = y_{\min} < \mu^{(-1)} < \mu = 35.5$

- 1.4 Si definiscano le relazioni esistenti rispettivamente tra le medie e le varianze di Y e di $W = 100 - 2Y$ e si calcolino quindi media e varianza di W .

relazioni: $M(W) = 100 - 2 \cdot M(Y)$ $\text{Var}(W) = 2^2 \cdot \text{Var}(Y)$

valori: $M(W) = 100 - 2 \cdot 35.5 = 29$ $\text{Var}(W) = 4 \cdot 111.75 = 447$

commenti si sono applicate le proprietà degli operatori media e varianza con riferimento a una trasformazione lineare.

2. Con riferimento ad n unità statistiche si sono raccolti alcuni dati in una tabella a doppia entrata:

- 2.1 Determinare, motivando le risposte, i valori delle frequenze assolute mancanti in modo tale che $\chi^2 = 0$.

$X \setminus Y$	y_1	y_2	y_3	$n_{i\bullet}$
x_1	29	10		
x_2			6	
$n_{\bullet j}$				45

commenti

dal momento che la somma delle frequenze congiunte risulta uguale a n , le frequenze mancanti possono assumere solo valore nullo, si realizza così una situazione di dipendenza funzionale di X da Y .
 È quindi impossibile che $\chi^2 = 0$.

$X \setminus Y$	y_1	y_2	y_3	$n_{i\bullet}$
x_1	20		30	
x_2	10	11	15	
$n_{\bullet j}$				

commenti

ponendo $n_{12} = 22$ le distribuzioni condizionate risultano somiglianti e quindi $\chi^2 = 0$.

$X \setminus Y$	y_1	y_2	y_3	$n_{i\bullet}$
x_1				
x_2	29	0	15	
$n_{\bullet j}$				

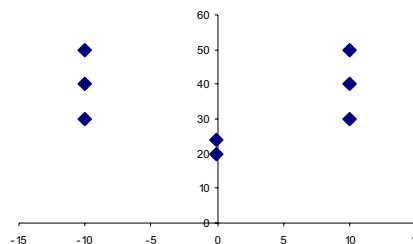
commenti

dal momento che la frequenza congiunta $n_{22} = 0$ si può escludere che $\chi^2 = 0$, ossia la presenza di indipendenza stocastica, in quanto non può essere verificata la condizione di fattorizzazione.

3. Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y :

X	-10	0	0	10	-10	-10	10	10
Y	30	20	24	50	40	50	40	30

3.1 Si rappresentino graficamente le coppie di dati.



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + bX \quad \text{e} \quad \text{II) } Y^* = c + dX^2.$$

$$a = 35.5$$

$$b = 0$$

$$c = 22$$

$$d = 0.18$$

3.3 Si indichi quale modello presenta il migliore adattamento.

$$\text{misura adattamento I) } \rho^2 = 0$$

$$\text{misura adattamento II) } \rho^2 = 0.5436$$

Dall'analisi del grafico e dai risultati numerici si evince la presenza di una situazione di indipendenza lineare. Il modello II) coincide con la funzione di regressione in quanto le medie condizionate sono su una parabola.

4. Con riferimento ad n unità statistiche si sono raccolti alcuni dati in una tabella a doppia entrata:

4.1 Determinare i valori delle frequenze mancanti e delle modalità di X e Y in modo che $\eta_{Y|X}^2 = 1$.

$X \setminus Y$			
	29		
			6

Impossibile, in quanto non è possibile costruire le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ entrambe degeneri.

4.2 Determinare i valori delle frequenze mancanti e delle modalità di X e Y in modo che $0 < \rho^2 = \eta_{Y|X}^2 < 1$.

$X \setminus Y$			
	29		
			6

$\rho^2 = \eta_{Y|X}^2$ se le medie condizionate $M(Y|x)$ sono allineate e nel caso in esame le due medie $M(Y|x_1)$ e $M(Y|x_2)$ sono sempre allineate. Per quanto attiene la scelta delle modalità e delle frequenze è sufficiente definire una qualsiasi configurazione tale che $x_1 < x_2$; $y_1 < y_2 < y_3$ e $n_{ij} \geq 0$ ricordandosi di escludere la situazione $M(Y|x_1) = M(Y|x_2)$ per la quale $\rho^2 = \eta_{Y|X}^2 = 0$.

5. Si considerino due urne, indicate con U_1 e U_2 . L'urna U_1 contiene 14 palline rosse e 6 palline blu; l'urna U_2 contiene 48 palline rosse e 12 palline blu. Si estraggono 3 palline da ciascuna urna.

5.1 Nell'ipotesi di estrazioni senza reimmissione si calcoli la probabilità che le 3 palline estratte dall'urna U_1 siano blu e le altre di colore qualsiasi.

$$P = 0.0175$$

procedimento Indicando con i pedici 1, 2, 3 e 4, 5, 6 rispettivamente le 3 estrazioni dalle urne U_1 e U_2 si ha:

$$P\{(B_1 \cap B_2 \cap B_3) \cap [(R_4 \cup B_4) \cap (R_5 \cup B_5) \cap (R_6 \cup B_6)]\} = P\{(B_1 \cap B_2 \cap B_3)\} \cdot 1 = P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 \cap B_2) = \frac{6}{20} \cdot \frac{5}{19} \cdot \frac{4}{18}$$

5.2 Nell'ipotesi di estrazioni con reimmissione si calcoli la probabilità che 2 delle 6 palline estratte siano blu.

procedimento Siano X_1 e X_2 le variabili casuali (binomiali) che descrivono il numero di palline blu nelle 3 estrazioni con reimmissione dalle urne U_1 e U_2 : $X_1 \sim \text{Bin}(3, \frac{6}{20}=0.3)$ e $X_2 \sim \text{Bin}(3, \frac{12}{60}=0.2)$. L'evento in oggetto ha probabilità:

$$P\{[(X_1=2) \cap (X_2=0)] \cup [(X_1=1) \cap (X_2=1)] \cup [(X_1=0) \cap (X_2=2)]\} = P(X_1=2)P(X_2=0) + P(X_1=1)P(X_2=1) + P(X_1=0)P(X_2=2) = \binom{3}{2}0.3^2 \cdot 0.7^1 \cdot \binom{3}{0}0.2^0 \cdot 0.8^3 + \binom{3}{1}0.3^1 \cdot 0.7^2 \cdot \binom{3}{1}0.2^1 \cdot 0.8^2 + \binom{3}{0}0.3^0 \cdot 0.7^3 \cdot \binom{3}{2}0.2^2 \cdot 0.8^1 = 0.033 + 0.169 + 0.097 = 0.299.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 12.02.2009 (Tema 249)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. In un'indagine condotta su **250** studenti sono state rilevate la mutabile statistica X = "valutazione del servizio fotocopisteria" e la variabile statistica Y = "numero di servizi richiesti in un anno". Nel seguente prospetto sono riportate le distribuzioni delle frequenze condizionate relative $Y|x_i, i = 1, 2$:

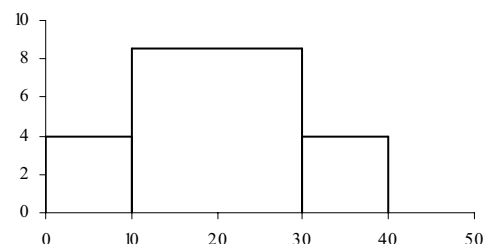
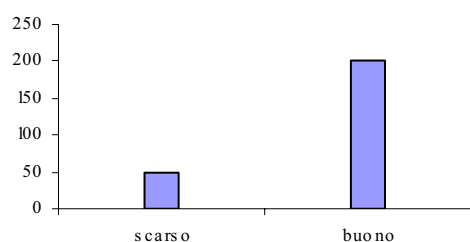
$X \setminus Y$	0 - 10	10 - 30	30 - 40	
scarso	0.4	0.2	0.4	1
buono	0.1	0.8	0.1	1

- 1.1 Si ricostruisca la tabella a doppia entrata delle frequenze congiunte sapendo che gli studenti che hanno dato giudizio "scarso" sono 50.

$X \setminus Y$	0 - 10	10 - 30	30 - 40	
scarso	20	10	20	50
buono	20	160	20	200
	40	170	40	250

commenti $n_{ij} = f.c.r.(y_j|x_i) \cdot n(x_i)$

- 1.2 Si dia una rappresentazione grafica delle distribuzioni marginali X e Y .



- 1.3 Si calcolino opportuni indici di variabilità per le variabili marginali X e Y .

X : indice di variabilità Frosini valore 0.4 Y : indice di variabilità c.v. valore 0.4243

commenti la mutabile X presenta un basso livello di mutabilità.

- 1.4 Si dica perché X e Y non sono stocasticamente indipendenti e si calcoli una misura della loro connessione.

Le distribuzioni condizionate relative $Y|x_i$, non sono somiglianti.

indice χ_{NORM} valore 0.5135

commenti il livello di connessione tra X e Y è medio

2. Si riporta l'evoluzione temporale dei prezzi e delle quantità scambiate di 2 prodotti a e b , tra il 2000 il 2002.

	a		b	
	$a p_t$	$a q_t$	$b p_t$	$b q_t$
2000	330	10	80	10
2001	345	$a q_{2001}$	80	$b q_{2001}$
2002	390	10	85	10

- 2.1 Si calcoli l'indice dei prezzi di Laspeyres al tempo 2002 con base 2000.

$${}^{2000}_P I^L_{2002} = 1.1585$$

- 2.2 Supponendo che l'indice dei prezzi di Fisher, ${}^{2000}_P I^F_{2001}$, al tempo 2001 con base 2000, sia risultato pari a 1.033, valore inferiore a quello dell'indice di Laspeyres, si indichi l'intervallo dei possibili valori dell'indice di Paasche.

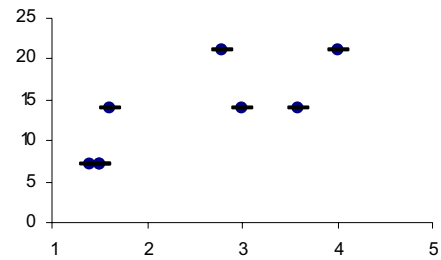
$1 < {}^{2000}_P I^P_{2001} < 1.033$ Commenti: ${}^{2000}_P I^P_{2001}$ è una media potenziata dei prezzi relativi ed è quindi compreso tra il prezzo relativo minimo e quello massimo; ${}^{2000}_P I^F_{2001}$ è la media geometrica tra ${}^{2000}_P I^P_{2001}$ e ${}^{2000}_P I^L_{2001}$ e quindi ${}^{2000}_P I^P_{2001} < {}^{2000}_P I^F_{2001}$.

3. Su un campione di 8 famiglie è stato rilevato il reddito mensile medio, R , in migliaia di euro e il numero di giorni di vacanza effettuati nello stesso anno, V :

R	1.5	1.6	1.5	2.8	3.0	3.6	4.0	1.4
V	7	14	7	21	14	14	21	7

- 3.1 Si rappresentino graficamente le coppie di punti e la funzione di regressione.

commento a ogni valore della variabile R corrisponde un unico valore della variabile V ; sussiste, quindi, dipendenza funzionale di V da R



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$\text{I) } V^* = a + b R \quad \text{e} \quad \text{II) } V^* = c.$$

$$a = 2.5907 \quad b = 4.3441 \quad c = 13.125$$

- 3.3 Si indichi quale modello presenta il migliore adattamento.

$$\text{misura adattamento I) } \rho^2 = 0.6142 \quad \text{misura adattamento II) } \rho^2 = 0$$

commenti il modello retta presenta ovviamente un adattamento migliore rispetto al modello con la sola costante

- 3.4 Senza effettuare calcoli si indichi l'intervallo dei possibili valori per $\eta_{R|V}^2$, motivando la risposta data.

$$0.6142 = \rho^2 < \eta_{R|V}^2 < 1$$

commenti dal grafico si evince che le medie condizionate $M(R|v=7)$, $M(R|v=14)$ e $M(R|v=21)$ non sono allineate; non sussiste dipendenza funzionale di R da V in quanto al valore $r = 14$ corrispondono i tre valori distinti 1.6, 3 e 3.6 di V

4. Due dadi vengono truccati in modo che non si presentino mai le facce 1 e 2 del primo e la 4 del secondo.

- 4.1 Si calcoli la probabilità dell'evento A = "somma dei valori ≥ 10 ".

$$P(A) = 0.25$$

procedimento applicando la formula di Laplace $\frac{\# \text{ casi favorevoli}}{\# \text{ casi possibili}}$ si ottiene $\frac{5}{20} = 0.25$

- 4.2 Si calcoli la probabilità che su 11 lanci l'evento A si presenti almeno 2 volte.

$$P = 0.8029$$

procedimento $X \sim \text{Bin}(n=11, p=0.25)$;

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X=0) + P(X=1)] = 1 - \binom{11}{0} \cdot 0.25^0 \cdot 0.75^{11} - \binom{11}{1} \cdot 0.25^1 \cdot 0.75^{10} = 1 - 0.0422 - 0.1549$$

- 4.3 Si calcoli la probabilità che su 103 lanci l'evento A si presenti almeno 30 volte.

$$P = 0.1660$$

procedimento $X \sim \text{Bin}(n=103, p=0.25)$; $M(X) = 25.75$; $\text{Var}(X) = 19.3125$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=25.75, \sigma^2=19.3125)$ e

$$P(X \geq 30) = P\left(\frac{X - \mu}{\sigma} \geq \frac{30 - \mu}{\sigma}\right) = P\left(Z \geq \frac{30 - 25.75}{4.3946}\right) = P(Z \geq 0.97) = 1 - P(Z \leq 0.97) = 1 - 0.8340.$$

Prova scritta di STATISTICA del 04.06.2009 (Tema 250)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

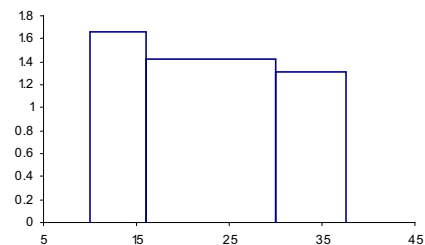
Studente _____ matricola _____

1. È data la seguente seriazione X :

i	$h_{i-1} \rightarrow h_i$	n_i
1	10 $\rightarrow$ 16	10
2	16 $\rightarrow$ 30	20
3	30 $\rightarrow$ h_3	10

1.1 Si calcoli il valore da attribuire ad h_3 , sapendo che la densità di frequenza (assoluta) della terza classe risulta pari a 1.31579 e si dia una rappresentazione grafica della distribuzione X .

$h_3 = 37.6$ commenti ricordando che $d_3 = \frac{n_3}{a_3} \rightarrow a_3 = \frac{n_3}{d_3}$



1.2 Si calcolino media aritmetica, varianza e coefficiente di variazione di X .

$M(X) = 23.2$ $Var(X) = 54.12$ $cv(X) = 0.3171$

1.3 Si indichi, almeno approssimativamente, il valore assunto dalla media potenziata di X di ordine -1000 .

$\mu_X^{(-1000)} \cong 13$

commenti in base al Teorema sulle medie potenziate se l'ordine $r \rightarrow -\infty$ allora $\mu_X^{(r)} \rightarrow x_{\min}$

1.4 Si indichi, senza effettuare calcoli e giustificando la risposta, se la distribuzione di X è simmetrica.

La distribuzione di X non è simmetrica; dall'esame del grafico sembrerebbe esserci asimmetria positiva

2. Si completino le seguenti tabelle in modo che siano soddisfatte, se possibile, le condizioni indicate:

(a)

$X \setminus Y$	y_1	y_2	y_3
x_1	18		
x_2		8	
x_3			4

$\chi^2_{\text{NORM}} = 1$ e $\chi^2_{\text{MAX}} = 90$

(b)

$X \setminus Y$	y_1	y_2	y_3
x_1	18	8	4
x_2	18	8	4
x_3	18	8	4

$\chi^2_{\text{NORM}} = 0$ e $\chi^2_{\text{MAX}} = 180$

(c)

$X \setminus Y$	$y_1 = y_2 - a$	y_2	$y_3 = y_2 + a$
$x_1 = x_2 - c$	18	0	4
x_2	0	8	0
$x_3 = x_2 + c$	18	0	4

$\eta^2_{X|Y} = 0$ e $\eta^2_{Y|X} \neq 0$

(a) È impossibile; ponendo $n_{12} = n_{13} = n_{21} = n_{23} = n_{31} = n_{32} = 0$ si avrebbe $\chi^2_{\text{MAX}} = 60$.

(b) $\chi^2_{\text{NORM}} = 0$ solo se sussiste indipendenza stocastica. Se $\chi^2_{\text{MAX}} = 180$ segue che $n = 90$; si devono costruire delle distribuzioni condizionate somiglianti (nel caso in esame risultano identiche).

(c) $\eta^2_{X|Y} = 0$ solo se sussiste indipendenza in media di X da Y ; i.e. $M(X|y_1) = M(X|y_2) = M(X|y_3) = M(X)$;

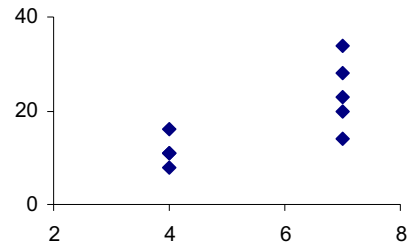
$\eta^2_{Y|X} \neq 0$ sicuramente non sussiste indipendenza in media di Y da X ; quindi almeno due delle $M(Y|x_i)$, $i = 1, 2, 3$ devono essere tra loro diverse.

Nella situazione proposta $M(X|y_1) = M(X|y_2) = M(X|y_3) = M(X) = x_2$; e $M(Y|x_1) = M(Y|x_3) < M(Y|x_2) = y_2$.

3. Su un campione di 9 unità statistiche sono state rilevate le seguenti coppie di osservazioni:

x_i	4	7	7	7	4	7	4	7	4
y_i	8	14	28	20	11	34	11	23	16

3.1 Si rappresentino graficamente le coppie di punti



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

I) $Y^* = a + bX$ e II) $Y^* = c$.

$a = -4.9$

$b = 4.1$

$c = 18.3333$

3.3 Si indichi quale modello presenta il migliore adattamento.

misura adattamento I) $\rho^2 = 0.5585$

misura adattamento II) $\rho^2 = 0$

commenti Il modello retta presenta sicuramente un adattamento migliore rispetto al modello con la sola costante.

3.4 Si confronti l'adattamento del modello migliore con quello della funzione di regressione.

$\rho^2 = \eta^2_{Y|X} = 0.5585$.

Essendo presenti due soli valori della X si hanno due distribuzioni condizionate $Y|x_i$ e due sole medie condizionate $M(Y|x_i)$; quindi la funzione di regressione coincide con il modello retta.

4. Con riferimento alla seguente tabella riportante le valutazioni di $n=90$ consumatori su soddisfazione (SAT) e fedeltà (LOY) riguardo ad un determinato servizio:

LOY	SAT	<i>bassa</i>	<i>alta</i>
<i>bassa</i>		7	14
<i>alta</i>		16	53

si valuti, con un opportuno indice di connessione, quanto la fedeltà dipende dalla soddisfazione.

$\chi_{\text{NORM}} = 0.0985$

con riferimento alle 90 unità statistiche esaminate si può concludere che la fedeltà non dipende dalla soddisfazione

5. Un gioco consiste nel lanciare una moneta e un dado; il giocatore vince se ottiene *testa* (evento T) e un punteggio del dado maggiore di 4 (evento D).

5.1 Si calcoli $P(D|T)$.

T	•	•	•	•	•	•
C	•	•	•	•	•	•
	1	2	3	4	5	6

$P(D|T) = 0.3333$

$P(D|T) = \frac{P(D \cap T)}{P(T)} = \frac{2/12}{6/12} = \frac{2}{6}$.

5.2 Si calcoli la probabilità di vincita.

$P = P(D \cap T) = \frac{2}{12} = 0.1667$.

5.3 Si calcoli la probabilità che su 120 tentativi un giocatore vinca almeno 22 volte.

$P = 0.3121$

procedimento $X \sim \text{Bin}(n=120, p=0.1667)$; $M(X) = 20$; $\text{Var}(X) = 16.6667$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=20, \sigma^2=16.6667)$ e

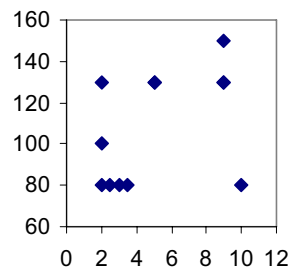
$P(X \geq 22) = P\left(\frac{X - \mu}{\sigma} \geq \frac{22 - \mu}{\sigma}\right) = P\left(Z \geq \frac{22 - 20}{4.0825}\right) = P(Z \geq 0.49) = 1 - P(Z \leq 0.49) = 1 - 0.6879 = 0.3121$.

c) le medie condizionate devono essere diverse tra loro e non allineate; occorre assicurarsi che $M(Y|X=1) \neq 2$.

3. Volendo studiare Y in funzione di Z :

Y	80	130	100	80	80	150	130	80	130	80	130	130
Z	3	9	2	2	2.5	9	2	10	5	3.5	9	5

3.1 Si rappresentino graficamente le coppie dei dati;



3.2 si dica, senza effettuare calcoli, giustificando teoricamente la risposta, se $\eta^2_{Y|Z} = 1$ oppure $\eta^2_{Y|Z} = 0$:

$0 < \eta^2_{Y|Z} < 1$: le medie condizionate sono sicuramente diverse tra loro e non sussiste dipendenza funzionale di Y da Z .

3.3 si calcolino con il metodo dei minimi quadrati i parametri dei modelli di regressione:

$$\text{I) } Y^* = a + bZ \quad \text{e} \quad \text{II) } Y^* = cZ^2.$$

$$a = 89.2942 \quad b = 3.685 \quad c = 1.6367$$

3.4 Si indichi quale modello presenta il migliore adattamento.

misura adattamento I) RMS = 553.628

misura adattamento II) RMS = 5440.06

commenti: Il modello retta completa presenta un miglior adattamento rispetto al modello parabolico vincolato

4. Date due differenti urne, U_1 e U_2 , contenenti palline numerate da 1 a 5, un esperimento casuale consiste nell'estrarre una pallina da ciascuna urna. Sapendo che nell'urna U_1 vi è una pallina per ogni numero e che la composizione dell'urna U_2 è la seguente: $U_2 = (1, 1, 2, 3, 3, 3, 4, 5, 5, 5)$

4.1 Si costruisca la distribuzione di probabilità delle variabili X = "risultato estrazione urna U_1 " e Y = "risultato estrazione urna U_2 ".

x	$P(X=x)$
1	0.2
2	0.2
3	0.2
4	0.2
5	0.2

y	$P(Y=y)$
1	0.2
2	0.1
3	0.3
4	0.1
5	0.3

Ω	5	•	•	•	•	•	•	•	•	•	•
4	•	•	•	•	•	•	•	•	•	•	•
3	•	•	•	•	•	•	•	•	•	•	•
2	•	•	•	•	•	•	•	•	•	•	•
1	•	•	•	•	•	•	•	•	•	•	•
		1	1	2	3	3	3	4	5	5	5

4.2 Si calcoli la probabilità che la somma dei punteggi delle 2 palline estratte sia pari a 3.

$$P = 0.06$$

$$P[(X=1 \cap Y=2) \cup (X=2 \cap Y=1)] = P(X=1 \cap Y=2) + P(X=2 \cap Y=1) = P(X=1) \cdot P(Y=2) + P(X=2) \cdot P(Y=1) = 0.2 \cdot 0.1 + 0.2 \cdot 0.2$$

$$\text{oppure } P(X=1 \cap Y=2) + P(X=2 \cap Y=1) = \frac{1}{50} + \frac{2}{50}.$$

4.3 Indicati con D_1 e D_2 gli eventi "estrazione di numero dispari" rispettivamente da U_1 e U_2 , si calcolino $P(D_1)$, $P(D_2|D_1)$ e $P(D_1 \cap D_2)$.

$$P(D_1) = 0.2 + 0.2 + 0.2 = 0.6 \quad P(D_2|D_1) = 0.8$$

$$P(D_1 \cap D_2) = 0.48$$

$$P(D_2|D_1) = \frac{P(D_1 \cap D_2)}{P(D_1)} = \frac{P(D_1) \cdot P(D_2)}{P(D_1)} = P(D_2) = 0.8 \quad \text{oppure} \quad P(D_2|D_1) = \frac{P(D_1 \cap D_2)}{P(D_1)} = \frac{\frac{24}{50}}{\frac{30}{50}} = 0.8;$$

$$P(D_1 \cap D_2) = P(D_1) \cdot P(D_2) = 0.6 \cdot 0.8 = 0.48 \quad \text{oppure} \quad P(D_1 \cap D_2) = \frac{24}{50}.$$

4.4 Considerando ora solo l'urna U_2 , si calcoli la probabilità che, estraendo 100 palline con reinserimento, almeno 21 siano col numero 3.

$$P = 0.975$$

procedimento $X \sim \text{Bin}(n=100, p=0.3)$; $M(X) = 30$; $\text{Var}(X) = 21$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=30, \sigma^2=21)$ e

$$P(X \geq 21) = P\left(\frac{X - \mu}{\sigma} \geq \frac{21 - \mu}{\sigma}\right) = P\left(Z \geq \frac{21 - 30}{4.5826}\right) = P(Z \geq -1.96) = P(Z \leq 1.96) = 0.975.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 02.07.2009 (Tema 252)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

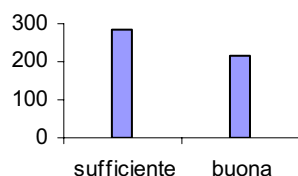
Studente _____ matricola _____

1. Nella seguente tabella è riportata la distribuzione congiunta relativa all'età (Y) e alla conoscenza della matematica (X) di 500 ragazzi:

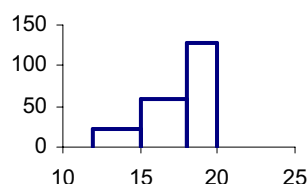
$X \quad Y$	12-15	15-18	18-20
<i>sufficiente</i>	61	128	94
<i>buona</i>	6	51	160

- 1.1 Si dia una rappresentazione grafica delle distribuzioni marginali X e Y .

X	$n_{i\bullet}$	$N_{i\bullet}$
$x_1 = \textit{sufficiente}$	283	283
$x_2 = \textit{buona}$	217	500
	500	



Y	$n_{\bullet j}$	$a_{\bullet j}$	$d_{\bullet j}$
12-15	67	3	22.33
15-18	179	3	59.67
18-20	254	2	127
	500		



- 1.2 Si descriva con opportuni indici sintetici di posizione e di variabilità la distribuzione marginale X .

indice di posizione: moda e/o mediana

valore *sufficiente*

indice di variabilità: Indice di Frosini

valore 0.8696

In base al solo indice di posizione sembrerebbe essere presente un livello sufficiente di conoscenza; l'indice di eterogeneità conferma la presenza di una elevata mutabilità (il 43.4% dei ragazzi ha infatti una buona conoscenza).

- 1.3 Si valuti la connessione esistente tra le due variabili mediante l'impiego di un opportuno indice e si commenti il risultato ottenuto.

indice di connessione χ_{NORM}

valore 0.4201

Il livello di connessione tra le due variabili è medio/basso

2. Un mini-market decide di effettuare un'indagine su 10 clienti, rilevando il titolo di studio X , la spesa media mensile Y per prodotti biologici (in Euro) e l'età Z (in anni compiuti). Nella seguente tabella sono riportate le medie e le varianze delle distribuzioni condizionate $Y|x_i$ e la distribuzione marginale X .

x_i	$n(x_i)$	$M(Y x_i)$	$\text{Var}(Y x_i)$
media inferiore	2	10	0
diploma	5	18	50
laurea	3	20	76.6667

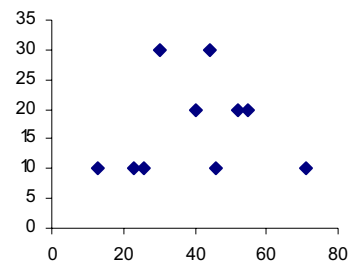
- 2.1 Sapendo che la media aritmetica e la varianza del carattere Y sono pari a $M(Y) = 17$ e $\text{Var}(Y) = 61$ si stabilisca, giustificando la risposta, quanto la spesa mensile Y dipende in media dal titolo di studio X .

$\eta^2_{Y|X} = 0.2131$ Esiste, quindi, un basso livello di dipendenza della spesa media mensile per prodotti biologici rispetto al titolo di studio.

Ricordando che $M(Y) = 17$, $\text{Var}(Y) = 61$ e sapendo che per i caratteri Y e Z sono state rilevate le seguenti osservazioni:

y_i	10	30	10	20	10	30	10	20	20	10
z_i	71	30	26	55	13	44	46	40	52	23

2.2 Si rappresentino graficamente le coppie dei dati;



2.3 si calcolino con il metodo dei minimi quadrati i parametri dei modelli di regressione:

$$\text{I) } Y^* = b Z \quad \text{e} \quad \text{II) } Y^* = c + d \cdot \ln(Z).$$

$$b = 0.3717$$

$$c = 2.7346$$

$$d = 3.9756$$

2.4 Si indichi quale modello presenta il migliore adattamento.

$$\text{misura adattamento I) } \text{RMS} = 91.6426$$

$$\text{misura adattamento II) } \text{RMS} = 57.4207$$

Il secondo modello presenta un adattamento migliore rispetto al primo.

2.5 Sulla base del modello migliore si preveda la spesa media mensile di un cliente di 40 anni.

$$Y^*(40)_{\text{previsione}} = 2.7346 + 3.9756 \ln(40) = 17.40$$

3. Il titolare di un'azienda conduce un'indagine sui suoi 100 dipendenti. Indicati con F l'evento "il soggetto è fumatore" e con S l'evento "il soggetto consuma abitualmente *snack*", si ha la seguente tabella a doppia entrata:

	F	$\bar{F}$
S	26	40
$\bar{S}$	25	9

3.1 Si calcoli la probabilità che, estraendo a caso un dipendente, questo: A) sia un fumatore; B) non fumi e non consumi abitualmente *snack*; C) fumi sapendo che non consuma abitualmente *snack*.

$$P(F) = \frac{51}{100} = 0.51$$

$$P(\bar{F} \cap \bar{S}) = \frac{9}{100} = 0.09$$

$$P(F|\bar{S}) = \frac{P(F \cap \bar{S})}{P(\bar{S})} = \frac{\frac{25}{100}}{\frac{34}{100}} = 0.7353$$

3.2 Calcolare infine la probabilità che, considerati i 10 dipendenti più anziani, ci siano 5 fumatori e 5 no.

$$P = 0.2588$$

Il problema corrisponde al calcolo della probabilità di ottenere 5 elementi di un certo tipo estraendo senza reimmissione 10 elementi da un'urna contenente due tipi di elementi.

Può essere, quindi, indifferentemente risolto come $\binom{10}{5} \cdot P(F_1 \cap F_2 \cap \dots \cap F_5 \cap \bar{F}_6 \cap \bar{F}_7 \cap \dots \cap \bar{F}_{10}) =$

$$= \binom{10}{5} \cdot P(F_1) \cdot P(F_2|F_1) \cdot \dots \cdot P(F_5|F_1 \cap \dots \cap F_4) \cdot P(\bar{F}_6|F_1 \cap \dots \cap F_5) \cdot \dots \cdot P(\bar{F}_{10}|F_1 \cap \dots \cap F_5 \cap \bar{F}_6 \cap \dots \cap \bar{F}_9) =$$

$$= \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \cdot \frac{51}{100} \cdot \frac{50}{99} \cdot \frac{49}{98} \cdot \frac{48}{97} \cdot \frac{47}{96} \cdot \frac{49}{95} \cdot \frac{48}{94} \cdot \frac{47}{93} \cdot \frac{46}{92} \cdot \frac{45}{91} \text{ oppure mediante la variabile casuale ipergeometrica } P(X=5) = \frac{\binom{51}{5} \cdot \binom{49}{5}}{\binom{100}{10}}$$

4. Con riferimento alla seguente serie storica si costruisca la serie dei numeri indici a base mobile e si calcoli ${}_2I_4$.

t	X_t	$NIBM$
1	2.8	--
2	3.2	1.1429
3	2.8	0.8750
4	2.3	0.8214
5	2.9	1.2609

${}_2I_4 = \frac{2.3}{3.2} = 0.8750 \cdot 0.8214 = 0.7188$; quindi tra l'istante temporale 2 e l'istante temporale 4 si è verificata una variazione percentuale del -28.12%

Studente _____ matricola _____

1. In un'indagine sull'inquinamento atmosferico nella città di Sondrio sono state effettuate 10 misurazioni del livello Y di PM10, espresso in $\mu\text{g}/\text{m}^3$ e della temperatura X in $^\circ\text{C}$. I valori osservati sono riportati nella seguente tabella.

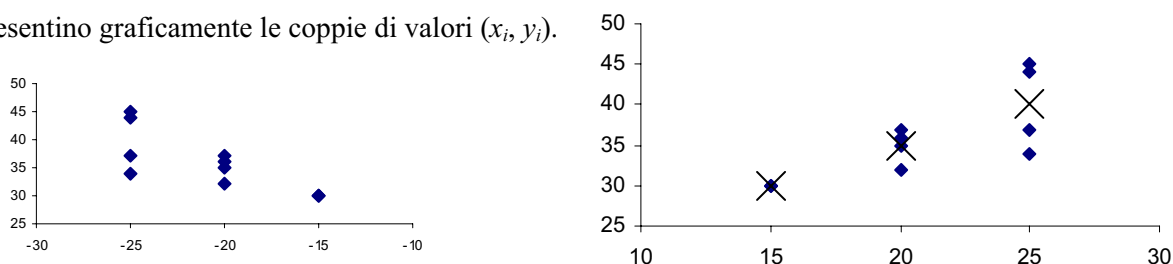
x_i	15	20	25	20	20	15	25	25	25	20
y_i	30	35	44	36	37	30	37	34	45	32

- 1.1 Si valuti la correlazione lineare fra le variabili X e Y mediante un opportuno indice, commentando opportunamente il risultato ottenuto. Si dica come cambia il valore dell'indice se le modalità della variabile X vengono moltiplicate per -1 .

$\rho_{X,Y} = 0.7638$ commenti: esiste correlazione positiva tra X e Y (tra le variabili esiste una relazione diretta)

$\rho_{-X,Y} = -0.7638$ commenti: $-X$ e Y sono legate da una relazione inversa (si veda la figura a sinistra)

- 1.2 Si rappresentino graficamente le coppie di valori (x_i, y_i) .



- 1.3 Si stimino, attraverso il criterio dei minimi quadrati, i parametri dei modelli di regressione:

$$\text{I) } Y^* = a + bX \quad \text{e} \quad \text{II) } Y^* = cX.$$

$$\hat{a} = 15$$

$$\hat{b} = 1$$

$$\hat{c} = 1.6923$$

- 1.4 Si indichi quale modello presenta il migliore adattamento.

misura adattamento I) RMS = 10

misura adattamento II) RMS = 16.923

Il primo modello presenta un adattamento migliore rispetto al secondo.

- 1.5 Dopo aver rappresentato graficamente la funzione di regressione di Y rispetto a X , si dica, senza effettuare calcoli e motivando teoricamente la risposta, che valore assume il rapporto di correlazione.

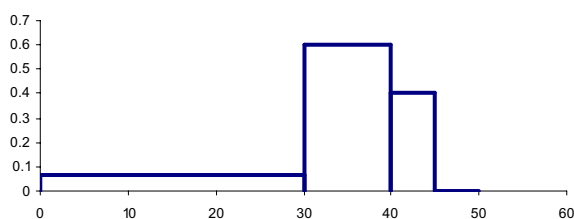
$\eta^2_{Y|X} = 0.5833$ commenti: la funzione di regressione è identificata dalle medie condizionate $M(Y|x=15) = 30$, $M(Y|x=20) = 35$ e $M(Y|x=25) = 40$, che sono allineate; quindi $\eta^2_{Y|X} = \rho^2$.

- 1.6 Sulla base del modello migliore si preveda il livello del PM10 in corrispondenza di una temperatura di 23°C .

previsione $Y^*(23) = \hat{a} + \hat{b} \cdot 23 = 15 + 1 \cdot 23 = 38$

- 1.7 Dopo aver raccolto le modalità della variabile Y nelle seguenti classi chiuse a destra $0 \rightarrow 30$; $30 \rightarrow 40$; $40 \rightarrow 45$; $45 \rightarrow 50$, si rappresenti graficamente la distribuzione di frequenza della variabile Y .

Y	n_j	a_j	d_j
$0 \rightarrow 30$	2	30	0.0667
$30 \rightarrow 40$	6	10	0.6
$40 \rightarrow 45$	2	5	0.4
$45 \rightarrow 50$	0	5	0
	10		



- 1.8 Si calcolino moda, media aritmetica e coefficiente di variazione di Y riclassificata.

Moda = classe $30 \rightarrow 40$ Media aritmetica = 32.5 Coefficiente di variazione = 0.2837

Dal grafico risulta evidente un'asimmetria negativa della distribuzione di frequenza di Y .

2. Si consideri la seguente tabella delle frequenze congiunte del carattere quantitativo X e del carattere qualitativo ordinato Y

$X \backslash Y$	y_1	y_2	y_3	
1	7	n_{12}	n_{13}	$n_{1\bullet}$
x_2	n_{21}	n_{22}	n_{23}	$n_{2\bullet}$
	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet 3}$	34

- 2.1 Si completi la tabella (modalità e frequenze) in modo tale che $\chi^2 = \chi^2_{\text{MAX}}$ e che per la variabile Y si verifichi la condizione $n_{\bullet 1} = n_{\bullet 3}$.

$X \backslash Y$	y_1	y_2	y_3	
1	7	20	0	27
$x_2 > 1$	0	0	7	7
	7	20	7	34

$\chi^2 = \chi^2_{\text{MAX}}$ solo se è presente una situazione di dipendenza funzionale, che nel caso in esame può verificarsi solo considerando X quale variabile dipendente. È sufficiente imporre che le distribuzioni condizionate $X|y_j$ siano degeneri.

- 2.2 Si completi la tabella (modalità e frequenze) in modo che contemporaneamente ci sia indipendenza in media di X da Y e che la moda di Y sia y_2 .

$X \backslash Y$	y_1	y_2	y_3	
1	7	9	1	
$x_2 > 1$	7	9	1	
	14	18	2	34

se $n_{\bullet 2} > n_{\bullet 1}$ e $n_{\bullet 2} > n_{\bullet 3}$ la moda di Y risulta y_2 . Affinché sussista indipendenza in media di X da Y le medie condizionate $M(X|y_j)$, $j = 1, 2, 3$, devono essere tutte uguali. Nel caso in esame, dal momento che X ha due sole modalità, tale condizione risulta verificata solo se le distribuzioni $X|y_j$ sono somiglianti.

3. Un test è costituito da 56 domande le cui 3 possibili risposte sono codificate con A, B, C. Se la risposta è corretta il punteggio assegnato è +1 mentre se la risposta è errata il punteggio è -0.25; se invece la risposta non viene data il punteggio è 0. Per superare il test bisogna raggiungere il punteggio minimo di 10.

- 3.1 Dire, motivando la risposta, se risulta più conveniente tentare tutte le risposte a caso o non rispondere alle domande.

La risposta a una domanda non influenza la risposta alle altre; quindi le domande possono considerarsi indipendenti. Risulta conveniente tentare di rispondere alla singola domanda (rischiare, quindi, una posta di -0.25) solo se il corrispondente “gioco” non è equo, ma è favorevole a chi deve, eventualmente, rispondere. Si deve calcolare il valore atteso del punteggio nelle due ipotesi di non rispondere alla singola domanda, oppure di rispondere.

Se non si dà risposta si consegue 0 con probabilità 1, quindi il valore atteso del punteggio è 0.

Rispondendo, il punteggio conseguibile è descritto dalla seguente variabile casuale, caratterizzata da $M(X) = 0.16 > 0$.

x	$P(X=x)$
-0.25	2/3
+1	1/3
	1

“Il gioco non è equo” ma favorevole; conviene, quindi, parteciparvi.

- 3.2 Supponendo ora di aver deciso di rispondere a caso a tutte le 56 domande:

- si calcoli la probabilità di rispondere correttamente ad almeno 20 domande.

$$P = 0.352$$

Indicata con X la variabile casuale “numero di risposte corrette” alle 56 domande risulta $X \sim \text{Bin}(n=56, p=1/3)$.

$M(X) = 18.6667$; $\text{Var}(X) = 12.4444$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=18.6667, \sigma^2=12.4444)$ e

$$P(X \geq 20) = P\left(\frac{X - \mu}{\sigma} \geq \frac{20 - \mu}{\sigma}\right) = P\left(Z \geq \frac{20 - 18.6667}{3.5277}\right) = P(Z \geq 0.38) = 1 - P(Z \leq 0.38) = 0.352.$$

- si calcoli la probabilità di rispondere correttamente ad almeno 2 domande tra le prime 6.

$$P = 0.6488$$

Indicata con X la variabile casuale “numero di risposte corrette” alle prime 6 domande, risulta $X \sim \text{Bin}(n=6, p=1/3)$.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X=0) - P(X=1) = 1 - \binom{6}{0} \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^6 - \binom{6}{1} \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^5 = 1 - 0.0878 - 0.2634 = 0.6488.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 17.09.2009 (Tema 254)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. E' stata svolta un'indagine su 100 studenti delle scuole superiori rilevando la quantità di tempo che dedicano allo studio, Y , e se svolgono o meno attività extrascolastiche, X . Si riportano le densità di frequenza delle distribuzioni condizionate $Y|x_1$ ="Svolgono attività" e $Y|x_2$ ="Non svolgono attività":

$Y x_1$	densità
0 → 2	5
2 → 4	19
4 → 5	9

$Y x_2$	densità
0 → 2	4
2 → 4	5
4 → 5	25

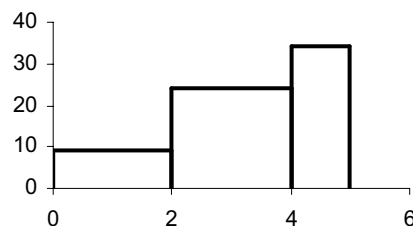
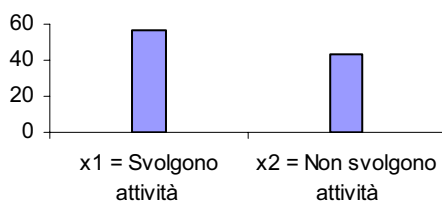
- 1.1 Si ricostruisca la tabella a doppia entrata delle frequenze congiunte.

$X \backslash Y$	0 → 2	2 → 4	4 → 5	
x_1 = Svolgono attività	10	38	9	57
x_2 = Non svolgono attività	8	10	25	43
	18	48	34	100

- 1.2 Si rappresentino graficamente le distribuzioni marginali.

x_i	$n_{i\bullet}$
x_1 = Svolgono attività	57
x_2 = Non svolgono attività	43

Y	$n_{\bullet j}$	$d_{\bullet j}$
0 → 2	18	9
2 → 4	48	24
4 → 5	34	34



- 1.3 Si calcolino la moda e la media aritmetica di Y .

Moda = classe 4 → 5 Media aritmetica = 3.15

- 1.4 Si confronti la variabilità delle due distribuzioni condizionate $Y|x_i$.

misura variabilità c.v. valore $Y|x_1 = 0.3540$ valore $Y|x_2 = 0.3836$

commenti la seconda distribuzione condizionata risulta leggermente più variabile della prima.

2. Dati due caratteri quantitativi X ed Y rilevati congiuntamente indicare, giustificando la risposta data, che valori assume la covarianza nelle seguenti ipotesi:

2.1 $\rho_{XY} = 0$ commenti: $\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$ quindi $\rho_{XY} = 0$ se e solo se $\text{Cov}(X,Y) = 0$

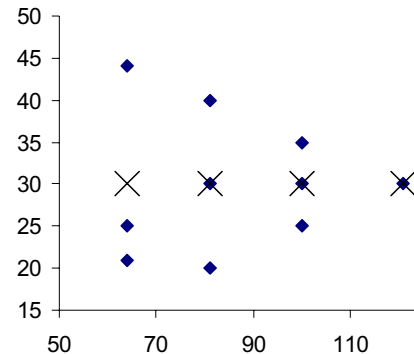
2.2 $\chi^2 = 0$ commenti: $\chi^2 = 0$ solo se sussiste indipendenza stocastica, che implica la non correlazione, quindi anche $\text{Cov}(X,Y) = 0$.

2.3 $\rho_{XY} = -1$, $\sigma_X = 3$ e $\sigma_Y = 6$ commenti: $\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \rightarrow \text{Cov}(X,Y) = \rho \cdot \sigma_X \cdot \sigma_Y = -1 \cdot 3 \cdot 6 = -18$

3. Di un campione di 10 alberghi, della medesima categoria, sono riportati il prezzo medio per notte (X) e quello di un pasto nel ristorante interno (Y):

x_i	64	81	100	81	64	121	100	64	81	100
y_i	44	20	25	40	21	30	30	25	30	35

3.1 Si rappresentino graficamente le coppie di valori (x_i, y_i) .



3.2 Si stimino, attraverso il criterio dei minimi quadrati, i parametri dei modelli di regressione:

$$\text{I) } Y^* = a + bX \quad \text{e} \quad \text{II) } Y^* = c + d\sqrt{X}.$$

$$a = 30$$

$$b = 0$$

$$c = 30$$

$$d = 0$$

3.3 Si indichi quale modello presenta il migliore adattamento.

$$\text{misura adattamento I) } \rho^2 = 0$$

$$\text{misura adattamento II) } \rho^2 = 0$$

I due modelli non sono di alcuna utilità al fine di interpretare il comportamento in media di Y in funzione di X .

3.4 Dopo aver rappresentato graficamente la funzione di regressione di Y rispetto a X , si determini, giustificando la risposta e senza effettuare calcoli, il valore di $\eta^2_{Y|X}$.

$\eta^2_{Y|X} = 0$; $M(Y|x=64) = M(Y|x=81) = M(Y|x=100) = M(Y|x=121) = M(Y) = 30$: sussiste indipendenza in media di Y da X .

4. Esaminando i risultati raccolti attraverso un'analisi condotta sui lavoratori di una impresa milanese emerge che abitualmente il 36% degli intervistati utilizza l'auto, il 24% utilizza il treno, il 10% la metropolitana ed il restante autobus o tram. Calcolare le seguenti probabilità:

4.1 scegliendo a caso un intervistato, che questo usi l'auto;

$$P = 0.36$$

procedimento: coincide con la frazione di casi favorevoli all'interno dell'insieme dei lavoratori dell'azienda.

4.2 scegliendo a caso quattro diversi intervistati, che due utilizzino il treno;

$$P = 0.1996$$

Indicata con X la variabile casuale "numero di intervistati che utilizzano il treno" risulta (supponendo che la dimensione dell'azienda, in termini di addetti, sia sufficientemente grande) $X \sim \text{Bin}(n=4, p=0.24)$.

$$P(X=2) = \binom{4}{2} 0.24^2 \cdot 0.76^2 = 0.1996.$$

4.3 scegliendo a caso tre diversi intervistati, che almeno uno si rechi al lavoro con autobus o tram.

$$P = 0.657$$

Indicata con X la variabile casuale "numero di intervistati che utilizzano autobus o tram" risulta (supponendo sempre che la dimensione dell'azienda, in termini di addetti, sia sufficientemente grande) $X \sim \text{Bin}(n=3, p=0.3)$, dove $p=1-0.36-0.24-0.10=0.3$.

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \binom{3}{0} 0.3^0 \cdot 0.7^3 = 1 - 0.343 = 0.657.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 14.1.2010 (Tema 255)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Con riferimento a 80 soggetti sono stati rilevati il livello di conoscenza artistica, X , (con modalità “esperto” e “visitatore occasionale”) e il tempo in minuti, Y , dedicato alla visita di una mostra:

X	
x_i	n_i
esperto	30
visitatore occasionale	50

Y	
$h_{i-1} \rightarrow h_i$	n_i
5 → 10	50
10 → 20	14
20 → 25	16

- 1.1 si rappresentino graficamente X e Y e li si riassume con opportuni indici di posizione e di variabilità;

Grafico X

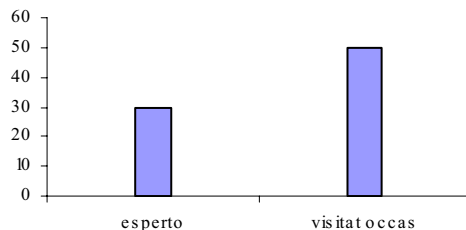
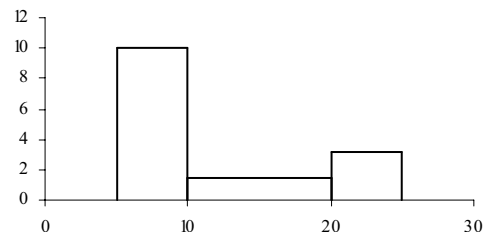


Grafico Y



X : Indice di posizione Moda o mediana valore vis.occ. Indice di variabilità Indice di Frosini valore 0.75
 Y : Indice di posizione Media valore 11.8125 Indice di variabilità c.v. valore 0.5097

Pur essendo la maggior parte delle unità statistiche visitatori occasionali, la distribuzione del livello di conoscenza artistica presenta un livello medio/alto di eterogeneità. Gran parte dei visitatori ha dedicato un tempo ridotto alla visita.

- 1.2 si costruisca la distribuzione congiunta di X e Y nell'ipotesi che tra i due caratteri sussista dipendenza funzionale almeno in un verso (“ X in funzione di Y ” oppure “ Y in funzione di X ”);

$X \backslash Y$	5-10	10-20	20-25	
esperto	0	14	16	30
visitatore occasionale	50	0	0	50
	50	14	16	80

- 1.3 si costruisca la distribuzione congiunta di X e Y nell'ipotesi di indipendenza stocastica;

$X \backslash Y$	5-10	10-20	20-25	
esperto	18.75	5.25	6	30
visitatore occasionale	31.25	8.75	10	50
	50	14	16	80

- 1.4 si riassumano nell'ipotesi di indipendenza stocastica le distribuzioni condizionate $Y|x$ = “esperto” e $Y|x$ = “visitatore occasionale” mediante opportuni indici di posizione e di variabilità.

indice di posizione Media valori: $M(Y|x = \text{“esperto”}) = M(Y|x = \text{“vis. occasionale”}) = M(Y) = 11.8125$
 indice di variabilità c.v. valori: $c.v.(Y|x = \text{“esperto”}) = c.v.(Y|x = \text{“vis. occasionale”}) = c.v.(Y) = 0.5097$

Se vale l'ipotesi di indipendenza stocastica le distribuzioni condizionate sono somiglianti alla marginale e, quindi, sono anche caratterizzate dagli stessi momenti

2. Sapendo che per un aggregato macroeconomico il numero indice riferito al tempo 2003 con base 2000 è pari a ${}_{2000}I_{2003} = 1.49$ e che il numero indice riferito al tempo 2003 con base 2002 è pari a ${}_{2002}I_{2003} = 1.21$ si determinino:

- 2.1 la variazione relativa subita dall'aggregato tra il 2000 e il 2002;

Dal momento che ${}_{2000}I_{2003} = {}_{2000}I_{2002} \cdot {}_{2002}I_{2003}$ segue che ${}_{2000}I_{2002} = \frac{{}_{2000}I_{2003}}{{}_{2002}I_{2003}} = \frac{1.49}{1.21} = 1.2314$, corrispondente a una variazione relativa in termini percentuali del 23.14%.

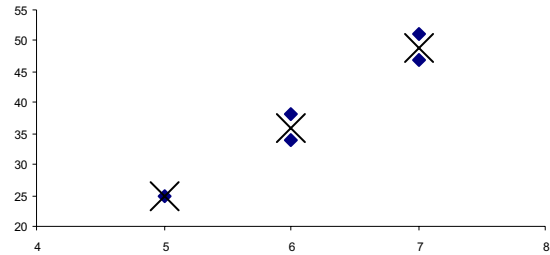
- 2.2 il tasso medio annuo di variazione tra il 2000 e il 2003.

Se ${}_{2000}I_{2003} = 1.49$ il valore relativo medio riferito al triennio considerato risulta $\sqrt[3]{1.49} = 1.1422$, da cui segue un tasso medio di variazione del 14.22%.

3. Il prospetto riporta i valori rilevati su 5 unità statistiche con riferimento a due caratteri quantitativi X e Y :

X	5	6	6	7	7
Y	25	34	38	47	51

3.1 Si rappresentino graficamente le coppie di dati e la funzione di regressione di Y in funzione di X .



N.B. Le medie condizionate risultano:

$$M(Y|x=5) = 25, \quad M(Y|x=6) = 36, \quad M(Y|x=7) = 49.$$

3.2 Calcolare la varianza di Y e determinare, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento del seguente modello di regressione: $Y^* = a + bX^2$.

$$\text{Var}(Y) = 86 \quad a = 0 \quad b = 1 \quad \text{misura adattamento: } \rho_{X^2, Y}^2 = 0.9628$$

3.3 Si indichi, senza effettuare calcoli e giustificando teoricamente la risposta, il valore del rapporto $\eta_{Y|X}^2$.

$$\eta_{Y|X}^2 = \rho_{X^2, Y}^2 = 0.9628$$

Si osservi che: $M(Y|x=5) = 25 = 5^2$, $M(Y|x=6) = 36 = 6^2$ e $M(Y|x=7) = 49 = 7^2$, vale a dire $M(Y|x) = x^2$. Le medie condizionate sono su una parabola e, quindi, il modello studiato coincide con la funzione di regressione.

4. E' stata studiata la distribuzione teorica del tempo, in minuti, necessario per la visita di una mostra.

Si assume che i tempi dei visitatori seguono una distribuzione normale, X , con media 21 e varianza 2.

Si calcoli, nell'ipotesi che i visitatori si comportino in maniera indipendente, la probabilità che:

4.1 la durata della visita del generico visitatore sia superiore a 20 minuti;

$$P = 0.7611$$

$$P(X \geq 20) = P\left(\frac{X - \mu}{\sigma} \geq \frac{20 - \mu}{\sigma}\right) = P\left(Z \geq \frac{20 - 21}{1.4142}\right) = P(Z \geq -0.71) = P(Z \leq 0.71) = 0.7611.$$

4.2 su 3 visitatori la durata della visita di almeno 2 sia superiore a 20 minuti.

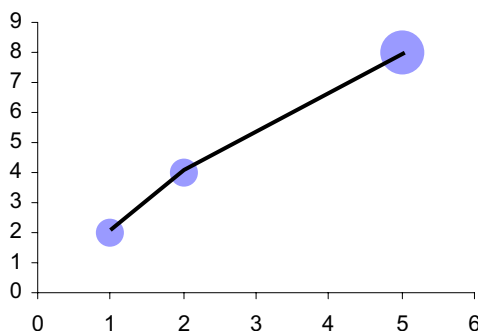
$$P = 0.8561$$

Sia Y la variabile casuale (binomiale) che descrive il numero di visitatori tra i 3 considerati che ha dedicato alla visita un tempo superiore a 20 minuti: $Y \sim \text{Bin}(n=3, p=0.7611)$. La probabilità dell'evento è: $P(Y \geq 2) = P[(Y=2) \cup (Y=3)] = P(Y=2) + P(Y=3) = \{(1 - P(Y=0) - P(Y=1))\} = \binom{3}{2} \cdot 0.7611^2 \cdot 0.2389^1 + \binom{3}{3} \cdot 0.7611^3 \cdot 0.2389^0 = 0.4151 + 0.4410 = 0.8561$.

5. Si completino ove possibile le seguenti tabelle, motivando la risposta, in modo tale che:

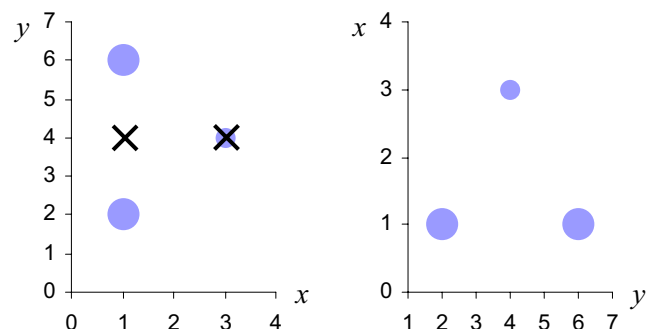
$X \backslash Y$	2	4	8	
1		0	0	
2	0		0	
5	0	0		
				4

$$\rho^2 = \eta_{Y|X}^2 = 1$$



$X \backslash Y$	2	4	6	
1		0		
3	0		0	
				5

$$\eta_{Y|X}^2 = 0 \text{ e contemporaneamente } \eta_{X|Y}^2 > 0$$



E' impossibile verificare entrambe le condizioni con riferimento alla tabella proposta: anche inserendo delle frequenze congiunte diverse da zero si ottengono delle distribuzioni condizionate degeneri, ma le medie condizionate non sono allineate, quindi $\rho^2 < 1$.

Viene richiesto di realizzare una situazione di indipendenza in media di Y da X ; le medie condizionate $M(Y|x=1)$ e $M(Y|x=2)$ sono uguali tra loro (4) se $n_{11} = n_{13} = 2$ e $n_{22} = 1$ oppure se $n_{11} = n_{13} = 1$ e $n_{22} = 3$. In queste situazioni si avrà dipendenza funzionale di X da Y e, quindi, $\eta_{X|Y}^2 = 1 > 0$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 28.1.2010 (Tema 256)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Un'azienda milanese svolge un'indagine su 20 dipendenti di cui: 12 dichiarano di avere un livello di soddisfazione del lavoro basso, 6 dichiarano di avere un livello di soddisfazione del lavoro medio e 2 dichiarano di avere un livello di soddisfazione del lavoro alto. Della variabile Y , numero di annunci di lavoro a cui hanno risposto i dipendenti negli ultimi 3 mesi, sono riportati nella tabella seguente le medie e le varianze dei 3 gruppi:

LIVELLO DI SODDISFAZIONE	μ_i	σ_i^2
basso	39	4
medio	12	25
alto	26	16

- 1.1 Si ricavino la media e lo scarto quadratico medio del numero di annunci per il complesso dei 20 dipendenti.

$$M(Y): \text{procedimento } \mu = \frac{\mu_1 n_1 + \mu_2 n_2 + \mu_3 n_3}{n_1 + n_2 + n_3} \quad \text{valore } 29.6$$

$$\text{Var}(Y): \text{procedimento } \sigma^2 = \sigma_{\text{Between}}^2 + \sigma_{\text{Within}}^2 \quad \text{valore } 158.74 = 147.24 + 11.5 \quad \text{valore } \text{sqm}(Y) = 12.5992$$

- 1.2 Mediante il calcolo di un opportuno indice si valuti quanto il numero di annunci è legato al livello di soddisfazione.

$$\text{Indice utilizzato } \eta_{Y|\text{LIVELLO SODDISFAZIONE}}^2 = 0.9276$$

Esiste un elevato livello di dipendenza del numero di annunci di lavoro a cui i dipendenti hanno risposto negli ultimi 3 mesi dal livello di soddisfazione del lavoro.

2. La rilevazione congiunta su 50 studenti di una scuola media superiore del numero di insufficienze alla fine del I semestre (X) e del giudizio finale attribuito a ciascuno (Y) ha fornito i seguenti risultati:

$X \backslash Y$	promosso	bocciato
0	15	1
1	9	6
2	3	16

- 2.1 Si rappresentino graficamente le distribuzioni delle marginali X e Y .

Grafico X

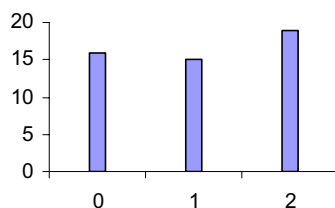
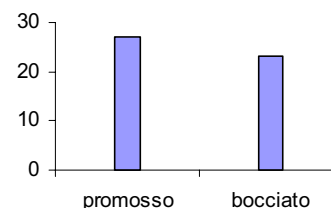


Grafico Y



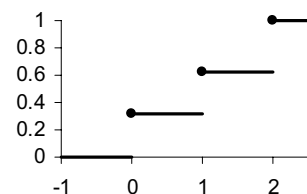
- 2.2 Si descriva con opportuni indici sintetici di posizione e di variabilità la distribuzione marginale Y .

Indice di posizione Moda valore promosso Indice di variabilità Indice di Frosini valore 0.92

La variabile Y presenta un elevato livello di eterogeneità, quindi il valore modale non è rappresentativo.

- 2.3 Si calcoli la mediana di X e si rappresenti graficamente la funzione di ripartizione di X .

Mediana $X = 1$



- 2.4 Si valuti la connessione esistente tra le variabili X e Y mediante l'impiego di un opportuno indice.

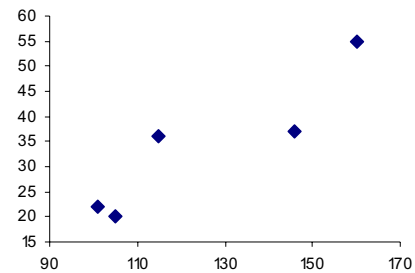
Indice $\chi_N = 0.6567$

Esiste un discreto livello di connessione tra le variabili in gioco.

3. Un pediatra svolge un'indagine su 5 pazienti dai 7 ai 15 anni, rilevando le variabili X , statura in cm, e Y , peso in kg. I risultati sono riportati nella seguente tabella:

X	101	105	115	146	160
Y	22	20	36	37	55

3.1 Si rappresentino graficamente le coppie di dati.



3.2 Si stimino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + bX$$

$$\text{II) } Y^* = cX$$

$$a = -27.5746 \quad b = 0.4910$$

$$c = 0.2785$$

3.3 Si indichi quale modello presenta il migliore adattamento.

$$\text{misura adattamento I) RMS} = 26.6158$$

$$\text{misura adattamento II) RMS} = 52.2319$$

3.3 Si determini, utilizzando il modello ritenuto migliore, il peso teorico di un bambino alto 140 cm.

$$Y^*(140) = -27.5746 + 0.4910 \cdot 140 = 41.17$$

4. Il titolare di un ristorante ha condotto un'indagine su 50 clienti osservando che 29 di essi amano i dolci mentre 11 sono vegetariani. Sapendo che tra i vegetariani, 8 amano i dolci:

4.1 si rappresenti l'insieme degli intervistati attraverso un diagramma di Venn.

	D	$\bar{D}$	
V	8	3	11
$\bar{V}$	21	18	39
	29	21	50

4.2 Si calcoli la probabilità che un cliente non sia vegetariano.

$$P(\bar{V}) = \frac{39}{50} = 0.78$$

4.3 Si calcoli la probabilità che un cliente sia vegetariano oppure non ami i dolci.

$$P(V \cup \bar{D}) = P(V) + P(\bar{D}) - P(V \cap \bar{D}) = \frac{11}{50} + \frac{21}{50} - \frac{3}{50} = \frac{29}{50} = 0.58$$

4.4 Si calcoli la probabilità che un cliente sia vegetariano, sapendo che lo stesso non ama i dolci.

$$P(V|\bar{D}) = \frac{P(V \cap \bar{D})}{P(\bar{D})} = \frac{\frac{3}{50}}{\frac{21}{50}} = \frac{3}{21} = 0.1429$$

4.5 Supponendo di estrarre a caso 3 diversi clienti, si calcoli la probabilità che solo uno di essi sia vegetariano.

$$3 \cdot P(V_1) \cdot P(\bar{V}_2|V_1) \cdot P(\bar{V}_3|V_1 \cap \bar{V}_2) = 3 \cdot \frac{11}{50} \cdot \frac{39}{49} \cdot \frac{38}{48} = \text{oppure} = \frac{\binom{11}{1} \cdot \binom{39}{2}}{\binom{50}{3}} = 0.4159 \quad (\text{le estrazioni sono senza reimmissione}).$$

5. Si indichi, senza effettuare calcoli e motivando le risposte, il valore (o l'intervallo di valori) assunto dagli indici:

$X \backslash Y$	1	2	3
4	9	0	0
6	0	8	6

$\eta_{Y|X}^2$: $0 < \eta_{Y|X}^2 < 1$; la distribuzione condizionata $Y|X=6$ non è degenere $\rightarrow \eta_{Y|X}^2 < 1$; $M(Y|X=4) = 1 < 2 < M(Y|X=6) \rightarrow \eta_{Y|X}^2 > 0$

$\eta_{X|Y}^2 = 1$ le distribuzioni condizionate $X|Y$ sono degeneri, quindi sussiste dipendenza funzionale di X da Y

$\chi_N^2 = 1$ sussiste dipendenza funzionale di X da Y .

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO
Prova scritta di STATISTICA del 11.2.2010 (Tema 257)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1) Si vogliono studiare i caratteri X , con modalità a e b , e Y , con modalità y_1, y_2 e y_3 , disponendo dei dati raccolti nella seguente tabella di distribuzioni condizionate $X|y_j$

	$y_1=1$	$y_2=2$	$y_3=3$
a	1/3	2/3	1/4
b	2/3	1/3	3/4
	1	1	1

- 1.1) Sapendo che la numerosità totale è $n = 120$ e che la marginale Y presenta il 30% di unità statistiche per y_1 e anche per y_2 , ricostruire la tabella delle frequenze congiunte.

$X \backslash Y$	$y_1=1$	$y_2=2$	$y_3=3$	
a	12	24	12	48
b	24	12	36	72
	36	36	48	120

- 1.2) Confrontare la variabilità delle distribuzioni $Y|x=a$ e $Y|x=b$.

Indice di variabilità c.v. valore $Y|x=a$ 0.3536 valore $Y|x=b$ 0.4142

La distribuzione condizionata $Y|x=b$ presenta un livello di variabilità superiore a quello della distribuzione $Y|x=a$.

- 1.3) Calcolare un opportuno indice di connessione tra X e Y .

Indice $\chi_N = 0.3632$

Le variabili X e Y presentano un basso livello di connessione.

- 2) Il consumo pro-capite di gelato in Italia ha presentato, negli ultimi 5 anni, le seguenti variazioni % annuali

2004	2005	2006	2007	2008
+1.2%	+8%	+8%	+2%	-8%

- 2.1) Scelto come base il 2003 (=100) ricostruire la serie dei numeri indici a base fissa.

t	$NIBM$	$NIBF$ (2003=100)
2003	--	100.00
2004	1.012	101.20
2005	1.08	109.30
2006	1.08	118.04
2007	1.02	120.40
2008	0.92	110.77

- 2.2) Sapendo che nel 2003 si è manifestato un consumo pro-capite di 33 kg, indicare il consumo del 2008.

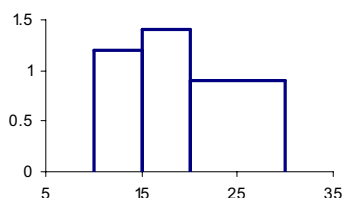
$X_{2008} = 33 \cdot 1.1077 = 36.55$; tra il 2003 e il 2008 si è osservato un incremento complessivo dei consumi del 10.77%.

- 2.3) Indicare l'aumento % complessivo registratosi tra il 2006 e il 2008.

${}_{2006}I_{2008} = {}_{2006}I_{2007} \cdot {}_{2007}I_{2008} = 1.02 \cdot 0.92 = 0.9384$ valore relativo, da cui segue una variazione relativa del -6.16%.

- 3) Data la seguente tabella della v.s. W fornirne la rappresentazione grafica e calcolarne la media geometrica.

10-15	15-20	20-30
6	7	9

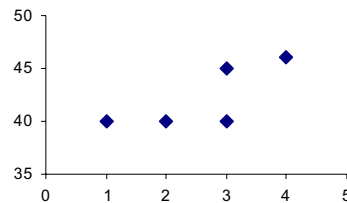


media geometrica $\mu^{(0)} = 18.4737$

4) Date le seguenti 5 coppie di osservazioni

x_k	1	2	3	3	4
y_k	40	40	40	45	46

4.1) Fornire una rappresentazione grafica della nuvola dei punti.



4.2) Confrontare i seguenti 2 modelli (calcolando preventivamente i parametri incogniti)

I) $Y^* = a + bX$

II) $Y^* = 45$

$a = 37$

$b = 2$

misura adattamento: RMS I) 3.2 II) $\frac{1}{5} \cdot [(40-45)^2 + (40-45)^2 + (40-45)^2 + (45-45)^2 + (46-45)^2] = 15.2$

5) Se possibile e giustificando teoricamente le risposte, completare le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

		X		
		3	4	?
Y	4	0	8	?
	7	12	0	?

$\chi_N^2 = 1 \quad \rho = 0$

		X		
		3	4	?
Y	4	0	8	?
	7	12	0	?

$\eta_{Y|X}^2 = \eta_{X|Y}^2 = 1$

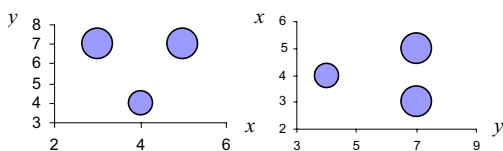
		X		
		3	4	?
Y	4	0	8	?
	7	12	0	?

$\rho^2 = \eta_{Y|X}^2 = \eta_{X|Y}^2 = 0$

		X		
		3	4	5
Y	4	0	8	0
	7	12	0	12

IMPOSSIBILE

IMPOSSIBILE



Per realizzare la dipendenza funzionale di Y da X si dovrebbe inserire una sola frequenza non nulla nella terza colonna, ma in questo modo non si avrebbe dipendenza funzionale di X da Y .

$M(Y|X=3) = 7 \neq M(Y|X=4) = 4$
quindi $\eta_{Y|X}^2 > 0$

dipendenza funzionale di Y da $X \rightarrow \chi_N^2 = 1$;
 $M(X|Y=4) = M(X|Y=7) = 4, \eta_{X|Y}^2 = 0 \rightarrow \rho = 0$

6) Una classe è composta da 11 femmine e 16 maschi, dei quali il 50% ha i capelli neri.

6.1) Calcolare la probabilità che, interrogando a caso 3 differenti alunni, questi siano maschi con i capelli neri.

$P = 0.0191$

Si effettuano 3 estrazioni senza reimmissione dall'urna $\Omega = \{11F, 8M \cap N, 8M \cap \bar{N}\}$; indicato con $A = M \cap N$ si ottiene:

$P = P(A_1 \cap A_2 \cap A_3) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) = \frac{8}{27} \cdot \frac{7}{26} \cdot \frac{6}{25} = 0.0191.$

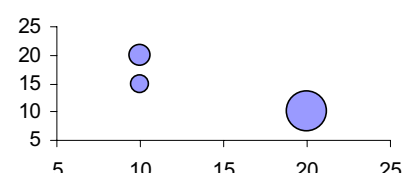
6.2) Calcolare la probabilità che, nel caso sia stato estratto un maschio, questo abbia i capelli neri.

$P(N|M) = \frac{P(N \cap M)}{P(M)} = \frac{\frac{8}{27}}{\frac{16}{27}} = \frac{8}{16} = 0.5$

6.3) Calcolare la probabilità che, estraendo ogni giorno uno studente a caso dall'elenco completo degli iscritti alla classe, solo al 5° giorno venga interrogata una studentessa.

$P = 0.0502$

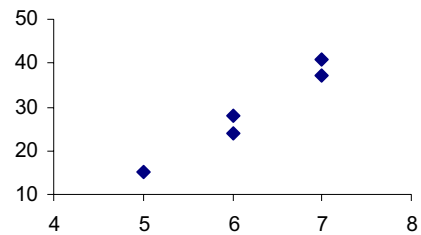
$P(M_1 \cap M_2 \cap M_3 \cap M_4 \cap F_5) = P(M_1) \cdot P(M_2|M_1) \cdot P(M_3|M_1 \cap M_2) \cdot P(M_4|M_1 \cap M_2 \cap M_3) \cdot P(F_5|M_1 \cap M_2 \cap M_3 \cap M_4)$; estrazioni dall'elenco completo quindi con reimmissione: $P = P(M_1) \cdot P(M_2) \cdot P(M_3) \cdot P(M_4) \cdot P(F_5) = [P(M)]^4 \cdot P(F) = \left(\frac{16}{27}\right)^4 \cdot \frac{11}{27} = 0.0502.$



3. Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y :

X	5	6	6	7	7
Y	15	24	28	37	41

- 3.1 Si rappresentino graficamente le coppie di dati.



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$;

I) $Y^* = cX$;

I) $Y^* = d$.

$a = -46.2857$ $b = 12.1429$ misura di adattamento: RMS = 3.4286

$c = 4.7846$ misura di adattamento: RMS = 34.1908

$d = 29$ misura di adattamento: RMS = Var(Y) valore 86

- 3.3 Si indichi, giustificando teoricamente la risposta, l'intervallo di valori che può assumere l'indice di connessione χ^2 .

$\chi^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y .

5. E' stata studiata la distribuzione teorica del tempo, in minuti, necessario per la visita di una mostra. Si assume che i tempi dei visitatori seguano la seguente variabile casuale W .

$w_{i-1} \rightarrow w_i$	$P(w_{i-1} < W \leq w_i)$
5 → 10	0.43
10 → 15	0.31
15 → 20	0.16
20 → 25	0.09
25 → 30	0.01

- 5.1 Si calcoli il valore atteso di W ;

$$M(W) = E(W) = 12.2$$

Si calcoli, nell'ipotesi che i visitatori si comportino in maniera indipendente, la probabilità che:

- 5.2 la durata della visita del generico visitatore sia superiore a 20 minuti;

$$P = 0.10$$

$$P(W > 20) = 1 - P(W \leq 20) = 1 - [P(5 < W \leq 10) + P(10 < W \leq 15) + P(15 < W \leq 20)] = 1 - (0.43 + 0.31 + 0.16) = 1 - 0.90$$

- 5.3 su 3 visitatori, la durata della visita di 2 sia superiore a 20 minuti e di uno sia inferiore a 15 minuti.

$$P = 0.0222$$

Indicati con V e Q gli eventi durata della visita superiore a 20 minuti e inferiore a 15 minuti" si ha:

$$P = P[(V_1 \cap V_2 \cap Q_3) \cup (V_1 \cap Q_2 \cap V_3) \cup (Q_1 \cap V_2 \cap V_3)] = 3 \cdot P(V \cap V \cap Q) = 3 \cdot [P(V)]^2 \cdot P(Q) = 3 \cdot 0.10^2 \cdot 0.74.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 1.7.2010 (Tema 258-2)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1** La tabella seguente raccoglie le distribuzioni delle frequenze relative condizionate, $Y|G$, ottenute con i dati delle prestazioni Y (in metri) di 100 atlete e 150 atleti di salto con l'asta, distinte per genere G (maschi e femmine).

$Y G$	3.0 ÷ 3.5	3.5 ÷ 4.0	4.0 ÷ 4.5	
maschi	0.2	0.5	0.3	1
femmine	0.4	0.5	0.1	1

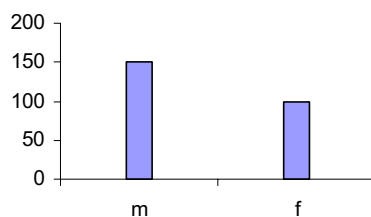
- 1.1** Si ricostruisca la tabella delle frequenze congiunte n_{ij} .

$G \backslash Y$	3.0 ÷ 3.5	3.5 ÷ 4.0	4.0 ÷ 4.5	
maschi	30	75	45	150
femmine	40	50	10	100
	70	125	55	250

- 1.2** Si rappresentino graficamente le distribuzioni marginali.

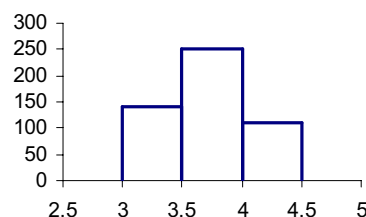
G	
maschi	150
femmine	100
	250

Grafico G



Y	$n_{\bullet j}$	$d_{\bullet j}$
3.0 ÷ 3.5	70	140
3.5 ÷ 4.0	125	250
4.0 ÷ 4.5	55	110
	250	

Grafico Y



- 1.3** Si calcolino opportuni indici di posizione e variabilità di Y e G .

Y : Indice di posizione $M(Y)$ _____ valore 3.72 _____ Indice di variabilità $Var(Y)$ _____ valore 0.1241

G : Indice di posizione moda _____ valore maschi _____ Indice di variabilità Frosini _____ valore 0.8

- 1.4** si calcoli con un opportuno indice il livello di connessione esistente tra le variabili G e Y .

Indice χ_N _____ valore 0.2791.

Le due variabili presentano un livello di connessione basso.

- 1.5** si calcoli con un opportuno indice il livello di dipendenza in media di $Y|G$.

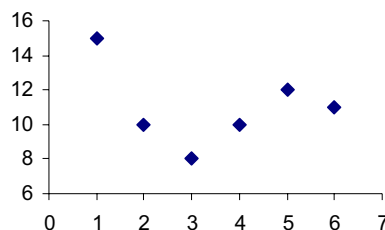
Indice $\eta^2_{Y|G}$ _____ valore 0.0774

Le due variabili presentano un livello di dipendenza in media molto basso

$(M(Y|_{G=m}) = 3.8, M(Y|_{G=f}) = 3.6; \text{ var.sp.} = 0.0096) \quad (Var(Y|_{G=m}) = 0.1225, Var(Y|_{G=f}) = 0.1025; \text{ var.res.} = 0.1145)$

2. Si dia una rappresentazione grafica della seguente serie storica (stock) W_t ($t = 0, 1, \dots, 6$).

t	W_t
1	15
2	10
3	8
4	10
5	12
6	11



2.1 Scelta come base $t = 3$ ($=100$) si ricostruisca la serie dei numeri indici a base fissa.

t	W_t	$NIBF$ ($3=100$)
1	15	187.50
2	10	125.00
3	8	100.00
4	10	125.00
5	12	150.00
6	11	137.50

2.2 Si calcoli il tasso medio di variazione della serie storica tra il tempo $t = 3$ e il tempo $t = 6$.

$$\text{tasso medio di variazione} = 100 \cdot \left(\sqrt[3]{\frac{11}{8}} - 1 \right) \% = 11.199\%.$$

2.3 Si calcolino i parametri della retta di regressione che esprime W_t in funzione di t e si calcoli l'adattamento.

$$a = 12.2 \quad b = -0.3429 \quad \text{misura di adattamento: formula } p^2 \quad \text{valore } 0.0735$$

2.4 Si calcoli il valore teorico futuro W_7 .

$$W_7 = 9.8$$

2.5 Si indichi il valore dell'indice di adattamento della funzione di regressione di $W|t$.

$$\eta_{W|t}^2 = 1$$

Sussiste dipendenza funzionale di W da t

3. In un gioco a premi i concorrenti devono effettuare delle estrazioni, senza re-immissione, da un'urna contenente 13 gettoni ROSSI e 17 BIANCHI. Se la prima estrazione è ROSSA il concorrente ha diritto ad una ulteriore estrazione; se invece la prima è BIANCA esso ha diritto a 2 ulteriori estrazioni. Si vince se, al termine della procedura, sono stati estratti 2 gettoni ROSSI.

3.1 Si calcoli la probabilità di vincere.

$$P = 0.2882$$

$$P = P[(R_1 \cap R_2) \cup (B_1 \cap R_2 \cap R_3)] = P(R_1)P(R_2|R_1) + P(B_1)P(R_2|B_1)P(R_3|B_1 \cap R_2) = \frac{13}{30} \cdot \frac{12}{29} + \frac{17}{30} \cdot \frac{13}{29} \cdot \frac{12}{28}$$

3.2 Nell'ipotesi che il giocatore G1 estragga il primo gettone ROSSO e il giocatore G2 BIANCO, si determini chi dei due ha più probabilità di vincere.

$$P(V|R_1) = \frac{P(V \cap R_1)}{P(R_1)} = \frac{P(R_1)P(R_2|R_1)}{P(R_1)} = P(R_2|R_1) = \frac{12}{29}$$

$$P(V|B_1) = \frac{P(V \cap B_1)}{P(B_1)} = \frac{P(B_1)P(R_2|B_1)P(R_3|B_1 \cap R_2)}{P(B_1)} = P(R_2|B_1)P(R_3|B_1 \cap R_2) = \frac{13}{29} \cdot \frac{12}{28} < \frac{12}{29} \quad \text{quindi G1}$$

3.3 Considerando ora l'estrazione con re-immissione di $n=90$ gettoni, si calcoli la probabilità di ottenere non più di 34 gettoni ROSSI.

$$P = 0.1446$$

Indicata con X la variabile casuale "numero di gettoni ROSSI" nelle 90 estrazioni risulta $X \sim \text{Bin}(n=90, p=0.4333)$.

$M(X) = 39$; $\text{Var}(X) = 22.1$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=39, \sigma^2=22.1)$ e

$$P(X \leq 34) = P\left(\frac{X - \mu}{\sigma} \leq \frac{34 - \mu}{\sigma}\right) = P\left(Z \leq \frac{34 - 39}{\sqrt{22.1}}\right) = P(Z \leq -1.06) = 1 - P(Z \leq 1.06) = 1 - 0.8554.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 15.7.2010 (Tema 258-3)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1** Con riferimento a 80 soggetti si riportano le distribuzioni del tempo in minuti, Y , dedicato alla visita di una mostra, condizionate rispetto al livello di conoscenza artistica, X , (con modalità “esperto” e “visitatore occasionale”):

$Y|X=\text{“esperto”}$

$h_{i-1} \rightarrow h_i$	freq. cond. rel.
5 → 10	0.4
10 → 20	0.4
20 → 25	0.2

$Y|X=\text{“visitatore occasionale”}$

$h_{i-1} \rightarrow h_i$	freq. cond. rel.
5 → 10	0.76
10 → 20	0.08
20 → 25	0.16

- 1.1** si rappresentino graficamente le distribuzioni condizionate $Y|X=\text{“esperto”}$ e $Y|X=\text{“visitatore occasionale”}$ e le si confronti mediante opportuni indici di posizione e di variabilità;

Grafico $Y|X=\text{“esperto”}$

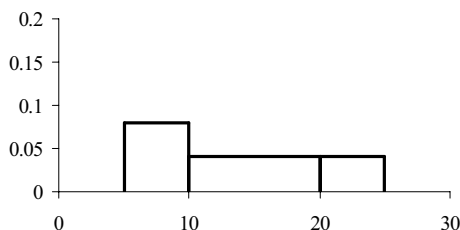
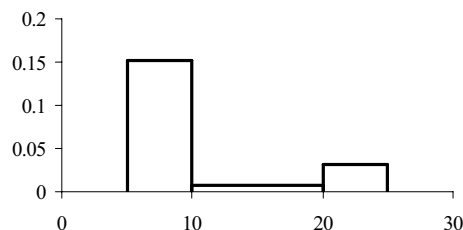


Grafico $Y|X=\text{“visitatore occasionale”}$



$Y|X=\text{“esperto”}$: Indice di posizione $M(Y|X=\text{esp})$ valore 13.5 _____ Indice variabilità $cv(Y|X=\text{esp})$ _ valore 0.4157

$Y|X=\text{“visit. occ.”}$: Indice posizione $M(Y|X=\text{v.o.})$ valore 10.5 _____ Indice variabilità $cv(Y|X=\text{v.o.})$ _ valore 0.5345

I visitatori esperti dedicano in media 3 minuti in più alla visita rispetto ai visitatori occasionali, la cui durata della visita presenta un livello di variabilità superiore.

- 1.2** sapendo che la media di Y , riferita a tutte le 80 unità statistiche, è pari a 11.625 e che il numero di soggetti esperti è pari a 30 si calcoli, applicando il risultato della scomposizione della varianza, il valore della varianza della variabile Y riferita a tutte le 80 unità statistiche;

$$\text{Var}(Y) = 33.6094$$

$$\text{Var}(Y) = \text{Var}_X[M(Y|X)] + M_X[\text{Var}(Y|X)] = \text{var.between} + \text{var.within} = 2.1094 + 31.5$$

- 2.** Con riferimento alla seguente serie dei prezzi di un prodotto tra il 1998 il 2002 si costruiscano le serie dei numeri indici a base fissa (2000=100) e a base mobile e sulla base di quest'ultima si calcoli il tasso medio di variazione del prezzo di tale prodotto tra il 1999 e il 2001.

	p_t	NIBF (2000=100)	NIBM
1998	130	92.86	--
1999	135	96.43	103.85
2000	140	100.00	103.70
2001	145	103.57	103.57
2002	150	107.14	103.45

tasso medio di variazione 3.64% _____

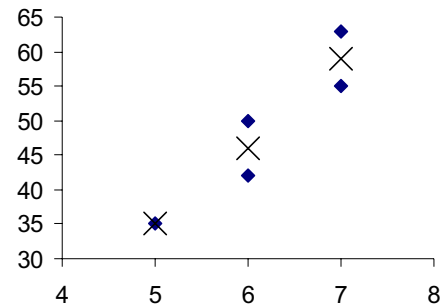
$${}_{1999}I_{2001} = {}_{1999}I_{2000} \cdot {}_{2000}I_{2001} = 1.0370 \cdot 1.0357 = 1.0364 \text{ valore relativo medio, da cui si ricava } 100 \cdot (1.0364 - 1)\%$$

3. Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y :

X	5	6	6	7	7
Y	35	42	50	55	63

- 3.1 Si rappresentino graficamente le coppie di dati e la funzione di regressione di Y in funzione di X .

x	$M(Y x)$
5	35
6	46
7	59



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento del seguente modello di regressione:

$$I) Y^* = b X^2.$$

$b = 1.2432$ _____ misura di adattamento: formula RMS _____ valore 17.9627 _____

- 3.3 Si indichi, giustificando teoricamente la risposta, l'intervallo di valori che può assumere il coefficiente di correlazione lineare ρ .

$0 < \rho < 1$; non sussistono né indipendenza lineare né dipendenza funzionale lineare.

- 3.4 Si indichi, giustificando teoricamente la risposta, l'intervallo di valori che può assumere il rapporto di correlazione $\eta_{X|Y}^2$ (studiando X in funzione di Y).

$\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y

4. Dall'elenco degli 80 visitatori (30 esperti e 50 occasionali) si estraggano 5 nominativi senza reimmissione. Si calcoli la probabilità che:

- 4.1 tra i 5 nominativi estratti ve ne siano solo 2 esperti;

$$P = 0.3547$$

$$\binom{5}{2} \cdot P(E_1) \cdot P(E_2|E_1) \cdot P(\bar{E}_3|E_1 \cap E_2) \cdot P(\bar{E}_4|E_1 \cap E_2 \cap \bar{E}_3) \cdot P(\bar{E}_5|E_1 \cap E_2 \cap \bar{E}_3 \cap \bar{E}_4) = 10 \cdot \frac{30}{80} \cdot \frac{29}{79} \cdot \frac{50}{78} \cdot \frac{49}{77} \cdot \frac{48}{76} \text{ oppure } \frac{\binom{30}{2} \cdot \binom{50}{3}}{\binom{80}{5}}.$$

- 4.2 tra i 5 nominativi estratti ve ne sia almeno 1 esperto;

$$P = 0.9119$$

$$1 - P(0 \text{ esperti}) = 1 - P(\bar{E}_1 \cap \bar{E}_2 \cap \bar{E}_3 \cap \bar{E}_4 \cap \bar{E}_5) = 1 - \frac{50}{80} \cdot \frac{49}{79} \cdot \frac{48}{78} \cdot \frac{47}{77} \cdot \frac{46}{76} = 1 - 0.0881.$$

- 4.3 supponendo di estrarre 100 nominativi con reimmissione, vi siano almeno 36 esperti.

$$P = 0.6217$$

Indicata con X la variabile casuale "numero di esperti" nelle 100 estrazioni risulta $X \sim \text{Bin}(n=100, p=0.375)$.

$M(X) = 37.5$; $\text{Var}(X) = 23.4375$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=37.5, \sigma^2=23.4375)$ e

$$P(X \geq 36) = P\left(\frac{X - \mu}{\sigma} \geq \frac{36 - \mu}{\sigma}\right) = P\left(Z \geq \frac{36 - 37.5}{4.8412}\right) = P(Z \geq -0.31) = P(Z \leq 0.31) = 0.6217.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO
Prova scritta di STATISTICA del 2.9.2010 (Tema 259)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1** Per uno studio sull'edilizia abitativa si sono intervistate 100 famiglie e rilevato il numero di mq dell'abitazione, Y , e la classe di reddito X (con modalità: bassa, alta).

$X \backslash Y$	60 → 80	80 → 100	100 → 150
bassa	11	4	9
alta	12	24	40

- 1.1** Si rappresentino graficamente le distribuzioni marginali.

Grafico X

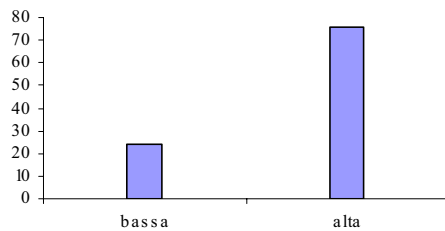
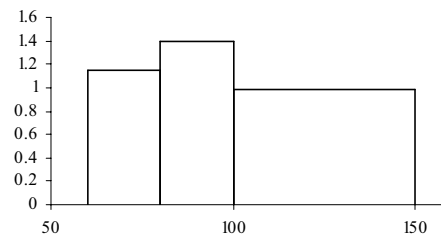


Grafico Y



- 1.2** Si calcolino opportuni indici di posizione e variabilità di X e Y .

X : Indice di posizione Moda valore alta Indice variabilità Frosini valore 0.48

Y : Indice posizione $M(Y)$ valore 102.55 Indice variabilità c.v. valore 0.2255

La maggior parte delle famiglie (76%) ha un reddito alto.

- 1.3** Si dica perché non esiste indipendenza stocastica tra X e Y e si calcoli con un opportuno indice il livello di connessione esistente tra le due variabili.

Non esiste indipendenza stocastica tra X e Y perché le distribuzioni condizionate non sono somiglianti

Indice χ_N valore 0.3076

Tra le variabili X e Y sussiste un livello di connessione basso.

- 2.** Sia data la seguente distribuzione di frequenze relative della variabile statistica X :

x_i	11	30	40	a	b
f_i	0.25	f_2	0.4	f_4	f_5

- 2.1** Si determinino i valori di a, b, f_2, f_4, f_5 in modo che risulti Moda=Media=40.

$$a = 50$$

$$b = 69$$

$$f_2 = 0.05$$

$$f_4 = 0.05$$

$$f_5 = 0.25$$

Se la distribuzione è simmetrica la media aritmetica coincide con la moda, se quest'ultima esiste.

- 2.2** Si dica che valore assume l'indice di asimmetria di X .

Indice di asimmetria γ_1 valore 0

Se la distribuzione è simmetrica l'indice di asimmetria vale 0.

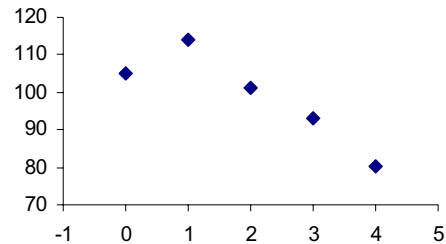
- 2.3** Si dica che valore può assumere la media potenziata di X di ordine $r = 1000$.

$\mu_X^{(1000)} \cong 69 = x_{\max}$ in base al Teorema sul comportamento delle medie potenziate.

3. Nella sottostante tabella è riportato il prezzo medio annuo in euro di una data obbligazione in cinque anni successivi:

anno	2005	2006	2007	2008	2009
Y	105	114.2	101.4	93.1	80.2

- 3.1 Indicando con X il numero di anni trascorsi dal 2005 (0,1,2,3,4), si rappresentino le coppie di valori (x_i, y_i) .



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$; II) $Y^* = \alpha$;

$a = 112.92$ $b = -7.07$ misura di adattamento: formula p^2 valore 0.7564

$\alpha = 98.78$ misura di adattamento: formula Indice di adattamento valore 0

- 3.3 In base al modello migliore si fornisca la previsione del prezzo teorico per l'anno 2010.

$$Y_{2010}^* = 112.92 - 7.07 \cdot 5 = 77.57$$

- 3.4 Si costruisca la serie dei numeri indici con base 2005 e la serie dei numeri indici a base mobile (base = 100).

	Y	NIBF (2005=100)	NIBM
2005	105	100.00	--
2006	114.2	108.76	108.76
2007	101.4	96.57	88.79
2008	93.1	88.67	91.81
2009	80.2	76.38	86.14

4. Una società che gestisce campi da golf effettua un'indagine sui suoi 300 dipendenti, di cui 170 sono maschi. E' emerso che 84 donne non giocano a golf, e che il 70% degli uomini gioca a golf. Si calcoli:

- 4.1 la probabilità che scegliendo a caso un dipendente questo sia giocatore di golf;

$$P = 0.55$$

	Ω		
	g	$\bar{g}$	
m	119	51	170
f	46	84	130
	165	135	300

- 4.2 la probabilità che sia maschio e giocatore di golf;

$$P = \frac{119}{300} = 0.3967$$

- 4.3 la probabilità che estraendone in blocco 10, la metà di questi siano femmine giocatrici di golf.

$$P = 0.008301$$

$$\binom{10}{5} \cdot P(E_1) \cdot P(E_2|E_1) \cdot P(E_3|E_1 \cap E_2) \cdot \dots \cdot P(\bar{E}_{10}|E_1 \cap \dots \cap E_5 \cap \bar{E}_6 \cap \dots \cap \bar{E}_9) = 252 \cdot \frac{46}{300} \cdot \frac{45}{299} \cdot \frac{44}{298} \cdot \frac{43}{297} \cdot \frac{42}{296} \cdot \frac{254}{295} \cdot \frac{253}{294} \cdot \frac{252}{293} \cdot \frac{251}{292} \cdot \frac{250}{291}$$

oppure $\frac{\binom{46}{5} \cdot \binom{254}{5}}{\binom{300}{10}}$, indicato con E l'evento femmina giocatrice di golf.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 16.9.2010 (Tema 260)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1** Vengono intervistati 16 iscritti in una piscina comunale, rilevando le variabili X = “tipologia di abbonamento” (a=ordinario, b=studenti, c=over 65) e Y = “numero medio di accessi mensili”:

X	a	a	b	a	b	c	a	b	c	b	c	b	b	c	b	b
Y	4	6	10	6	10	4	10	6	10	6	6	6	4	6	6	4

- 1.1** Si rappresentino graficamente le distribuzioni delle variabili X e Y .

Grafico X

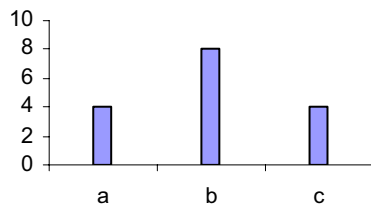
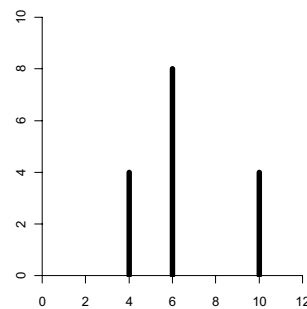


Grafico Y



- 1.2** Dopo aver costruito la tabella delle frequenze congiunte, dire, motivando la risposta, se esiste indipendenza stocastica fra X e Y . Calcolare poi un opportuno indice di connessione.

$X \backslash Y$	4	6	10	$n_{i\bullet}$
a	1	2	1	4
b	2	4	2	8
c	1	2	1	4
$n_{\bullet j}$	4	8	4	16

Esiste indipendenza stocastica tra X e Y ? ☒ NO perché le distribuzioni condizionate sono somiglianti.

Indice χ^2 valore 0

Se sussiste indipendenza stocastica la misura della connessione è nulla.

- 1.3** Confrontare con opportuni indici di posizione e variabilità le variabili $Y|X=a$ e $Y|X=b$.

$Y|X=a$: Indice di posizione $M(Y|X=a)$ valore 6.5 Indice variabilità $c.v.(Y|X=a)$ valore 0.3353

$Y|X=b$: Indice di posizione $M(Y|X=b)$ valore 6.5 Indice variabilità $c.v.(Y|X=b)$ valore 0.3353

Dal momento che le distribuzioni condizionate sono somiglianti saranno caratterizzate dagli stessi momenti.

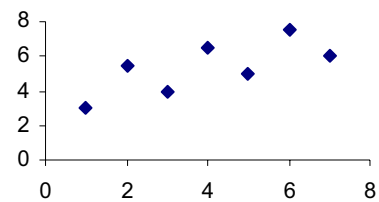
- 1.4** Sia $Z = 10 + Y$ una nuova variabile statistica ottenuta mediante trasformazione di Y . Quale tra i due caratteri presenta maggiore variabilità?

☒ ☐ perché $c.v.(Y) < c.v.(Z)$ infatti: $c.v.(Z) = \frac{\sigma_Z}{\mu_Z} = \frac{\sigma_Y}{10 + \mu_Y} < \frac{\sigma_Y}{\mu_Y} = c.v.(Y)$.

3. La seguente tabella riporta i valori della variabile Y rilevati in 7 successivi istanti temporali.

t	1	2	3	4	5	6	7
y_t	3	5.5	4	6.5	5	7.5	6

3.1 Rappresentare graficamente le coppie di valori (t, y_t) .



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y_t^* = a + b t$; II) $Y_t^* = c t$;

$a = 3.3571$ $b = 0.5$ misura di adattamento: formula RMS valore 0.9796

$c = 1.1714$ misura di adattamento: formula RMS valore 3.2337

3.3 Con riferimento ai valori y_t riportati in tabella si costruisca le serie dei numeri indici a base mobile.

t	Y_t	NIBM
1	3	--
2	5.5	1.8333
3	4	0.7273
4	6.5	1.6250
5	5	0.7692
6	7.5	1.5000
7	6	0.8000

3.4 Si calcoli ${}_3I_6$ e si interpreti il valore ottenuto.

$${}_3I_6 = \frac{Y_6}{Y_3} = \frac{7.5}{4} = 1.875$$

Tra l'istante temporale 3 e l'istante temporale 6 la variabile Y ha subito un incremento dell'87.5%.

3.5 Si calcoli il tasso medio di variazione tra gli istanti $t=3$ e $t=6$, commentando il risultato.

$$\text{tasso medio} = 100 \cdot (\sqrt[3]{1.875} - 1) \% = 20.51\%$$

Tra l'istante temporale 3 e l'istante temporale 6 la variabile Y ha subito un incremento medio del 20.51%.

4. Sapendo che $P(A) = 0.6$, $P(A-B) = 0.33$, $P(A \cap C) = 0.2$, $P(B \cap C) = 0$:

4.1 calcolare $P(A \cap B)$, $P(C|A)$ e $P(A-C)$;

$$P(A-B) = P(A) - P(A \cap B) \rightarrow P(A \cap B) = P(A) - P(A-B) = 0.6 - 0.33 = 0.27;$$

$$P(C|A) = \frac{P(C \cap A)}{P(A)} = \frac{0.2}{0.6} = 0.3333;$$

$$P(A-C) = P(A) - P(A \cap C) = 0.6 - 0.2 = 0.4$$

4.2 dire se B e C possono ritenersi stocasticamente indipendenti;

Gli eventi B e C sono disgiunti, quindi non possono ritenersi stocasticamente indipendenti.

4.3 calcolare la probabilità di ottenere meno di 20 successi in 51 lanci indipendenti, sapendo che la probabilità di successo è $p = 0.333$.

$$P = \underline{\hspace{2cm}}$$

Indicata con X la variabile casuale "numero di successi" nei 51 lanci indipendenti risulta $X \sim \text{Bin}(n=51, p=0.333)$.

$M(X) = 16.983$; $\text{Var}(X) = 11.3277$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=16.983, \sigma^2=11.3277)$ e

$$P(X < 20) = P\left(\frac{X - \mu}{\sigma} < \frac{20 - \mu}{\sigma}\right) = P\left(Z < \frac{20 - 16.983}{3.3657}\right) = P(Z < 0.90) = 0.8159.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 13.1.2011 (Tema 261)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Al fine di svolgere un'indagine sull'utilizzo di internet da parte dei propri studenti, una scuola media rileva il tempo di connessione giornaliero (in minuti) degli studenti delle sezioni A e B dell'ultimo anno. Sapendo che per le due sezioni il numero di studenti, la media e la varianza del tempo di connessione sono rispettivamente:

$$n_A = 25, \quad \mu_A = 28, \quad \sigma_A^2 = 49 \quad \text{e} \quad n_B = 25, \quad \mu_B = 25, \quad \sigma_B^2 = 36$$

- 1.1 si calcolino media e varianza del tempo di connessione, X , per il complesso delle due sezioni della scuola.

media formula: $\mu = \frac{\mu_A n_A + \mu_B n_B}{n_A + n_B}$ valore 26.5 varianza formula: $\sigma^2 = \sigma_{Between}^2 + \sigma_{Within}^2$ valore $2.25 + 42.5 = 44.75$

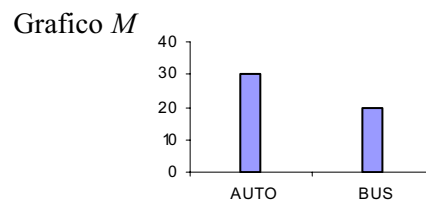
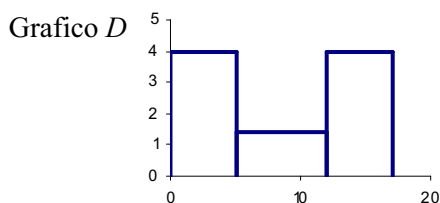
- 1.2 Supponendo che il costo per la connessione, Y , sia legato al tempo di connessione, X , dalla relazione $Y = 105 - 3X$, si calcolino media e varianza di Y .

$M(Y)$ formula: $M(Y) = 105 - 3 \cdot M(X)$ valore 25.5 $Var(Y)$ formula: $Var(Y) = 3^2 \cdot Var(X)$ valore 402.75

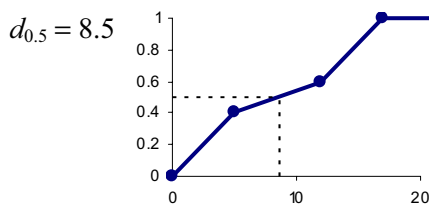
2. Al fine di studiare la relazione esistente tra il mezzo, M , utilizzato per raggiungere l'aeroporto e la durata prevista del viaggio aereo (D , in ore), è stato intervistato un campione di 50 passeggeri in partenza da un dato aeroporto del Nord Italia. I dati raccolti sono riassunti nella seguente tabella:

$D \backslash M$	auto	bus
0-5	7	13
5-12	7	3
12-17	16	4

- 2.1 Si rappresentino graficamente le distribuzioni marginali D e M .



- 2.2 Si individui su un'opportuno grafico la mediana di D e si calcoli l'indice di eterogeneità di M .



Eterogeneità M Indice di Frosini = 0.8

- 2.3 Si valuti l'asimmetria della variabile D . valore 0 commenti distribuzione simmetrica

- 2.4 Si valuti il livello di connessione esistente tra M e D mediante l'impiego di un opportuno indice e si commenti il risultato ottenuto.

indice χ_N valore 0.4233 commenti basso livello di connessione

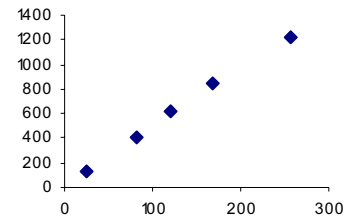
- 2.5 Si costruisca una tabella di massima connessione tra D e M con la distribuzione marginale di M assegnata.

$D \backslash M$	auto	taxi	treno	Bus	
0-5	9	0	0	0	9
5-12	0	20	9	0	29
12-17	0	0	0	12	12
	9	20	9	12	50

3. Con riferimento a un comprensorio sciistico composto da 5 impianti di risalita, si considerano i valori delle variabili “durata della risalita” (Y , in secondi) e “lunghezza dell’impianto” (X , in hm) riportati nella seguente tabella:

X	25	81	121	169	256
Y	138	402	618	854	1218

3.1 Si rappresentino graficamente le coppie di dati.



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = c X; \quad \text{II) } Y^* = a + b \sqrt{X};$$

$$c = 4.89 \quad \text{misura di adattamento: RMS} = 576.31$$

$$a = -423.33 \quad b = 99.01 \quad \text{misura di adattamento: RMS} = 2872.96$$

commenti Il modello che si adatta meglio ai dati è $Y^* = 4.89 X$

4. Si completino ove possibile le seguenti tabelle, motivando la risposta, in modo tale che:

$$0 < \rho^2 = \eta_{Y|X}^2 < 1$$

$X \backslash Y$	1	5	
10	10	0	10
20	20	20	40
$x_3 > 20$	0	5	5
	30	25	55

$$\chi^2 = 55$$

$X \backslash Y$	1	5	
10	20	0	20
20	30	0	30
$x_3 > 20$	0	5	5
	50	5	55

$$\rho^2 = 0$$

$X \backslash Y$	1	5	
10		0	
20			
	0		
			55

$\rho^2 = \eta_{Y|X}^2 \rightarrow$ medie condiz. allineate: $M(X|Y_1)$ e $M(X|Y_2)$ sono allineate.

Se $\chi^2 = 55 = n \cdot \min(h-1, k-1) = \chi_{\text{MAX}}^2$
 $\chi_{\text{NORM}}^2 = 1$; quindi $x_3 > 20$ e distribuzioni $Y|X_i$ degeneri.

Se $\rho^2 = 0$ indipendenza lineare (coefficiente angolare mod retta nullo). IMPOSSIBILE: qualsiasi combinazione di frequenze comporta $\rho^2 > 0$ (provare per via grafica).

5. L’ufficio *controllo qualità* di una data azienda rileva che la probabilità che un pezzo prodotto su una determinata linea sia difettoso è pari a 0.1875. Si indichi con X la variabile casuale “numero di pezzi difettosi rilevati” in 10 estrazioni indipendenti.

5.1 Si determini la probabilità che su 10 pezzi estratti a caso se ne presentino al massimo 2 difettosi.

$$P = 0.7152$$

Risulta $X \sim \text{Bin}(n=10, p=0.1875)$.

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = \binom{10}{0} 0.1875^0 \cdot 0.8125^{10} + \binom{10}{1} 0.1875^1 \cdot 0.8125^9 + \binom{10}{2} 0.1875^2 \cdot 0.8125^8 = 0.7152.$$

5.2 Supponendo che i pezzi estratti siano 120, si calcoli la probabilità di trovare meno di 25 pezzi difettosi.

$$P = 0.352$$

Indicata con X la variabile casuale “numero di pezzi difettosi” nelle 120 estrazioni risulta $X \sim \text{Bin}(n=120, p=0.1875)$.
 $M(X) = 22.5$; $\text{Var}(X) = 18.2813$; utilizzando l’approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=22.5, \sigma^2=18.2813)$ e

$$P(X \leq 25) = P\left(\frac{X - \mu}{\sigma} \leq \frac{25 - \mu}{\sigma}\right) = P\left(Z \leq \frac{25 - 22.5}{4.2757}\right) = P(Z \leq 0.58) = 0.7190.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 3.2.2011 (Tema 262)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

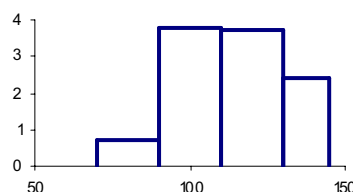
1. In un'indagine condotta su **200** studenti sono state rilevate la variabile statistica Y = "punteggio a un test per la misurazione del Q.I." e la mutabile statistica X = "sesso". Nel seguente prospetto sono riportate le distribuzioni delle frequenze relative condizionate ($Y|X$):

$Y X$	70 → 90	90 → 110	110 → 130	130 → 145	
maschi	0.10	0.35	0.40	0.15	1
femmine	0.05	0.40	0.35	0.20	1

- 1.1 Si ricostruisca la tabella a doppia entrata delle frequenze congiunte sapendo che i maschi sono 80.

$X \backslash Y$	70 → 90	90 → 110	110 → 130	130 → 145	
maschi	8	28	32	12	80
femmine	6	48	42	24	120
	14	76	74	36	200

- 1.2 Si rappresenti graficamente la distribuzione marginale Y .



- 1.3 Si confrontino i punteggi per i gruppi dei maschi e delle femmine con un opportuno indice di posizione e con un opportuno indice di variabilità, commentando i risultati ottenuti.

$Y|X=\text{maschi}$ Indice di posizione: $M(Y|X=\text{maschi}) = 111.625$ Indice di variabilità: $cv(Y|X=\text{maschi}) = 0.1488$

$Y|X=\text{femmine}$ Indice di posizione: $M(Y|X=\text{femmine}) = 113.5$ Indice di variabilità: $cv(Y|X=\text{femmine}) = 0.1418$

Con riferimento ai 200 studenti oggetto di indagine, le femmine presentano un QI medio leggermente superiore ai maschi, la variabilità è quasi identica nei due gruppi.

- 1.4 Si dica perché X e Y non sono stocasticamente indipendenti, giustificando la risposta data.

Le distribuzioni condizionate $Y|X$ non sono somiglianti.

- 1.5 Si calcoli un opportuno indice di connessione, commentando il risultato ottenuto.

indice χ_{Norm} valore 0.1229 commenti Livello di connessione molto basso.

2. Con riferimento ai prezzi di un prodotto nel periodo 2004–2008 sono disponibili alcuni elementi della serie dei numeri indici a base fissa 2004=100 e della serie dei numeri indici a base fissa 2006=100.

Anni	$NIBF(2004=100)$	$NIBF(2006=100)$	$NIBM$
2004	100	94.34	--
2005	102.5	96.70	1.025
2006	106	100	1.0341
2007	109.18	103	1.03
2008	115.54	109	1.0583

- 2.1 Si ricostruiscano le due serie complete dei NIBF dei prezzi e si costruisca la serie dei NI a base mobile.

$${}_{t-1}I_t = \frac{p_t}{p_{t-1}} = \frac{{}_0I_t}{{}_0I_{t-1}}, \quad {}_{2004}I_{2008} = {}_{2004}I_{2007} \cdot {}_{2007}I_{2008} \quad \text{ovvero} \quad {}_{2004}I_{2008}:106=109:100, \quad {}_{2006}I_{2004}:100=100:106, \quad {}_{2006}I_{2005}:100=102.5:106$$

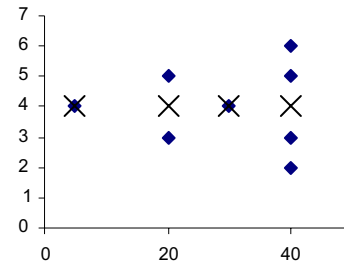
- 2.2 Si calcoli il tasso d'incremento medio dei prezzi tra il 2004 e il 2007.

$$\sqrt[3]{109.18/100} = 1.0297, \text{ valore relativo medio, cui corrisponde una variazione relativa media del 2.97\%}.$$

3. Di un gruppo di 8 studenti di un corso universitario sono state rilevate le variabili X = “numero di ore di corso seguite in aula” e Y = “numero di quesiti sbagliati alla prova d’esame”.

X	5	20	20	30	40	40	40	40
Y	4	5	3	4	3	2	6	5

- 3.1 Si rappresentino graficamente le coppie di dati e la funzione di regressione.



- 3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = k$; II) $Y^* = a + bX$ III) $Y^* = cX$;

$k = 4$ misura di adattamento: RMS = 1.5

$a = 4$ $b = 0$ misura di adattamento: RMS = 1.5

$c = 0.1157$ misura di adattamento: RMS = 3.9061

Dal grafico si evince una situazione di indipendenza in media di Y da X .

4. Si completino se possibile le seguenti tabelle, motivando teoricamente la risposta, in modo tale che:

$\eta^2_{Y|X} = 0$ e $\rho^2 = 1$

$X \backslash Y$	1	2	3	
1	1			
3		1		
				10

IMPOSSIBILE: $\rho^2 = 1$ implicherebbe $\eta^2_{Y|X} = 1$

$\rho^2 = 0$

$X \backslash Y$	1	2	3	
1	1	3	1	5
3	2	1	2	5
	3	4	3	10

Se $\eta^2_{Y|X} = 0$ anche $\rho^2 = 0$;
nel caso in esame $M(Y|X=1) = M(Y|X=3) = 2$

5. Sapendo che i tre eventi A , B , C costituiscono una partizione dello spazio probabilistico Ω , che $P(A) = 0.25$ e che $P(B)$ è il doppio di $P(C)$, si calcoli:

5.1 $P(A \cap B) = 0$ $A \cap B = \emptyset$ in quanto A e B sono disgiunti

5.2 $P(B - C) = P(B) - P(B \cap C) = P(B) = 0.5$ (B e C sono disgiunti)

5.3 $P(A \cup \Omega) = P(\Omega) = 1$

- 5.4 Si ipotizzi che $P(A)$ sia la probabilità di vincere un premio a un gioco: ripetendo il gioco 105 volte qual è la probabilità di vincere al massimo 30 volte?

$P = 0.8023$

Indicata con X la variabile casuale “numero di vincite” nelle 105 ripetizioni del gioco risulta $X \sim \text{Bin}(n=105, p=0.25)$. $M(X) = 26.25$; $\text{Var}(X) = 19.6875$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=26.25, \sigma^2=19.6875)$ e

$$P(X \leq 30) = P\left(\frac{X - \mu}{\sigma} \leq \frac{30 - \mu}{\sigma}\right) = P\left(Z \leq \frac{30 - 26.25}{4.4371}\right) = P(Z \leq 0.85) = 0.8023.$$

A	B
C	

Ω

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 17.2.2011 (Tema 263)

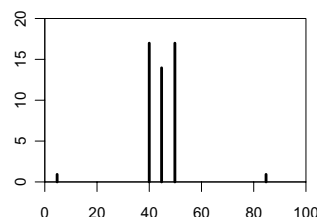
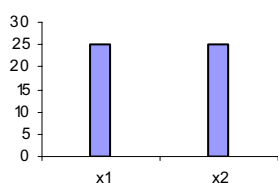
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. In un'indagine condotta su 50 unità statistiche sono state rilevate la mutabile statistica X e la variabile statistica discreta Y . Nel seguente prospetto è riportata la distribuzione delle frequenze congiunte:

$X \backslash Y$	5	40	45	50	85
x_1	0	12	7	5	1
x_2	1	5	7	12	0

- 1.1 Si rappresentino graficamente le distribuzioni marginali X e Y .



- 1.2 Si calcoli il coefficiente di variazione della marginale Y .

Varianza: 81

coefficiente di variazione: 0.2

La distribuzione Y è simmetrica rispetto al valore 50 (media); i valori 5 e 85 sembrano ritenersi dati anomali.

- 1.3 Si indichi senza effettuare calcoli l'intervallo dei possibili valori che può assumere la media geometrica di Y .

$$5 = y_{\min} < \mu^{(0)} < \mu^{(1)} = M(Y) = 50$$

- 1.4 Si calcolino i valori della media aritmetica e della mediana per le distribuzioni condizionate $Y|x_i$, commentando i risultati ottenuti.

$Y|x_1$ media aritmetica: $M(Y|x_1) = 45.2$ mediana: 45

$Y|x_2$ media aritmetica: $M(Y|x_2) = 44.8$ mediana: 45

La mediana è un indicatore più robusto della media aritmetica, pertanto non risente della presenza dei valori anomali. La distribuzione $Y|x_1$ presenta asimmetria positiva, mentre $Y|x_2$ asimmetria negativa.

- 1.5 Si calcoli una misura di mutabilità e si costruisca una tabella di minima mutabilità per la marginale X .

Indice: Frosini (normalizzato) valore 1

La variabile X è caratterizzata da equidistribuzione delle frequenze, presenta quindi massima eterogeneità

x_i	n_i
x_1	50
x_2	0

2. Si consideri la serie dei numeri indici a base mobile dei prezzi del prodotto A nel periodo 2004–2007.

anno	$NIBM_A$	$NIBF_A$ (2004=100)	$NIBM_B$
2004	-	100.00	-
2005	105	105.00	105
2006	105	110.25	105
2007	105	115.76	105

- 2.1 Si ricostruisca la serie dei numeri indici a base fissa (2004=100).

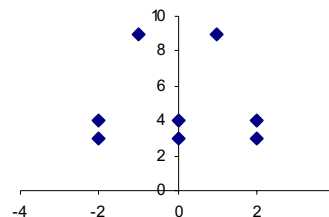
- 2.2 Si costruisca la serie dei numeri indici a base mobile per i prezzi del prodotto B , sapendo che nel periodo considerato i prezzi di B sono costantemente risultati il doppio dei prezzi di A .

I numeri indici sono valori relativi; non risentono quindi di eventuali trasformazioni di scala.

3. Il seguente prospetto riporta le rilevazioni della temperatura, X , (in gradi centigradi) e delle precipitazioni nevose, Y , (in cm) nei primi 8 giorni di gennaio in una certa località.

i	1	2	3	4	5	6	7	8
x_i	-2	-1	-2	1	2	2	0	0
y_i	3	9	4	9	4	3	4	3

3.1 Si rappresentino graficamente le coppie di dati (x_i, y_i) .



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$

II) $Y^* = c + dX^2$

$a = 4.875$

$b = 0$

misura di adattamento: $\rho^2 = 0$

$c = 6.0882$

$d = -0.5392$

misura di adattamento: $\rho^2 = 0.1582$

Il modello che si adatta meglio ai dati è $Y^* = 6.0882 - 0.5392 X^2$

3.3 Si indichi l'intervallo dei valori che può assumere il rapporto di correlazione η_{YX}^2 .

$0.1582 < \eta_{YX}^2 < 1$. Anche il secondo modello è confrontabile con la funzione di regressione (modello passante per le medie condizionate, nel caso in esame anche rappresentabile come un polinomio di quarto grado).

4. Si completino se possibile le seguenti tabelle, riportando modalità e frequenze e motivando teoricamente la risposta, in modo tale che:

$\eta_{YX}^2 = 1$, $\eta_{XY}^2 = 0$ e $\rho > 0$

$X \backslash Y$	6	$y_2 =$
5		
6		
7		

$\eta_{YX}^2 = 1$ e $\rho < 0$

$X \backslash Y$	6	$y_2 > y_1$
5	0	$n_{12} \geq 5$
6	4	0
7	0	5
8	$n_{41} \geq 4$	0

IMPOSSIBILE se $\eta_{XY}^2 = 0$ segue che $\rho^2 = 0$

Si è riportata una possibile soluzione in corrispondenza della quale si ottiene $M(X|y_1) \geq 7$ e $M(X|y_2) \leq 6$.

5. L'urna U1 contiene 10 palline bianche e 40 palline rosse. L'urna U2 contiene 14 palline bianche, 20 rosse e 15 nere.

5.1 Si calcoli la probabilità che estraendo senza reimmissione 5 palline dall'urna U1 ce ne siano almeno 4 rosse.

$P = 0.7419$

Può essere, quindi, indifferentemente risolto come $\binom{5}{4} \cdot P(R_1 \cap R_2 \cap R_3 \cap R_4 \cap \bar{R}_5) + \binom{5}{5} \cdot P(R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5) =$

$= \binom{5}{4} \cdot P(R_1) \cdot P(R_2|R_1) \cdot P(R_3|R_1 \cap R_2) \cdot P(R_4|R_1 \cap R_2 \cap R_3) \cdot P(\bar{R}_5|R_1 \cap R_2 \cap R_3 \cap R_4) + \binom{5}{5} \cdot P(R_1 \cap R_2 \cap R_3 \cap R_4 \cap R_5) =$

$= 5 \cdot \frac{40}{50} \cdot \frac{39}{49} \cdot \frac{38}{48} \cdot \frac{37}{47} \cdot \frac{10}{46} + 1 \cdot \frac{40}{50} \cdot \frac{39}{49} \cdot \frac{38}{48} \cdot \frac{37}{47} \cdot \frac{36}{46}$ o mediante la v.c. ipergeometrica $P(X=4) + P(X=5) = \frac{\binom{40}{4} \cdot \binom{10}{1}}{\binom{50}{5}} + \frac{\binom{40}{5} \cdot \binom{10}{0}}{\binom{50}{5}}$.

5.2 Si estraiga una pallina dall'urna U1 e la si metta nell'urna U2. Si calcoli la probabilità che estraendo con reimmissione 5 palline dall'urna U2 si ottengano 2 palline bianche.

$P = 0.2958$

$P(2B) = P(2B \cap (U_{1B} \cup U_{1R})) = P((2B \cap U_{1B}) \cup (2B \cap U_{1R})) = P(U_{1B})P(2B|U_{1B}) + P(U_{1R})P(2B|U_{1R}) =$

$\frac{10}{50} \binom{5}{2} \left(\frac{15}{50}\right)^2 \left(\frac{35}{50}\right)^3 + \frac{40}{50} \binom{5}{2} \left(\frac{14}{50}\right)^2 \left(\frac{36}{50}\right)^3 = 0.0617 + 0.2341 = 0.2958$, avendo determinato $P(2B|U_{1B})$ e $P(2B|U_{1R})$ facendo

rispettivamente ricorso a vv.cc. Binomiali con parametri $\left(n=5, p=\frac{15}{50}\right)$ e $\left(n=5, p=\frac{14}{50}\right)$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 9.6.2011 (Tema 264_1)

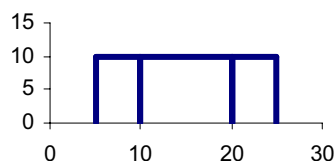
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. In un'indagine condotta su 200 unità statistiche è stata rilevata la variabile statistica X di cui si conoscono le densità di frequenza:

$a_{i-1} \rightarrow a_i$	d_i
16 $\rightarrow$ 21	10
21 $\rightarrow$ 26	10
26 $\rightarrow$ 36	10

- 1.1 Si rappresenti graficamente la variabile X .



- 1.2 Si calcolino media aritmetica e coefficiente di variazione di X .

$M(X) = 26$

coefficiente di variazione: 0.204

- 1.3 Si indichi senza effettuare calcoli l'intervallo dei possibili valori che può assumere la media quadratica di X .

$26 < \mu^{(2)} < 31$ in base alle proprietà delle medie potenziate.

- 1.4 Sia $Y = 9 + 4X$ una nuova variabile statistica. Si indichi commentando la risposta se le seguenti affermazioni sono corrette.

$M(Y) > M(X)$ VERO in quanto $M(Y) = 9 + 4 M(X) = 9 + 4 \cdot 26 > 26$

$\text{Var}(Y) > \text{Var}(X)$ VERO in quanto $\text{Var}(Y) = 4^2 \text{Var}(X) > \text{Var}(X)$

2. Si consideri la seguente distribuzione congiunta dei caratteri X e Y .

$X \backslash Y$	y_1	y_2	y_3
x_1	24	26	28
x_2	12	13	14

- 2.1 Si calcoli un opportuno indice di connessione, commentando il risultato ottenuto.

Indice $\chi_N^2 = 0$ Sussiste indipendenza stocastica tra le variabili X e Y .

- 2.2 Si indichi l'intervallo teorico di valori che può assumere, nel caso in esame, l'indice χ^2 assoluto.

$\chi^2 = 0$ in quanto sussiste indipendenza stocastica tra le variabili X e Y .

3. Il seguente prospetto riporta i prezzi e le quantità scambiate per i prodotti A e B negli anni 1990, 2000 e 2010.

anno	AP_t	AQ_t	BP_t	BQ_t
1990	100	111	200	101
2000	150	142		92
2010	260	123	600	83

- 3.1 Si indichi quale valore deve assumere il prezzo del prodotto B nel 2000 affinché l'indice dei prezzi di Laspeyres al tempo 2000 con base 1990 risulti pari a $\frac{P_{2000}^L}{P_{1990}^L} = 1.5$.

$BP_{2000} = 300$ dal momento che $\frac{AP_{2000}}{AP_{1990}} = 1.5$ segue che $\frac{BP_{2000}}{BP_{1990}} = 1.5$

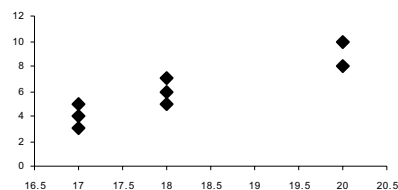
- 3.2 Si calcoli l'indice dei prezzi di Paasche al tempo 2010 con base 1990, commentando il risultato ottenuto.

$\frac{P_{2010}^P}{P_{1990}^P} = 2.83$. Tra il 1990 e il 2010 si è verificato un incremento dei prezzi del 183%.

4. Si considerino le seguenti 8 coppie di valori (x_i, y_i) .

x_i	17	17	17	18	18	18	20	20
y_i	3	4	5	5	6	7	8	10

4.1 Si rappresentino graficamente le coppie di dati (x_i, y_i) .



4.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$

II) $Y^* = c;$

$a = -24$

$b = 1.6552$

misura di adattamento: formula $\rho^2 = 0.8276$

$c = 6$

misura di adattamento: formula $\rho^2 = 0$

Il modello che si adatta meglio ai dati è $Y^* = -24 + 1.6552X$

4.3 Si indichi l'intervallo dei valori che può assumere il rapporto di correlazione $\eta_{Y|X}^2$.

$0.8276 = \rho^2 < \eta_{Y|X}^2 < 1.$

5. Si completino se possibile le seguenti tabelle, riportando modalità e frequenze e motivando teoricamente la risposta, in modo tale che:

$\eta_{Y|X}^2 = 0, \eta_{X|Y}^2 > 0$

$X \backslash Y$	2	5	8
x_1	7	9	7
x_2		9	7

$\rho^2 = 0$

$X \backslash Y$	y_1	y_2	y_3
x_1	7	9	7
x_2		9	7

Impossibile.

$\eta_{Y|X}^2 = 0 \rightarrow M(Y|x_1) = M(Y|x_2) = M(Y) \rightarrow n_{21} = 7$

se $n_{21} = 7$ le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono somiglianti, sussiste indipendenza stocastica e $\eta_{X|Y}^2 = 0$.

$n_{21} = 7$, le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono somiglianti, sussiste indipendenza stocastica e $\rho^2 = 0$.

6. Si considerino gli eventi A e B . Sapendo che $P(B|A) = 0.8$ e che $P(A \cap B) = 0.2$:

6.1 si calcoli $P(A)$ e si indichi l'intervallo dei valori che può assumere $P(B)$.

$P(A) = 0.25$

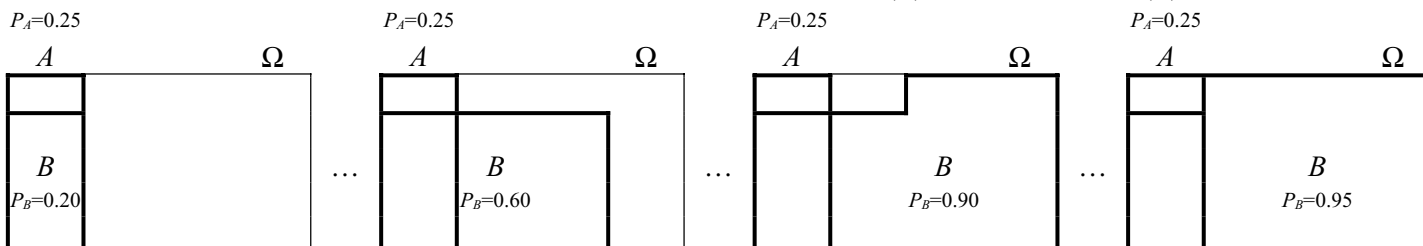
$P(B|A) = \frac{P(A \cap B)}{P(A)} \rightarrow P(A) = \frac{P(A \cap B)}{P(B|A)} = \frac{0.2}{0.8}$

$0.2 \leq P(B) \leq 0.95$

$0.25 = P(A) \leq P(A \cup B) = P(A) + P(B) - P(A \cap B) \leq 1$

$0.25 \leq 0.25 + P(B) - 0.2 \leq 1$

$0.25 \leq 0.05 + P(B) \leq 1 \rightarrow 0.2 \leq P(B) \leq 0.95$



6.2 La probabilità di rispondere correttamente a un certo test è pari a 0.82. Considerando le ripetizioni del test eventi indipendenti, si calcoli la probabilità che replicandolo 6 volte lo si superi almeno 5 volte.

$P = 0.7044$

La variabile X numero di risposte corrette nelle sei replicazioni è una variabile casuale $\text{Bin}(n=6, p=0.82)$.

$P(X \geq 5) = P(X = 5) + P(X = 6) = \binom{6}{5} 0.82^5 \cdot 0.18^1 + \binom{6}{6} 0.82^6 \cdot 0.18^0 = 0.4004 + 0.3040 = 0.7044.$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 23.6.2011 (Tema 264_2)

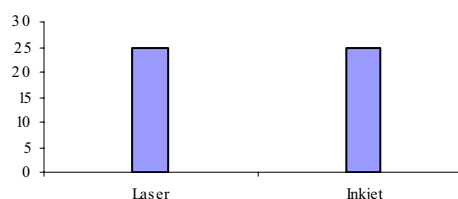
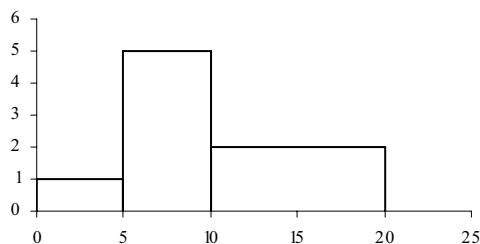
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. La seguente tabella riporta la distribuzione di frequenze congiunte riguardanti il numero di pagine stampate al minuto (Y) da 50 stampanti utilizzate in un centro di calcolo distinte per modalità di stampa (X).

$X \backslash Y$	0-5	5-10	10-20
Laser	2	13	10
Inkjet	3	12	10

- 1.1 Si rappresentino graficamente le distribuzioni marginali Y e X .



- 1.2 Si confrontino con opportuni indici di posizione e di variabilità le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.

$Y|x_1$ indice di posizione: $M(Y|x_1)$ valore 10.1 indice di variabilità: $cv(Y|x_1)$ valore 0.4170

$Y|x_2$ indice di posizione: $M(Y|x_2)$ valore 9.9 indice di variabilità: $cv(Y|x_2)$ valore 0.4488

Nel centro di stampa considerato le stampanti Laser sono leggermente più veloci, in media, di quelle Inkjet, la cui velocità presenta un livello di variabilità moderatamente più elevato.

- 1.3 Si calcoli l'indice di mutabilità della variabile X .

indice di mutabilità: Indice di Frosini valore 1

Sussiste massima eterogeneità.

- 1.4 si calcoli un opportuno indice di connessione tra X e Y .

$\chi_N = 0.0693$ Le variabili X e Y presentano un livello di connessione quasi nullo.

2. E' data la serie storica dei numeri indici con base mobile (su scala centesimale) del prezzo di un kg di pane:

anno	$_{t-1}I_t$	p_t	NIBF(2006=100)
2006	--	0.75	100
2007	106	0.7950	106.00
2008	104	0.8268	110.24
2009	102	0.8433	112.44
2010	101	0.8518	113.57

- 2.1 Si ricostruisca la serie storica dei prezzi sapendo che nel 2006 il pane costava 0.75 euro.

- 2.2 Si ricostruisca la serie storica dei numeri indici a base fissa (2006 = 100).

- 2.3 Si determini il tasso medio di variazione del prezzo del pane tra il 2006 e il 2010.

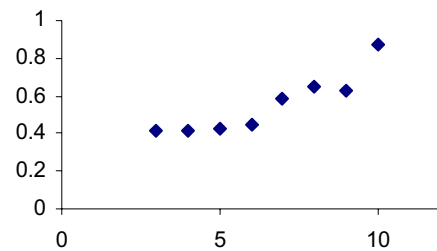
tasso medio di variazione = +3.23%

$$\sqrt[4]{_{2006}I_{2007} \cdot _{2007}I_{2008} \cdot _{2008}I_{2009} \cdot _{2009}I_{2010}} = \sqrt[4]{_{2006}I_{2010}} = \sqrt[4]{1.1357} = 1.0323$$
 valore relativo medio, da cui segue una variazione relativa media del +3.23%.

3. La tabella seguente riporta l'andamento della media mensile dei valori Euribor tra marzo e ottobre del 2010:

t	3	4	5	6	7	8	9	10
Y	0.41	0.41	0.43	0.45	0.58	0.65	0.63	0.87

3.1 Si rappresentino graficamente le coppie di dati.



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + b t$$

$$\text{II) } Y^* = c + d t^2;$$

$$a = 0.1583 \quad b = 0.0608 \quad \text{misura di adattamento: formula } \rho^2 = 0.8485$$

$$c = 0.3261 \quad d = 0.0048 \quad \text{misura di adattamento: formula } \rho^2 = 0.9112$$

Il modello migliore risulta $Y^* = 0.3261 + 0.0048 t^2$

3.3 Si determini, utilizzando il modello ritenuto migliore, il valore del tasso Euribor per il mese di gennaio 2011.

$$Y^*(13) = 0.3261 + 0.0048 \cdot 13^2 = 1.1361$$

3.4 Si indichi l'intervallo dei valori che può assumere il rapporto di correlazione η_{YX}^2 .

$\eta_{YX}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X.

4. Il coefficiente di correlazione lineare, ρ , fra due caratteri X ed Y è pari a 0.6.

Sapendo inoltre che $M(X) = 10$, $M(Y) = 20$, $\sigma_X = 1.5$ e $\sigma_Y = 2$, si determinino:

4.1 la covarianza tra X ed Y;

$$\rho = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y} \rightarrow \text{Cov}(X,Y) = \rho \sigma_X \sigma_Y = 0.6 \cdot 1.5 \cdot 2 = 1.8$$

4.2 i parametri della retta di regressione $Y = a + bX$.

$$b = \frac{\text{Cov}(X,Y)}{\sigma_X^2} = \frac{1.8}{1.5^2} = 0.80 \quad a = M(Y) - bM(X) = 20 - 0.8 \cdot 10 = 12$$

5. Si consideri l'esperimento di lancio di due monete regolari.

5.1 Si costruisca lo spazio degli eventi elementari.

Ω	C		
	T	•	•
	T	•	•
		T	C

5.2 Si determini la probabilità che in almeno 2 su 7 prove si ottenga la stessa faccia in entrambe le monete.

$$P = 0.9375$$

Sia $p = P((T \cap T) \cup (C \cap C)) = 0.5^2 + 0.5^2 = 0.5 = \frac{2}{4}$ la probabilità dell'evento 'stessa faccia in entrambe le monete'.

La variabile casuale X numero successi nelle 7 prove risulta distribuita secondo una v.c. binomiale $\text{Bin}(n=7, p=0.5)$.

$$P(X \geq 2) = 1 - P(X=0) - P(X=1) = 1 - \binom{7}{0} 0.5^0 \cdot 0.5^7 - \binom{7}{1} 0.5^1 \cdot 0.5^6 = 1 - 0.0078 - 0.0547 = 0.9375.$$

5.3 Supponendo ora che le prove siano 50, determinare la probabilità che in almeno 30 si ottenga la stessa faccia in entrambe le monete.

$$P = 0.0793$$

Indicata con X la variabile casuale "numero di successi" nelle 50 prove indipendenti risulta $X \sim \text{Bin}(n=50, p=0.5)$.

$M(X) = 25$; $\text{Var}(X) = 12.5$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=25, \sigma^2=12.5)$ si ottiene

$$P(X > 30) = P\left(\frac{X - \mu}{\sigma} > \frac{30 - \mu}{\sigma}\right) = P\left(Z > \frac{30 - 25}{3.5355}\right) = P(Z > 1.41) = 1 - P(Z < 1.41) = 1 - 0.9207 = 0.0793.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 7.7.2011 (Tema 264_3)

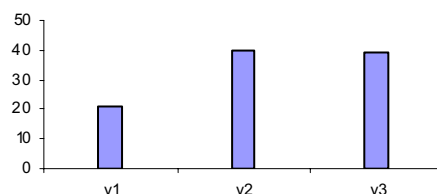
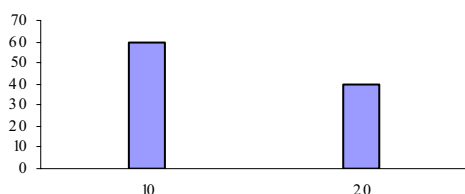
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

- 1.** In un'indagine condotta su 100 unità statistiche sono state rilevate la variabile statistica discreta X e la mutabile statistica Y . Nel seguente prospetto è riportata la distribuzione delle frequenze congiunte:

$X \backslash Y$	y_1	y_2	y_3
10	11	20	29
20	10	20	10

- 1.1** Si rappresentino graficamente le distribuzioni marginali X e Y .



- 1.2** Si confrontino con opportuni indici di posizione e variabilità le distribuzioni condizionate $Y|x_i$, commentando i risultati ottenuti.

$Y x_1$ indice di posizione: Moda	valore y_3	indice di variabilità: Frosini	valore 0.7402
$Y x_2$ indice di posizione: Moda	valore y_2	indice di variabilità: Frosini	valore 0.75

Entrambe le distribuzioni presentano un livello medio/alto di mutabilità

- 1.2** Si calcolino media e varianza di X .

Media: 14 Varianza: 24

- 1.3** Si calcoli un opportuno indice di connessione, commentando il risultato ottenuto.

$\chi_N = 0.2352$ Le variabili X e Y presentano un livello di connessione basso.

- 2.** Si consideri la serie storica dei prezzi del prodotto A nel periodo 2001–2004.

anno	p_A	$NIBF_A$ (2003=100)
2001	20	100
2002	25	125
2003	20	100
2004	30	150
2005	40	200

- 2.1** Si ricostruisca la serie dei numeri indici a base fissa (2003=100).

- 2.2** Si calcolino, interpretandone il significato, i seguenti numeri indici.

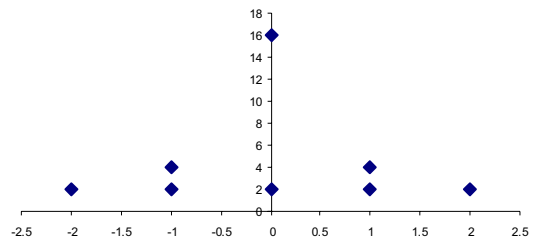
${}_{2002}I_{2001} = 0.80$ Il prezzo nel 2001 è risultato il 20% inferiore rispetto al prezzo nel 2002

${}_{2003}I_{2005} = 2$ Tra il 2003 e il 2005 si è avuto un incremento del 100%

3. Si considerino le seguenti 8 coppie di valori (x_i, y_i) .

x_i	-2	-1	-1	0	0	1	1	2
y_i	2	2	4	2	16	4	2	2

3.1 Si rappresentino graficamente le coppie di dati (x_i, y_i) .



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = \alpha$ II) $Y^* = c + dX^2$;

$\alpha = 4.25$ misura di adattamento: formula $\rho^2 = 0$

$c = 6.1667$ $d = -1.2778$ misura di adattamento: formula $\rho^2 = 0.1797$

Il modello migliore risulta $Y^* = 6.1667 - 1.2778 X^2$

3.3 Si indichi l'intervallo dei valori che può assumere il rapporto di correlazione η_{YX}^2 .

$0.1797 < \eta_{YX}^2 < 1$

4. Si completino se possibile le seguenti tabelle, riportando modalità e frequenze e motivando teoricamente la risposta, in modo tale che:

$\eta_{YX}^2 < \eta_{XY}^2 < 1$

$X \backslash Y$	1	2	3
2	2	5	2
5	10	0	10
8	10	10	10

Una soluzione consiste nel fissare $\eta_{YX}^2 = 0$;
 $M(Y|X=2) = M(Y|X=5) = M(Y|X=8) = M(Y) = 2 \rightarrow \eta_{YX}^2 = 0$
 $M(X|Y=1) = M(X|Y=3) \neq M(X|Y=2) \rightarrow \eta_{XY}^2 > 0$.

$\eta_{YX}^2 = \eta_{XY}^2$

$X \backslash Y$	1	2	3
2	2	5	2
5	10	0	10
8	2	5	2

Una soluzione consiste nel fissare $\eta_{YX}^2 = \eta_{XY}^2 = 0$;
 $M(Y|X=2) = M(Y|X=5) = M(Y|X=8) = M(Y) = 2 \rightarrow \eta_{YX}^2 = 0$
 $M(X|Y=1) = M(X|Y=3) = M(X|Y=2) = M(X) = 5 \rightarrow \eta_{XY}^2 = 0$.

5. Un'urna contiene 14 palline bianche e 36 palline rosse.

5.1 Si calcoli la probabilità che estraendo senza reimmissione 7 palline dall'urna ce ne siano almeno 5 rosse.

$P = 0.7001$

La variabile X numero di palline rosse nelle 7 estrazioni risulta distribuita come una variabile casuale ipergeometrica.

$$P(X \geq 5) = P(X=5) + P(X=6) + P(X=7) = \frac{\binom{36}{5} \binom{14}{2}}{\binom{50}{7}} + \frac{\binom{36}{6} \binom{14}{1}}{\binom{50}{7}} + \frac{\binom{36}{7} \binom{14}{0}}{\binom{50}{7}} = 0.3435 + 0.2730 + 0.0836 = 0.7001.$$

5.2 Si calcoli la probabilità che estraendo con reimmissione 500 palline dall'urna si ottengano almeno 127 palline bianche.

$P = 0.9015$

Indicata con X la variabile casuale "numero di successi" nelle 500 prove indipendenti risulta $X \sim \text{Bin}(n=500, p=0.28)$.
 $M(X) = 140$; $\text{Var}(X) = 100.8$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=140, \sigma^2=100.8)$ si ottiene

$$P(X > 127) = P\left(\frac{X - \mu}{\sigma} > \frac{127 - \mu}{\sigma}\right) = P\left(Z > \frac{127 - 140}{10.0399}\right) = P(Z > -1.29) = P(Z < 1.29) = 0.9015.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO
Prova scritta di STATISTICA del 8.9.2011 (Tema 265)

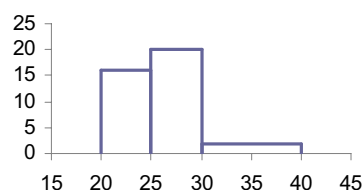
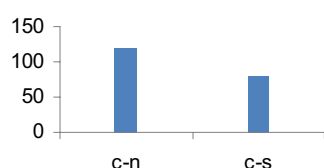
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

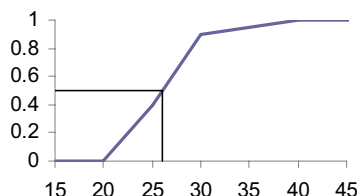
1. La seguente tabella doppia riporta le frequenze congiunte riguardanti la rilevazione del numero di sigarette fumate ogni giorno (Y) da 200 fumatori, distinti per zona geografica (X).

$X \backslash Y$	20 → 25	25 → 30	30 → 40
x_1 =centro-nord	45	60	15
x_2 =centro-sud	35	40	5

- 1.1 Si rappresentino graficamente le distribuzioni marginali X e Y .



- 1.2 Si individui, mediante una opportuna rappresentazione grafica la mediana della distribuzione marginale Y .



- 1.3 Si confrontino con opportuni indici di posizione e variabilità le distribuzioni condizionate $Y|x_i$, commentando i risultati ottenuti.

$Y|x_1$ indice di posizione: $M(Y|x_1)$ valore 26.5625 indice di variabilità: $cv(Y|x_1)$ valore 0.1483

$Y|x_2$ indice di posizione: $M(Y|x_2)$ valore 25.7813 indice di variabilità: $cv(Y|x_2)$ valore 0.1315

$Y|x_1$ presenta un livello medio e un coefficiente di variazione leggermente superiori a $Y|x_2$.

- 1.4 Si calcoli un opportuno indice di connessione, commentando il risultato ottenuto.

indice $\chi_N = 0.1082$. Tra le variabili X e Y sussiste un livello di connessione molto basso.

2. Il prospetto seguente riporta la produzione di energia elettrica in TWh per le fonti energetiche rinnovabili nel periodo 2002–2007:

t	2002	2003	2004	2005	2006	2007
rinnovabili	49.0	48.0	55.7	49.9	52.2	49.4

- 2.1 Si ricostruisca la serie dei numeri indici a base mobile.

t	2002	2003	2004	2005	2006	2007
rinnovabili	--	0.9796	1.1604	0.8959	1.0461	0.9464

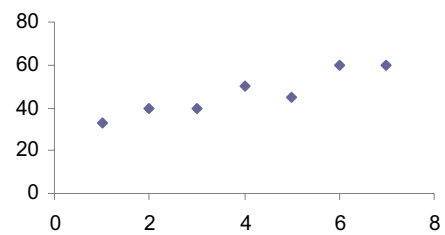
- 2.2 Si calcoli il tasso medio di crescita nel periodo considerato.

$\sqrt[5]{49.4/49} = 1.001627$, valore relativo a cui corrisponde una variazione relativa dello 0.16%. L'andamento è quasi costante

3. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 7 valori della X :

y_i	33	40	40	50	45	60	60
x_i	1	2	3	4	5	6	7

3.1 Si rappresentino graficamente le coppie di dati (x_i, y_i) .



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + b X^2$

II) $Y^* = c X$

$a = 36.0043$ $b = 0.5426$ misura di adattamento: $M(E^2) = \text{var.res.}$ valore 13.2065

$c = 10.2714$ misura di adattamento: $M(E^2)$ valore 177.6694

Il modello migliore è $Y^* = 36.0043 + 0.5426 X^2$

3.3 Si confronti con le precedenti la misura di adattamento del seguente modello di regressione, ottenuta secondo il metodo dei minimi quadrati:

$$Y^* = 30 + X^2/2$$

misura di adattamento: $M(E^2)$ valore 60.7143

Il residuo quadratico medio relativo al modello (parabola simmetrica rispetto all'origine con parametri assegnati) risulta superiore a quello del modello I) i cui parametri sono stati stimati secondo il criterio dei minimi quadrati.

4. Si completino se possibile le seguenti tabelle, motivando teoricamente la risposta, in modo tale che:

$$\eta_{X|Y}^2 = \rho^2 = 0$$

$X \backslash Y$	10	20	30
1	0	11	
2		0	

$$\eta_{Y|X}^2 = \rho^2 \neq 0$$

$X \backslash Y$	10	20	30
1		11	0
2	0	9	

Dipendenza funzionale lineare di X da Y ;
quindi $M(X|Y)$ allineate IMPOSSIBILE

$M(Y|X)$ allineate e distribuzioni $Y|x_i$ non degeneri.
Situazione realizzata per qualsiasi combinazione $n_{11} \geq 0$ e $n_{23} \geq 0$ tranne $n_{11} = n_{23} = 0$.

5. Un gioco consiste nel lanciare 1 dado equilibrato e, subito dopo, una moneta equilibrata, tante volte quante il risultato ottenuto nel dado.

5.1 Si calcoli la probabilità dell'evento $A = \{\text{numero teste} = 4\}$.

Indicati con D_i , T_n gli eventi associati a risultato i nel lancio del dado e 4 teste in n lanci della moneta si ha:

$P(A) = P(D_1 \cap T_1) + \dots + P(D_6 \cap T_6) = P(D_1)P(T_1|D_1) + \dots + P(D_6)P(T_6|D_6) = P(D_4)P(T_4|D_4) + P(D_5)P(T_5|D_5) + P(D_6)P(T_6|D_6)$ essendo $P(T_n|D_n) = 0$, $n = 1, 2, 3$. Ricordando che T_n può essere ricavato avvalendosi della v.c.

$\text{Bin}(n, 0.5)$ si ha: $P(A) = \frac{1}{6} \binom{4}{4} 0.5^4 \cdot 0.5^0 + \frac{1}{6} \binom{5}{4} 0.5^4 \cdot 0.5^1 + \frac{1}{6} \binom{6}{4} 0.5^4 \cdot 0.5^2 = 0.01042 + 0.02604 + 0.03906 = 0.07552$.

5.2 Si calcoli la probabilità che, eseguendo il gioco 100 volte, l'evento A si presenti almeno 6 volte.

Indicata con X la variabile casuale "numero volte evento A " nelle 100 prove indipendenti risulta $X \sim \text{Bin}(n=100, p=0.07552)$.

$M(X) = 7.552$; $\text{Var}(X) = 6.9817$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=7.552, \sigma^2=6.9817)$ si ottiene

$$P(X > 6) = P\left(\frac{X - \mu}{\sigma} > \frac{6 - \mu}{\sigma}\right) = P\left(Z > \frac{6 - 7.552}{2.6423}\right) = P(Z > -0.59) = P(Z < 0.59) = 0.7224.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 22.9.2011 (Tema 265_1)

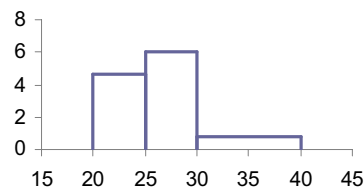
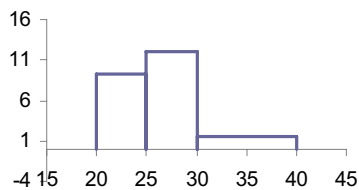
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. La seguente tabella doppia riporta le frequenze congiunte riguardanti la rilevazione del numero di sigarette fumate ogni giorno (Y) da 183 fumatori, distinti per zona geografica (X).

$X \backslash Y$	20 - 25	25 - 30	30 - 40
x_1 =centro-nord	46	60	16
x_2 =centro-sud	23	30	8

- 1.1 Si rappresentino graficamente le distribuzioni condizionate $Y|x_i$.



- 1.2 Si riassumano mediante opportuni indici di posizione e variabilità le distribuzioni marginali.

X indice di posizione: Moda valore centro-nord indice di variabilità: Indice di Frosini valore 0.6665
 Y indice di posizione: $M(Y)$ valore 26.5984 indice di variabilità: c.v. valore 0.1503

- 1.4 Si misuri il livello di dipendenza in media di Y da X .

$\eta^2_{Y|X} = 0$; le distribuzioni condizionate sono somiglianti, sussiste indipendenza stocastica, quindi anche indep. in media.

2. Il prospetto seguente riporta la serie dei numeri indice a base mobile per una determinata grandezza macroeconomica X :

t	${}_{t-1}I_t$	${}_{t-1}X_t$
2002	--	77
2003	0.8545	65.80
2004	0.7173	47.20
2005	0.7585	35.80
2006	0.9441	33.80
2007	0.6775	22.90

- 2.1 Sapendo che il valore della grandezza X nel 2002 era 77 si ricostruiscano i valori della serie storica.

- 2.2 Si calcoli il tasso medio di variazione della serie tra il 2003 e il 2006.

$\sqrt[3]{33.8/65.8} = \sqrt[3]{0.7173 \cdot 0.7585 \cdot 0.9441} = 0.8010$, valore relativo medio, cui corrisponde una variazione relativa media del -19.9%.

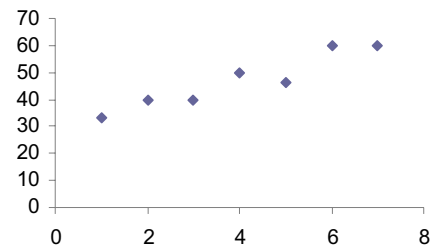
- 2.3 Si indichi il valore mediano delle variazioni relative subite dalla grandezza X negli anni considerati.

Occorre considerare l'elemento che occupa la terza posizione tra i numeri indici a base mobile ordinati: 0.7585, la corrispondente variazione relativa risulta -24.15%.

3. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 7 valori della X :

y_i	33	40	40	50	46	60	60
x_i	1	2	3	4	5	6	7

3.1 Si rappresentino graficamente le coppie di dati (x_i, y_i) .



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$ II) $Y^* = c$;

$a = 28.8571$ $b = 4.5357$ misura di adattamento: $\rho^2 = 0.8973$

$c = 47$ misura di adattamento: $\rho^2 = 0$

Il modello migliore è $Y^* = 28.8571 + 4.5357X$

4. Si completino se possibile le seguenti tabelle, motivando teoricamente la risposta, in modo tale che:

$\eta_{X|Y}^2 = \rho^2 = 0$

$X \backslash Y$	10	20	30
1	0	12	0
2	$a > 0$	0	$a > 0$
4	0	6	0

$\eta_{Y|X}^2 = \rho^2 \neq 0$

$X \backslash Y$	10	20	30
1	$a > 0$	12	0
2	0	9	0
3	0	12	$a > 0$

Indipendenza in media di X da Y .

$\eta_{X|Y}^2 = 0 \rightarrow \rho^2 = 0$

$M(X|Y_1) = M(X|Y_3) = 2 = M(X|Y_2) = \frac{1 \cdot 12 + 4b}{12+b} \rightarrow b = 6$

Medie condizionate $M(Y|x_i)$ allineate.

No indipendenza in media.

Si riporta una possibile soluzione.

5. Un'urna contiene 20 palline bianche e 20 palline rosse e 10 nere.

5.1 Si calcoli la probabilità che estraendo senza reimmissione dall'urna 4 palline se ne ottengano almeno 3 rosse.

$P = 0.1695$

$P(\text{almeno } 3R) = P(3R) + P(4R) = \binom{4}{3} \frac{20}{50} \frac{19}{49} \frac{18}{48} \frac{30}{47} + \binom{4}{4} \frac{20}{50} \frac{19}{49} \frac{18}{48} \frac{17}{47}$

o mediante la v.c. ipergeometrica $P(X=3) + P(X=4) = \frac{\binom{20}{3} \cdot \binom{30}{1}}{\binom{50}{4}} + \frac{\binom{20}{4} \cdot \binom{30}{0}}{\binom{50}{4}}$.

5.2 Si calcoli la probabilità che estraendo con reimmissione dall'urna 4 palline se ne ottengano 1 bianca, 2 rosse e 1 nera.

$P = 0.1536$

n. sequenze $\cdot P(\text{generica sequenza}) = \frac{4!}{1! \cdot 2! \cdot 1!} \cdot P(B_1 \cap R_2 \cap R_3 \cap N_4) = \frac{4!}{1! \cdot 2! \cdot 1!} \cdot \frac{20}{50} \frac{20}{50} \frac{20}{50} \frac{10}{50}$.

5.2 Si calcoli la probabilità che, estraendo con reimmissione 100 palline se ne ottengano almeno 40 bianche.

$P = 0.5$

$X \sim \text{Bin}(n=100, p=0.4)$; $M(X) = 40$; $\text{Var}(X) = 24$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=40, \sigma^2=24)$ si ottiene $P(X \geq 40) = P(X \geq \mu) = 0.5$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 12.01.2012 (Tema 266)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Il seguente prospetto riporta la rilevazione del carattere X con riferimento a 10 unità statistiche appartenenti ai gruppi A e B .

Gruppo A	x_{iA}	4	5	6	7	8
Gruppo B	x_{iB}	6	7	8	9	10

- 1.1 Si calcolino per i due gruppi media aritmetica e varianza, confrontando i risultati ottenuti.

Gruppo A $\mu_A = 6$ $\sigma_A^2 = 2$

Gruppo B $\mu_B = 8$ $\sigma_B^2 = 2$

Il livello medio di B è superiore a quello di A e meno variabile ($c.v.(B) = 0.1768$, $c.v.(A) = 0.2357$)

Distribuzioni simmetriche, inoltre $x_{iB} = x_{iA} + 2$ (stessa varianza).

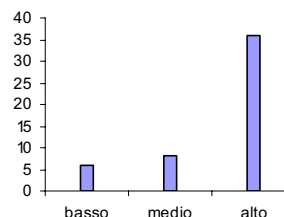
- 1.2 Considerando le 10 unità statistiche come gruppo unico, si calcolino media e varianza (globali) di X in funzione degli indici parziali calcolati al punto precedente, riportando anche le relazioni teoriche esistenti.

$M(X)$ relazione teorica $\mu = \frac{\mu_A n_A + \mu_B n_B}{n_A + n_B} = 7$

$Var(X)$ relazione teorica $\sigma^2 = \sigma_{Between}^2 + \sigma_{Within}^2 = 1 + 2 = 3$

2. Si dia una rappresentazione grafica della seguente mutabile statistica Y e la si riassume mediante opportuni indici di posizione e di mutabilità.

y_i	n_i
<i>basso</i>	6
<i>medio</i>	8
<i>alto</i>	36



Indice di posizione: Moda/Mediana valore *alto*

Indice di mutabilità: Indice di Frosini valore 0.419

3. Si riporta l'evoluzione temporale del prezzo di un prodotto tra il 2001 il 2005.

	p_t	${}_{t-1}I_t$
2001	330	--
2002	342	1.0364
2003	390	1.1404
2004	405	1.0385
2005	420	1.0370

- 3.1 Si costruisca la serie dei numeri indici a base mobile per i prezzi del prodotto.

- 3.2 Si interpreti il valore assunto dal numero indice ${}_{2002}I_{2003}$.

${}_{2002}I_{2003} = 1.1404$. Il prezzo del prodotto ha subito un incremento del 14.04% tra il 2002 e il 2003

- 3.3 Si calcoli il tasso medio di variazione dei prezzi tra il 2002 e il 2005.

$\sqrt[3]{\frac{{}_{2002}I_{2005}}{342}} = \sqrt[3]{\frac{420}{342}} = \sqrt[3]{1.2281} = 1.0709$, valore relativo medio, a cui corrisponde una variazione relativa media del 7.09%.

4. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 5 valori della X :

x_i	8	8	10	12	12
y_i	4	6	y_3	8	10

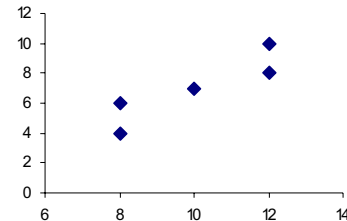
4.1 Si indichi un valore da assegnare a y_3 affinché $\rho^2 < \eta_{Y|X}^2$ e $\eta_{X|Y}^2 = 1$.

y_3 qualsiasi valore tranne 4, 6, 7, 8, 10

$M(Y|8) = 5$, $M(Y|12) = 9$; affinché $\rho^2 < \eta_{Y|X}^2$ le medie condizionate non devono essere allineate, quindi $M(Y|10) = y_3 \neq 7$. Inoltre affinché $\eta_{X|Y}^2 = 1$ a ogni valore distinto y_i deve corrispondere una sola x_i , quindi $y_3 \notin \{4, 6, 8, 10\}$

4.2 Si assuma, nel seguito, $y_3 = 7$.

Si rappresentino graficamente le 5 coppie (x_i, y_i) e la funzione di regressione



4.3 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e gli indici di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + bX$$

$$\text{II) } Y^* = cX^2$$

$$a = -3$$

$$b = 1$$

$$\text{misura di adattamento: } M(E^2) = 0.8$$

$$c = 0.0659$$

$$\text{misura di adattamento: } M(E^2) = 1.1744$$

Il modello migliore è $Y^* = -3 + X$

4.4 Senza effettuare calcoli e motivando teoricamente la risposta, si indichi il valore assunto dall'indice di adattamento della funzione di regressione.

$\eta_{Y|X}^2 = \rho^2 = 0.8$ in quanto le medie condizionate $M(Y|8) = 5$, $M(Y|10) = 7$, $M(Y|12) = 9$ sono allineate.

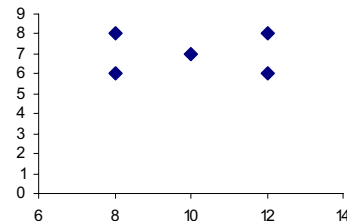
4.5 Si indichino i valori che dovrebbero assumere y_1 e y_5 affinché risulti $\eta_{Y|X}^2 = 0$.

$$y_1 = 8$$

$$y_5 = 6$$

In questo caso le medie condizionate coincidono:

$$M(Y|8) = M(Y|10) = M(Y|12) = 7$$



5. Un'urna contiene 2 palline bianche, 1 pallina rossa e 2 nere.

5.1 Si calcoli la probabilità che estraendo dall'urna con reimmissione 90 palline se ne ottengano almeno 39 bianche.

$X \sim \text{Bin}(n=90, p=0.4)$; $M(X) = 36$; $\text{Var}(X) = 21.6$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=36, \sigma^2=21.6)$ si ottiene:

$$P(X \geq 39) = P\left(\frac{X - \mu}{\sigma} \geq \frac{39 - \mu}{\sigma}\right) = P\left(Z \geq \frac{39 - 36}{\sqrt{21.6}}\right) = P(Z \geq 0.65) = 1 - P(Z \leq 0.65) = 1 - 0.7422 = 0.2578.$$

5.2 Si calcoli la probabilità che estraendo senza reimmissione dall'urna 4 palline, di queste una sola sia nera.

$$P = 0.4$$

Può essere, quindi, indifferentemente risolto come $\binom{4}{1} \cdot P(N_1 \cap \bar{N}_2 \cap \bar{N}_3 \cap \bar{N}_4) = 4 \cdot \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2}$ o mediante la v.c.

$$\text{ipergeometrica } P(X=1) = \frac{\binom{2}{1} \cdot \binom{3}{3}}{\binom{5}{4}}.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 26.01.2012 (Tema 267)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

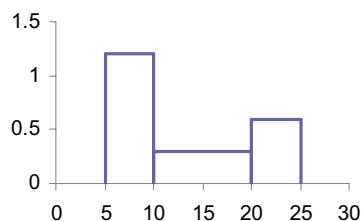
1. Date le seguenti 12 osservazioni del carattere X se ne calcoli la mediana.

x_i	25	8	14	9	15	21	20	24	9	10	8	7
-------	----	---	----	---	----	----	----	----	---	----	---	---

1.1 Mediana = $(10+14)/2=12$

1.2 Si costruisca la distribuzione di frequenze con classi $5 \rightarrow 10$, $10 \rightarrow 20$ e $20 \rightarrow 25$ e la si rappresenti graficamente.

$h_{i-1} \rightarrow h_i$	n_i	d_i
$5 \rightarrow 10$	6	1.2
$10 \rightarrow 20$	3	0.3
$20 \rightarrow 25$	3	0.6



1.3 Si calcolino media aritmetica e coefficiente di variazione dei dati riclassificati.

Media aritmetica 13.125

Coefficiente di variazione 0.4738

2. Per studiare la distribuzione per genere (X) dei clienti di un determinato servizio A si sono estratti 3 campioni di pari numerosità dalle 3 zone Nord, Centro, Sud-Isole (carattere Y), ottenendo la seguente tabella di frequenze condizionate relative:

$X Y$	Nord	Centro	Sud-Isole
m	0.8	0.5	0.4
f	0.2	0.5	0.6
	1	1	1

2.1 Sapendo che il totale dei selezionati è pari ad $n=180$, si ricostruisca la tabella a doppia entrata (X,Y) delle frequenze congiunte assolute.

$X \backslash Y$	Nord	Centro	Sud-Isole	
M	48	30	24	102
F	12	30	36	78
	60	60	60	180

2.2 Si calcoli un opportuno indice di connessione tra X e Y , commentando il risultato ottenuto.

Indice $\chi_N = 0.3429$

Tra le variabili sussiste un livello basso di connessione.

3. Con riferimento alla seguente serie storica W_t

anno	W_t	NIBF (2008=100)	NIBM
2008	800	100.00	--
2009	850	106.25	1.0625
2010	900	112.50	1.0588
2011	1000	125.00	1.1111

3.1 Si costruisca la serie dei numeri indici a base fissa, con base 2008=100.

3.2 Si interpreti l'ultimo valore della serie sopra calcolata.

${}_{2008}I_{2011} = 1.25$ Tra il 2008 e il 2011 la serie ha subito un incremento percentuale del 25%.

3.3 Si costruisca la serie dei numeri indici a base mobile.

3.4 Si calcoli quale dovrebbe essere il valore W_{2012} in modo che il tasso medio di crescita nel periodo 2008-2012 risulti pari al 6%.

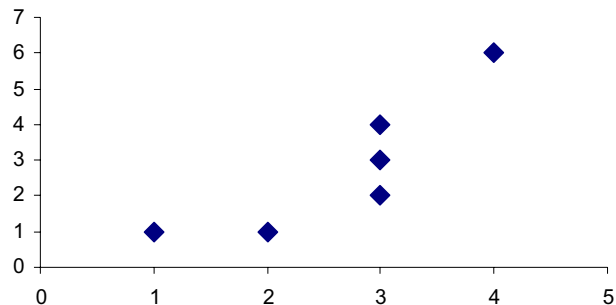
A un tasso medio di crescita (variazione relativa media) del 6% corrisponde un valore relativo medio pari 1.06, quindi

$$W_{2012} = 800 \cdot 1.06^4, \quad \text{o, anche, in maniera formale: } 1.06 = \sqrt[4]{{}_{2008}I_{2012}} = \sqrt[4]{\frac{W_{2012}}{W_{2008}}} = \sqrt[4]{\frac{W_{2012}}{800}}.$$

4. Date le seguenti coppie di dati

x_i	1	2	3	3	3	4
y_i	1	1	2	3	4	6

4.1 Si rappresentino graficamente le coppie (x_i, y_i)



4.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$

II) $Y^* = \beta X$

III) $Y^* = \alpha$

$a = -1.5$ $b = 1.625$ misura di adattamento: $M(E^2) = 0.7917$

$\beta = 1.125$ misura di adattamento: $M(E^2) = 1.0417$

$\alpha = 2.833$ misura di adattamento: $M(E^2) = 3.1389$

Il modello migliore è $Y^* = -1.5 + 1.625X$

4.3 Si calcoli il valore assunto dall'indice di adattamento della funzione di regressione.

$\eta_{Y|X}^2 = 0.8938$

5. Se possibile e giustificando teoricamente le risposte, si completino le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

$X \backslash Y$	10	27.143	30
5	1	0	6
10	0	5	0

$\eta_{X|Y}^2 = 1$ $\eta_{Y|X}^2 = 0$

Dipendenza funzionale di X da Y e indipendenza in media di Y da X .

Distribuzioni $X|y_j$ degeneri e $M(Y|x_1) = M(Y|x_2) = y_2$

$X \backslash Y$	10	10 < y_2 < 30	30
5	1	5	6
10	1	5	6

$\eta_{X|Y}^2 = 0$ $\chi^2 = 0$

$\chi^2 = 0$ indep. stocastica $\rightarrow \eta_{X|Y}^2 = 0$
Distribuzioni condizionate somiglianti

$X \backslash Y$	10		30
5	1		6
10		5	

$\eta_{X|Y}^2 = 1$ $\eta_{Y|X}^2 = 1$

Dipendenza funzionale di X da Y e contemporaneamente di Y da X .
Impossibile: la distribuzione condizionata $Y|X_1$ non è degenera

6. Due dadi vengono truccati, in modo tale che in uno compaiano solo i numeri pari e nell'altro solo dispari.

6.1 Si calcoli la probabilità che, in un generico lancio, la somma dei risultati sia ≥ 8 (evento A).

$P = \frac{3}{9} = 0.3333$

Ω	5	•	•	•
3	•	•	•	•
1	•	•	•	•
2	4	6		

6.2 Si calcoli la probabilità condizionata che, essendo uscito 5 in uno dei due dadi, nell'altro si presentino 4.

L'esito del secondo dado è indipendente da quello del primo, quindi $P = 1/3$.

Oppure $P(D_1=5|D_2=4) + P(D_2=5|D_1=4) = 0 + \frac{P(D_2=5 \cap D_1=4)}{P(D_1=4)} + 0 = \frac{1/9}{3/9} = \frac{1}{3}$

6.3 Si calcoli la probabilità che su 90 lanci l'evento A (somma ≥ 8) si presenti almeno 24 volte.

$X \sim \text{Bin}(n=90, p=1/3)$; $M(X) = 30$; $\text{Var}(X) = 20$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=30, \sigma^2=20)$ si ottiene:

$P(X \geq 24) = P\left(\frac{X - \mu}{\sigma} \geq \frac{24 - \mu}{\sigma}\right) = P\left(Z \geq \frac{24 - 30}{\sqrt{20}}\right) = P(Z \geq -1.34) = P(Z \leq 1.34) = 0.9099.$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 9.02.2012 (Tema 268)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Sia Y il consumo di gas per il riscaldamento (rapportato alla volumetria condominiale) rilevato in 10 condomini di diversa classe energetica X , con modalità (a, b) .

x_i	a	a	b	b	a	b	b	b	b	a
y_i	15	14	29	20	19	18	20	21	20	23

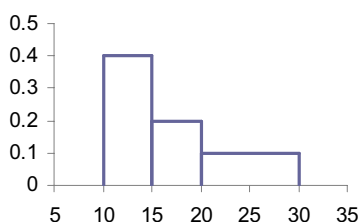
- 1.1 Si costruisca la distribuzione delle frequenze congiunte, suddividendo i valori di Y nelle seguenti classi $10 \rightarrow 15$, $15 \rightarrow 20$ e $20 \rightarrow 30$.

$X \backslash Y$	10 → 15	15 → 20	20 → 30	
a	2	1	1	4
b	0	4	2	6
	2	5	3	10

- 1.2 Senza effettuare calcoli e motivando la risposta, dire se può esistere indipendenza stocastica tra X e Y .

Non sussiste indipendenza stocastica in quanto le distribuzioni condizionate non sono somiglianti (è presente una frequenza congiunta nulla).

- 1.3 Si rappresenti graficamente la distribuzione condizionata $Y|X=a$. Se ne calcoli, poi, la media armonica.



media armonica = 15.5556

- 1.4 Si calcoli un opportuno indice di variabilità per la marginale X .

Indice di Frosini = 0.8. Sussiste un livello elevato di eterogeneità.

- 1.5 Si calcoli il coefficiente di variazione della variabile $W = 4Y$ in funzione del coefficiente di variazione di Y , esplicitandone la relazione.

$$c.v.(Y) = 0.2404 \quad c.v.(W) = \frac{\sigma_W}{\mu_W} = \frac{4\sigma_Y}{4\mu_Y} = \frac{\sigma_Y}{\mu_Y} = 0.2404$$

- 1.6 Si calcoli un opportuno indice di connessione tra X e Y commentando il risultato ottenuto.

Indice $\chi_N = 0.6236$

Tra le variabili sussiste un livello medio di connessione.

2. Si consideri la seguente serie dei numeri indice a base fissa riguardante l'andamento della serie Z_t .

anno	$NIBF$	$NIBM$
2008	1	--
2009	1.1	1.1
2010	0.9	0.8182
2011	1.2	1.3333

- 2.1 Si costruisca la serie dei numeri indici a base mobile.

- 2.2 Si interpreti l'ultimo valore della serie sopra calcolata.

${}_{2010}I_{2011} = 1.3333$ Tra il 2010 e il 2011 la serie ha subito un incremento percentuale del 33.33%.

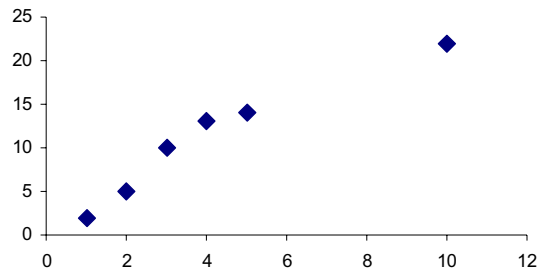
- 2.3 Si calcoli il valore Z_{2011} sapendo che $Z_{2009} = 47$.

$$Z_{2011} = Z_{2009} \cdot {}_{2009}I_{2010} \cdot {}_{2010}I_{2011} = 47 \cdot 0.8182 \cdot 1.3333 = 51.2727.$$

3. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 6 valori della X :

x_i	1	2	3	4	5	10
y_i	2	5	10	13	14	22

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

I) $Y^* = aX$ II) $Y^* = b + c/X$

$a = 2.4774$ misura di adattamento: formula $M(E^2) = 4.4452$

$b = 18.6394$ $c = -19.2321$ misura di adattamento: formula $M(E^2) = 9.5192$

Il modello migliore è $Y^* = 2.4774 X$

3.3 Si calcolino i valori assunto dai rapporti di correlazione $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

$\eta_{Y|X}^2 = 1$. Sussiste dipendenza funzionale di Y da X .

$\eta_{X|Y}^2 = 1$. Sussiste dipendenza funzionale di X da Y .

4. Se possibile e giustificando teoricamente le risposte, si completino le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

$X \backslash Y$	3	6	9
2	3	c	b ≠ 3
4	a	3	a
6	b ≠ 3	c	3

$$0 < \rho^2 = \eta_{X|Y}^2 < 1$$

Medie condizionate allineate. No dipendenza funzionale. No indipendenza in media.

Si riporta una possibile soluzione.

$X \backslash Y$	20	$20 < y_2 < 40$	40
10	3	0	0
$10 < x_2 < 20$	0	0	11
20	0	1	0

$$\chi^2 = 30 \text{ e } \eta_{Y|X}^2 = 1$$

$$\eta_{Y|X}^2 = 1 \text{ dipendenza funzionale} \rightarrow \chi_N^2 = 1.$$

$$\chi^2 = 30 = \chi_{\max}^2 = n \cdot \min(3-1, 3-1) \rightarrow n = 15$$

5. Gli eventi A, B, C e D costituiscono una partizione dello spazio campionario, inoltre:

$$P(A) = 0.4, P(B) \text{ è 4 volte } P(D) \text{ e } P(\Omega \cap C) = 0.2.$$

5.1 Si calcolino $P(D) = 0.08$ $P(D|A) = 0$ $P(B-C) = P(B) = 0.32$

$$A \cup B \cup C \cup D \equiv \Omega \quad \text{e} \quad A, B, C, D \text{ non hanno elementi in comune.}$$

$$P(\Omega \cap C) = P(C) = 0.2, \quad 1 = P(\Omega) = P(A) + P(B) + P(C) + P(D) = 0.4 + 4P(D) + 0.2 + P(D) \rightarrow P(D) = 0.08$$

5.2 Si stabilisca se A e D sono tra loro indipendenti.

$$P(A \cap D) = 0 \neq P(A) \cdot P(D), \text{ quindi gli eventi non sono indipendenti.}$$

5.3 Si consideri un'urna contenente $N = 35$ palline, di cui alcune bianche e le altre nere. Si calcoli la probabilità che su 4 estrazioni senza ripetizione si ottenga almeno 1 pallina nera, sapendo che la probabilità di estrarre una pallina nera al primo tentativo è pari a $P(C)$.

$$P = 0.6090$$

L'urna è composta da 7 palline nere ($35 \cdot 0.2$) e 28 palline bianche.

$$P(\text{almeno 1 pallina nera}) = 1 - P(0 \text{ palline nere}) = 1 - P(B_1 \cap B_2 \cap B_3 \cap B_4) = P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 \cap B_2) \cdot$$

$$P(B_4|B_1 \cap B_2 \cap B_3) = 1 - \frac{28}{35} \cdot \frac{27}{34} \cdot \frac{26}{33} \cdot \frac{25}{32} = 1 - 0.3910 = 0.6090.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 7.06.2012 (Tema 269_1)

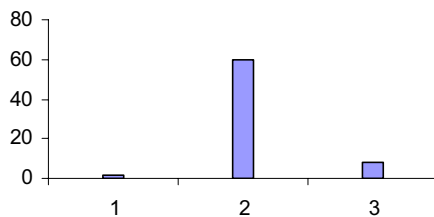
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la distribuzione congiunta delle variabili X e Y .

$X \setminus Y$	y_1	y_2
1	2	73
2	60	20
3	8	37

1.1 Si rappresenti graficamente la distribuzione condizionata $X|Y=y_1$ e si calcoli la mediana della marginale X .



mediana = 2

1.2 Senza effettuare calcoli e motivando la risposta, dire se può esistere indipendenza stocastica tra X e Y .

Non sussiste indipendenza stocastica in quanto le distribuzioni condizionate non sono somiglianti.

1.3 Si calcoli un opportuno indice di connessione tra X e Y commentando il risultato ottenuto.

Indice $\chi_N = 0.695$

Tra le variabili sussiste un livello medio/alto di connessione.

1.4 Si riassumano con opportuni indici di posizione e di variabilità le distribuzioni condizionate $X|Y=y_i$, $i = 1, 2$.

$X|Y=y_1$ indice di posizione: $M(X|Y=y_1)$ valore 2.0857 indice di variabilità: $cv(X|Y=y_1)$ valore 0.1765

$X|Y=y_2$ indice di posizione: $M(X|Y=y_2)$ valore 1.7231 indice di variabilità: $cv(X|Y=y_2)$ valore 0.5091

2. Si consideri la seguente serie dei numeri indice a base fissa (2008) riguardante l'andamento della serie Z_t .

anno	$NIBF(2008)$	$NIBF(2010)$
2008	1	1.1111
2009	1.1	1.2222
2010	0.9	1.0000
2011	1.2	1.3333

2.1 Si costruisca la serie dei numeri indici con base fissa 2010.

2.2 Si interpreti il primo valore della serie sopra calcolata.

${}_{2010}I_{2008} = 1.1111$. Nel 2008 il valore della serie Z era superiore dell'11.11% rispetto al 2010.

2.3 Si calcoli il tasso medio di variazione tra il 2008 e il 2011.

$\sqrt[3]{{}_{2008}I_{2011}} = \sqrt[3]{1.2} = 1.0627$ valore relativo medio, a cui corrisponde una variazione relativa media del 6.27%.

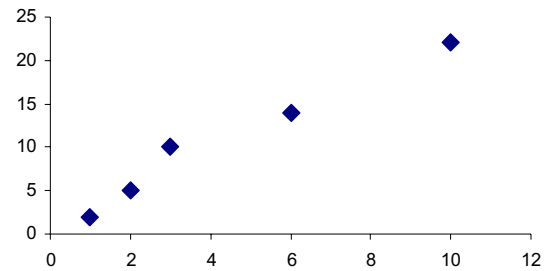
2.4 Si calcoli il valore Z_{2007} sapendo che $Z_{2008} = 47$.

Z_{2007} = impossibile calcolarlo, occorrerebbe conoscere ${}_{2007}I_{2008}$.

3. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 5 valori della X :

x_i	1	2	3	6	10
y_i	2	5	10	14	22

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$

II) $Y^* = c$

$a = 1.2707$

$b = 2.1203$

misura di adattamento: formula ρ^2

valore 0.9675

$c = 10.6$

misura di adattamento: formula ρ^2

valore 0

Il modello migliore è $Y^* = 1.2707 + 2.1203X$

3.3 Si fornisca, in base al modello migliore, la previsione in corrispondenza di $X = 8$.

$Y^*(8) = 1.2707 + 2.1203 \cdot 8 = 18.2331$.

4. Se possibile e giustificando teoricamente le risposte, si completino le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

$X \backslash Y$	3	6	9
2	3	10	3
4	0	3	0
6	3	1	3

$0 < \rho^2 < \eta_{X|Y}^2 < 1$

Non sussistono né indipendenza in media né dipendenza funzionale e le medie condizionate non sono allineate.

Si riporta una possibile soluzione.

$X \backslash Y$	20		40
10	3		
20		1	

$\rho = -0.30$ e $\eta_{Y|X}^2 = 0$

Impossibile: $\eta_{Y|X}^2 = 0 \rightarrow \rho^2 = 0$.

5. Si consideri l'esperimento di estrazione con reimmissione di 10 palline da un'urna contenente 10 palline rosse, 23 palline bianche e 17 palline nere.

5.1 Si definisca la natura della variabile casuale X = "numero di palline rosse estratte".

X variabile casuale binomiale. $X \sim \text{Bin}(n=10, p=0.2)$ $P(X=x) = \binom{10}{x} 0.2^x 0.8^{10-x}$, $x = 0, 1, \dots, 10$.

5.2 Si calcoli la probabilità di ottenere almeno 8 palline bianche nelle 10 estrazioni.

$Y \sim \text{Bin}(n=10, p=0.46)$ $P(Y=y) = \binom{10}{y} 0.46^y 0.54^{10-y}$, $y = 0, 1, \dots, 10$.

$P(Y \geq 8) = P(Y=8) + P(Y=9) + P(Y=10) = 0.02631 + 0.004980 + 0.0004242 = 0.03171$.

5.3 Si consideri l'esperimento di estrazione con reimmissione di 100 palline dalla medesima urna; si calcoli la probabilità di ottenere almeno 23 palline rosse.

$P = 0.2266$

$X \sim \text{Bin}(n=100, p=0.2)$; $M(X) = 20$; $\text{Var}(X) = 16$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=20, \sigma^2=16)$ si ottiene:

$P(X \geq 23) = P\left(\frac{X-\mu}{\sigma} \geq \frac{23-\mu}{\sigma}\right) = P\left(Z \geq \frac{23-20}{4}\right) = P(Z \geq 0.75) = 1 - P(Z \leq 0.75) = 1 - 0.7734 = 0.2266$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 28.06.2012 (Tema 269_2)

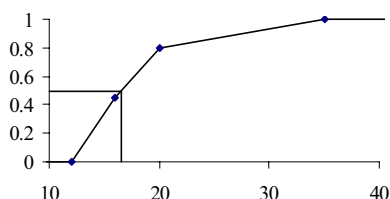
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. E' stata svolta un'indagine sull'uso internet da parte dei giovani. Nella tabella che segue è riportata la distribuzione congiunta dei caratteri uso dei social network (X) ed Età (Y):

$X \backslash Y$	12 - 16	16 - 20	20 - 35
Facebook	35	13	20
Twitter	10	22	0

- 1.1 Si rappresenti graficamente la funzione di ripartizione della marginale Y , si individuino sulla stessa l'età modale e l'età mediana. Si calcoli l'età media del campione.



età media = 18.1

la classe modale è la prima in quanto caratterizzata dalla maggiore pendenza della f. di ripartizione

- 1.2 Si confronti la variabilità delle distribuzioni condizionate $X|Y=y_1$ e $X|Y=y_3$.

$X|Y=y_1$ indice di variabilità: Indice di Frosini valore 0.4445

$X|Y=y_3$ indice di variabilità: Indice di Frosini valore 0

- 1.3 Senza effettuare calcoli e motivando la risposta, dire se può esistere indipendenza stocastica tra X e Y .

Non sussiste indipendenza stocastica in quanto le distribuzioni condizionate non sono somiglianti (è presente una frequenza congiunta nulla).

- 1.4 Si calcoli un opportuno indice di connessione tra X e Y commentando il risultato ottenuto.

Indice $\chi_N = 0.5167$

Tra le variabili sussiste un livello medio di connessione.

2. Si consideri la serie storica delle quotazioni di un generico titolo Q_t .

Anno	Q_t	$NIBM$
1	114	--
2	86	0.7544
3	101	1.1744
4	113	1.1188
5	86	0.7611

- 2.1 Si costruisca la serie dei numeri indici con base mobile.

- 2.2 Si calcoli e interpreti il valore del numero indice al tempo 5 con base 1.

${}_1I_5 = 0.7544$ Tra l'anno 1 e l'anno 5 la quotazione del titolo ha subito un calo del 24.56%.

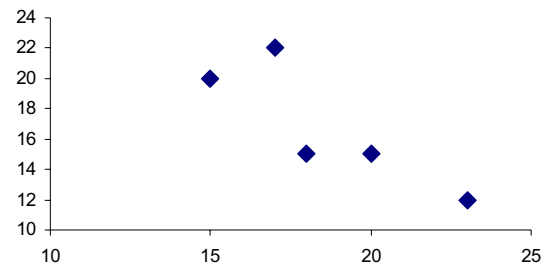
- 2.3 Si calcoli il tasso medio di variazione per l'intero periodo considerato.

$\sqrt[4]{{}_1I_5} = 0.9320$ valore relativo medio, a cui corrisponde una variazione relativa media del -6.8%.

3. Si desidera valutare la relazione esistente fra il livello di un certo parametro fisico (Y) ed il numero di ore mensili di attività fisica (X). Nella seguente tabella vengono riportati i dati rilevati su $n=5$ atleti.

x_i	15	17	18	20	23
y_i	20	22	15	15	12

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Indicare, senza calcolarlo e giustificando teoricamente la risposta, se l'indice ρ assume un valore negativo, positivo o nullo.

Il coefficiente di correlazione lineare ρ è in questo caso negativo (tra le variabili sussiste una relazione inversa).

3.3 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + bX$$

$$\text{II) } Y^* = cX$$

$$a = 38$$

$$b = -1.1398$$

misura di adattamento: formula $M(E^2)$

$$\text{valore } 3.6946$$

$$c = 0.8602$$

misura di adattamento: formula $M(E^2)$

$$\text{valore } 34.0946$$

Il modello migliore è $Y^* = 38 + 0.8602 X$

4. Dati due caratteri Z e W , indicare che valore assume l'indice $\eta_{Z|W}^2$ nelle seguenti ipotesi:

4.1 l'indice di connessione χ^2 assume valore zero;

$\eta_{Z|W}^2 = 0$. Se $\chi^2 = 0$ sussiste indipendenza stocastica che implica l'indipendenza in media.

4.2 la media delle varianze condizionate è pari a 3 volte la varianza delle medie condizionate;

$$\eta_{Z|W}^2 = \frac{\text{Var}_W[M(Z|W)]}{\text{Var}(Z)} = \frac{\text{Var}_W[M(Z|W)]}{\text{Var}_W[M(Z|W)] + M_W[\text{Var}(Z|W)]} = \frac{\text{Var}_W[M(Z|W)]}{\text{Var}_W[M(Z|W)] + 3\text{Var}_W[M(Z|W)]} = \frac{1}{4} = 0.25.$$

4.3 la distribuzione delle medie condizionate e delle varianze condizionate è la seguente

	w_1	w_2	w_3
$M(Z W)$	2	4	6
$\sigma_{Z W}^2$	7	7	7
frequenze marginali di W	20	10	20

$$\eta_{Z|W}^2 = \frac{\text{Var}_W[M(Z|W)]}{\text{Var}(Z)} = \frac{\text{Var}_W[M(Z|W)]}{\text{Var}_W[M(Z|W)] + M_W[\text{Var}(Z|W)]} = \frac{3.2}{3.2 + 7} = 0.3137.$$

5. In un dado perfettamente bilanciato le sei facce sono colorate nel seguente modo: 2 blu e 4 rosse.

5.1 Il dado viene lanciato due volte: un giocatore perde 3 euro se i due lanci danno luogo a un colore diverso e vince 5 euro se danno luogo allo stesso colore. Si fornisca la distribuzione della variabile casuale X che descrive la vincita del giocatore e si calcoli la vincita media attesa.

x	$P(X=x)$
-3	0.4444
5	0.5556

$$P(X = -3) = P[(B_1 \cap N_2) \cup (N_1 \cap B_2)] = P(B_1 \cap N_2) + P(N_1 \cap B_2) = \frac{2}{6} \cdot \frac{4}{6} + \frac{4}{6} \cdot \frac{2}{6} = 0.4444$$

$$P(X = 5) = P[(B_1 \cap B_2) \cup (N_1 \cap N_2)] = P(B_1 \cap B_2) + P(N_1 \cap N_2) = \frac{2}{6} \cdot \frac{2}{6} + \frac{4}{6} \cdot \frac{4}{6} = 0.5556$$

La vincita media attesa risulta 1.4444, pertanto il gioco non è equo.

5.2 Si supponga di lanciare 70 volte il dado sopra descritto; calcolare la probabilità di ottenere almeno 47 volte il colore rosso.

$X \sim \text{Bin}(n=70, p=0.6667)$; $M(X) = 46.6667$; $\text{Var}(X) = 15.5556$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=46.6667, \sigma^2=15.5556)$ si ottiene:

$$P(X \geq 47) = P\left(\frac{X - \mu}{\sigma} \geq \frac{47 - \mu}{\sigma}\right) = P\left(Z \geq \frac{47 - 46.6667}{3.9441}\right) = P(Z \geq 0.0845) = 1 - P(Z \leq 0.0845) = 1 - 0.5337 = 0.4663.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 12.07.2012 (Tema 269_3)

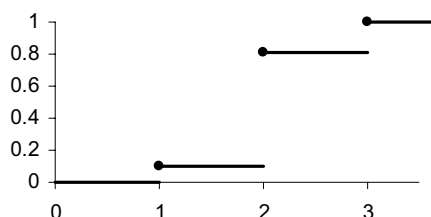
Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la distribuzione congiunta delle variabili X e Y .

$X \setminus Y$	y_1	y_2
1	7	68
2	50	20
3	13	42

1.1 Si rappresenti graficamente la Funzione di Ripartizione della distribuzione condizionata $X|Y=y_1$ e la si riassume con opportuni indici di posizione e di variabilità.



Indice di posizione: $M(X|Y=y_1)$ valore 2.0857

Indice di variabilità: $M(X|Y=y_1)$ valore 0.2530

1.2 Si riassume con opportuni indici di posizione e di mutabilità la distribuzione marginale Y .

Y indice di posizione: Moda valore y_2 indice di variabilità: Indice di Frosini valore 0.7

1.3 Si calcoli il rapporto di correlazione η^2_{XY} commentando il risultato ottenuto.

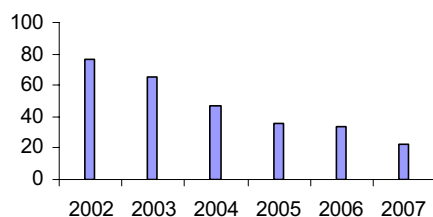
$\eta^2_{XY} = 0.029$. Il livello di dipendenza di X da Y è quasi nullo.

2. Il prospetto seguente riporta la serie dei numeri indice a base mobile, NIBM, per una determinata grandezza macroeconomica X (serie storica di flusso)

t	${}_{t-1}I_t$	X_t
2002	--	77.0087
2003	0.8545	65.8039
2004	0.7173	47.2012
2005	0.7585	35.8021
2006	0.9441	33.8007
2007	0.6775	22.90

2.1 Sapendo che il valore della grandezza X nel 2007 era 22.90 si ricostruiscano i valori della serie storica.

2.2 Si rappresenti graficamente la serie storica X .



2.3 Si calcoli il tasso medio di variazione della serie tra il 2003 e il 2006.

$\sqrt[3]{\frac{I_{2004} \cdot I_{2005} \cdot I_{2006}}{I_{2003}}} = \sqrt[3]{\frac{0.7173 \cdot 0.7585 \cdot 0.9441}{0.8545}} = \sqrt[3]{0.5137} = 0.8009$ valore relativo medio, da cui segue una variazione relativa media del -19.91% .

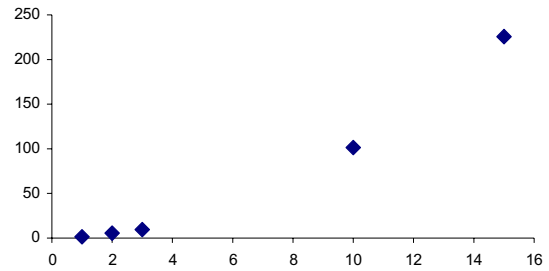
2.4 Si indichi la peggiore variazione relativa subita dalla grandezza X negli anni considerati.

-32.25% corrispondente al minimo valore relativo (NIBM) 0.6775

3. La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 5 valori della X :

x_i	1	2	3	10	15
y_i	2	5	10	101	226

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + bX^2$$

$$\text{II) } Y^* = cX^2$$

$a = 1$ $b = 1$ misura di adattamento: formula $M(E^2)$ valore 0

$c = 1.0056$ misura di adattamento: formula $M(E^2)$ valore 0.6215

Il modello migliore è $Y^* = 1 + X^2$

3.3 Si indichi l'intervallo di valori che può assumere il coefficiente di correlazione lineare ρ .

$$0 < \rho < 1$$

4. Se possibile e giustificando teoricamente le risposte, si completino le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

$X \setminus Y$	3	$3 < y_2 < 9$	9
2	8	8	8
$x_2 > 2$	8	8	8

$$\rho^2 = \eta_{X|Y}^2$$

Si è costruita tabella di indipendenza stocastica:

$$\chi^2 = 0 \rightarrow \eta_{X|Y}^2 = 0 \rightarrow \rho^2 = 0$$

$X \setminus Y$	3	$3 < y_2 < 9$	9
2	8	4	0
$x_2 > 2$	0	8	2

$$\rho^2 = \eta_{Y|X}^2$$

$\rho^2 = \eta_{Y|X}^2 \rightarrow$ medie condiz. allineate: $M(X|y_1)$ e $M(X|y_2)$ sono allineate (si verifica per qualsiasi combinazione di frequenze $n_{12}, n_{13}, n_{21}, n_{23}$)

5. Si consideri il seguente gioco: si estraggono senza reimmissione 4 palline da un'urna contenente 18 palline bianche e 22 palline nere; pagando una posta di 1 € si vincono 4 € se il numero di palline bianche estratte è superiore a quello delle palline nere.

5.1 Si calcoli la probabilità di ottenere 4 palline bianche nelle 4 estrazioni.

$$P = 0.0335$$

Può essere indifferentemente risolto come $P(B_1 \cap B_2 \cap B_3 \cap B_4) = P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 \cap B_2) \cdot P(B_4|B_1 \cap B_2 \cap B_3) =$

$$= \frac{18}{40} \frac{17}{39} \frac{16}{38} \frac{15}{37} \text{ o mediante la v.c. ipergeometrica } P(X=4) = \frac{\binom{18}{4} \binom{22}{0}}{\binom{40}{4}}.$$

5.2 Si costruisca la variabile casuale X = "guadagno del giocatore".

x	$P(X=x)$
-1	0.7701
3	0.2299

$$P(X=3) = P(4B) + P(3B e 1N) = 0.0335 + 0.1964.$$

5.3 Si stabilisca se il gioco in oggetto è equo.

Il gioco non è equo in quanto il valore atteso (media) della variabile casuale guadagno risulta $-0.0804 \neq 0$.

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 06.09.2012 (Tema 270)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la variabile X_t = numerosità della popolazione italiana, in milioni di unità, misurata tra il 1861 e il 2001.

Anno	X_t	$NIBM$
1861	22	--
1881	28	1.2727
1901	33	1.1786
1921	39	1.1818
1941	43	1.1026
1961	51	1.1860
1981	56	1.0980
2001	58	1.0179

1.1 Si costruisca la serie dei numeri indici con base mobile.

1.2 Si interpreti il valore dell'ultimo degli indici sopra calcolati.

${}_{1981}I_{2001} = 1.0179$ Tra il 1981 e il 2001 la popolazione è cresciuta del 1.79%

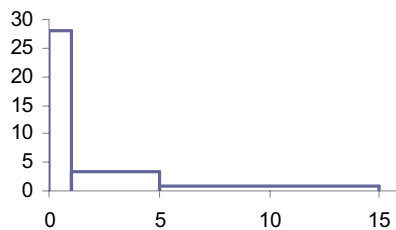
1.3 Si calcoli il tasso di crescita medio della popolazione italiana tra il 1901 e il 2001.

tmc +11.55% tasso medio di crescita (riferito a un ventennio ipotetico tra il 1901 e il 2001)

2. La tabella seguente contiene la distribuzione delle scadenze in anni (A) dei 108 titoli di stato detenuti in portafoglio da 2 banche italiane (B):

$B \setminus A$	0 → 1	1 → 5	5 → 15	
$b_1 = \text{Intesa San Paolo}$	21	18	19	58
$b_2 = \text{Unicredit}$	28	13	9	50
	49	31	28	108

2.1 Si rappresenti graficamente la distribuzione $A|Unicredit$.



2.2 Si confronti la variabilità delle distribuzioni condizionate $A|B=b_1$ e $A|B=b_2$.

$A|B=b_1$ indice di variabilità: c.v. = 0.9226

$A|B=b_2$ indice di variabilità: c.v. = 1.2263

2.3 Si calcoli l'indice $\eta_{A|B}^2 = 0.0385$ Sussiste un livello di dipendenza in media quasi nullo.

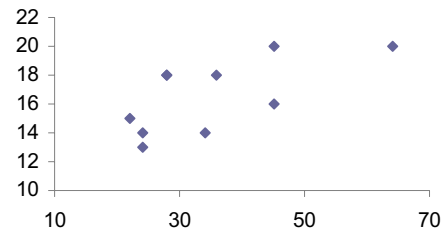
2.4 Si calcoli un opportuno indice di connessione tra A e B commentando il risultato ottenuto.

indice $\chi_N = 0.2112$. Tra le variabili A e B sussiste un livello di connessione molto basso.

3. Si considerino le seguenti 10 coppie di osservazioni delle variabili X e Y .

x_i	24	28	24	22	45	34	36	45	64	28
y_i	14	18	13	15	20	14	18	16	20	18

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

I) $Y^* = a + bX$

II) $Y^* = cX$

$a = 12.0897$ $b = 0.1289$ misura di adattamento: formula $M(E^2) = \text{var.res.}$ valore 3.2627

$c = 0.4354$ misura di adattamento: formula $M(E^2)$ valore 19.6981

Il modello migliore è: $Y^* = 12.0897 + 0.1289X$

4. Un gioco consiste nel lanciare 4 volte una moneta regolare scommettendo, a ogni lancio, 1 euro sull'evento *Testa*.

4.1 Si stabilisca se il gioco in oggetto è equo.

Occorre verificare se la variabile casuale guadagno ha valore atteso nullo.

G	$P(G=g)$
-4	0.0625
-2	0.2500
0	0.3750
2	0.2500
4	0.0625

La distribuzione di probabilità di G è simmetrica rispetto a 0 quindi $M(G) = 0$.

Ipotizzando di disporre di 100 euro all'inizio del gioco, calcolare la probabilità che:

4.2 dopo i primi 4 lanci ci si ritrovi esattamente con ancora 100 euro;

La proposizione è equivalente ad affermare che nei 4 lanci si vince 2 volte, quindi, indicata con X la variabile casuale "numero di successi nelle 4 prove", si ha che $X \sim \text{Bin}(n=4, p=0.5)$ e $P(X=2) = \binom{4}{2} 0.5^2 \cdot 0.5^2 = 0.375$.

4.3 dopo i primi 4 lanci il proprio capitale sia superiore ai 100 euro iniziali;

La proposizione è equivalente ad affermare che nei 4 lanci si vince almeno 3 volte, quindi: $P(X=3) + P(X=4) = \binom{4}{3} 0.5^3 \cdot 0.5^1 + \binom{4}{4} 0.5^4 \cdot 0.5^0 = 0.3125$.

4.4 Calcolare la probabilità che in 50 lanci si siano ottenuti almeno 26 eventi *Testa*.

Indicata con X la variabile casuale "numero di teste" nei 50 lanci $X \sim \text{Bin}(n=50, p=0.5)$.

$M(X) = 25$; $\text{Var}(X) = 12.5$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=25, \sigma^2=12.5)$ si ottiene

$$P(X > 26) = P\left(\frac{X - \mu}{\sigma} > \frac{26 - \mu}{\sigma}\right) = P\left(Z > \frac{26 - 25}{3.5355}\right) = P(Z > 0.28) = 1 - P(Z \leq 0.28) = 0.3897.$$

UNIVERSITÀ CATTOLICA DEL SACRO CUORE - MILANO
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
LAUREA IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO

Prova scritta di STATISTICA del 20.09.2012 (Tema 270_1)

Svolgere per esteso sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

1. Si consideri la variabile X_t = numerosità della popolazione, in milioni di unità, misurata tra il 1861 e il 2001.

Anno	X_t	$NIBF$ (1861=1)
1861	22	1.0000
1881	28	1.2727
1901	33	1.5000
1921	39	1.7727
1941	43	1.9545
1961	51	2.3182
1981	56	2.5455
2001	58	2.6364

1.1 Si costruisca la serie dei numeri indici con base fissa 1861.

1.2 Si interpreti il valore del numero indice al tempo 1881 con base 1861.

${}_{1861}I_{1881} = 1.2727$ Tra il 1861 e il 1881 la popolazione è cresciuta del 27.27%

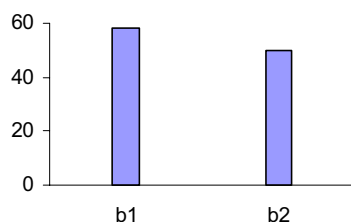
1.3 Si calcoli il tasso di crescita medio annuo della popolazione tra il 1861 e il 1881.

tmc +1.21%

2. La tabella seguente contiene la distribuzione delle scadenze in anni (A) dei 108 titoli di stato detenuti in portafoglio da 2 banche italiane (B):

$B \setminus A$	0 → 1	1 → 5	5 → 15	
$b_1 = \text{Intesa San Paolo}$	21	18	19	58
$b_2 = \text{Unicredit}$	28	13	9	50
	49	31	28	108

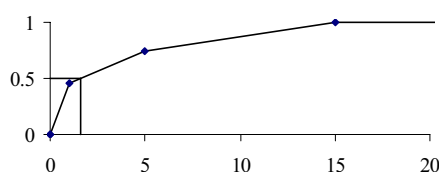
2.1 Si rappresenti graficamente la distribuzione marginale della variabile B .



2.2 Si calcolino media aritmetica, moda e mediana della distribuzione marginale B .

media aritmetica: non definita (carattere QLT) moda b_1 mediana non definita (carattere QLT sconnesso)

2.3 Si calcolino media, moda della distribuzione marginale A e si identifichi la mediana su un opportuno grafico.



media: 3.6806

moda: classe 0 → 1

2.4 Si riassumano con opportuni indici di variabilità le distribuzioni marginali A e B .

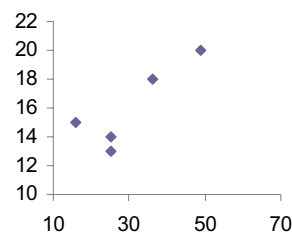
A indice di variabilità: varianza = 15.0762 c.v. = 3.8828

B indice di variabilità: Indice di Frosini valore 0.9258

3. Si considerino le seguenti 5 coppie di osservazioni delle variabili X e Y .

x_i	25	36	25	16	49
y_i	14	18	13	15	20

3.1 Si rappresentino graficamente le coppie (x_i, y_i)



3.2 Si calcolino, secondo il metodo dei minimi quadrati, i parametri e le misure di adattamento dei seguenti modelli di regressione:

$$\text{I) } Y^* = a + b\sqrt{X}$$

$$\text{II) } Y^* = c$$

$$a = 4.5769 \quad b = 2.1154 \quad \text{misura di adattamento: formula } \rho^2 = 0.6844$$

$$c = 16 \quad \text{misura di adattamento: formula } \rho^2 = 0$$

Il modello migliore è: $Y^* = 4.5769 + 2.1154\sqrt{X}$

4. Se possibile e giustificando teoricamente le risposte, si completino le seguenti tabelle in modo da soddisfare le condizioni sotto indicate.

$X \setminus Y$	3	6	9
2	2	a	2
4	b	2	b
6	2	a	2

$X \setminus Y$	20		40
10	2		

$$\eta_{Y|X}^2 = \eta_{X|Y}^2$$

$$\rho = 1 \text{ e } \chi^2 = 0$$

Le soluzioni immediatamente ottenibili sono:

$$\eta_{Y|X}^2 = \eta_{X|Y}^2 = 0 \text{ oppure } \eta_{Y|X}^2 = \eta_{X|Y}^2 = 1.$$

In tabella si riporta una possibile configurazione relativa alla prima opzione.

Per quanto attiene alla seconda opzione: $2 < x_2 < 6$, $3 < y_2 < 9$, $n_{12} = n_{13} = n_{21} = n_{23} = n_{31} = n_{32} = 0$, $n_{33} > 0$

Impossibile in quanto $\chi^2 = 0$ se sussiste indipendenza stocastica, che implica anche l'indipendenza lineare, vale a dire $\rho^2 = 0$.

5. Si supponga di lanciare un dado regolare. Si consideri il seguente gioco: si scommette, a ogni lancio, 1 euro e si vincono 2 euro se esce il numero 2 e 4 euro se esce il numero 4.

5.1 Si stabilisca se il gioco in oggetto è equo.

Occorre verificare se la variabile casuale guadagno ha valore atteso nullo.

G	$P(G=g)$
-1	$4/6 = 2/3$
1	$1/6$
3	$1/6$

Nel caso in oggetto $M(G) = 0$, quindi il gioco è equo.

Calcolare la probabilità che:

5.2 In 3 lanci esca sempre il numero 4;

$$P(4_1 \cap 4_2 \cap 4_3) = P(4_1) \cdot P(4_2|4_1) \cdot P(4_3|4_1 \cap 4_2) = P(4_1) \cdot P(4_2) \cdot P(4_3) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{216} = 0.004630.$$

5.3 In 90 lanci si vinca almeno 30 volte.

Indicata con X la variabile casuale "numero di vincite" nei 90 lanci $X \sim \text{Bin}(n=90, p=2/6)$.

$M(X) = 30$; $\text{Var}(X) = 20$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu=30, \sigma^2=20)$ si ottiene

$$P(X > 30) = P(X > \mu) = 0.5 \quad \left[= P\left(\frac{X - \mu}{\sigma} > \frac{30 - \mu}{\sigma}\right) = P\left(Z > \frac{30 - 30}{4.4721}\right) = P(Z > 0) = 1 - P(Z \leq 0) = 0.5. \right]$$

Università Cattolica del Sacro Cuore - Milano
FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA
CORSO DI STUDIO IN ECONOMIA E GESTIONE DEI BENI CULTURALI E DELLO SPETTACOLO
 Prova scritta di STATISTICA del 10.01.2013 (Tema 271)
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

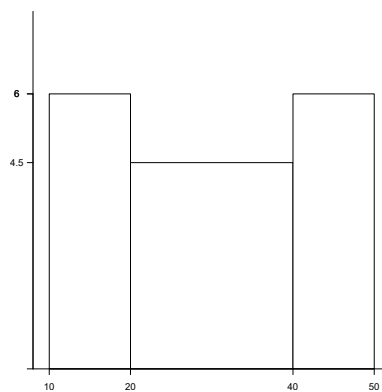
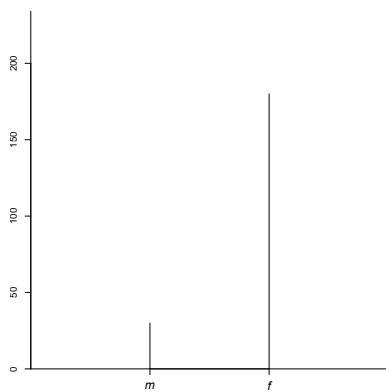
ESERCIZIO 1.

Con riferimento a 210 soggetti sono stati rilevati il genere, X , (con modalità m e f) e la spesa mensile, Y ripartita in classi. I seguenti prospetti riportano la distribuzione delle frequenze assolute di X e delle densità di frequenza di Y :

x_i	n_i
m	30
f	180

$h_{i-1} \rightarrow h_i$	d_i
10 $\rightarrow$ 20	6
20 $\rightarrow$ 40	4.5
40 $\rightarrow$ 50	6

1. Si rappresentino graficamente le distribuzioni delle frequenze di X e Y e si riassumano le due variabili con opportuni indici di posizione e di variabilità.



X : Indice di posizione: $\text{Moda}(X)=f$
 Y : Indice di posizione: $M(Y) = 30$

Indice di variabilità: Indice di Frosini: $F_N(X) = 0.2857$
 Indice di variabilità: $CV = 0.3780$

2. Si calcolino la media armonica e una misura di asimmetria per Y .

Y : Media armonica: $\mu^{(-1)} = 25.2$ Indice di asimmetria: 0 (la distribuzione è simmetrica)

ESERCIZIO 2.

Si consideri la seguente serie storica delle vendite di un'azienda tra il 2006 e il 2010:

t	x_t	$NIBM$
2006	205	—
2007	290	1.4146
2008	315	1.0862
2009	340	1.0794
2010	225	0.6618

1. Si costruisca la corrispondente serie dei numeri indici a base mobile.

2. Si calcoli il tasso medio di variazione tra il 2008 e il 2010, commentando il risultato ottenuto.

$\sqrt[2]{_{2008}I_{2010}} = \sqrt[2]{\frac{225}{315}} = \sqrt[2]{0.71} = 0.8452 = \sqrt[2]{_{2008}I_{2009} \cdot _{2009}I_{2010}} = \sqrt[2]{1.0794 \cdot 0.6618}$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 15.48%.

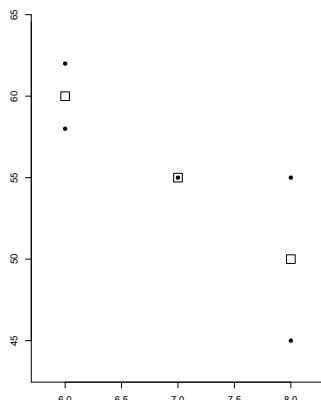
Le vendite hanno subito un decremento medio annuo del 15.48% tra il 2008 e il 2010.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	6	6	7	8	8
y_i	62	58	55	45	55

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione di Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + bX \quad II) Y^* = cX$$

$$\hat{a} = 90, \hat{b} = -5, M(\hat{E}^2) = 11.6 \quad \hat{c} = 7.6506, M(\hat{E}^2) = 141.7205 \quad \text{Il modello migliore è } Y^* = 90 - 5X.$$

3. Si determini il valore del rapporto di correlazione $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = 0.6329$. Le medie condizionate sono allineate; la retta coincide con la funzione di regressione, $\rho^2 = \eta_{Y|X}^2$

ESERCIZIO 4.

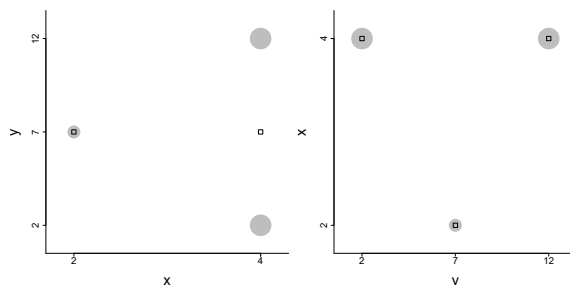
Si completino, giustificando opportunamente la risposta, le seguenti tabelle in modo tale che:

a) $\rho^2 = 1$ e $\eta_{X|Y}^2 = 0$

	$y_1 = 2$	$y_2 = ______$	$y_3 = 12$
$x_1 = 2$		3	
$x_2 = 4$	5		5

b) $\rho^2 = \eta_{Y|X}^2 = 0$ e $\eta_{X|Y}^2 = 1$

	$y_1 = 2$	$y_2 = ______$	$y_3 = 12$
$x_1 = 2$	0	3	0
$x_2 = 4$	5	0	5



Impossibile in quanto $\rho^2 \leq \eta_{X|Y}^2$

Dipendenza funzionale di X da $Y \rightarrow X|y_j$ degeneri

Indipendenza in media di Y da $X \rightarrow$

$$\rightarrow M(Y|x_1) = M(Y|x_2) = M(Y) = y_2$$

ESERCIZIO 5.

Si consideri l'esperimento di estrazione di 4 palline da un'urna contenente 21 palline rosse e 29 palline blu.

1. Si calcoli la probabilità di ottenere almeno 3 palline blu nell'ipotesi che le 4 estrazioni siano effettuate senza reimmissione.

$$P(\text{almeno } 3 \text{ B}) = P(3B) + P(4B) = \binom{4}{3} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{21}{47} + \binom{4}{4} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{26}{47} = 0.3332 + 0.1031 = 0.4363$$

o mediante la v.c. ipergeometrica

$$P(X = 3) + P(X = 4) = \frac{\binom{29}{3} \binom{21}{1}}{\binom{50}{4}} + \frac{\binom{29}{4} \binom{21}{0}}{\binom{50}{4}} = 0.4363$$

2. Si calcoli la probabilità di ottenere almeno 36 palline blu nell'ipotesi che siano effettuate 60 estrazioni con reimmissione. Indicata con X la variabile casuale 'numero di successi' nelle 60 prove indipendenti risulta $X \sim \text{Bin}(n = 60, p = 0.58)$. $M(X) = 34.8$; $\text{Var}(X) = 14.6160$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 34.8, \sigma^2 = 14.6160)$ si ottiene

$$P(X > 36) = P\left(\frac{X - \mu}{\sigma} > \frac{36 - \mu}{\sigma}\right) = P\left(Z > \frac{36 - 34.8}{3.8231}\right) = P(Z > 0.31) = 1 - P(Z \leq 0.31) = 1 - 0.6217 = 0.3783.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 24.01.2013 (Tema 272)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

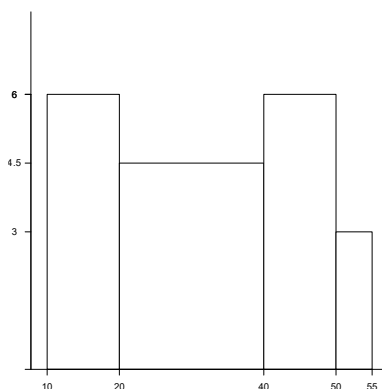
Studente _____ matricola _____

ESERCIZIO 1.

Su 225 soggetti è stata rilevata la variabile Y , ripartita in classi, le cui densità di frequenza sono riportate nel prospetto.

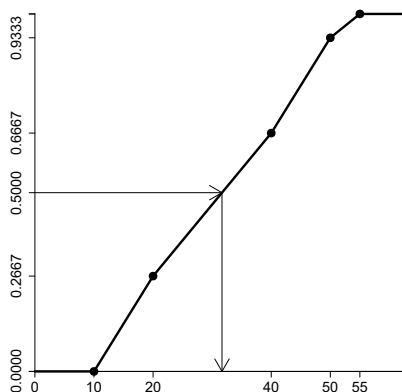
$h_{i-1} \rightarrow h_i$	d_i	f_i
10 → 20	6	0.2667
20 → 40	4.5	0.4000
40 → 50	6	0.2667
50 → 55	3	0.0667

1. Si ricostruisca la tabella delle frequenze relative e si rappresenti graficamente la distribuzione di Y .



2. Si calcolino moda, media quadratica e coefficiente di variazione. Si individui su un opportuno grafico la mediana.

Moda(X) = 10 è bimodale $\mu^{(2)} = 33.8194$ $CV = 0.3907$



ESERCIZIO 2.

Della serie storica X_t delle vendite di un'azienda tra il 2006 e il 2010 abbiamo le seguenti informazioni:

t	$NIBM$	$NIBF$
2006		100
2007	113	
2008		116
2009	106	
2010		131

t	$NIBM$	$NIBF$
2006	—	100
2007	113	113
2008	102.65	116
2009	106	122.96
2010	106.54	131

1. Si riempiano le caselle vuote della tabella e sapendo che $x_{2010}=650$ si calcoli x_{2006} .

$${}_{2010}I_{2006} = 1.31 = \frac{x_{2010}}{x_{2006}} \rightarrow x_{2006} = \frac{x_{2010}}{{}_{2010}I_{2006}} = \frac{650}{1.31} = 496.1832$$

2. Si calcoli il tasso medio di variazione tra il 2007 e il 2010, commentando il risultato ottenuto.

$$\sqrt[3]{{}_{2007}I_{2010}} = \sqrt[3]{\frac{131}{113}} = \sqrt[3]{1.1593} = 1.0505 = \sqrt[3]{{}_{2007}I_{2008} \cdot {}_{2008}I_{2009} \cdot {}_{2009}I_{2010}} = \sqrt[3]{1.0265 \cdot 1.0600 \cdot 1.0654}, \text{ numero indice a base mobile medio, a cui corrisponde una variazione relativa media del } +5.05\%.$$

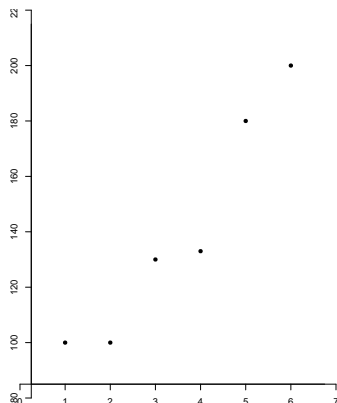
Le vendite hanno subito un incremento medio annuo del 5.05% tra il 2007 e il 2010.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 6 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	3	4	5	6
y_i	100	100	130	133	180	200

1. Si rappresentino graficamente le coppie di dati.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + bX \quad II) Y^* = c + dX^2$$

$$\hat{a} = 66.2, \hat{b} = 21.2286, \rho^2 = 0.9227 \quad \hat{c} = 94.5254, \hat{d} = 3.0313, \text{ indice adattamento } \rho^2 = 0.9620$$

Il modello migliore è $Y^* = 94.5254 + 3.0313X^2$.

3. Si determinino i valori dei rapporti di correlazione $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

$$\eta_{Y|X}^2 = 1 \text{ in quanto sussiste dipendenza funzionale di } Y \text{ da } X; \quad \eta_{X|Y}^2 = 0.9714.$$

ESERCIZIO 4.

Si completi la tabella, giustificando opportunamente la risposta, in modo che $\rho^2 < \eta_{Y|X}^2$

	$y_1 = 2$	$y_2 = \text{—}$	$y_3 = 12$
$x_1 = 2$		3	
$x_2 = 4$	5		5

Impossibile, si hanno due distribuzioni condizionate $Y|x = 2$ e $Y|x = 4$, le 2 medie condizionate sono allineate e quindi $\rho^2 = \eta_{Y|X}^2$

ESERCIZIO 5.

Si consideri l'esperimento di lancio di 10 monete equilibrate e 6 dadi equilibrati.

1. Si calcoli la probabilità di ottenere 5 teste e 5 facce con numero pari.

Siano X la variabile casuale 'n. di teste nel lancio delle 10 monete' e

Y la variabile casuale 'n. di facce con numero pari nel lancio dei 6 dadi'.

$$X \sim \text{Bin}(n = 10, p = 0.5) \text{ e } Y \sim \text{Bin}(n = 6, p = 0.5).$$

Il risultato del lancio dei dadi non dipende da quello delle monete, X e Y sono stocasticamente indipendenti, quindi

$$P\{(X = 5) \cap (Y = 5)\} = P(X = 5) \cdot P(Y = 5) = \binom{10}{5} 0.5^5 0.5^5 \cdot \binom{6}{5} 0.5^5 0.5 = 0.2461 \cdot 0.0938 = 0.0231.$$

2. Ipotizzando di lanciare 100 monete e 100 dadi, si calcoli la probabilità di ottenere almeno 50 teste e almeno 39 numeri pari.

Siano X la variabile casuale 'n. di teste nel lancio delle 100 monete' e

Y la variabile casuale 'n. di facce con numero pari nel lancio dei 100 dadi'.

$$X \sim \text{Bin}(n = 100, p = 0.5) \text{ e } Y \sim \text{Bin}(n = 100, p = 0.5).$$

$$M(X) = M(Y) = 50 \quad \text{Var}(X) = \text{Var}(Y) = 25.$$

Utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 50, \sigma^2 = 25) \text{ e } Y \sim N(\mu = 50, \sigma^2 = 25)$$

Il risultato del lancio dei dadi non dipende da quello delle monete, X e Y sono stocasticamente indipendenti, quindi

$$P\{(X \geq 50) \cap (Y \geq 39)\} = P(X \geq 50) \cdot P(Y \geq 39) = 0.5 \cdot 0.9861 = 0.4930.$$

essendo:

$$P(X > 50) = P\left(\frac{X - \mu}{\sigma} > \frac{50 - \mu}{\sigma}\right) = P\left(Z > \frac{50 - 50}{5}\right) = P(Z > 0) = 0.5$$

$$P(Y > 39) = P\left(\frac{Y - \mu}{\sigma} > \frac{39 - \mu}{\sigma}\right) = P\left(Z > \frac{39 - 50}{5}\right) = P(Z > -2.2) = P(Z \leq 2.2) = 0.9861.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 07.02.2013 (Tema 273)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

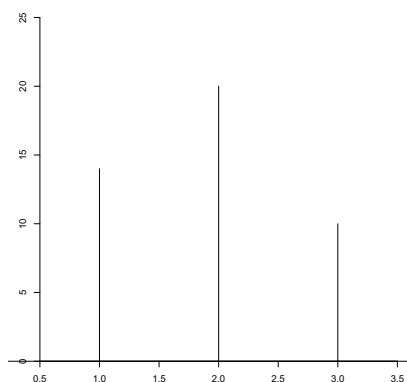
Studente _____ matricola _____ **q = 1**

ESERCIZIO 1.

Su 50 famiglie si sono rilevate le variabili X 'nazionalità capofamiglia' (I =italiana, S =straniera) e Y 'n. figli'.

$X \setminus Y$	1	2	3
I	11	19	5
S	3	7	5

1. Si rappresenti graficamente il carattere Y e se ne calcolino moda e mediana.



$$\text{Moda}(Y) = 2 \quad y_{0.5} = \frac{y_{(25)} + y_{(26)}}{2} = \frac{2+2}{2} = 2$$

2. Si confrontino con opportuni indici di posizione e di variabilità le variabili condizionate $Y|X = I$ e $Y|X = S$.

$Y|X = I$: Indice di posizione: $M(Y|X = I) = 1.8286$ Indice di variabilità: $CV(Y|X = I) = 0.3577$

$Y|X = S$: Indice di posizione: $M(Y|X = S) = 2.1333$ Indice di variabilità: $CV(Y|X = S) = 0.3366$

$Y|X = S$ presenta un valore medio più elevato ma un minore livello di variabilità rispetto a $Y|X = I$.

3. Si calcoli un opportuno indice di variabilità di X .

$F_N = 0.6$ X presenta un livello medio di eterogeneità.

ESERCIZIO 2.

Dati i seguenti $NIBM$ delle vendite di un'azienda tra il 2001 e il 2005:

t	$NIBM$	$NIBF$
2001		1
2002	0.94	0.9400
2003	1.12	1.0528
2004	1.06	1.1160
2005	1.11	1.2387

1. Si calcolino i numeri indice a base fissa (base 2001).

2. Si calcoli il tasso medio di variazione tra il 2001 e il 2005 esprimendolo in valore percentuale.

$\sqrt[4]{2001I_{2005}} = \sqrt[4]{1.2387} = 1.0550 = \sqrt[4]{2001I_{2002} \cdot \dots \cdot 2004I_{2005}} = \sqrt[4]{0.94 \cdot 1.12 \cdot 1.06 \cdot 1.11}$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 5.5%.

Le vendite hanno subito un incremento medio annuo del 5.5% tra il 2001 e il 2005.

3. Per ottenere un tasso medio di variazione tra il 2001 e il 2006 del 7% che valore dovrebbe avere $2001I_{2006}$?

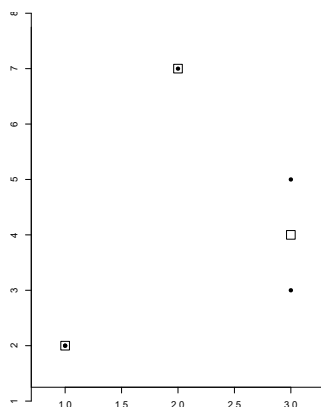
$$2001I_{2006} = 1.07^5 = 1.4026$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	2	3	3
y_i	2	2	7	5	3

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + bX \quad II) Y^* = c$$

$\hat{a} = 1.8, \hat{b} = 1, \rho^2 = 0.2128 \quad \hat{c} = 3.8, R^2 = 0$ Il modello migliore è $Y^* = 1.8 + X$.

3. Utilizzando il modello ritenuto migliore si determini y quando $x = 5$.

$$Y^*(5) = 1.8 + 5 = 6.8$$

4. Si dica, giustificando teoricamente la risposta, se è verificata la relazione $\rho^2 = \eta_{Y|X}^2$.

La relazione non è verificata in quanto le medie condizionate non sono allineate.

ESERCIZIO 4.

Si completino le seguenti tabelle, giustificando opportunamente la risposta, in modo tale che :

a) $\chi^2 = 62$ e $M(X) = 2$

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7	0	0	7
$x_2 = 2$	0	0	17	17
$x_3 = 3$	0	7	0	7
				$n = 31$

Se $\chi^2 = 62$ allora $\chi_N^2 = 1$

sussiste dipendenza funzionale (biuttiva)

X deve essere simmetrica affinché $M(X) = 2$

b) $\chi^2 = 0$ e $M(Y) = 4$

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7	n_{12}	7	
$x_2 = 2$	7c	$n_{12}c$	7c	
$x_3 = 3$	7d	$n_{12}d$	7d	
				$n =$

Se $\chi^2 = 0$ allora sussiste indipendenza stocastica

→ distribuzioni condizionate somiglianti

n_{12}, c, d numeri interi positivi

Y deve essere simmetrica affinché $M(Y) = 2$

ESERCIZIO 5.

Siano A, B, C e D quattro eventi che costituiscono una partizione dello spazio campionario, tali che $P(A) = P(B) = 0.2$ e $P(D) = 2 \cdot P(C)$.

1. Si calcolino $P(A|B)$, $P(B \cup D)$ e $P(\overline{A \cup B})$.

A, B, C, D costituiscono una partizione di Ω se $A \cup B \cup C \cup D = \Omega$ e A, B, C e D non hanno elementi in comune.

$$1 = P(A) + P(B) + P(C) + P(D) = 0.2 + 0.2 + P(C) + 2 \cdot P(C) \rightarrow P(C) = 0.2, P(D) = 0.4$$

$$P(A|B) = 0 \text{ in quanto } A \cap B = \emptyset,$$

$$P(B \cup D) = P(B) + P(D) = 0.2 + 0.4 = 0.6$$

$$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - \{P(A) + P(B)\} = 1 - 0.2 - 0.2 = 0.4 = P(C \cup D)$$

2. Si dica se A e C possono ritenersi indipendenti.

A e C non possono ritenersi indipendenti in quanto $A \cap C = \emptyset$, quindi $P(A \cap C) = 0 \neq P(A) \cdot P(C)$

3. Si calcoli la probabilità di ottenere meno di 6 successi in 80 prove indipendenti con probabilità di successo $p = P(B)$.

Indicata con X la variabile casuale 'numero di successi' nelle 80 prove indipendenti risulta $X \sim \text{Bin}(n = 80, p = 0.2)$.

$M(X) = 16$; $\text{Var}(X) = 12.8$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 16, \sigma^2 = 12.8)$ si ottiene

$$P(X < 6) = P\left(\frac{X - \mu}{\sigma} < \frac{6 - \mu}{\sigma}\right) = P\left(Z < \frac{6 - 16}{3.5777}\right) = P(Z < -2.80) = 1 - P(Z \leq 2.80) = 1 - 0.9974 = 0.0026.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 06.06.2013 (Tema 274.1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

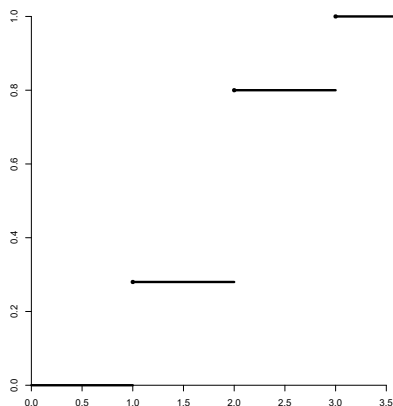
Studente _____ matricola _____

ESERCIZIO 1.

Su 50 famiglie si sono rilevate le variabili X 'nazionalità capofamiglia' (I =italiana, S =straniera) e Y 'n. figli'.

$X \backslash Y$	1	2	3
I	11	19	5
S	3	7	5

1. Si rappresenti graficamente la funzione di ripartizione del carattere Y e se ne calcolino media e coefficiente di variazione.



$$M(Y) = 1.9200 \quad CV(Y) = 0.3584$$

2. Si calcoli un opportuno indice di connessione tra X e Y commentando il risultato ottenuto.

$$\chi_N = 0.2241.$$

Tra le variabili sussiste un livello molto basso di connessione.

3. Si calcoli la mediana di X .

Impossibile calcolare la mediana in quanto il carattere X è qualitativo sconnesso.

ESERCIZIO 2.

Dati i seguenti $NIBF$ (base 2001) delle vendite di un'azienda tra il 2001 e il 2005:

t	$NIBF$	$NIBM$
2001	1.00	—
2002	0.94	0.9400
2003	1.12	1.1915
2004	1.06	0.9464
2005	1.11	1.0472

1. Si calcolino i numeri indice a base mobile.

2. Si calcoli il tasso medio di variazione tra il 2003 e il 2005 esprimendolo in valore percentuale.

$\sqrt[2]{2003I_{2005}} = \sqrt[2]{\frac{1.11}{1.12}} = \sqrt[2]{0.9911} = 0.9955 = \sqrt[2]{2003I_{2004} \cdot 2004I_{2005}} = \sqrt[2]{0.9464 \cdot 1.0472}$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del -0.45% .

Le vendite hanno subito un decremento medio annuo del 0.45% tra il 2003 e il 2005.

3. Sapendo che tra il 2005 e il 2006 vi è stato un calo delle vendite del 10% si determinino $2005I_{2006}$ e il valore delle vendite nel 2006.

$$2005I_{2006} = 0.90$$

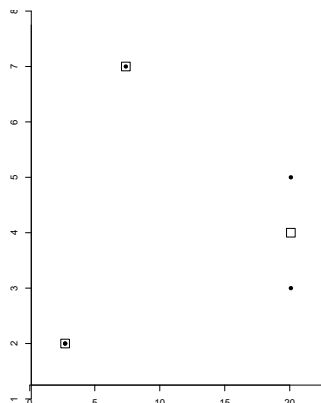
Non è possibile determinare il valore x_{2006} in quanto non si conosce x_{2005} .

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2.7183	2.7183	7.3891	20.0855	20.0855
y_i	2	2	7	5	3

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot \ln X \quad II) Y^* = c + dX$$

$$\hat{a} = 1.8, \hat{b} = 1, \text{ indice adattamento } \rho^2 = 0.2128 \quad \hat{c} = 3.0621, \hat{d} = 0.0696, \rho^2 = 0.0811$$

Il modello migliore è $Y^* = 1.8 + \ln X$.

ESERCIZIO 4.

Si completi la seguente tabella, giustificando opportunamente la risposta, in modo tale che :

$$a) \chi_N^2 = \rho^2 = \eta_{Y|X}^2$$

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7			
$x_2 = 2$				
$x_3 = 3$				
				$n = 120$

È possibile realizzare (senza effettuare calcoli) la situazione $\chi_N^2 = \rho^2 = \eta_{Y|X}^2$ se:

$$\chi_N^2 = \rho^2 = \eta_{Y|X}^2 = 0$$

situazione di indipendenza stocastica

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7	8	9	
$x_2 = 2$	14	16	18	
$x_2 = 2$	14	16	18	
				$n = 120$

$$\chi_N^2 = \rho^2 = \eta_{Y|X}^2 = 1$$

situazione di dipendenza funzionale lineare

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7	0	0	
$x_2 = 2$	0	n ₂₂	0	
$x_3 = 3$	0	0	120 - (7 + n ₂₂)	
				$n = 120$

ESERCIZIO 5.

L'urna $U1$ contiene 3 palline nere e 1 pallina rossa. L'urna $U2$ contiene 4 palline nere e 5 palline rosse.

1. Si calcoli la probabilità che estraendo con reimmissione 5 palline dall'urna $U1$ ce ne siano almeno 4 nere. Sia X la variabile casuale 'n. di palline nere nelle 5 estrazioni con reimmissione'.

$$X \sim \text{Bin}(n = 5, p = 0.75).$$

$$P(X \geq 4) = P(X = 4) + P(X = 5) = \binom{5}{4} 0.75^4 0.25^1 + \binom{5}{5} 0.75^5 0.25^0 = 0.3955 + 0.2373 = 0.6328.$$

2. Si estraiga una pallina dall'urna $U1$ e la si metta nell'urna $U2$. Si calcoli la probabilità che estraendo senza reimmissione 3 palline dall'urna $U2$ si ottengano 2 palline nere.

Siano N l'evento 'estrazione di una pallina nera dall'urna $U1$ ', $X|N$ e $X|\bar{N}$ le variabile casuale 'n. di palline nere nelle 3 estrazioni senza reimmissione dall'urna $U2$ ' condizionate al verificarsi degli eventi N e $\bar{N}$.

$$\begin{aligned} P(2 \text{ nere in 3 estraz. senza reimmissione}) &= P(N) \cdot P\{(X|N) = 2\} + P(\bar{N}) \cdot P\{(X|\bar{N}) = 2\} = \\ &= \frac{3}{4} \cdot \frac{\binom{5}{2} \binom{1}{1}}{\binom{10}{3}} + \frac{1}{4} \cdot \frac{\binom{4}{2} \binom{6}{1}}{\binom{10}{3}} = 0.75 \cdot 0.4167 + 0.25 \cdot 0.3 = 0.3875. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 27.06.2013 (Tema 274)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

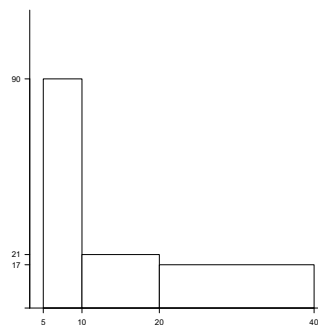
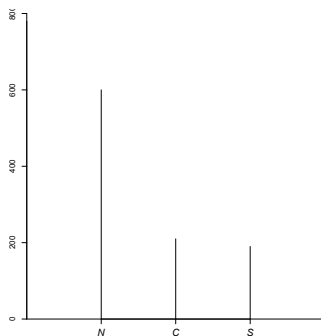
In un'indagine su 1000 cittadini sono stati rilevati l'area di residenza X (N = Nord; C = Centro; S = Sud Isole) e la spesa media mensile per libri Y . Si riportano la distribuzione di frequenze relative di X e la densità di frequenza di Y .

x_i	f_i	y_i	d_i
N	0.60	5 - 10	90
C	0.21	10 - 20	21
S	0.19	20 - 40	17

1. Si costruiscano le distribuzioni delle frequenze assolute di X e di Y .

x_i	n_i	y_i	n_i
N	600	5 - 10	450
C	210	10 - 20	210
S	190	20 - 40	340

2. Si rappresentino graficamente le due variabili e si riassumano le stesse con opportuni indici di posizione e di variabilità.



X : Indice di posizione: $\text{Moda}(X) = N$

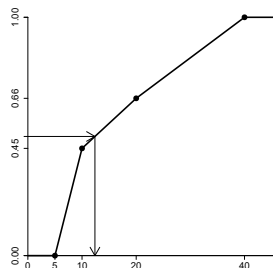
Y : Indice di posizione: $M(Y) = 16.725$

Indice di variabilità: Indice di Frosini: $F_N(X) = 0.5996$

Indice di variabilità: $CV(Y) = 0.5944$

3. Si calcoli la media geometrica e si individui su un opportuno grafico la mediana della variabile Y .

$$\mu_{(0)} = 13.8988$$



ESERCIZIO 2.

Si consideri la serie NIBF (base 2006) relativa alla spesa delle Amministrazioni Pubbliche rilevata tra il 2005 e il 2009:

t	$NIBF$	$NIBM$
2005	0.9705	—
2006	1.0000	1.0304
2007	1.0164	1.0164
2008	0.7607	0.7484
2009	1.0951	1.4396

1. Si calcolino i numeri indice a base mobile.

2. Si calcoli il tasso medio di variazione tra il 2006 e il 2009 esprimendolo in valore percentuale.

$\sqrt[3]{2006I_{2009}} = \sqrt[3]{\frac{1.0951}{1}} = \sqrt[3]{1.0951} = 1.0307$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del 3.07%.

3. Si calcoli il Numero Indice del 2006 con base l'anno 2009 interpretando il valore ottenuto.

$$2009I_{2006} = \frac{x_{2006}}{x_{2009}} = \frac{1}{\frac{1}{2006I_{2009}}} = \frac{1}{1.0951} = 0.9132$$

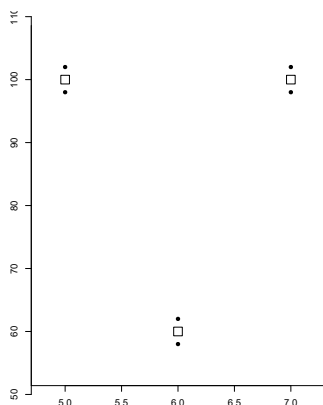
La spesa delle Amministrazioni Pubbliche nel 2006 è stata il 91.32% della spesa nel 2009.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 6 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	5	5	6	6	7	7
y_i	102	98	62	58	102	98

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 86.6667, \hat{b} = 0, M(\hat{E}^2) = 359.5556 \quad \hat{c} = 14.1818, M(\hat{E}^2) = 496.1212 \quad \text{Il modello migliore è } Y^* = 86.6667.$$

3. Si fornisca la previsione in corrispondenza di $x = 8$

$$Y^*(8) = 86.6667.$$

4. Si determini il valore del rapporto di correlazione $\eta_{Y|X}^2$.

$$\eta_{Y|X}^2 = 0.9889$$

ESERCIZIO 4.

Un macchinario presenta un tasso di difettosità del 6 %.

1. Scelti a caso 4 pezzi (con reimmissione) dal flusso produttivo si calcoli la probabilità che nessuno sia difettoso. Sia X la variabile casuale 'n. di pezzi difettosi nelle 4 prove indipendenti'.

$$X \sim \text{Bin}(n = 4, p = 0.06).$$

$$P(X = 0) = \binom{4}{0} 0.06^0 0.94^4 = 0.7807.$$

2. Nell'ipotesi in cui siano effettuate 60 estrazioni si calcoli la probabilità che vi sia almeno un pezzo difettoso. Indicata con X la variabile casuale 'numero pezzi difettosi nelle 60 prove indipendenti' risulta

$$X \sim \text{Bin}(n = 60, p = 0.06).$$

quindi

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{60}{0} 0.06^0 0.94^{60} = 1 - 0.02442 = 0.97558.$$

Si osserva come non valgano le condizioni per ottenere un risultato affidabile utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale, infatti $np = 60 \cdot 0.06 = 3.6 \not\geq 5$.

Si riportano, per completezza, i risultati che si avrebbero utilizzando tale approssimazione:

- Ricordando che $M(X) = 3.6$ e $\text{Var}(X) = 3.3840$ abbiamo

$$X \sim N(\mu = 3.6, \sigma^2 = 3.3840)$$

quindi

$$P(X \geq 1) = P\left(\frac{X - \mu}{\sigma} \geq \frac{1 - \mu}{\sigma}\right) = P\left(Z \geq \frac{1 - 3.6}{1.8396}\right) = P(Z \geq -1.41) = P(Z \leq 1.41) = 0.9207.$$

- tenendo conto della correzione per la continuità si ottiene

$$P(X \geq 1 - 0.5) = P\left(\frac{X - \mu}{\sigma} \geq \frac{1 - 0.5 - \mu}{\sigma}\right) = P\left(Z \geq \frac{1 - 0.5 - 3.6}{1.8396}\right) = P(Z \geq -1.69) = P(Z \leq 1.69) = 0.9545.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 11.07.2013 (Tema 274.2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

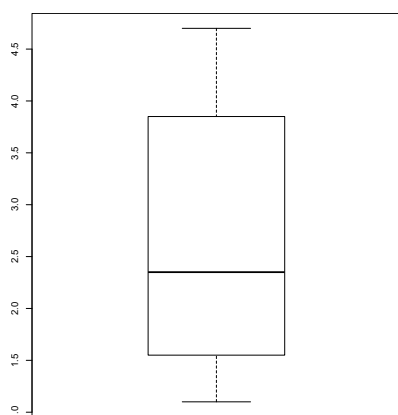
ESERCIZIO 1.

Nel seguente prospetto sono riportati i valori assunti dalla variabile X 'stipendio mensile' riferiti a 16 soggetti, appartenenti alle categorie $i =$ 'impiegato' e $d =$ 'dirigente'.

d	3.6	4.1	3.1	4.6	4.7	4.3					
i	1.6	2.1	2.6	1.3	2.1	2.5	1.1	1.4	2.2	1.5	

1. Si rappresenti graficamente, mediante un grafico a scatola e baffi (*box & whiskers plot*), la variabile X (per tutte le unità statistiche), commentando i risultati ottenuti.

Rappr. grafica variabile X



x_i	n_i	N_i	F_i
1.1	1	1	0.0625
1.3	1	2	0.1250
1.4	1	3	0.1875
1.5	1	4	0.2500
1.6	1	5	0.3125
2.1	2	7	0.4375
2.2	1	8	0.5000
2.5	1	9	0.5625
2.6	1	10	0.6250
3.1	1	11	0.6875
3.6	1	12	0.7500
4.1	1	13	0.8125
4.3	1	14	0.8750
4.6	1	15	0.9375
4.7	1	16	1.0000
16			

Elementi per la costruzione del *box & whiskers plot*

$$\begin{aligned}
 x_{0.25} &= \frac{1.5+1.6}{2} = 1.55 \\
 x_{0.5} &= \frac{2.2+2.5}{2} = 2.35 \\
 x_{0.75} &= \frac{3.6+4.1}{2} = 3.85 \\
 x_{\min} &= x_{(1)} = 1.1 \\
 x_{\max} &= x_{(16)} = 4.7 \\
 \text{lunghezza teorica baffi} &= 3.45 \\
 \text{baffo inferiore} &= 1.1 \\
 \text{baffo superiore} &= 4.7
 \end{aligned}$$

2. Si confrontino con opportuni indici di posizione e di variabilità le distribuzioni delle variabili $X|i$ e $X|d$ per le due categorie lavorative. Si commentino i risultati ottenuti.

$Y|X = i$: Indice di posizione: $M(Y|X = i) = 1.84$

Indice di variabilità: $CV(Y|X = i) = 0.2709$

$Y|X = d$: Indice di posizione: $M(Y|X = d) = 4.0667$

Indice di variabilità: $CV(Y|X = d) = 0.1381$

Lo stipendio dei soggetti dirigenti presenta un valore medio più elevato ma un minore livello di variabilità.

3. Sulla base dei valori delle medie e varianze parziali, ottenute al punto precedente, si ricavino i valori della media e varianza della variabile X riferiti all'intera popolazione.

$$M(Y) = 2.6750, \text{Var}(Y) = 1.4356$$

$$M(Y) = M_X\{M(Y|X)\}, \text{Var}(Y) = \text{Var}_X\{M(Y|X)\} + M_X\{\text{Var}(Y|X)\}$$

ESERCIZIO 2.

Si consideri la serie NIBM relativa alla spesa delle Amministrazioni Pubbliche rilevata tra il 2005 e il 2009:

t	$NIBM$
2005	
2006	1.0304
2007	1.0164
2008	0.7484
2009	1.4397

1. Sapendo che $x_{2009} = 334$ si calcoli il valore di x_{2005} .

$${}_{2005}I_{2009} = {}_{2005}I_{2006} \cdot {}_{2006}I_{2007} \cdot {}_{2007}I_{2008} \cdot {}_{2008}I_{2009} = 1.0304 \cdot 1.0164 \cdot 0.7484 \cdot 1.4397 = 1.1284$$

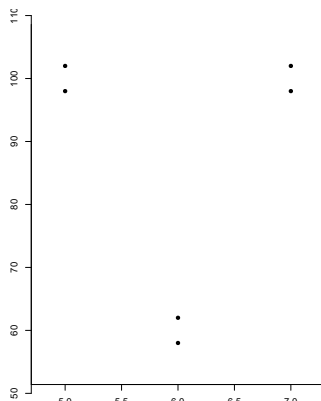
$${}_{2005}I_{2009} = \frac{x_{2009}}{x_{2005}} = 1.1284 \text{ da cui } x_{2005} = \frac{334}{1.1284} = 296$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 6 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	5	5	6	6	7	7
y_i	102	98	62	58	102	98

1. Si rappresentino graficamente le coppie di dati.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot (X - 6)^2 \quad II) Y^* = c$$

$$\hat{a} = 60, \hat{b} = 40, \rho^2 = 0.9889 \quad \hat{c} = 86.6667, R^2 = 0 \quad \text{Il modello migliore è } Y^* = 60 + 40X.$$

ESERCIZIO 4.

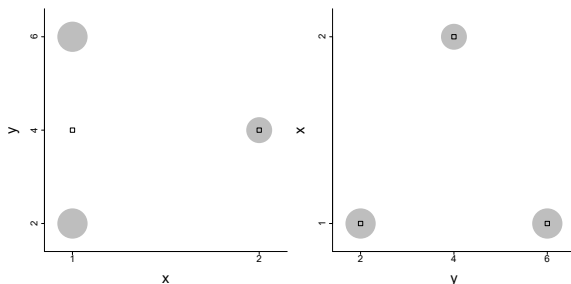
Si completino le seguenti tabelle, giustificando opportunamente la risposta, in modo tale che :

a) $\rho^2 = 1$ ed $\eta_{Y|X}^2 = 0$

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7			
$x_2 = 2$				
				$n = 20$

b) $\eta_{X|Y}^2 = 1$ ed $\eta_{Y|X}^2 = 0$

	$y_1 = 2$	$y_2 = 4$	$y_3 = 6$	
$x_1 = 1$	7	0	7	
$x_2 = 2$	0	6	0	
				$n = 20$



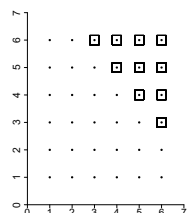
Impossibile in quanto $\rho^2 \leq \eta_{Y|X}^2$

Dipendenza funzionale di X da $Y \rightarrow X|y_j$ degeneri
Indipendenza in media di Y da $X \rightarrow$
 $\rightarrow M(Y|x_1) = M(Y|x_2) = M(Y) = y_2$

ESERCIZIO 5.

Per il seguente gioco è prevista una posta di 2 Euro: si lanciano due dadi e se la somma dei numeri ottenuti è almeno pari a 9 si vincono 6 Euro.

1. Si stabilisca se il gioco è equo.



$$P(\text{vincita}) = \frac{10}{36} = 0.2778$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
posta	-2	0.7222	-1.4444
vincita - posta	4	0.2778	1.1111
		1	-0.3333

Il valore atteso della variabile casuale guadagno è negativo, quindi il gioco non è equo.

2. Nell'ipotesi di ripetere il gioco 5 volte si calcoli la probabilità di vincere almeno 2 volte.

Sia X la variabile casuale 'n. di vincite nelle 5 prove indipendenti'. $X \sim \text{Bin}(n = 5, p = 0.2778)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - \{P(X = 0) + P(X = 1)\} = \\ &= 1 - \binom{5}{0} 0.2778^0 0.7222^5 - \binom{5}{1} 0.2778^1 0.7222^4 = 1 - 0.1965 - 0.3779 = 0.4256. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 05.09.2013 (Tema 275)

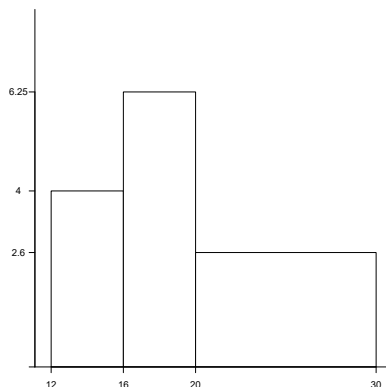
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Nella tabella si riportano i risultati di un'indagine sull'uso dei mezzi di trasporto da parte dei giovani nella città di Milano.

Mezzo di trasporto (T) \ Età (E)	12 + 16	16 + 20	20 + 30
bicicletta	3	6	9
mezzo pubblico	10	8	9
motorino	3	11	8

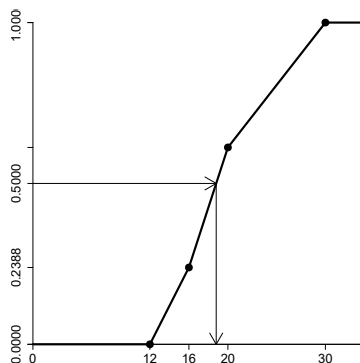
1. Si rappresenti graficamente la variabile marginale E .



2. Si calcolino, se possibile, moda, media aritmetica e mediana per la variabile T . Si calcolino, se possibile, moda, media aritmetica e si individui su un opportuno grafico la mediana per la variabile E .

Essendo il carattere T qualitativo sconnesso risulta possibile il solo calcolo della moda, che risulta 'mezzo pubblico'.

$$\begin{aligned} \text{Moda}(E) &= 16 + 20 \\ M(E) &= 19.7612 \end{aligned}$$



3. Si calcoli per la variabile T un opportuno indice di variabilità e se ne commenti il risultato.

$F_N = 0.8834$. Sussiste un elevato livello di eterogeneità.

4. Si calcoli un opportuno indice di connessione.

$\chi_N = 0.2042$. Il livello di connessione esistente tra le due variabili è basso.

ESERCIZIO 2.

Si consideri la seguente serie storica degli indici a base mobile relativa ai costi di un'azienda tra il 2008 e il 2012:

t	$NIBM$	$NIBF$
2008		1.0000
2009	1.20	1.2000
2010	0.55	0.6600
2011	0.85	0.5610
2012	1.05	0.5891

1. Si costruisca la serie dei numeri indici a base fissa con base 2008.

2. Si calcoli l'indice a base fissa: ${}_{2010}I_{2012}$ e si commenti il risultato.

${}_{2010}I_{2012} = 0.8925$ Tra il 2010 e il 2012 vi è stato un decremento complessivo dei costi del 10.75%.

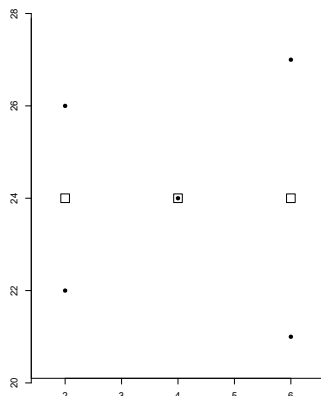
3. Si calcoli il tasso medio di variazione tra il 2010 e il 2012 e lo si confronti con l'indice calcolato al precedente punto. tasso medio di variazione -0.0553 . Tra il 2010 e il 2012 vi è stato un decremento medio annuo dei costi del 5.53%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	2	4	6	6
y_i	22	26	24	27	21

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 24, \hat{b} = 0, \rho^2 = 0 \quad \hat{c} = 24, R^2 = 0$ I due modelli coincidono.

3. Si fornisca la previsione in corrispondenza di $x = 10 \quad Y^*(10) = 24$.

4. Si determini il valore dei rapporti di correlazione $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

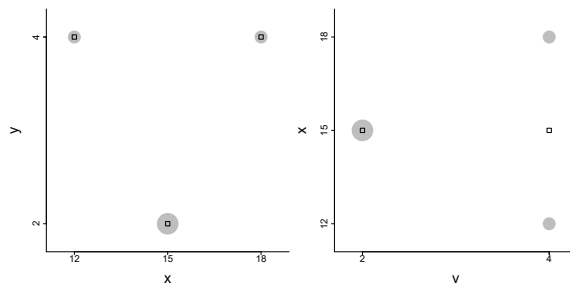
$\eta_{Y|X}^2 = 0$ sussiste indipendenza in media in quanto le medie condizionate coincidono.

$\eta_{X|Y}^2 = 1$ sussiste dipendenza funzionale di X da Y (a ogni valore di Y corrisponde un solo valore distinto di X).

ESERCIZIO 4.

1. Si completi la seguente tabella, giustificando opportunamente la risposta, in modo tale che $\chi^2 = 44$ e $\rho^2 = 0$

	$x_1 = 12$	$x_2 = \underline{\quad}$	$x_3 = 18$	
$y_1 = 2$		20		
$y_2 = 4$				
				$n = 44$



Dal momento che $\chi^2 = \chi_{\max}^2$, si richiede una situazione di dipendenza funzionale e assenza di correlazione lineare.

Ora, con riferimento alla tabella in oggetto, la situazione può essere realizzata imponendo $\eta_{Y|X}^2 = 1$, distribuzioni $Y|X$ degeneri, ed $\eta_{X|Y}^2 = 0$, indipendenza in media di Y da X ($\eta_{X|Y}^2 = 0 \rightarrow \rho^2 = 0$). $n_{11} = n_{13} = n_{22} = 0$ e $n_{21} = n_{23} = 12$.

2. Si calcoli il valore di $\eta_{Y|X}^2$. $\eta_{Y|X}^2 = 1$ in quanto le distribuzioni condizionate $Y|x_i$ sono degeneri

ESERCIZIO 5.

Si consideri l'esperimento di estrazione di 3 palline da un'urna contenente 16 palline rosse, 19 gialle e 5 blu.

1. Si calcoli la probabilità di estrarre zero palline rosse nell'ipotesi che le estrazioni vengano fatte senza reimmissione.

$$P(0 R) = P(\bar{R}_1 \cap \bar{R}_2 \cap \bar{R}_3) = P(\bar{R}_1) \cdot P(\bar{R}_2|\bar{R}_1) \cdot P(\bar{R}_3|\bar{R}_1 \cap \bar{R}_2) = \frac{24}{40} \frac{23}{39} \frac{22}{38} = 0.2049$$

oppure indicando con X la v.c. ipergeometrica 'n. palline rosse nelle 3 estrazioni senza reimmissione'

$$P(X = 0) = \frac{\binom{24}{3} \binom{16}{0}}{\binom{40}{3}} = 0.2049.$$

2. Si consideri, ora, l'esperimento di 50 estrazioni con reinserimento. Si calcoli la probabilità di ottenere almeno 31 palline di colore blu.

Indicata con X la variabile casuale 'numero di palline blu' nelle 50 estrazioni con reimmissione (prove indipendenti) risulta $X \sim \text{Bin}(n = 50, p = 0.125)$. $M(X) = 6.25$; $\text{Var}(X) = 5.4688$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 6.25, \sigma^2 = 5.4688)$ si ottiene

$$P(X \geq 31) = P\left(\frac{X - \mu}{\sigma} \geq \frac{31 - \mu}{\sigma}\right) = P\left(Z \geq \frac{31 - 6.25}{2.3385}\right) = P(Z \geq 10.58) = 0.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 16.09.2013 (Tema 275-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

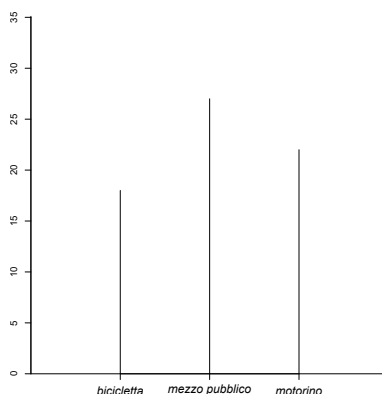
Studente _____ matricola _____

ESERCIZIO 1.

Nella tabella sottostante vengono riportati i risultati di un'indagine sull'uso dei mezzi di trasporto da parte dei giovani nella città di Milano.

Mezzo di trasporto (T) \ Età (E)	12 ÷ 16	16 ÷ 20	20 ÷ 30
bicicletta	3	6	9
mezzo pubblico	10	8	9
motorino	3	11	8

1. Si rappresenti graficamente la variabile marginale T .



2. Si confrontino con opportune misure di posizione e variabilità le distribuzioni condizionate $T|E = 12 \div 16$ e $T|E = 20 \div 30$.

$T|_{E=12 \div 16}$ Indice di posizione: $\text{Moda}(T|_{E=12 \div 16}) = \text{'mezzo pubblico'}$ Indice di variabilità: $F_N = 0.5625$

$T|_{E=20 \div 30}$ Indice di posizione: $\text{Moda}(T|_{E=20 \div 30})$ è bimodale Indice di variabilità: $F_N = 0.9613$

La fascia di età 20 ÷ 30 presenta un livello di eterogeneità nell'utilizzo dei mezzi di trasporto molto più elevato di quello della fascia di età 12 ÷ 16.

3. Si calcoli il rapporto di correlazione $\eta_{E|T}^2$ commentando il risultato.

$$\eta_{E|T}^2 = \frac{\text{Var}_T[M(E|T)]}{\text{Var}(E)} = \frac{0.6608}{19.734} = 0.0335.$$

ESERCIZIO 2.

Si consideri la seguente serie storica dei costi di un'azienda tra il 2008 e il 2012:

t	x_t	$NIBM$
2008	200	
2009	240	1.2000
2010	110	0.4583
2011	170	1.5455
2012	210	1.2353

1. Si costruisca la serie dei numeri indici a base mobile.

2. Si calcoli l'indice a base fissa: ${}_{2012}I_{2010}$ e si commenti il risultato.

${}_{2012}I_{2010} = 0.5238$ L'ammontare dei costi nel 2010 era circa il 52% di quello registrato nel 2012

3. Si calcoli il tasso medio di variazione tra il 2009 e il 2012 commentando il risultato ottenuto.

tasso medio di variazione -0.0435 . Tra il 2009 e il 2012 si è registrato un decremento medio annuo del 4.35%.

4. Sapendo che ${}_{2012}I_{2013} = 1$ si calcolino i costi del 2013.

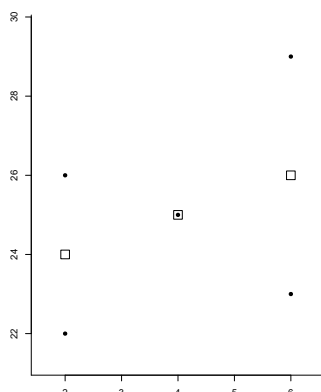
$x_{2013} = x_{2012} = 210$ dal momento che non vi è stata alcuna variazione tra il 2012 e il 2013

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	2	4	6	6
y_i	22	26	25	29	23

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = cX$$

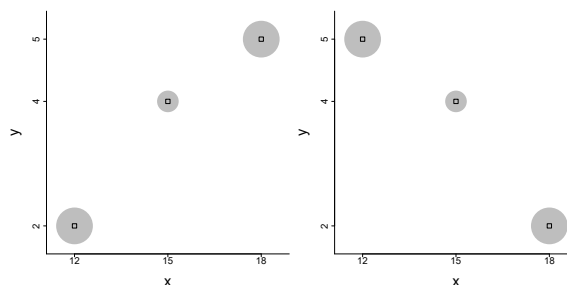
$$\hat{a} = 23, \hat{b} = 0.5, M(\hat{E}^2) = 5.2 \quad \hat{c} = 5.2917, M(\hat{E}^2) = 93.3667 \quad \text{Il modello migliore è } Y^* = 23 + 0.5X.$$

3. Si fornisca la previsione in corrispondenza di $x = 10$ $Y^*(10) = 28$

ESERCIZIO 4.

1. Si completi la seguente tabella, giustificando opportunamente la risposta, in modo tale che $\chi^2 = 88$ e $\rho^2 = 1$

	$x_1 = 12$	$x_2 = 15$	$x_3 = 18$	
$y_1 = 2$				
$y_2 = 4$		20		
$y_2 = 5$				
				$n = 44$



Impossibile, in quanto $\rho^2 = 1$ implica una situazione di dipendenza funzionale lineare, compatibile con $\chi^2 = 88$ ma che non può realizzarsi in quanto non è possibile disporre le medie condizionate su una retta.

ESERCIZIO 5.

Si consideri l'esperimento di estrazione di 9 palline da un'urna contenente 16 palline rosse, 19 gialle e 5 blu.

1. Si calcoli la probabilità di estrarre 2 palline rosse, 1 gialla e 6 blu nell'ipotesi che le estrazioni vengano fatte senza reimmissione.

$P(2R \cap 1G \cap 6B) = 0$ in quanto non è possibile estrarre 6 palline blu se le estrazioni vengono effettuate senza reimmissione; l'evento è quindi impossibile.

2. Si calcoli la probabilità di estrarre 3 palline rosse, 4 gialle e 2 blu nell'ipotesi che le estrazioni vengano fatte con reimmissione.

$$P(3R \cap 4G \cap 2B) = \text{n. sequenze} \cdot P(\text{generica sequenza}) = \frac{9!}{3! \cdot 4! \cdot 2!} \left(\frac{16}{40}\right)^3 \left(\frac{19}{40}\right)^4 \left(\frac{5}{40}\right)^2 = 0.0641.$$

3. Si calcoli la probabilità di ottenere almeno 12 palline di colore blu in 100 estrazioni con reimmissione.

Indicata con X la variabile casuale 'numero di palline blu' nelle 100 estrazioni con reimmissione (prove indipendenti) risulta $X \sim \text{Bin}(n = 100, p = 0.125)$. $M(X) = 12.5$; $\text{Var}(X) = 10.9375$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 12.5, \sigma^2 = 10.9375)$ si ottiene

$$P(X \geq 12) = P\left(\frac{X - \mu}{\sigma} \geq \frac{12 - \mu}{\sigma}\right) = P\left(Z \geq \frac{12 - 12.5}{3.3072}\right) = P(Z \geq -0.15) = P(Z \leq 0.15) = 0.5596.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 16.01.2014 (Tema 276)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

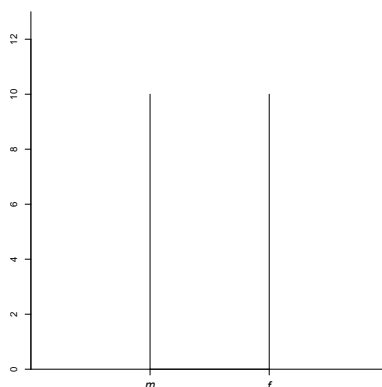
La tabella seguente riporta le misure delle altezze al garrese (H) di 20 renne adulte, distinte per genere ($G = m, f$).

m	m	m	m	m	m	m	m	m	m	f	f	f	f	f	f	f	f	f	f
158	162	165	170	180	182	184	184	188	190	153	155	158	160	160	168	170	180	180	182

1. Si costruisca una tabella a doppia entrata, usando per le altezze le seguenti classi ($150 \div 160, 160 \div 180, 180 \div 190$).

$G \setminus H$	$150 \div 160$	$160 \div 180$	$180 \div 190$	
m	1	4	5	10
f	5	4	1	10
	6	8	6	20

2. Si rappresenti graficamente la variabile marginale G .



3. Si individui il valore della mediana della marginale H .

La distribuzione H è simmetrica rispetto a 170, quindi $h_{0.5} = 170$.

4. Si fornisca un indice di variabilità di G .

La mutabile G presenta massima eterogeneità, quindi $F_N = G_N = 1$.

5. Si calcoli un opportuno indice di connessione tra H e G .

$\chi_N = 0.5164$. Il livello di connessione esistente tra le due variabili è medio.

ESERCIZIO 2.

Con riferimento alla seguente serie di Numeri Indice a Base Mobile:

t	$NIBM$	x_t
2008		102.9748
2009	0.92	94.7368
2010	0.95	90
2011	1.05	94.5
2012	1.08	102.06

1. Si ricostruisca la serie storica dei dati x_t sapendo che $x_{2010} = 90$.

2. Si calcoli il tasso medio di variazione nel periodo 2008-2012 interpretando il risultato ottenuto.

tasso medio di variazione -0.0022 . Tra il 2008 e il 2012 vi è stato un decremento medio annuo dei costi dello 0.22%.

3. Si calcoli il numero indice al tempo 2012 con base 2010 e si commenti il risultato.

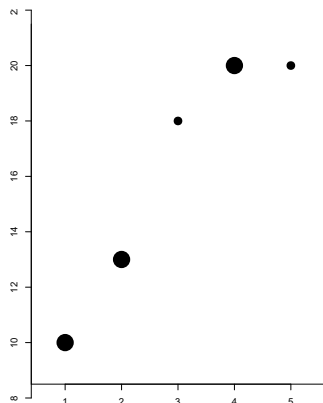
${}_{2010}I_{2012} = 1.134$. Tra il 2010 e il 2012 vi è stato un incremento dei costi del 13.4%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	2	2	3	4	4	5
y_i	10	10	13	13	18	20	20	20

1. Si rappresenti graficamente Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = \alpha + \beta \sqrt{X}$$

$$\hat{a} = 7.5161, \hat{b} = 2.9032, \rho^2 = 0.9332 \quad \hat{\alpha} = 0.5377, \hat{\beta} = 9.3539, \rho^2 = 0.9569$$

3. Si individui il modello migliore Il modello migliore è $Y^* = 0.5377 + 9.3539\sqrt{X}$.

4. Si motivi perché $\eta_{Y|X}^2 > \eta_{X|Y}^2$

Sussiste dipendenza funzionale di Y da X , quindi $\eta_{Y|X}^2 = 1$. $0 < \eta_{X|Y}^2 < 1$ in quanto non sussiste indipendenza in media e nemmeno dipendenza funzionale di X da Y (X assume entrambi i valori 4 e 5, in corrispondenza di $Y = 20$).

ESERCIZIO 4.

Si completino le seguenti tabelle, giustificando la risposta, in modo che siano soddisfatte le condizioni indicate.

	$y_1 = 10$	$y_2 = 15$
$x_1 = 1$	11	
$x_2 = 2$		10

a) $\eta_{Y|X}^2 = \rho^2 > 0$

$$\eta_{Y|X}^2 = \rho^2 \rightarrow \text{medie condiz. allineate:}$$

$M(Y|x_1)$ e $M(Y|x_2)$ sono allineate
qualsiasi siano le frequenze assegnate.

Occorre escludere la combinazione

$n_{12} = 11, n_{21} = 10$, in quanto si avrebbe

$$M(Y|x_1) = M(Y|x_2) \text{ con}$$

$$\eta_{Y|X}^2 = \rho^2 = 0.$$

	$y_1 = 10$	$y_2 = 15$
$x_1 = 1$	11	
$x_2 = 2$	5	

b) $\chi^2 = 0$

$\chi^2 = 0$ indep. stocastica

$$n_{12} = 11c, n_{22} = 5c$$

con $c > 0$ intero

	$y_1 = 10$	$y_2 = 15$
$x_1 = 1$	11	
$x_2 = 2$		

c) $\rho = -1$

$\rho = -1$ dip. funz. lineare inversa
tra X e Y

impossibile; si verificherebbe con
 $n_{11} = 0, n_{12} > 0, n_{21} > 0, n_{22} = 0$

ESERCIZIO 5.

Vengono lanciati 3 dadi regolari.

1. Si calcoli la probabilità che, essendosi presentati 3 numeri uguali, questi siano i numeri da 1 a 4.

Si indichino con A e B gli eventi $A =$ 'si presentano i numeri da 1 a 4' e $B =$ 'si presentano 3 numeri uguali'.

L'evento $(A \cap B)$ corrisponde a 3 numeri 1 oppure 3 numeri 2, ..., 3 numeri 4 e ha probabilità $P(A \cap B) = 4 \left(\frac{1}{6}\right)^3$. Invece $P(B) = 6 \left(\frac{1}{6}\right)^3$. Applicando, ora, la formula della probabilità condizionata si ottiene

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{4 \left(\frac{1}{6}\right)^3}{6 \left(\frac{1}{6}\right)^3} = \frac{4}{6} = 0.6667$$

2. Si calcoli la probabilità che su 51 lanci in almeno 10 tutti e 3 i dadi presentino un numero pari.

L'evento tutti e 3 i dadi presentino un numero pari in un singolo lancio ha probabilità $p = \left(\frac{3}{6}\right)^3 = 0.5^3 = 0.125$.

Indicata con X la variabile casuale 'numero di volte che tutti e 3 i dadi presentino un numero pari' nei 51 lanci (prove indipendenti) risulta $X \sim \text{Bin}(n = 51, p = 0.125)$. $M(X) = 6.375$; $\text{Var}(X) = 5.5781$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 6.375, \sigma^2 = 5.5781)$ si ottiene

$$P(X \geq 10) = P\left(\frac{X - \mu}{\sigma} \geq \frac{10 - \mu}{\sigma}\right) = P\left(Z \geq \frac{10 - 6.375}{2.3618}\right) = P(Z \geq 1.53) = 1 - P(Z < 1.53) = 1 - P(Z \leq 1.53) = 0.063.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 30.01.2014 (Tema 277)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

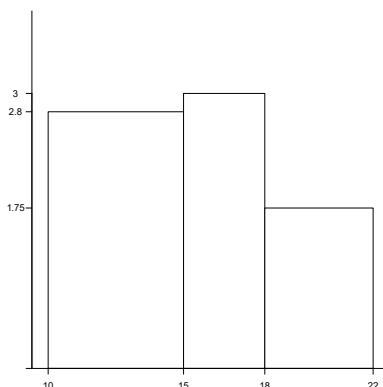
Studente _____ matricola _____

ESERCIZIO 1.

Si riporta la valutazione della qualità (Q) e il prezzo (P) di uno stesso piatto proposto da 30 ristoranti italiani:

$Q \setminus P$	10 - 15	15 - 18	18 - 22
<i>discreta</i>	9	2	0
<i>buona</i>	5	7	5
<i>eccellente</i>	0	0	2

1. Fornire la rappresentazione grafica della distribuzione marginale P .



2. Calcolare la mediana e una misura di variabilità della marginale Q .

$$q_{0.5} = \textit{buona}$$

Gli elementi che occupano le posizioni 15 e 16 nella serie dei dati ordinati corrispondono entrambi alla categoria *buona*.

$F_N = 0.5641$ La mutabile Q presenta un livello medio di eterogeneità.

3. Dopo aver valutato, senza fare calcoli, se Q e P sono stocasticamente indipendenti, calcolare un indice di connessione.

Q e P non sono stocasticamente indipendenti in quanto la presenza di frequenze congiunte nulle esclude che le distribuzioni condizionate siano somiglianti.

$\chi_N = 0.4937$. Il livello di connessione esistente tra le due variabili è medio.

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi medi annui di 2 beni di uso quotidiano nel periodo 2002 – 2006:

t	<i>bene A</i>	<i>bene B</i>	$NIBF(2003)$
2002	99.4	77.0	0.8474
2003	117.3	65.8	1
2004	129.8	47.2	1.1066
2005	149.3	35.8	1.2728
2006	158.1	33.8	1.3478

1. Costruire la serie dei numeri indici a base fissa con base 2003 per il bene A .

2. Calcolare il tasso medio di variazione per il prezzo del bene B tra il 2002 e il 2005, interpretando il risultato.

tasso medio di variazione -0.2253 . Tra il 2008 e il 2012 vi è stato un decremento medio annuo dei costi del 22.53%.

3. Note le quantità $Aq_{2003} = 1500$ e $Bq_{2003} = 8500$, calcolare l'indice di Laspeyres dei prezzi al 2006 con base 2003.

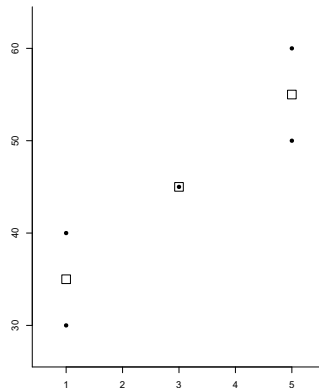
$^P_{2003}I^L_{2006} = 0.7133$. Tra il 2003 e il 2006 si è registrato una variazione negativa del 28.67%.

ESERCIZIO 3.

La tabella seguente riporta le osservazioni della variabile Y in corrispondenza di 5 osservazioni della X :

x_i	1	1	3	5	5
y_i	30	40	45	50	60

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = \alpha \cdot X^2$$

$$\hat{a} = 30, \hat{b} = 5, M(\hat{E}^2) = 20 \quad \hat{\alpha} = 2.4194, M(\hat{E}^2) = 564.52 \quad \text{Il modello migliore è } Y^* = 30 + 5X.$$

3. Si indichi il valore assunto dal rapporto di correlazione $\eta_{Y|X}^2$, commentando il risultato.

Si osserva che le medie condizionate $M(Y|x_i)$ sono allineate; quindi $\eta_{Y|X}^2 = \rho^2 = 0.8$.

ESERCIZIO 4.

Si completino le seguenti tabelle, giustificando la risposta, in modo che siano soddisfatte le condizioni indicate.

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$	0	11	
$x_2 = 2$		0	

$$a) \eta_{X|Y}^2 = 1, \rho^2 = 0$$

$\eta_{X|Y}^2 = 1, \rho^2 = 0$ dip. funz. di X da Y e indep. lin.

ricordando che $\eta_{Y|X}^2 = 0 \rightarrow \rho^2 = 0$

$$n_{13} = 0, n_{21} = n_{23} > 0$$

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$		11	
$x_2 = 2$		11	

$$b) \eta_{Y|X}^2 = \rho^2 \neq 0$$

$\eta_{Y|X}^2 = \rho^2 \rightarrow$ medie condiz. allineate.

Le due medie condizionate saranno allineate qualsiasi siano le frequenze assegnate. Occorre escludere una delle combinazioni $n_{11} = n_{13}$, $n_{21} = n_{23}$. Nel caso fossero entrambe verificate, $Y|x_1$ e $Y|x_2$ risulterebbero simmetriche con $M(Y|x_1) = M(Y|x_2)$ ed $\eta_{Y|X}^2 = \rho^2 = 0$.

ESERCIZIO 5.

Date due differenti urne, U_1 e U_2 , contenenti palline numerate, un esperimento casuale consiste nell'estrarre una pallina da ciascuna urna. Sapendo che nell'urna U_1 vi sono 5 palline con numero pari (2,4,6,8,10) e che nell'urna U_2 vi sono 2 palline con il numero 1, 2 con numero 3 e 2 con il numero 5:

1. costruire la distribuzione di probabilità delle variabili casuali $X = \text{"risultato estrazione urna } U_1\text{"}$ e $Y = \text{"risultato estrazione urna } U_2\text{"}$;

x	$P(X = x)$	y	$P(Y = y)$
2	0.2	1	1/3
4	0.2	3	1/3
6	0.2	5	1/3
8	0.2		
10	0.2		

2. calcolare $P(\{X = 10\} \cap \{Y = 1\})$, ovvero la probabilità che sia estratto il numero 10 da U_1 e il numero 1 da U_2 ; Dal momento che le estrazioni dalle due urne avvengono in maniera indipendente abbiamo che

$$P(\{X = 10\} \cap \{Y = 1\}) = P(\{X = 10\}) \cdot P(\{Y = 1\}) = 0.2 \cdot 0.3333 = 0.06667.$$

3. calcolare la probabilità che, su 100 estrazioni con reinserimento dalla sola urna U_2 , almeno 21 presentino il numero 3. Indicata con X la variabile casuale 'numero di volte che si presenta il numero 3' nelle 100 estrazioni con reinserimento (prove indipendenti) risulta $X \sim \text{Bin}(n = 100, p = 0.3333)$. $M(X) = 33.3333$; $\text{Var}(X) = 22.2222$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 33.3333, \sigma^2 = 22.2222)$ si ottiene

$$P(X \geq 21) = P\left(\frac{X - \mu}{\sigma} \geq \frac{21 - \mu}{\sigma}\right) = P\left(Z \geq \frac{21 - 33.3333}{4.714}\right) = P(Z \geq -2.62) = P(Z \leq 2.62) = 0.9956.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 13.02.2014 (Tema 278)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

Con riferimento alla distribuzione congiunta delle variabili X e Y sono note le sole frequenze congiunte riportate nella seguente tabella a doppia entrata:

$X \setminus Y$	y_1	y_2	y_3	
x_1		11		
x_2		11		

1. Si completi la tabella in modo tale che le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ siano caratterizzate rispettivamente da minima e massima eterogeneità. Si indichi il numero totale delle unità statistiche.

$X \setminus Y$	y_1	y_2	y_3	
x_1	0	11	0	11
x_2	11	11	11	33
	11	22	11	44

2. Sapendo che $y_1 = 10$, $y_2 = 15$ e $y_3 = 20$, si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione, variabilità e asimmetria, commentando i risultati ottenuti.

$Y|x_1$: Indice di posizione: $M(Y|x_1) = 15$ Indice di variabilità: $CV(Y|x_1) = 0$

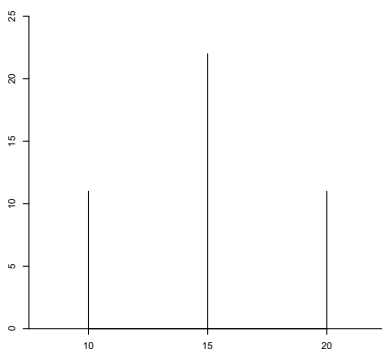
Indice asimmetria. La distribuzione è simmetrica, quindi gli indici di asimmetria assumono tutti valore nullo.

$Y|x_2$: Indice di posizione: $M(Y|x_2) = 15$ Indice di variabilità: $CV(Y|x_2) = 0.2722$

Indice asimmetria. La distribuzione è simmetrica, quindi gli indici di asimmetria assumono tutti valore nullo.

Le distribuzioni sono simmetriche, hanno eguale media; $Y|x_2$ presenta un livello di variabilità superiore a $Y|x_1$, distribuzione degenera.

3. Si rappresenti graficamente la distribuzione marginale Y .



ESERCIZIO 2.

Della seguente serie storica sono note solo alcune informazioni.

t	x_t	x_t	$NIBF$	$NIBM$
1	200	200	1.000	
2	x_2	221	1.105	1.1050
3	180	180	0.900	0.8145
4	x_4	171	0.855	0.9500
5	160	160	0.800	0.9357

1. Si ricostruisca la serie storica sapendo che tra il tempo 1 e il tempo 2 vi è stato un incremento di 21 e che tra $t = 3$ e $t = 4$ vi è stata una diminuzione del 5%.

2. Si costruiscano le serie dei numeri indici a base fissa, con base $t = 1$, e a base mobile.

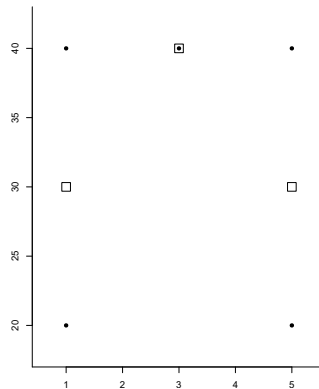
3. Si calcoli il tasso medio di variazione di x_t tra l'istante temporale 1 e l'istante 3, interpretando il risultato. tasso medio di variazione -0.0513 . Tra l'istante temporale 1 e l'istante 3 vi è stato un decremento medio del 5.13%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	3	5	5
y_i	20	40	40	40	20

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 32$, $\hat{b} = 0$, $\rho^2 = 0$ $\hat{c} = 32$, indice adattamento $R^2 = 0$ I due modelli sono equivalenti.

3. Si motivi perché $\eta_{Y|X}^2 > \eta_{X|Y}^2$.

Dal momento che $M(X|y = 20) = M(X|y = 40) = 3$, sussiste indipendenza in media di X da Y , quindi $\eta_{X|Y}^2 = 0$.
 $0 < \eta_{Y|X}^2 < 1$ in quanto non sussiste indipendenza in media e nemmeno dipendenza funzionale di Y da X .

ESERCIZIO 4.

Un'urna è composta da gettoni rossi e neri in proporzione 0.15 e 0.85. Ogni gettone rosso reca impresso il numero 1, i neri il numero 2. Si consideri l'esperimento casuale $\mathcal{G}_3$ = 'estrazione con reimmissione di 3 gettoni dall'urna'.

1. Si costruisca lo spazio probabilistico associato all'esperimento $\mathcal{G}_3$.

RRR	<i>oppure</i>	111
RRN, RNR, NRR		112, 121, 211
NNR, NRN, RNN		221, 212, 122
NNN		222

2. Si costruisca la variabile casuale X = 'somma dei valori impressi sui gettoni estratti' e si calcoli la probabilità che X assuma valori maggiori o uguali a 5.

x	$P_{\Omega}(X^{-1}(x))$	$P(X = x)$
3	$P(RRR)$	0.003375
4	$P(RRN \cup RNR \cup NRR)$	0.057375
5	$P(NNR \cup NRN \cup RNN)$	0.325125
6	$P(NNN)$	0.614125

$$P(X \geq 5) = 0.325125 + 0.614125 = 0.93925.$$

3. Si calcoli la probabilità di ottenere almeno due gettoni rossi nelle 3 estrazioni.

Sia X la variabile casuale 'n. di gettoni rossi nelle 3 estrazioni (con reimmissione) (prove indipendenti)'. $X \sim \text{Bin}(n = 3, p = 0.15)$.

$$P(X \geq 2) = \binom{3}{2} 0.15^2 0.85^1 + \binom{3}{3} 0.15^3 0.85^0 = 0.003375 + 0.057375 = 0.06075.$$

4. Si calcoli la probabilità di ottenere non più di un gettone rosso in 26 estrazioni con reimmissione.

Sia X la variabile casuale 'n. di gettoni rossi nelle 26 estrazioni (con reimmissione) (prove indipendenti)'. $X \sim \text{Bin}(n = 26, p = 0.15)$.

$$P(X \leq 1) = \binom{26}{0} 0.15^0 0.85^{26} + \binom{26}{1} 0.15^1 0.85^{25} = 0.01462 + 0.06707 = 0.08169.$$

Si osserva come non valga l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale in quanto $np = 3.9 < 5$.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 05.06.2014 (Tema 279-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

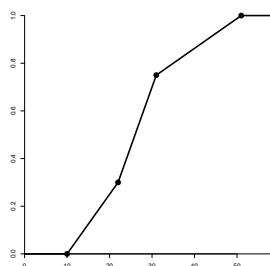
Studente _____ matricola _____

ESERCIZIO 1.

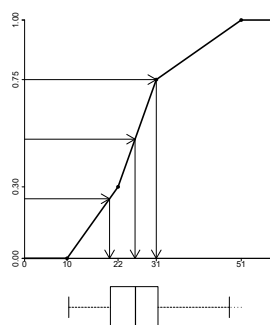
Si riportano le classi e le densità di frequenza della variabile statistica X :

$h_{i-1} \rightarrow h_i$	d_i	f_i	F_i
10 $\rightarrow$ 22	0.025	0.3	0.3
22 $\rightarrow$ 31	0.05	0.45	0.75
31 $\rightarrow$ 51	0.0125	0.25	1

1. Si ricostruisca la distribuzione delle frequenze cumulate relative di X e se ne dia una rappresentazione grafica.



2. Si rappresenti X mediante un grafico Box & Whiskers Plot.



$$x_{0.25} = 20, x_{0.50} = 26, x_{0.75} = 31, \text{lunghezza teorica baffo} = 1.5 \cdot (31 - 20) = 16.5, \text{baffo inferiore} = 10, \text{baffo superiore} = 47.5$$

3. Si calcolino opportuni indici di posizione e di variabilità per la variabile statistica X , al fine di confrontarla con la variabile Y della quale sono noti solo i seguenti indici sintetici $M(Y) = 50$ e $Var(Y) = 100$.

indice di posizione: $M(X) = 26.975$ indice di variabilità $CV(X) = 0.3426$

indice di posizione: $M(Y) = 50$ indice di variabilità $CV(Y) = 0.2$

Il livello medio di X è inferiore a quello di Y . X presenta un livello di variabilità superiore a Y .

ESERCIZIO 2.

Sia data la seguente serie storica.

t	x_t
1	200
2	221
3	180
4	171
5	160

1. Si calcoli il numero indice al tempo 4 con base 3, commentando il risultato ottenuto.

${}_3I_4 = 0.95$ Tra l'istante 3 e l'istante 4 la serie ha subito un calo del 5%.

2. Si calcoli il tasso medio di variazione di x_t tra l'istante temporale 2 e l'istante 5, interpretando il risultato.

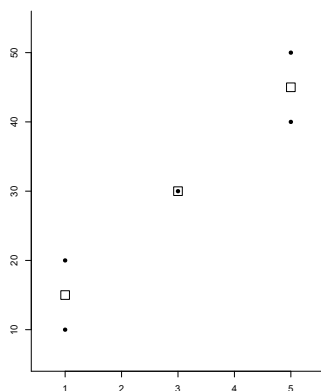
tasso medio di variazione -0.01021 . Tra l'istante 2 e l'istante 5 la serie ha subito un calo medio del 10.21%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	3	5	5
y_i	10	20	30	40	50

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = cX$$

$$\hat{a} = 7.5, \hat{b} = 7.5, M(\hat{E}^2) = 20 \quad \hat{c} = 9.3443, \text{ indice adattamento } M(\hat{E}^2) = 34.754$$

Il modello migliore è $Y^* = 7.5 + 7.5X$.

3. Si indichino i valori assunti da $\eta_{Y|X}^2$ ed $\eta_{X|Y}^2$.

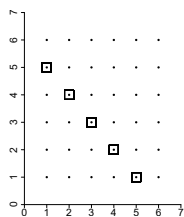
$\eta_{Y|X}^2 = \rho^2 = 0.9$ in quanto le medie condizionate $M(Y|X)$ sono allineate

$\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y .

ESERCIZIO 4.

Per il seguente gioco è prevista una posta di 2 Euro: si lanciano due dadi e se la somma dei numeri ottenuti è pari a 6 si vincono 6 Euro.

1. Si stabilisca se il gioco è equo.



$$P(\text{vincita}) = \frac{5}{36} = 0.1389$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
posta	-2	0.8611	-1.7222
vincita - posta	4	0.1389	0.5556
		1	-1.1667

Il valore atteso della variabile casuale guadagno è negativo, quindi il gioco non è equo.

2. Nell'ipotesi di ripetere il gioco 8 volte si calcoli la probabilità di vincere almeno 3 volte.

Sia X la variabile casuale 'n. di vincite nelle 8 prove indipendenti'. $X \sim \text{Bin}(n = 8, p = 0.1389)$.

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} = \\ &= 1 - \binom{8}{0} 0.1389^0 0.8611^8 - \binom{8}{1} 0.1389^1 0.8611^7 - \binom{8}{2} 0.1389^2 0.8611^6 = \\ &= 1 - 0.3023 - 0.3901 - 0.2202 = 0.0874. \end{aligned}$$

3. Nell'ipotesi di ripetere il gioco 200 volte si calcoli la probabilità di vincere almeno 30 volte.

Indicata con X la variabile casuale 'numero di volte che si presenta l'evento vincita nelle 200 ripetizioni del gioco (prove indipendenti) risulta $X \sim \text{Bin}(n = 200, p = 0.1389)$. $M(X) = 27.7778$; $\text{Var}(X) = 23.9198$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 27.7778, \sigma^2 = 23.9198)$ si ottiene

$$P(X \geq 30) = P\left(\frac{X - \mu}{\sigma} \geq \frac{30 - \mu}{\sigma}\right) = P\left(Z \geq \frac{30 - 27.7778}{4.8908}\right) = P(Z \geq 0.45) = 1 - P(Z \leq 0.45) = 1 - 0.6736 = 0.3264.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 26.06.2014 (Tema 279)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta i dati, relativi ad un anno solare, per un campione di 9 Stati UE classificati sulla base delle variabili X : area geografica, con modalità CN = Centro Nord e CS = Centro Sud e Y : PIL pro capite in migliaia di euro.

X	CN	CN	CN	CN	CN	CS	CS	CS	CS
Y	41.9	30.5	32.1	49.1	28.8	25.3	22.2	32.7	21.8

1. Si confronti la variabilità della variabile Y nei gruppi definiti dalla variabile X (CN e CS).

$$CV(Y|X = CN) = 0.2132$$

$$CV(Y|X = CS) = 0.1715$$

Il PIL dei paesi nell'area Centro Nord presenta una maggiore variabilità rispetto a quello dei paesi nell'area Centro Sud.

2. Si calcolino la varianza between e la varianza within. Si commentino i risultati ottenuti.

$$\sigma_{Between}^2 = \frac{1}{9} \left\{ (\mu_{CN} - \mu)^2 \cdot 5 + (\mu_{CS} - \mu)^2 \cdot 4 \right\} = \frac{1}{9} (\mu_{CN}^2 \cdot 5 + \mu_{CS}^2 \cdot 4) - \mu^2 = \frac{1}{9} (36.48^2 \cdot 5 + 25.5^2 \cdot 4) - 31.6^2 = 29.768$$

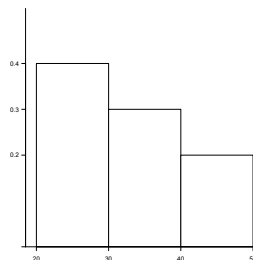
$$\sigma_{Within}^2 = \frac{1}{9} (\sigma_{CN}^2 \cdot 5 + \sigma_{CS}^2 \cdot 4) = \frac{1}{9} (60.5136 \cdot 5 + 19.115 \cdot 4) = 42.1142.$$

Il rapporto di correlazione $\eta_{Y|area\ geografica}^2$ risulta pari a 0.4141.

Sussiste, quindi, un livello di dipendenza medio basso del PIL rispetto all'area geografica di appartenenza dei paesi.

3. Si costruisca la distribuzione di frequenze della variabile Y con riferimento alle classi ($20 \rightarrow 30, 30 \rightarrow 40, 40 \rightarrow 50$) e si rappresenti graficamente tale distribuzione.

$h_{i-1} \rightarrow h_i$	n_i	a_i	d_i
$20 \rightarrow 30$	4	10	0.4
$30 \rightarrow 40$	3	10	0.3
$40 \rightarrow 50$	2	10	0.2



4. Costruita la tabella a doppia entrata (X, Y), con Y riclassificata come nel punto 3, si calcoli un opportuno indice di connessione tra X e Y .

$X \backslash Y$	$20 \rightarrow 30$	$30 \rightarrow 40$	$40 \rightarrow 50$
CN	1	2	2
CS	3	1	0

$\chi_N = 0.6021$. Il livello di connessione esistente tra le due variabili è medio.

ESERCIZIO 2.

Si consideri la seguente serie dei numeri indici a base fissa, con base 2009:

anno	$NIBF$	$NIBM$
2009	1	—
2010	0.90	0.9000
2011	0.93	1.0333
2012	1.02	1.0968
2013	1.11	1.0882

1. Si costruisca la corrispondente serie dei numeri indici a base mobile.

2. Sapendo che $x_{2012} = 265$, si calcoli x_{2013} .

$$x_{2013} = x_{2012} \cdot {}_{2012}I_{2013} = 265 \cdot 1.0882 = 288.3824$$

3. Si calcoli il tasso medio di variazione dal 2009 al 2013 e si commenti il risultato ottenuto.

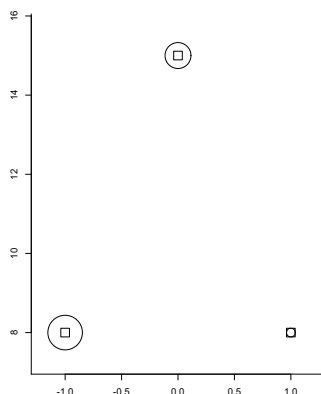
tasso medio di variazione 0.02643. Tra il 2009 e il 2013 si è osservato un incremento annuo medio del 2.64%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	0	-1	0	1	-1	0	-1	-1
y_i	15	8	15	8	8	15	8	8

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + bX^2 \quad II) Y^* = \alpha + \beta X^3$$

$$\hat{a} = 15, \hat{b} = -7, \text{ indice adattamento } \rho^2 = 1 \quad \hat{\alpha} = 11.3871, \hat{\beta} = 2.0323, \text{ indice adattamento } \rho^2 = 0.1742$$

Il modello migliore è $Y^* = 15 - 7X^2$.

3. Per il modello migliore, si fornisca la previsione in corrispondenza di $X = 2$.

$$Y^*(2) = 15 - 7 \cdot 2^2 = 15 - 7 \cdot 4 = 15 - 28 = -13.$$

4. Si determini il valore del rapporto di correlazione $\eta_{Y|X}^2$ motivando la risposta.

$\eta_{Y|X}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X .

ESERCIZIO 4.

Si considerino 500 studenti di una Facoltà ripartiti secondo tre diversi corsi di studio (A, B, C) e secondo lo stato (in corso = IC, fuori corso = FC):

	A	B	C
IC	200	130	100
FC	35	25	10

1. Si calcoli la probabilità che, estraendo uno studente a caso, sia fuori corso.

$$P(FC) = \frac{35+25+10}{500} = \frac{70}{500} = 0.14.$$

2. Si calcoli la probabilità che, estraendo uno studente a caso, sia iscritto al corso B.

$$P(B) = \frac{130+25}{500} = \frac{155}{500} = 0.31.$$

3. Si determini la probabilità che, estraendo uno studente a caso, sia fuori corso oppure sia iscritto al corso B.

$$P(FC \cup B) = P(FC) + P(B) - P(FC \cap B) = \frac{70}{500} + \frac{155}{500} - \frac{25}{500} = 0.14 + 0.31 - .05 = 0.4 \left(= \frac{130+35+25+10}{500} = \frac{200}{500} \right).$$

4. Si calcoli la probabilità che uno studente, estratto a caso, sia fuori corso sapendo che è iscritto al corso C.

$$P(FC|C) = \frac{P(FC \cap C)}{P(C)} = \frac{\frac{10}{500}}{\frac{110}{500}} = \frac{10}{110} = 0.0909.$$

5. Si calcoli la probabilità che, estratti senza reinserimento 12 studenti, la metà sia iscritta al corso A.

$$\begin{aligned} P(6 A) &= \binom{12}{6} P(A_1 \cap A_2 \cap \dots \cap A_6 \cap \bar{A}_7 \cap \bar{A}_8 \cap \dots \cap \bar{A}_{12}) = \\ &= \binom{12}{6} P(A_1) \cdot P(A_2|A_1) \cdot \dots \cdot P(A_6|A_1 \cap \dots \cap A_5) \cdot P(\bar{A}_7|A_1 \cap \dots \cap A_6) \cdot P(\bar{A}_8|A_1 \cap \dots \cap A_6 \cap \bar{A}_7) \cdot \dots \cdot P(\bar{A}_{12}|A_1 \cap \dots \cap A_6 \cap \bar{A}_7 \cap \dots \cap \bar{A}_{11}) = \\ &= \frac{12!}{6! \cdot 6!} \cdot \frac{235}{500} \cdot \frac{234}{499} \cdot \frac{233}{498} \cdot \frac{232}{497} \cdot \frac{231}{496} \cdot \frac{230}{495} \cdot \frac{265}{494} \cdot \frac{264}{493} \cdot \frac{263}{492} \cdot \frac{262}{491} \cdot \frac{261}{490} \cdot \frac{260}{489} = 0.2234 \end{aligned}$$

oppure indicando con X la v.c. ipergeometrica 'n. studenti iscritti al corso A nelle 12 estrazioni senza reimmissione'

$$P(X = 6) = \frac{\binom{235}{6} \binom{265}{6}}{\binom{500}{12}} = 0.2234.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 10.07.2014 (Tema 279-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta la distribuzione congiunta della variabile statistica doppia (X, Y) per un campione di 150 unità statistiche.

$X \setminus Y$	10 - 15	15 - 25	25 - 30	$n_{i\bullet}$
x_1	22	60	18	100
x_2	11	30	9	50
$n_{\bullet j}$	33	90	27	150

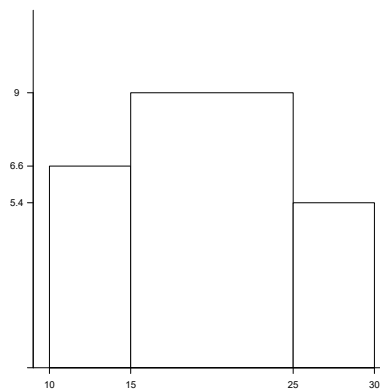
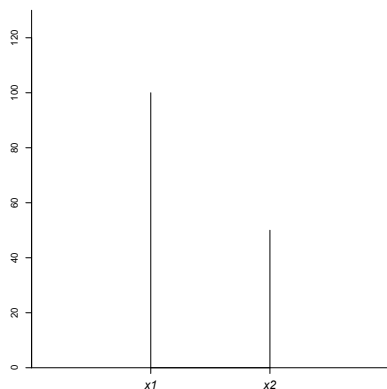
1. Si confrontino mediante opportuni indici di posizione e di variabilità le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 19.7$ Indice di variabilità: $CV(Y|x_1) = 0.2403$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 19.7$ Indice di variabilità: $CV(Y|x_2) = 0.2403$

Osservando la proporzionalità delle frequenze condizionate si evince che sussiste indipendenza stocastica; quindi le distribuzioni condizionate sono caratterizzate dalla stessa media e dallo stesso livello di variabilità.

2. Si rappresentino graficamente le distribuzioni marginali X e Y .



3. Si calcoli un opportuno indice di connessione tra X e Y .

$\chi_N = 0$ in quanto sussiste indipendenza stocastica.

ESERCIZIO 2.

Si consideri la seguente serie storica:

anno	x_t	NIBF	NIBM
2009	200	1	-
2010	180	0.90	0.90
2011	186	0.93	1.0333
2012	204	1.02	1.0968
2013	222	1.11	1.0882

1. Si costruiscano le serie dei numeri indici a base fissa (2009) e a base mobile.

2. Si calcoli ${}_{2010}I_{2013}$ commentando il risultato ottenuto.

${}_{2010}I_{2013} = 1.2333$ Tra il 2010 e il 2013 si è osservato un incremento del 23% della grandezza oggetto di studio.

3. Si calcoli il tasso medio di variazione dal 2010 al 2013 e si commenti il risultato ottenuto.

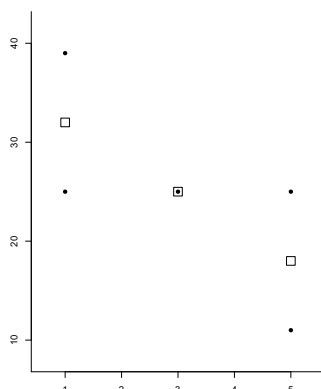
tasso medio di variazione 0.0724. Tra il 2010 e il 2013 si è osservato un incremento medio annuo del 7.24% della grandezza oggetto di studio.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	5	5	3	1	1
y_i	11	25	25	25	39

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + bX \quad II) Y^* = c$$

$$\hat{a} = 35.5, \hat{b} = -3.5, \text{ indice adattamento } \rho^2 = 0.5 \quad \hat{c} = 25, \text{ indice adattamento } R^2 = 0$$

Il modello migliore è $Y^* = 35.5 - 3.5X$.

3. Per il modello migliore, si fornisca la previsione in corrispondenza di $X = 3$.

$$Y^*(3) = 35.5 - 3.5 \cdot 3 = 25$$

ESERCIZIO 4.

Si completino le seguenti tabelle, giustificando la risposta, in modo che siano soddisfatte le condizioni indicate.

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$	0	11	
$x_2 = 2$		0	

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$		11	
$x_2 = 2$		11	

a) $\eta_{Y|X}^2 = \rho^2$ con $\rho > 0$

$\eta_{Y|X}^2$ risulta sempre uguale a ρ^2

in quanto abbiamo due sole medie condizionate che saranno sempre allineate.

Affinché $\rho > 0$ la retta deve avere pendenza crescente, vale a dire $M(Y|x_2) > M(Y|x_1)$.

Una possibile soluzione è $n_{13} = 0, 0 < n_{21} < n_{23}$.

b) $\rho^2 = 0$

Se $\eta_{Y|X}^2 = 0$ anche $\rho^2 = 0$

in quanto l'indipendenza in media implica l'indipendenza lineare.

$$n_{11} = n_{13}, n_{21} = n_{23}$$

rappresenta una possibile soluzione.

ESERCIZIO 5.

Un'urna contiene 10 palline nere e 20 palline rosse.

Si consideri il seguente esperimento: si toglie, in primo luogo, una pallina dall'urna; si estraggono, poi, 7 palline.

1. Si calcoli la probabilità che tra le 7 palline estratte ve ne siano 6 rosse nell'ipotesi che le 7 estrazioni avvengano con reimmissione.

Indicati con R l'evento 'la pallina tolta è rossa' e con A l'evento '6 palline rosse tra le 7 estratte con reimmissione' abbiamo:

$$\begin{aligned} P(A) &= P(R) \cdot P(A|R) + P(\bar{R}) \cdot P(A|\bar{R}) = \frac{20}{30} \cdot \binom{7}{6} \left(\frac{19}{29}\right)^6 \left(\frac{10}{29}\right)^1 + \frac{10}{30} \cdot \binom{7}{6} \left(\frac{20}{29}\right)^6 \left(\frac{9}{29}\right)^1 = \\ &= \frac{2}{3} \cdot 0.1909 + \frac{1}{3} \cdot 0.2337 = 0.2052. \end{aligned}$$

2. Si calcoli la probabilità che tra le 7 palline estratte ve ne siano 6 rosse nell'ipotesi che le 7 estrazioni avvengano senza reimmissione.

Indicati con R l'evento 'la pallina tolta è rossa' e con A l'evento '6 palline rosse tra le 7 estratte senza reimmissione' abbiamo:

$$\begin{aligned} P(A) &= P(R) \cdot P(A|R) + P(\bar{R}) \cdot P(A|\bar{R}) = \frac{20}{30} \cdot \frac{\binom{19}{6} \binom{10}{1}}{\binom{29}{7}} + \frac{10}{30} \cdot \frac{\binom{20}{6} \binom{9}{1}}{\binom{29}{7}} = \\ &= \frac{2}{3} \cdot 0.1738 + \frac{1}{3} \cdot 0.2235 = 0.1904. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 04.09.2014 (Tema 280-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

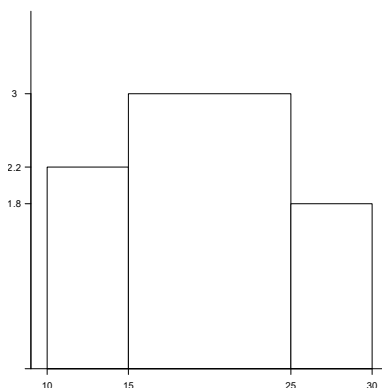
ESERCIZIO 1.

La seguente tabella riporta la distribuzione congiunta della variabile statistica doppia (X, Y) , con X carattere qualitativo, per un campione di 150 unità statistiche.

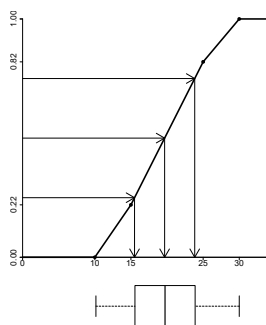
$X \setminus Y$	10 $\rightarrow$ 15	15 $\rightarrow$ 25	25 $\rightarrow$ 30	$n_{i\bullet}$
x_1	22	60	18	100
x_2	11	30	9	50
$n_{\bullet j}$	33	90	27	150

1. Si riassumano mediante opportuni indici di posizione e di variabilità le distribuzioni marginali X e Y .
 X Indice di posizione: $\text{Moda}(X) = x_1$ Indice di variabilità: Indice di Frosini $F_N(X) = 0.6667$
 Y Indice di posizione: $M(Y) = 19.7$ Indice di variabilità: $CV = 0.2403$

2. Si rappresenti graficamente la distribuzione condizionata $Y|x_2$.



3. Si costruisca il grafico Box & Whiskers plot per la marginale Y .



$$x_{0.25} = 15.5, x_{0.50} = 19.6667, x_{0.75} = 23.8333, \text{lung. teorica baffo} = 12.5, \text{baffo inferiore} = 10, \text{baffo superiore} = 30$$

ESERCIZIO 2.

Si consideri la seguente serie di numeri indici a base mobile:

anno	$NIBM$	$NIBF$	x_t
2009		1	200
2010	0.9000	0.90	180
2011	1.0333	0.93	186
2012	1.0968	1.02	204
2013	1.0882	1.11	222

1. Si costruiscano la serie dei numeri indici a base fissa (2009) e la serie dei dati iniziali sapendo che $x_{2009} = 200$.
2. Si calcoli il tasso medio di variazione dal 2011 al 2013 e si commenti il risultato ottenuto.

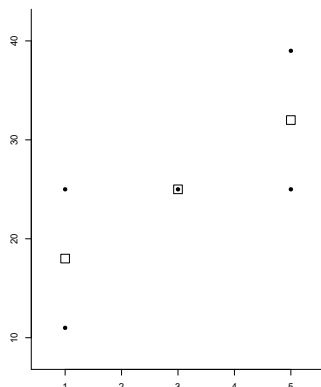
tasso medio di variazione 0.0925. Tra il 2011 e il 2013 si è osservato un incremento medio annuo del 9.25% della grandezza oggetto di studio.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	3	5	5
y_i	11	25	25	25	39

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + bX \quad II) Y^* = cX$$

$$\hat{a} = 14.5, \hat{b} = 3.5, \text{ indice adattamento } M(\hat{E}^2) = 39.2 \quad \hat{c} = 7.0656, \text{ indice adattamento } M(\hat{E}^2) = 94.348$$

Il modello migliore è $Y^* = 14.5 + 3.5X$.

3. Per il modello migliore, si fornisca la previsione in corrispondenza di $X = 3$.

$$Y^*(3) = 14.5 + 3.5 \cdot 3 = 25$$

ESERCIZIO 4.

Si completino le seguenti tabelle, giustificando la risposta, in modo che siano soddisfatte le condizioni indicate.

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$	0	11	
$x_2 = 2$		0	

$$a) \eta_{Y|X}^2 > \rho^2, \text{ con } \rho < 0$$

Impossibile: $\eta_{Y|X}^2$ risulta sempre uguale a ρ^2 in quanto figurano due sole medie condizionate che saranno sempre allineate.

	$y_1 = 10$	$y_2 = 20$	$y_3 = 30$
$x_1 = 1$		11	
$x_2 = 2$		11	

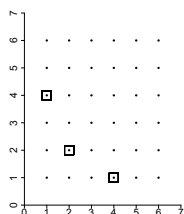
$$b) \eta_{Y|X}^2 = 1$$

Impossibile: $\eta_{Y|X}^2 = 1$ quando sussiste dipendenza funzionale di Y da X → distribuzioni condizionate $Y|x_1$ e $Y|x_2$ degeneri. Se si ponesse $n_{11} = n_{13} = n_{21} = n_{23} = 0$ le medie risulterebbero tutte identiche ($\eta_{Y|X}^2 = 0$) e la tabella non sarebbe comunque ammissibile.

ESERCIZIO 5.

Per il seguente gioco è prevista una posta di 2 Euro: si lanciano due dadi e se il prodotto dei numeri ottenuti è pari a 4 si vincono 6 Euro.

1. Si stabilisca se il gioco è equo.



$$P(\text{vincita}) = \frac{3}{36} = 0.08333$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
posta	-2	0.91667	-1.83333
vincita - posta	4	0.08333	0.33333
		1	-1.50000

Il valore atteso della variabile casuale guadagno è negativo, quindi il gioco non è equo.

2. Nell'ipotesi di ripetere il gioco 4 volte si calcoli la probabilità di vincere al massimo 1 volta.

Sia X la variabile casuale 'n. di vincite nelle 4 prove indipendenti'. $X \sim \text{Bin}(n = 4, p = 0.08333)$.

$$\begin{aligned} P(X < 1) &= P(X = 0) + P(X = 1) = \\ &= \binom{4}{0} 0.08333^0 \cdot 0.91667^4 + \binom{4}{1} 0.08333^1 \cdot 0.91667^3 = 0.7061 + 0.2568 = 0.9628. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 18.09.2014 (Tema 280)

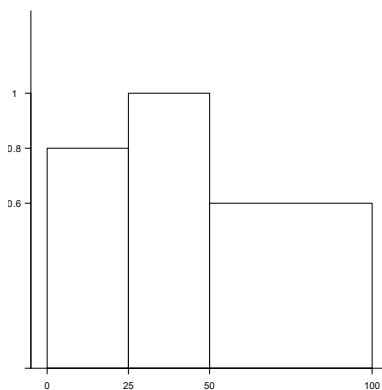
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Nella tabella sottostante vengono riportati i risultati di un'indagine sull'uso dei pesticidi nei terreni agricoli in Italia.

Regione (R) / Pesticidi in mg/ha (P)	0 - 25	25 - 50	50 - 100
nord	10	6	9
centro	5	8	9
sud	5	11	12

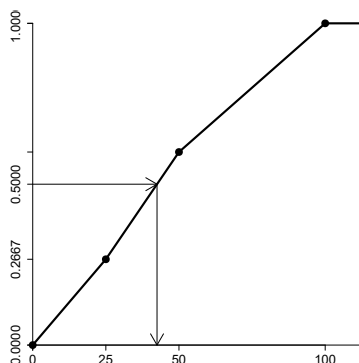
1. Si rappresenti graficamente la variabile marginale P .



2. Si calcolino, se possibile, moda, media aritmetica e mediana per la variabile R . Si calcolino, se possibile, moda, media aritmetica e si individui su un opportuno grafico la mediana per la variabile P .

Essendo il carattere R qualitativo sconnesso risulta possibile il solo calcolo della moda, che risulta 'sud'.

$$\begin{aligned} \text{Moda}(P) &= 25 - 50 \\ M(P) &= 45.833 \end{aligned}$$



3. Si calcoli per la variabile R un opportuno indice di variabilità e se ne commenti il risultato.

$F_N = 0.9307$. Sussiste un elevato livello di eterogeneità.

4. Si calcoli un opportuno indice di connessione.

$\chi_N = 0.1588$. Il livello di connessione esistente tra le due variabili è basso.

ESERCIZIO 2.

Si consideri la seguente serie di numeri indici a base mobile dei prezzi di un titolo azionario:

anno	$NIBM$	$NIBF$
2008	0.8	1
2009	1.5	1.5000
2010	1.5	2.2500
2011	1.5	3.3750
2012	1.5	5.0625

1. Si costruisca la corrispondente serie dei numeri indici a base fissa (anno base 2008).

2. Se nel 2008 fossero state acquistate 100 azioni a un prezzo di 10 Euro per azione, si calcoli il valore totale di queste azioni nel 2012.

$x_{2012} = 5062.50$. Da $x_{2008} = 100 \cdot 10 = 1000$ e da ${}_{2008}I_{2012} = \frac{x_{2012}}{x_{2008}} = 5.0625$ si ottiene $x_{2012} = x_{2008} \cdot 5.0625 = 1000 \cdot 5.0625$.

3. Si calcoli il tasso medio di variazione dal 2009 al 2012 e si commenti il risultato ottenuto.

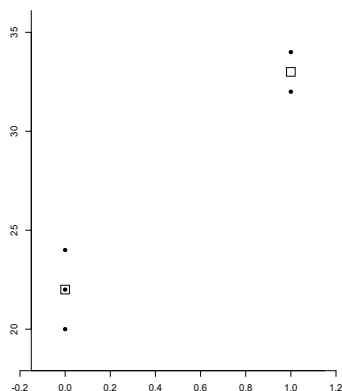
tasso medio di variazione 0.5. Tra il 2009 e il 2012 vi è stato un incremento medio annuo del prezzo del titolo del 50%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza del numero di bagnanti (Y) presenti in una piscina in 5 domeniche in funzione della situazione meteo rilevata con una variabile dicotomica X ($= 0$ se nuvoloso e $= 1$ se soleggiato)

x_i	0	0	0	1	1
y_i	22	20	24	32	34

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + bX \quad II) Y^* = c$$

$$\hat{a} = 22, \hat{b} = 11, M(\hat{E}^2) = 2 \quad \hat{c} = 26.4, M(\hat{E}^2) = 31.04.$$

Il modello migliore è $Y^* = 22 + 11X$.

3. Si determini, indicando il ragionamento, il valore dei rapporti di correlazione $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

$\eta_{Y|X}^2 = \rho^2 = 0.936$ in quanto le medie condizionate sono allineate.

$\eta_{X|Y}^2 = 1$ sussiste dipendenza funzionale di X da Y (a ogni valore di Y corrisponde un solo valore distinto di X).

ESERCIZIO 4.

Sia 0.06 la probabilità di colpire il centro del bersaglio lanciando una freccetta.

1. Si calcoli la probabilità di colpire il centro almeno una volta su 20 lanci.

Sia X la variabile casuale 'numero di volte che si colpisce il centro nelle 20 prove indipendenti'.

$$X \sim \text{Bin}(n = 20, p = 0.06).$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{20}{0} 0.06^0 0.94^{20} = 1 - 0.2901 = 0.7099.$$

2. Si calcoli la probabilità che, su 100 lanci, si colpisca almeno 6 volte il centro.

Indicata con X la variabile casuale 'numero di volte che si colpisce il centro' nelle 100 prove indipendenti risulta $X \sim \text{Bin}(n = 100, p = 0.06)$. $M(X) = 6$; $\text{Var}(X) = 5.64$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 6, \sigma^2 = 5.64)$ si ottiene

$$P(X \geq 6) = P(X \geq \mu) = 0.5.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 15.01.2015 (Tema 281)

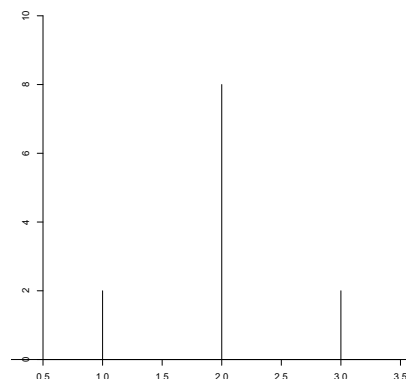
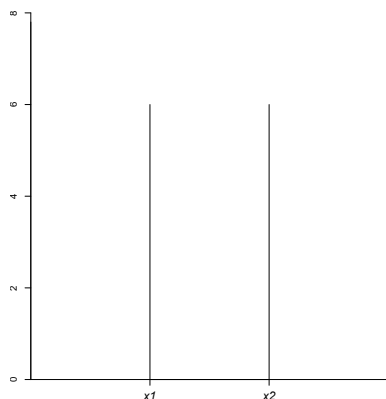
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Con riferimento alla distribuzione congiunta delle variabili quantitative discrete X e Y si consideri la seguente tabella a doppia entrata:

$X \setminus Y$	y_1	y_2	y_3
x_1	2	2	2
x_2	0	6	0

1. Dopo avere rappresentato graficamente le distribuzioni **marginali** X e Y , si stabilisca, senza effettuare calcoli e motivando teoricamente la risposta, quale delle due distribuzioni presenta una maggiore eterogeneità.



La distribuzione che presenta maggiore eterogeneità è la X in quanto caratterizzata da equidistribuzione delle frequenze (situazione di massima eterogeneità).

2. Sapendo che $y_1 = 1$, $y_2 = 2$ e $y_3 = 3$, si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e variabilità, commentando i risultati ottenuti.

$Y|x_1$: Indice di posizione: $M(Y|x_1) = 2$ Indice di variabilità: $CV(Y|x_1) = 0.4083$

$Y|x_2$: Indice di posizione: $M(Y|x_2) = 2$ Indice di variabilità: $CV(Y|x_2) = 0$

Le distribuzioni sono caratterizzate dallo stesso livello medio; $Y|x_2$ non presenta variabilità.

3. Si determini il grado di connessione tra le variabili X e Y con un opportuno indice normalizzato.

$\chi_N = 0.7071$. Il livello di connessione esistente tra le due variabili è medio alto.

4. Si indichi inoltre se e come potrebbero essere modificate le frequenze congiunte, a parità di numerosità totale, per ottenere una situazione di massima connessione.

$X \setminus Y$	y_1	y_2	y_3	
x_1	2	0	2	4
x_2	0	8	0	8
	2	8	2	12

Nel caso in esame è possibile realizzare una situazione di dipendenza funzionale di X da Y (distribuzioni condizionate $X|y_j$ degeneri).

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi medi annui di 2 beni di uso quotidiano nel periodo 2006 – 2010:

t	<i>bene A</i>	<i>bene B</i>	$NIBF(2007)$
2006	100	460	1.0204
2007	98	460	1.0000
2008	103	460	1.0510
2009	105	460	1.0714
2010	105	480	1.0714

1. Costruire la serie dei numeri indici a base fissa con base 2007 per il bene A .

2. Calcolare il tasso medio di variazione per il prezzo del bene B tra il 2006 e il 2009, interpretando il risultato. tasso medio di variazione 0. Nel periodo considerato la grandezza in oggetto è rimasta immutata.

3. Note le quantità $A_{q2007} = 1000$ e $B_{q2007} = 8500$, calcolare l'indice di Laspeyres dei prezzi al 2010 con base 2007.

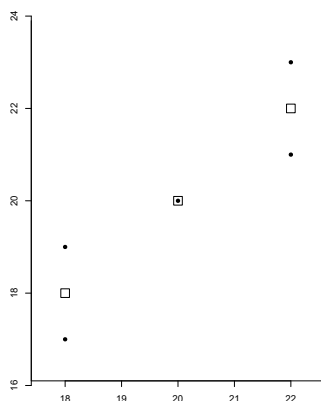
$\frac{P}{2007} I_{2010}^L = 1.0442$. La variazione relativa nel livello generale dei prezzi, secondo l'Indice di Laspeyres, è del 4.42%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	18	18	20	22	22
y_i	17	19	20	21	23

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 0, \hat{b} = 1, M(\hat{E}^2) = 0.8 \quad \hat{c} = 1, M(\hat{E}^2) = 0.8.$$

I modelli sono equivalenti; abbiamo $Y = X + E$.

3. Senza effettuare calcoli si individui il valore di $\eta_{Y|X}^2$. Si motivi perché $\eta_{Y|X}^2 < \eta_{X|Y}^2$.

$\eta_{Y|X}^2 = \rho^2$ Dalla rappresentazione grafica si osserva che le medie condizionate sono allineate.

$\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y .

ESERCIZIO 4.

Un'urna è composta da 26 gettoni rossi e 74 gettoni neri. Si consideri l'esperimento casuale $\mathcal{G}$ = 'estrazione **senza reimmissione** di 2 gettoni dall'urna, dei quali viene rilevato il colore'.

1. Si costruisca lo spazio probabilistico associato all'esperimento $\mathcal{G}$ e si calcoli la probabilità di ciascun evento elementare.

$$\Omega = \{RR, RN, NR, NN\}$$

$$P(R \cap R) = \frac{26}{100} \frac{25}{99} = 0.0657, P(R \cap N) = \frac{26}{100} \frac{74}{99} = 0.1943, P(N \cap R) = \frac{74}{100} \frac{26}{99} = 0.1943, P(N \cap N) = \frac{74}{100} \frac{73}{99} = 0.5457.$$

2. Si calcoli la probabilità che nelle due estrazioni si ottengano due gettoni del medesimo colore. $P(RR \cup NN) = P(RR) + P(NN) = 0.0657 + 0.5457 = 0.6113$.

3. Un giocatore puntando 1 Euro, ne vince 3 qualora i due gettoni estratti siano del medesimo colore. In caso contrario perde la somma puntata. Si stabilisca se il gioco è equo (guadagno atteso nullo).

$$P(\text{vincita}) = 0.6113,$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
<i>posta</i>	-1	0.3887	-0.3887
<i>vincita - posta</i>	2	0.6113	1.2226
		1	0.8339

Il valore atteso della variabile casuale guadagno è positivo, quindi il gioco non è equo ma favorevole al giocatore.

4. Si calcoli la probabilità di ottenere non più di 10 gettoni rossi in 51 estrazioni **con reimmissione**.

Indicata con X la variabile casuale 'numero di gettoni rossi' nelle 51 estrazioni con reimmissione (prove indipendenti) risulta

$$X \sim \text{Bin}(n = 51, p = 0.26).$$

$M(X) = 13.26$; $\text{Var}(X) = 9.8124$; utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 13.26, \sigma^2 = 9.8124)$$

si ottiene

$$\begin{aligned} P(X \leq 10) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{10 - \mu}{\sigma}\right) = P\left(Z \leq \frac{10 - 13.26}{3.1325}\right) = \\ &= P(Z \leq -1.04) = P(Z \geq 1.04) = \\ &= 1 - P(Z < 1.04) = 1 - 0.8508 = 0.1492. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 29.01.2015 (Tema 282)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

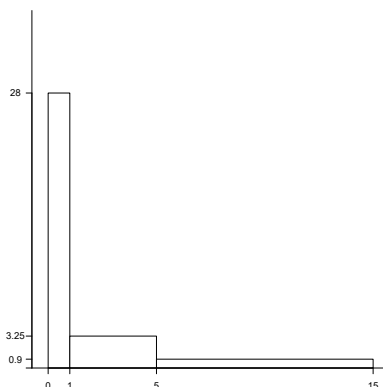
Studente _____ matricola _____

ESERCIZIO 1.

La tabella seguente contiene la distribuzione delle scadenze in anni (A) dei 108 titoli di stato detenuti in portafoglio da 2 banche italiane (B):

$B \setminus A$	0 + 1	1 + 5	5 + 15
$b_1 = \text{Intesa Sanpaolo}$	21	18	19
$b_2 = \text{Unicredit}$	28	13	9

1. Fornire la rappresentazione grafica della distribuzione $A|Unicredit$.



2. Confrontare opportunamente la variabilità delle due distribuzioni $A|b_i$ ($i = 1, 2$).

$A|b_1$: Indice di posizione: $M(A|b_1) = 4.3879$ Indice di variabilità: $CV(A|b_1) = 0.9226$

$A|b_2$: Indice di posizione: $M(A|b_2) = 2.86$ Indice di variabilità: $CV(A|b_1) = 1.2263$

La scadenza media dei titoli posseduti da *Intesa Sanpaolo* è superiore a quella dei titoli posseduti da *Unicredit*. La distribuzione delle scadenze dei titoli posseduti da *Unicredit* presenta un livello di variabilità più elevato.

3. Misurare la dipendenza di A da B con un opportuno indice di connessione.

$\chi_N = 0.2112$.

Tra le variabili sussiste un livello molto basso di connessione.

4. Indicare le condizioni nelle quali $\eta_{A|B}^2$ e χ^2 risultano nulli.

$\eta_{A|B}^2$ risulta nullo quando sussiste indipendenza in media di A da B , vale a dire le medie condizionate coincidono $M(A|b_1) = M(A|b_2)$.

χ^2 risulta nullo quando sussiste indipendenza stocastica tra A e B , vale a dire le distribuzioni condizionate sono somiglianti: $A|b_1 \sim A|b_2 \sim A$ e $B|A_1 \sim B|A_2 \sim B|A_3 \sim B$.

ESERCIZIO 2.

Si consideri la variabile X = vendite annuali di un prodotto, in migliaia di pezzi, misurata tra il 2008 e il 2012:

t	x_t	$NIBM$
2008	888	
2009	619	0.6971
2010	568	0.9176
2011	509	0.8961
2012	528	1.0373

1. Costruire la serie dei numeri indici a base mobile.

2. Interpretare il valore dell'ultimo degli indici sopra calcolati.

$_{2011}I_{2012} = 1.0373$ Tra il 2011 e il 2012 si è registrato un incremento delle vendite del 3.73%.

3. Calcolare il tasso medio di variazione delle vendite registrato tra il 2008 e il 2012, interpretando il risultato.

$\sqrt[4]{_{2008}I_{2012}} = \sqrt[4]{\frac{528}{888}} = \sqrt[4]{0.5946} = 0.8781$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del -12.19%.

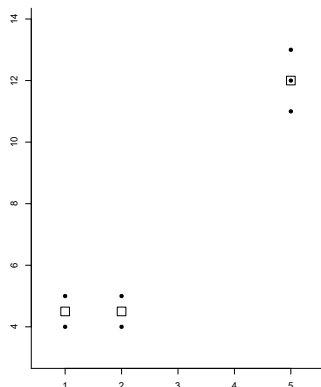
Le vendite hanno subito un decremento medio annuo del 12.19% tra il 2008 e il 2012.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolte le seguenti 7 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 7$

x_i	1	1	2	2	5	5	5
y_i	4	5	4	5	11	12	13

1. Fornire la rappresentazione grafica delle coppie di punti, insieme alla funzione di regressione.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = \alpha + \beta \cdot X^2$$

$$\hat{a} = 1.5779, \hat{b} = 2.0455, \text{ indice adattamento } \rho^2 = 0.9257 \quad \hat{\alpha} = 3.7082, \hat{\beta} = 0.3299, \rho^2 = 0.9599$$

Il modello migliore è $Y^* = 3.7082 + 0.3299X^2$.

3. Indicare l'intervallo dei valori possibili per $\eta_{Y|X}^2$.

$$0.9599 < \eta_{Y|X}^2 < 1.$$

Si osserva che le medie condizionate non stanno su una parabola dal momento che $M(Y|x=1) = M(Y|x=2) = 4.5$. Non sussiste nemmeno dipendenza funzionale di Y da X .

ESERCIZIO 4.

Una linea di produzione confeziona pacchetti di caramelle il cui peso complessivo è fissato al valore nominale di 250g. Sapendo che detti pesi seguono una v.c. $N(\mu = 250, \sigma^2 = 21)$ e che si ritengono difettosi i pacchetti con peso inferiore del 3% rispetto al valore nominale:

1. calcolare la probabilità che la linea produca pezzi difettosi;

Si ritengono difettosi i pacchetti con peso inferiore a

$$242.5 = 0.97 \cdot 250$$

Indicata con $X \sim N(\mu = 250, \sigma^2 = 21)$ la variabile casuale che descrive il peso di un pacchetto di caramelle, la probabilità che il peso sia inferiore a 242.5g risulta

$$\begin{aligned} P(X \leq 242.5) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{242.5 - \mu}{\sigma}\right) = P\left(Z \leq \frac{242.5 - 250}{4.5826}\right) = \\ &= P(Z \leq -1.64) = P(Z \geq 1.64) = \\ &= 1 - P(Z < 1.64) = 1 - 0.9495 = 0.0505. \end{aligned}$$

2. calcolare la probabilità che su 100 pacchetti estratti (con re-immissione) dal flusso produttivo, almeno il 5% risulti fuori norma.

Indicata ora con X la variabile casuale 'numero di pacchetti fuori norma' tra i 100 estratti con reimmissione (prove indipendenti) risulta

$$X \sim \text{Bin}(n = 100, p = 0.0505).$$

$$M(X) = 5.05; \text{Var}(X) = 4.795.$$

Dal momento che $np = 5.05 > 5$ ed $n(1-p) = 94.95 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 5.05, \sigma^2 = 4.795).$$

Si ottiene

$$\begin{aligned} P(X \geq 5) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{5 - \mu}{\sigma}\right) = P\left(Z \geq \frac{5 - 5.05}{2.1897}\right) = \\ &= P(Z \geq -0.02) = P(Z \leq 0.02) = 0.5080, \end{aligned}$$

essendo 5 il 5% di 100.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 12.02.2015 (Tema 283)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

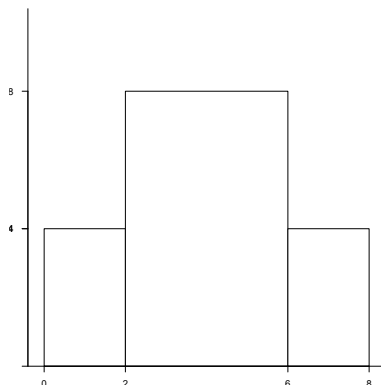
Studente _____ matricola _____

ESERCIZIO 1.

Di una v.s. doppia (X, Y) si dispone della seguente distribuzione delle densità di frequenza della condizionata $Y|x_1$:

$h_{j-1} \dashv h_j$	d_{1j}	n_{1j}
0 $\dashv$ 2	4	8
2 $\dashv$ 6	8	32
6 $\dashv$ 8	4	8

1. Si ricostruisca la distribuzione delle frequenze assolute, n_{1j} , di $Y|x_1$ e se ne dia una rappresentazione grafica.



2. Si calcolino la media aritmetica, la mediana e la varianza della variabile $Y|x_1$.

$M(Y|x_1) = 4$ $y_{0.5}|x_1 = 4$ $Var(Y|x_1) = 3$. Si osserva che la distribuzione è simmetrica

3. Sapendo che la distribuzione condizionata $Y|x_2$, riferita a $n_{2\bullet} = 16$ unità statistiche, è caratterizzata da media $M(Y|x_2) = 4$ e varianza $Var(Y|x_2) = 0$ si ricostruisca la distribuzione congiunta di X e Y e si misuri la dipendenza tra X e Y con un opportuno indice di connessione.

$X \setminus Y$	0 $\dashv$ 2	2 $\dashv$ 6	6 $\dashv$ 8	
x_1	8	32	8	48
x_2	0	16	0	16
	8	48	8	64

$\chi_N = 0.3333$. Sussiste un livello di connessione basso.

4. Ponendo ora $n_{2\bullet} = 48$, si indichi quali dovrebbero essere le frequenze della distribuzione condizionata $Y|x_2$ affinché sussista indipendenza stocastica tra X e Y .

$X \setminus Y$	0 $\dashv$ 2	2 $\dashv$ 6	6 $\dashv$ 8	
x_1	8	32	8	48
x_2	8	32	8	48
	16	64	16	96

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi di chiusura di 2 titoli azionari nei giorni dal 7 al 10 gennaio:

t	titolo A	titolo B
7	100	100
8	120	100
9	144	110
10	150	121

1. Si calcoli e interpreti il numero indice al tempo 7 con base 9 per il *titolo A*.

${}_9I_7 = \frac{x_7}{x_9} = \frac{100}{144} = 0.6944$. La grandezza al tempo 7 era inferiore del 30.56% rispetto alla grandezza al tempo 9.

2. Si calcoli il tasso medio di variazione registrato tra il giorno 7 e il giorno 9 per il *titolo A*, interpretando il risultato.

$\sqrt{7}I_9 = \sqrt{\frac{144}{100}} = 1.2$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 20%.

3. Supponendo che la media aritmetica calcolata a partire dai valori degli indici di Laspeyres, ${}_7^PI_{10}^L$, e di Paasche, ${}_7^PI_{10}^P$, sia risultata pari a 1.31, si indichi se l'indice di Fisher può assumere un valore maggiore di 1.31.

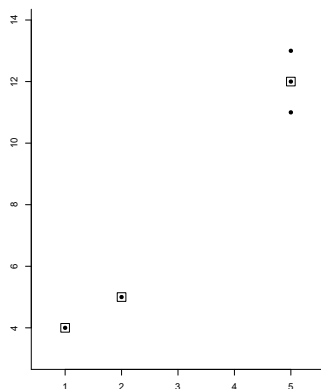
L'indice dei prezzi di Fisher è definito come media geometrica degli indici di Laspeyres e di Paasche. La media geometrica dei due indici risulterà inferiore o eguale alla media aritmetica degli stessi; l'eguaglianza si può realizzare solo nel caso in cui l'indice di Laspeyres coincida con quello di Paasche.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	5	5	5
y_i	4	5	11	12	13

1. Fornire la rappresentazione grafica delle coppie di punti, insieme alla funzione di regressione di Y dato X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 1.4211$, $\hat{b} = 2.1053$, indice adattamento $\rho^2 = 0.9624$ $\hat{c} = 9$, $R^2 = 0$
Il modello migliore è $Y^* = 1.4211 + 2.1053X$.

3. Si calcoli $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = 0.9714$. Sussiste, quindi, un livello di dipendenza molto alto.

ESERCIZIO 4.

Un'urna contiene 5 palline nere e 15 palline rosse.

1. Si calcoli la probabilità che, estraendo **senza reimmissione** 5 palline dall'urna, si ottengano 3 palline nere.

$$P(3N) = \binom{5}{3} \frac{5}{20} \frac{4}{19} \frac{3}{18} \frac{15}{17} \frac{14}{16} = 0.0677$$

o mediante la v.c. ipergeometrica

$$P(X = 3) = \frac{\binom{5}{3} \binom{15}{2}}{\binom{20}{5}} = 0.0677.$$

2. Indicata con Y la variabile casuale 'numero di palline nere' ottenute in 100 estrazioni **con reimmissione** dall'urna, si calcoli $P(Y \leq 6)$.

Risulta $Y \sim \text{Bin}(n = 100, p = 0.25)$. $M(X) = 25$; $\text{Var}(X) = 18.75$. Utilizzando l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 25, \sigma^2 = 18.75)$ si ottiene

$$P(X \leq 6) = P\left(\frac{X - \mu}{\sigma} \leq \frac{6 - \mu}{\sigma}\right) = P\left(Z \leq \frac{6 - 25}{4.3301}\right) = P(Z \leq -4.39) \simeq 0.$$

3. Sia ora $p = 0.8729$. Si determini il valore del percentile di ordine p della v.c. normale $X \sim N(\mu = 2, \sigma^2 = 1)$, vale a dire il valore x_p tale che $P(X \leq x_p) = p$.

$$\begin{aligned} P(X \leq x_p) &= 0.8729 \\ P\left(\frac{X - \mu}{\sigma} \leq \frac{x_p - \mu}{\sigma}\right) &= 0.8729 \\ P\left(Z \leq \frac{x_p - 2}{1}\right) &= 0.8729 \\ z_p = \frac{x_p - 2}{1} &= 1.14 \\ x_p &= 2 + 1.14 \cdot 1 = 3.14. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 04.06.2015 (Tema 284-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

Si completino le frequenze nella seguente tabella sapendo che il numero totale di unità statistiche sulle quali sono stati congiuntamente rilevati i caratteri X e Y è 11 e che $n_{\bullet 2} = 1$.

$X \setminus Y$	10 + 15	15 + 25	25 + 30	$n_{i\bullet}$
x_1	1	0	9	10
x_2	0	1	0	1
$n_{\bullet j}$	1	1	9	11

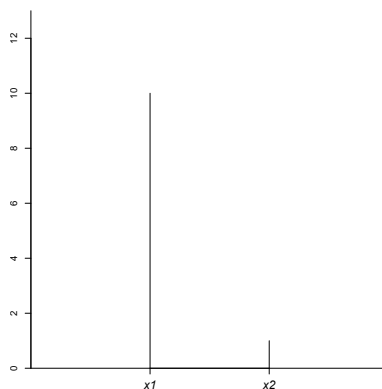
1. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 26$ Indice di variabilità: $CV(Y|x_1) = 0.1731$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 20$ Indice di variabilità: $CV(Y|x_2) = 0$

La distribuzione $Y|x_1$ presenta un livello medio superiore a quello della distribuzione $Y|x_2$. La distribuzione condizionata $Y|x_2$ presenta variabilità nulla.

2. Si rappresenti graficamente la marginale X .



3. Si riassuma la marginale qualitativa X con opportuni indici di posizione e di variabilità.

Indice di posizione: $\text{Moda}(X) = x_1$. Indice di variabilità: $F_N = 0.1818$. Sussiste un livello di eterogeneità basso.

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi di chiusura di 2 titoli azionari nei giorni dal 7 al 10 gennaio:

t	titolo A	titolo B	$NIBF(t = 7)$
7	100	100	1.00
8	120	100	1.00
9	144	110	1.10
10	150	121	1.21

1. Si calcoli e interpreti il numero indice al tempo 10 con base 7 per il *titolo A* e si costruisca la serie dei numeri indici a base fissa (al tempo 7) per il *titolo B*.

${}_7I_{10} = 1.50$. Tra il giorno 7 e il giorno 10 il prezzo del *titolo A* ha registrato un incremento del 50%.

2. Si calcoli il tasso medio di variazione registrato tra il giorno 7 e il giorno 10 per il *titolo B*, interpretando il risultato.

$\sqrt[3]{2007I_{2010}} = \sqrt[3]{\frac{121}{100}} = \sqrt[3]{1.21} = 1.0656$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 6.56%.

Il prezzo del *titolo B* ha registrato un incremento medio giornaliero del 6.56% tra il giorno 7 e il giorno 10.

3. Supponendo che la mediana calcolata a partire dai valori degli indici di Laspeyres, ${}_7I_{10}^L$, e di Paasche, ${}_7I_{10}^P$, sia risultata pari a 1.31, si indichi l'intervallo dei valori possibili per l'indice di Fisher.

Gli indici dei prezzi di Laspeyres e di Paasche sono medie potenziate dei prezzi relativi; saranno entrambi compresi tra il prezzo relativo minimo $1.21 = \frac{121}{100}$ e quello massimo $1.5 = \frac{150}{100}$.

L'indice dei prezzi di Fisher è definito come media geometrica degli indici dei prezzi di Laspeyres e di Paasche. La media geometrica dei due indici risulterà inferiore o eguale alla media aritmetica degli stessi che coincide con la mediana; l'eguaglianza si può realizzare solo nel caso in cui l'indice di Laspeyres coincida con quello di Paasche. Vale, quindi

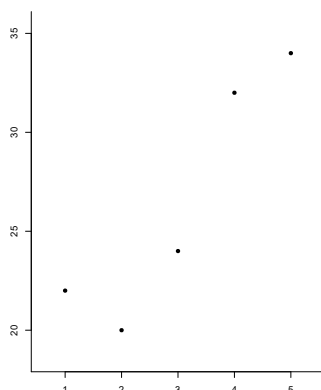
$$1.21 < {}_7I_{10}^F \leq 1.31.$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	3	4	5
y_i	22	20	24	32	34

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + bX \quad II) Y^* = cX$$

$$\hat{a} = 15.6, \hat{b} = 3.6, \text{ indice adattamento } M(\hat{E}^2) = 5.12 \quad \hat{c} = 7.8545, \text{ indice adattamento } M(\hat{E}^2) = 49.367$$

Il modello migliore è $Y^* = 15.6 + 3.6X$.

3. Per il modello migliore, si fornisca la previsione in corrispondenza di $X = 3$.

$$Y^*(3) = 15.6 + 3.6 \cdot 3 = 26.4.$$

4. Si determini il valore del rapporto di correlazione $\eta_{Y|X}^2$ motivando la risposta.

$$\eta_{Y|X}^2 = 1 \text{ in quanto sussiste dipendenza funzionale di } Y \text{ da } X.$$

ESERCIZIO 4.

Si consideri l'esperimento di estrazione di 4 palline da un'urna contenente 21 palline rosse e 29 palline blu.

1. Si calcoli la probabilità di ottenere almeno 3 palline blu nell'ipotesi che le 4 estrazioni siano effettuate senza reimmissione.

$$P(\text{almeno } 3 B) = P(3B) + P(4B) = \binom{4}{3} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{21}{47} + \binom{4}{4} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{26}{47} = 0.3332 + 0.1031 = 0.4363$$

o mediante la v.c. ipergeometrica

$$P(X = 3) + P(X = 4) = \frac{\binom{29}{3} \binom{21}{1}}{\binom{50}{4}} + \frac{\binom{29}{4} \binom{21}{0}}{\binom{50}{4}} = 0.4363$$

2. Si calcoli la probabilità di ottenere almeno 3 palline blu nell'ipotesi che le 4 estrazioni siano effettuate con reimmissione.

Sia X la variabile casuale 'n. di palline blu nelle 4 estrazioni (con reimmissione) (prove indipendenti)'. $X \sim \text{Bin}(n = 4, p = 0.58)$.

$$P(X \geq 3) = \binom{4}{3} 0.58^3 0.42^1 + \binom{4}{4} 0.58^4 0.42^0 = 0.3278 + 0.1132 = 0.4410.$$

3. Si calcoli la probabilità di ottenere almeno 36 palline blu nell'ipotesi che siano effettuate 60 estrazioni con reimmissione.

Indicata con X la variabile casuale 'numero di successi' nelle 60 prove indipendenti risulta $X \sim \text{Bin}(n = 60, p = 0.58)$. $M(X) = 34.8$; $\text{Var}(X) = 14.6160$. Dal momento che $np = 34.8 > 5$ ed $n(1 - p) = 25.2 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 34.8, \sigma^2 = 14.6160).$$

Si ottiene

$$P(X > 36) = P\left(\frac{X - \mu}{\sigma} > \frac{36 - \mu}{\sigma}\right) = P\left(Z > \frac{36 - 34.8}{3.8231}\right) = P(Z > 0.31) = 1 - P(Z \leq 0.31) = 1 - 0.6217 = 0.3783.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 25.06.2015 (Tema 284)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

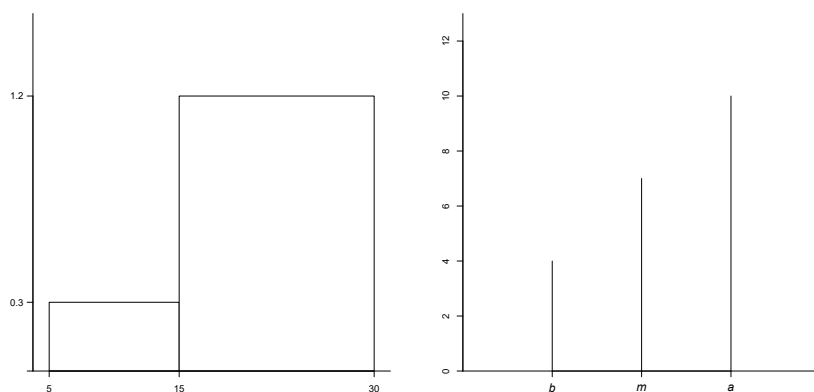
Studente _____ matricola _____

ESERCIZIO 1.

La Direzione di un sito destinato alla vendita online di libri ha estratto un campione di opere attualmente in vendita, rilevando le variabili G = giudizio dei clienti sull'opera (b = basso, m = medio, a = alto) e P = prezzo di vendita (in euro). I dati raccolti sono riportati nella seguente tabella a doppia entrata:

$P \setminus G$	b	m	a
$5 \div 15$	1	0	2
$15 \div 30$	3	7	8

1. Si rappresentino graficamente le distribuzioni marginali P e G .



2. Si calcolino la mediana e un opportuno indice di variabilità della marginale G .

$g_{0.5} = m$ (medio) variabilità(G), $F_N = 0.7526$. Sussiste un livello medio/alto di eterogeneità.

3. Si confronti la variabilità delle distribuzioni condizionate $P|G = m$ e $P|G = a$.

$P|g = m$: Indice di posizione: $M(P|g = m) = 22.5$ Indice di variabilità: $CV(P|g = m) = 0$

$P|g = a$: Indice di posizione: $M(P|g = a) = 20$ Indice di variabilità: $CV(P|g = a) = 0.25$

La distribuzione $P|g = m$ presenta un livello medio più elevato di $P|g = a$, e variabilità nulla.

4. Si determini il grado di connessione tra le variabili P e G mediante il calcolo di un opportuno indice.

$\chi_N = 0.2934$. Sussiste un livello di connessione basso.

ESERCIZIO 2.

Si consideri la seguente serie dei numeri indici a base fissa, con base 2010 = 1, della serie X_t dei prezzi del prodotto A :

t	$NIBF$	$NIBM$
2010	1	—
2011	0.92	0.92
2012	0.98	1.0652
2013	1.03	1.0510
2014	1.01	0.9806

1. Si costruisca la corrispondente serie dei numeri indici a base mobile.

2. Sapendo che $x_{2013} = 261$, si calcoli x_{2014} .

$$x_{2014} = x_{2013} \cdot {}_{2013}I_{2014} = 261 \cdot 0.9806 = 255.93$$

3. Si calcoli il tasso medio di variazione dal 2010 al 2014 e si commenti il risultato ottenuto.

$\sqrt[4]{{}_{2007}I_{2010}} = \sqrt[4]{1.01} = 1.0025$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media dello 0.25%.

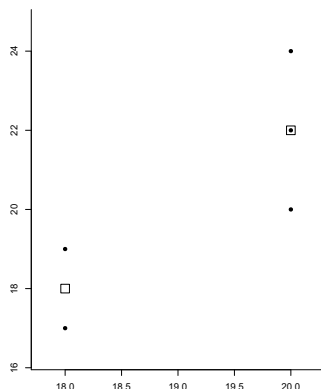
Il prezzo del prodotto A ha registrato un incremento medio giornaliero dello 0.25% tra il giorno 7 e il giorno 10.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	18	18	20	20	20
y_i	17	19	20	22	24

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$$\hat{a} = -18, \hat{b} = 2, \text{ indice adattamento } \rho^2 = 0.6575 \quad \hat{c} = 20.4, R^2 = 0$$

Il modello migliore è $Y^* = -18 + 2X$.

3. Si calcoli $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = \rho^2 = 0.6575$ in quanto le medie condizionate sono allineate.

ESERCIZIO 4.

Una compagnia assicurativa valuta che la probabilità che un suo assicurato richieda un risarcimento per danni provocati da eventi naturali sia pari all'1.1%.

1. Si calcoli la probabilità che, estratti in modo indipendente 10 clienti che hanno stipulato una polizza per danni provocati da eventi naturali, almeno due richiedano il risarcimento assicurativo.

Sia X la variabile casuale 'n. di clienti che richiedono il risarcimento assicurativo nelle 10 estrazioni indipendenti'. $X \sim \text{Bin}(n = 10, p = 0.011)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - (P(X = 0) + P(X = 1)) = \\ &= 1 - \left\{ \binom{10}{0} 0.011^0 0.989^{10} + \binom{10}{1} 0.011^1 0.989^9 \right\} = \\ &= 1 - (0.8953 + 0.0996) = 1 - 0.9949 = 0.0051. \end{aligned}$$

2. Si calcolino media e varianza della variabile casuale descritta dall'esperimento al punto precedente.

Essendo $X \sim \text{Bin}(n = 10, p = 0.011)$ abbiamo

$$\begin{aligned} M(X) &= np = 10 \cdot 0.011 = 0.11 \\ \text{Var}(X) &= np(1 - p) = 10 \cdot 0.011 \cdot (1 - 0.011) = 10 \cdot 0.011 \cdot 0.989 = 0.1088 \end{aligned}$$

3. Si calcoli la probabilità che, estratti in modo indipendente 1000 clienti che hanno stipulato una polizza per danni provocati da eventi naturali, al massimo 20 richiedano il risarcimento assicurativo.

Indicata con X la variabile casuale 'n. di clienti che richiedono il risarcimento assicurativo nelle 1000 estrazioni indipendenti' risulta $X \sim \text{Bin}(n = 1000, p = 0.011)$. $M(X) = 11$; $\text{Var}(X) = 10.879$. Dal momento che $np = 11 > 5$ ed $n(1 - p) = 989 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 11, \sigma^2 = 10.879).$$

Si ottiene

$$P(X \leq 20) = P\left(\frac{X - \mu}{\sigma} \leq \frac{20 - \mu}{\sigma}\right) = P\left(Z \leq \frac{20 - 11}{3.2983}\right) = P(Z \leq 2.73) = 0.9968.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 09.07.2015 (Tema 284-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

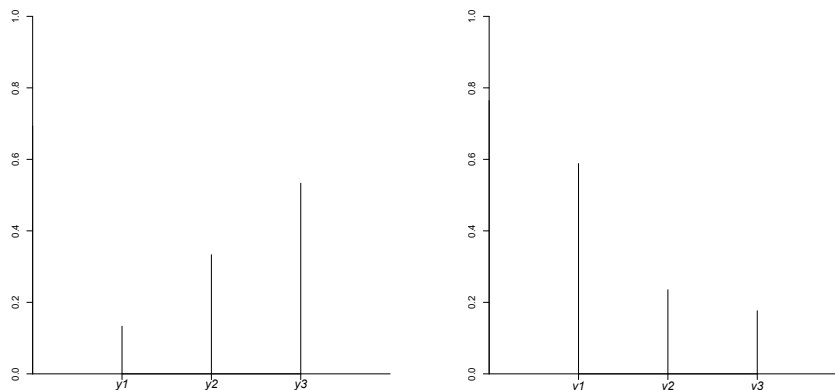
ESERCIZIO 1.

Il seguente prospetto riporta la distribuzione dei clienti presenti su un mercato dove opera l'agenzia A , riclassificati mediante le variabili X , con categorie $x_1 = \text{'clienti agenzia A'}$ e $x_2 = \text{'clienti della concorrenza'}$, e $Y = \text{'tipologia di viaggio'}$, con categorie $y_1 = \text{'low cost'}$, $y_2 = \text{'medium cost'}$ e $y_3 = \text{'high cost'}$.

Nella soluzione di questo esercizio si tenga presente che la clientela *target* dell'agenzia A è del tipo *'high cost'*.

$X \setminus Y$	y_1	y_2	y_3
x_1	200	500	800
x_2	5000	2000	1500

1. Si rappresentino graficamente le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.



2. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e di variabilità.

$Y|x_1$: Indice di posizione: $\text{Moda}(Y|x_1) = y_3$ Indice di variabilità: $F_N(Y|x_1) = 0.6536$

$Y|x_2$: Indice di posizione: $\text{Moda}(Y|x_2) = y_1$ Indice di variabilità: $F_N(Y|x_2) = 0.6143$

La maggior parte delle tipologie di viaggio scelte dai clienti dell'agenzia A (53%) sono di tipo *'high cost'*. La concorrenza è, invece, focalizzata su tipologie di viaggio *'low cost'* (59%). Entrambe le distribuzioni presentano un livello medio di eterogeneità, confermando la presenza di tutte e 3 le tipologie di viaggio nell'offerta sia dell'agenzia A che della concorrenza.

3. Si studi il posizionamento dell'agenzia A sul mercato utilizzando le contingenze assolute e le contingenze relative.

$X \setminus Y$	y_1	y_2	y_3
x_1	-580	125	455
x_2	580	-125	-455

$X \setminus Y$	y_1	y_2	y_3
x_1	-74.36%	+33.33%	+131.88%
x_2	+13.12%	-5.88%	-23.27%

Si rileva un ottimo posizionamento dell'agenzia A con riferimento alle mete di viaggio *'high cost'* con un *surplus* di clientela rispetto alla situazione di indipendenza stocastica di 455 unità in termini assoluti e del 131.88% in termini relativi. Con riferimento alle mete *'low cost'* il *deficit* di clientela è di 580 unità, circa il 74.36% in meno dei 780 clienti che si avrebbero nella situazione di indipendenza stocastica. In tale situazione il posizionamento dell'agenzia A sarebbe identico a quello della concorrenza. I surplus (131.88%) e deficit (-74.36%) di clientela rappresentano delle componenti distintive dell'agenzia A rispetto al comportamento medio generale del mercato.

4. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

$\chi_N = 0.3552$. Sussiste un livello di connessione basso.

ESERCIZIO 2.

Si consideri la seguente serie dei numeri indici a base mobile della serie X_t dei prezzi del prodotto A :

t	$NIBM$	$NIBF$
2010	--	1.000
2011	0.92	0.9200
2012	0.98	0.9016
2013	1.03	0.9286
2014	1.01	0.9379

1. Si costruisca la corrispondente serie dei numeri indici a base fissa, con base 2010 = 1.

2. Sapendo che $x_{2011} = 261$, si calcoli x_{2010} .

$$x_{2010} = \frac{x_{2011}}{I_{2010}^{2011}} = \frac{261}{0.92} = 283.6957$$

3. Si calcoli il tasso medio di variazione dal 2010 al 2014 e si commenti il risultato ottenuto.

$\sqrt[4]{I_{2014}^{2010}} = \sqrt[4]{0.92 \cdot 0.98 \cdot 1.03 \cdot 1.01} = \sqrt[4]{0.9379} = 0.9841$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del -1.59%.

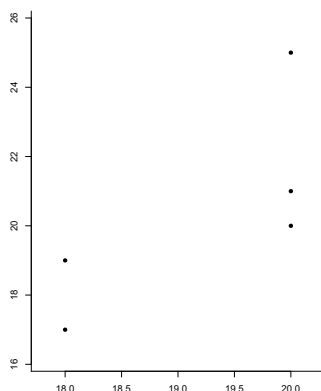
Il prezzo del prodotto A ha registrato un decremento medio annuo dell'1.59% tra l'anno 2010 e l'anno 2014.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	18	18	20	20	20
y_i	17	19	20	22	24

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = -18, \hat{b} = 2, \text{ indice adattamento } M(\hat{E}^2) = 2 \quad \hat{c} = 1.0649, \text{ indice adattamento } M(\hat{E}^2) = 2.8416$$

Il modello migliore è $Y^* = -18 + 2X$.

ESERCIZIO 4.

Si consideri il seguente esperimento. Si lancia una moneta (equilibrata). Se esce testa si effettuano 2 ulteriori lanci della moneta; nel caso esca croce si effettua 1 solo ulteriore lancio.

1. Si costruisca la variabile casuale $T =$ 'n. di teste ottenute alla fine dell'esperimento'.

t	$P(T = t)$
0	0.25
1	0.375
2	0.25
3	0.125
	1

2. Si calcolino media e varianza della variabile casuale descritta dall'esperimento al punto precedente.

$$M(T) = 1.25 \quad Var(T) = 0.9375$$

3. Si calcoli la probabilità che, lanciando 1000 volte la moneta si ottengano almeno 500 teste.

Indicata con X la variabile casuale 'n. di teste nei 1000 lanci (prove indipendenti)' risulta

$$X \sim Bin(n = 1000, p = 0.5).$$

$M(X) = 500$; $Var(X) = 250$. Dal momento che $np = 500 > 5$ ed $n(1-p) = 500 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 500, \sigma^2 = 250).$$

Si ottiene

$$P(X \geq 500) = P(X \geq \mu) = 0.5.$$

Università Cattolica del Sacro Cuore - Milano
FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 01.09.2015 (Tema 285-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

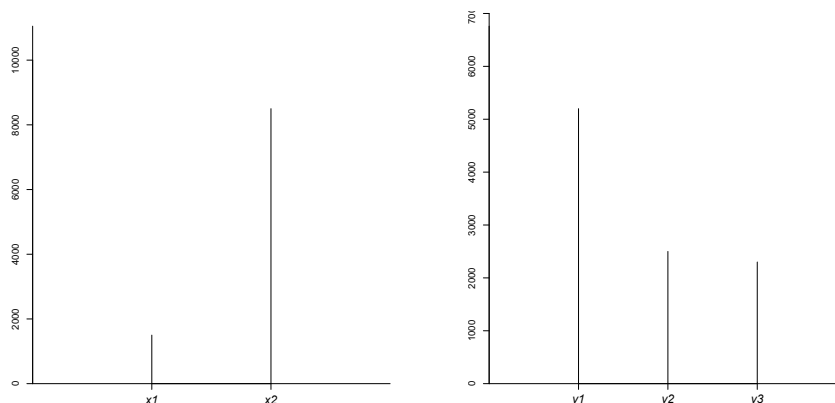
ESERCIZIO 1.

Il seguente prospetto riporta la distribuzione dei clienti presenti su un mercato dove opera l'agenzia A , riclassificati mediante le variabili X , con categorie $x_1 = \text{'clienti agenzia } A\text{'}$ e $x_2 = \text{'clienti della concorrenza'}$, e $Y = \text{'tipologia di viaggio'}$, con categorie $y_1 = \text{'low cost'}$, $y_2 = \text{'medium cost'}$ e $y_3 = \text{'high cost'}$.

Nella soluzione di questo esercizio si tenga presente che la clientela *target* dell'agenzia A è del tipo *'high cost'*.

$X \setminus Y$	y_1	y_2	y_3
x_1	780	375	345
x_2	4420	2125	1955

1. Si rappresentino graficamente le distribuzioni marginali X e Y .



2. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$, calcolando moda, mediana, media aritmetica, e una opportuna misura di variabilità.

$Y|x_1$: moda: $y_1 = \text{'low cost'}$ mediana: $y_1 = \text{'low cost'}$ media aritmetica: non è definita per un carattere qualitativo
Indice di variabilità: $F_N = 0.7195$

$Y|x_2$: moda: $y_1 = \text{'low cost'}$ mediana: $y_1 = \text{'low cost'}$ media aritmetica: non è definita per un carattere qualitativo
Indice di variabilità: $F_N = 0.7195$

Sia l'agenzia A che la concorrenza sono focalizzate su tipologie di viaggio *'low cost'*. Le distribuzioni delle frequenze condizionate relative, utilizzate per il calcolo dell'indice di Gini, sono identiche; si evince, quindi, che tra le variabili X e Y sussiste indipendenza stocastica. La struttura della clientela dell'agenzia A coincide con quella della concorrenza.

3. Si studi il posizionamento dell'agenzia A sul mercato utilizzando le contingenze assolute e le contingenze relative.

	$X \setminus Y$	y_1	y_2	y_3		$X \setminus Y$	y_1	y_2	y_3
contingenze assolute	x_1	0	0	0	contingenze relative	x_1	0	0	0
	x_2	0	0	0		x_2	0	0	0

Dal momento che sussiste indipendenza stocastica le frequenze congiunte n_{ij} coincidono con le frequenze teoriche $\hat{n}_{ij}$; quindi le contingenze assolute $n_{ij} - \hat{n}_{ij}$ sono nulle, i rapporti di contingenza $\frac{n_{ij}}{\hat{n}_{ij}}$ assumono valore unitario e le contingenze relative $\frac{n_{ij} - \hat{n}_{ij}}{\hat{n}_{ij}} = \frac{n_{ij}}{\hat{n}_{ij}} - 1$ sono pure nulle.

4. Si indichi se esiste connessione tra le variabili X e Y .

Dal momento che sussiste indipendenza stocastica non è presente connessione tra le variabili X e Y .

ESERCIZIO 2.

Si consideri la seguente serie X_t dei prezzi del prodotto A :

t	X_t	$NIBF$	$NIBM$
2010	90	1	--
2011	92	1.0222	1.0222
2012	98	1.0889	1.0652
2013	103	1.1444	1.0510
2014	101	1.1222	0.9806

1. Si costruiscano le corrispondenti serie dei numeri indici a base fissa, con base 2010 = 1, e a base mobile.

2. Sapendo che $x_{2016} = 261$, si calcoli il tasso medio di variazione tra il 2010 e il 2016 e si commenti il risultato ottenuto.

$\sqrt[6]{2010 I_{2016}} = \sqrt[6]{\frac{261}{90}} = \sqrt[6]{2.9} = 1.1942$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 19.42%.

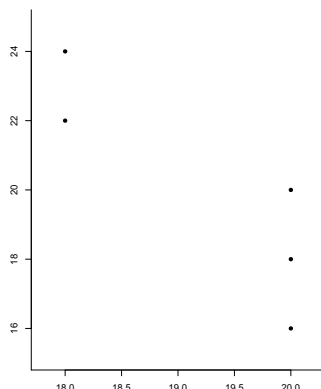
Il prezzo del prodotto A ha registrato un incremento medio annuo del 19.42% tra l'anno 2010 e l'anno 2016.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	18	18	20	20	20
y_i	24	22	20	18	16

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 68, \hat{b} = -2.5, \text{ indice adattamento } M(\hat{E}^2) = 2 \quad \hat{c} = 1.0325, \text{ indice adattamento } M(\hat{E}^2) = 14.0104$$

Il modello migliore è $Y^* = 68 - 2.5X$.

ESERCIZIO 4.

Per il seguente gioco è prevista una posta di 3 Euro.

Si lancia una moneta (equilibrata).

Se esce testa si lancia un dado con 6 facce; nel caso esca croce si lancia un dado con 4 facce.

Si vince un ammontare in Euro pari al numero che si è presentato sulla base del dado lanciato.

1. Si stabilisca se il gioco è equo.

Indicati con M , $D6$ e $D4$ i risultati del lancio della moneta e dei dadi a 6 e 4 facce, la variabile casuale guadagno risulta

x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
-2	$P(M = T) \cdot P(D6 = 1 M = T) + P(M = C) \cdot P(D4 = 1 M = C) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}$	$-\frac{10}{24}$
-1	$P(M = T) \cdot P(D6 = 2 M = T) + P(M = C) \cdot P(D4 = 2 M = C) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}$	$-\frac{5}{24}$
0	$P(M = T) \cdot P(D6 = 3 M = T) + P(M = C) \cdot P(D4 = 3 M = C) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}$	0
1	$P(M = T) \cdot P(D6 = 4 M = T) + P(M = C) \cdot P(D4 = 4 M = C) = \frac{1}{2} \cdot \frac{1}{6} + \frac{1}{2} \cdot \frac{1}{4} = \frac{1}{12} + \frac{1}{8} = \frac{5}{24}$	$\frac{5}{24}$
2	$P(M = T) \cdot P(D6 = 5 M = T) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} = \frac{2}{24}$	$\frac{2}{24}$
3	$P(M = T) \cdot P(D6 = 6 M = T) = \frac{1}{2} \cdot \frac{1}{6} = \frac{1}{12} = \frac{2}{24}$	$\frac{6}{24}$
	1	0

Il valore atteso della variabile casuale guadagno è nullo, quindi il gioco è equo.

2. Si calcoli la probabilità che, lanciando 9 volte la moneta si ottengano non più di 7 teste.

Sia X la variabile casuale 'n. di teste nelle 9 prove indipendenti', $X \sim \text{Bin}(n = 9, p = 0.5)$.

$$\begin{aligned} P(X \leq 7) &= 1 - P(X > 7) = 1 - [P(X = 8) + P(X = 9)] = \\ &= 1 - \left[\binom{9}{8} 0.5^8 \cdot 0.5^1 + \binom{9}{9} 0.5^9 \cdot 0.5^0 \right] = \\ &= 1 - (0.01758 + 0.001953) = 1 - 0.019533 = 0.980467. \end{aligned}$$

3. Si calcoli la probabilità che, lanciando 900 volte la moneta si ottengano non più di 450 teste.

Indicata con X la variabile casuale 'n. di teste nei 900 lanci (prove indipendenti)' risulta $X \sim \text{Bin}(n = 900, p = 0.5)$. $M(X) = 450$; $Var(X) = 225$. Dal momento che $np = 450 > 5$ ed $n(1-p) = 450 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 450, \sigma^2 = 225)$. Si ottiene

$$P(X \leq 450) = P(X \leq \mu) = 0.5$$

essendo la distribuzione Normale simmetrica.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 15.09.2015 (Tema 285)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

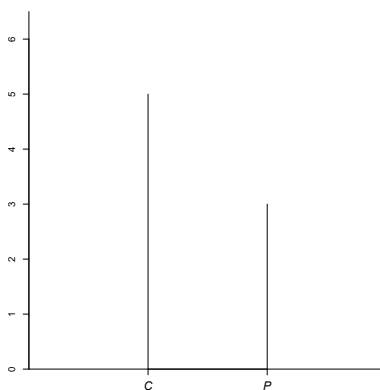
ESERCIZIO 1.

Per 8 sale cinematografiche vengono rilevate le seguenti variabili: F = fatturato medio settimanale (in migliaia di euro), S = spesa mensile per manutenzione sala e noleggio film (in migliaia di euro) e L = localizzazione della sala (C = centrale, P = periferica). I dati sono riportati di seguito:

F	15	6	11	16	18	14	16	12
S	8.2	3.4	6.2	12.5	13.2	7.2	11.2	5.2
L	C	C	P	C	P	C	C	P

1. Si rappresenti graficamente la distribuzione della mutabile L .

l_i	n_i
C	5
P	3



2. Si valuti, confrontando opportuni indici, la variabilità delle Spese e del Fatturato.

S Indice di posizione: $M(S) = 8.3875$ Indice di variabilità: $CV(S) = 0.3982$

F Indice di posizione: $M(F) = 13.5$ Indice di variabilità: $CV(F) = 0.2619$

Come ci si poteva attendere, le spese presentano un livello medio inferiore al fatturato.

Le spese presentano una maggiore variabilità rispetto al fatturato.

3. Per determinare il grado di connessione tra le variabili F e L si costruisca la corrispondente tabella a doppia entrata raggruppando le modalità di F in 3 classi, chiuse a destra, di ampiezza 7 a partire da 0 (limite inf della prima classe) e si calcoli un opportuno indice.

$L \setminus F$	$0 \div 7$	$7 \div 14$	$14 \div 21$
C	1	1	3
P	0	2	1

$\chi_N = 0.4944$. Il livello di connessione esistente tra le due variabili è medio.

ESERCIZIO 2.

Della serie storica X_t del fatturato medio nazionale delle sale cinematografiche tra il 2008 e il 2012 si hanno le seguenti serie incomplete di numeri indici espressi su scala centesimale (base 100):

t	$NIBM$	$NIBF$
2008		100
2009	114	
2010		116
2011	106	
2012		131

t	$NIBM$	$NIBF$
2008	—	100
2009	114	114
2010	101.75	116
2011	106	122.96
2012	106.54	131

1. Si completino le due serie di numeri indici in tabella.

2. Sapendo che $x_{2010} = 261$, si calcoli x_{2012} .

$$x_{2012} = x_{2010} \cdot {}_{2010}I_{2011} \cdot {}_{2011}I_{2012} = 261 \cdot 1.06 \cdot 1.0654 = 294.75.$$

3. Si calcoli il tasso medio di variazione dal 2010 al 2012 e si commenti il risultato ottenuto.

$\sqrt[2]{{}_{2010}I_{2012}} = \sqrt[2]{1.06 \cdot 1.0654} = \sqrt[2]{1.1293} = 1.0627$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 6.27%.

Il prezzo del prodotto A ha registrato un incremento medio annuo del 6.27% tra l'anno 2010 e l'anno 2012.

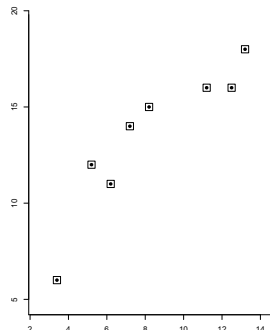
ESERCIZIO 3.

Con riferimento ai singoli dati relativi alle 8 sale cinematografiche (già utilizzati al punto 2 dell'Esercizio 1)

F	15	6	11	16	18	14	16	12
S	8.2	3.4	6.2	12.5	13.2	7.2	11.2	5.2
L	C	C	P	C	P	C	C	P

si vuole studiare la relazione tra Fatturato (variabile dipendente) e Spesa.

1. Si rappresentino graficamente le coppie di dati e la funzione di regressione di F dato S .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) F^* = a + b \cdot S \quad II) F^* = c \cdot \sqrt{S}$$

$$\hat{a} = 5.4413, \hat{b} = 0.9608, \text{ indice adattamento } M(\hat{E}^2) = 2.2014 \quad \hat{c} = 4.7966, \text{ indice adattamento } M(\hat{E}^2) = 1.7784$$

Il modello migliore è $Y^* = 4.7966\sqrt{X}$.

3. Sulla base del miglior modello si preveda il fatturato di una sala cinematografica con spese pari a 5 mila euro.

$$Y^*(5) = 4.7966\sqrt{5} = 10.7255.$$

4. Si determini il valore del rapporto di correlazione tra le due variabili $\eta_{F|S}^2$, commentando il valore trovato.

$\eta_{F|S}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X .

5. Descrivere come si potrebbe trattare la mutabile L per inserirla nel modello di regressione I).

Occorre creare una variabile indicatrice, o variabile *dummy*, X , definita a partire da L , ponendo in corrispondenza della i -ma unità statistica:

$$x_i = \begin{cases} 1 & \text{se } l_i = C \\ 0 & \text{se } l_i = P \end{cases} \quad \text{oppure} \quad x_i = \begin{cases} 1 & \text{se } l_i = P \\ 0 & \text{se } l_i = C \end{cases}$$

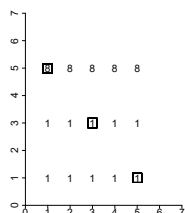
ESERCIZIO 4.

Date due differenti urne, U_1 e U_2 , contenenti palline numerate, un esperimento casuale consiste nell'estrarre una pallina da ciascuna urna. Sapendo che nell'urna U_1 vi sono 5 palline con numeri da 1 a 5 e che nell'urna U_2 vi sono 1 pallina con il numero 1, 1 pallina con il numero 3 e 8 palline con il numero 5:

1. costruire la distribuzione di probabilità della variabile casuale X = "risultato estrazione urna U_2 ".

x_i	$P(X = x_i)$
1	0.1
3	0.1
5	0.8

2. Calcolare la probabilità che la somma dei numeri delle 2 palline estratte dia 6.



$$P(\text{somma dei numeri} = 6) = \frac{8 + 1 + 1}{50} = \frac{10}{50} = 0.2$$

oppure

$$\begin{aligned} &= P([(U_1 = 1) \cap (U_2 = 5)] \cup [(U_1 = 3) \cap (U_2 = 3)] \cup [(U_1 = 5) \cap (U_2 = 1)]) \\ &= P(U_1 = 1) \cdot P(U_2 = 5) + P(U_1 = 3) \cdot P(U_2 = 3) + P(U_1 = 5) \cdot P(U_2 = 1) \\ &= \frac{1}{5} \cdot \frac{8}{10} + \frac{1}{5} \cdot \frac{1}{10} + \frac{1}{5} \cdot \frac{1}{10} = \frac{10}{50} = 0.2 \end{aligned}$$

3. Calcolare la probabilità che, su 10 estrazioni con reinserimento dall'urna U_2 , almeno 2 palline presentino il numero 1. Sia X la variabile casuale 'n. di palline con il numero 1 nelle 10 estrazioni con reinserimento (prove indipendenti)', $X \sim \text{Bin}(n = 10, p = 0.1)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = \\ &= 1 - \left[\binom{10}{0} 0.1^0 \cdot 0.9^{10} + \binom{10}{1} 0.1^1 \cdot 0.9^9 \right] = \\ &= 1 - (0.3487 + 0.3874) = 1 - 0.7361 = 0.2639. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano
FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 14.01.2016 (Tema 286)

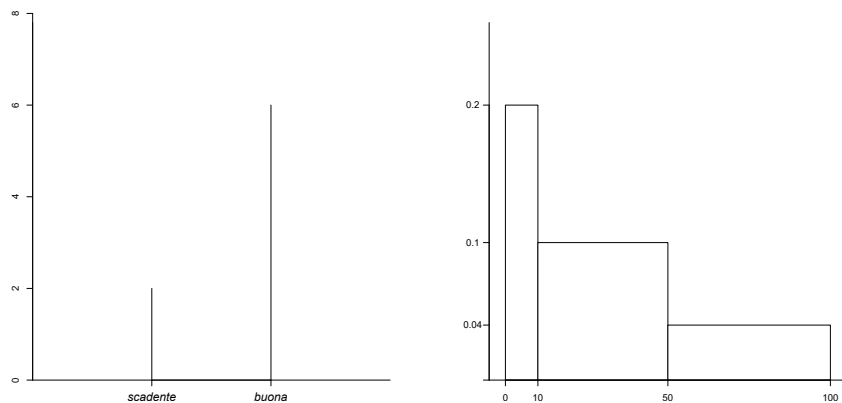
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Data la seguente tabella a doppia entrata della distribuzione congiunta dei caratteri X = 'quantità di neve caduta (in cm)' in un mese e Y = 'qualità della neve (scadente o buona)' in alcune località di montagna:

$Y \setminus X$	$0 \div 10$	$10 \div 50$	$50 \div 100$
<i>scadente</i>	0	2	0
<i>buona</i>	2	2	2

1. Fornire la rappresentazione grafica delle distribuzioni *marginali*.



2. Confrontare con un opportuno indice la variabilità delle distribuzioni condizionate della *qualità* della neve (variabile Y) rispetto alla *quantità* di neve caduta. Commentare i risultati.

Indice per confronto variabilità: Indice di Frosini normalizzato

valore $Y|x_1$ $F_N = 0$ valore $Y|x_2$ $F_N = 1$ valore $Y|x_3$ $F_N = 0$

Le distribuzioni condizionate $Y|x_1$ e $Y|x_3$ sono degeneri e quindi caratterizzate da assenza di eterogeneità. La distribuzione condizionata $Y|x_2$ è caratterizzata da equidistribuzione delle frequenze; presenta, quindi, massima eterogeneità.

3. Calcolare moda e mediana per la marginale X .

$\text{Moda}(X) = 0 \div 10$ $x_{0.5} = 30$

4. Misurare la dipendenza tra le due variabili con un opportuno indice di connessione.

$\chi_N = 0.5773$. Sussiste un livello di connessione medio.

ESERCIZIO 2.

Il seguente prospetto riporta i dati dei prezzi medi annui di 2 beni di largo consumo per 5 anni:

t	<i>bene A</i>	<i>bene B</i>	<i>NIBM</i>
2005	10	2	—
2006	11	3	1.5000
2007	10	4	1.3333
2008	12	4	1.0000
2009	14	3	0.7500

1. Calcolare, giustificando la risposta, la variazione percentuale del prezzo medio del bene A passando dal 2006 al 2009.

${}_{2006}I_{2009} = \frac{x_{2009}}{x_{2006}} = \frac{14}{11} = 1.2727$. Tra il tempo 2006 e il tempo 2009 si è registrato un incremento complessivo del 27.27% della grandezza oggetto di studio.

2. Calcolare, mostrandone il procedimento, la serie dei NIBM del bene B .

3. Calcolare il tasso medio di variazione del prezzo del bene B nell'intervallo 2005-2008, commentandone il risultato.

$\sqrt[3]{{}_{2005}I_{2008}} = \sqrt[3]{\frac{4}{2}} = \sqrt[3]{2} = 1.2599$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 25.99%.

Il prezzo medio del prodotto B ha registrato un incremento medio annuo del 25.99% tra l'anno 2005 e l'anno 2008.

4. Sapendo che le quantità dei due beni vendute nel 2005 sono state rispettivamente 100 e 300, calcolare, se possibile, l'indice di Laspeyres dei prezzi al 2009 con base 2005.

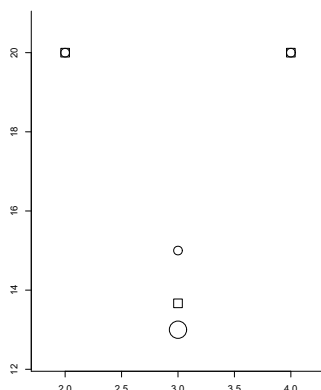
${}_{2005}^p I_{2009}^L = 1.5$. La variazione relativa nel livello generale dei prezzi, secondo l'Indice di Laspeyres, è del 50%.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolte le seguenti 5 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 5$

x_i	2	3	3	3	4
y_i	20	13	13	15	20

1. Fornire la rappresentazione grafica delle coppie di punti e della funzione di regressione di Y dato X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione.

$$I) Y^* = a + b \cdot X \quad II) Y^* = \alpha + \beta \cdot (X - 3)^2$$

$\hat{a} = 16.2$, $\hat{b} = 0$, indice adattamento $\rho^2 = 0$ $\hat{\alpha} = 13.6667$, $\hat{\beta} = 6.3333$, indice adattamento $\rho^2 = 0.9475$
Il modello migliore è $Y^* = 13.6667 + 6.3333(X - 3)^2$.

3. Confrontare gli indici di adattamento, commentando le eventuali differenze.

Cfr. punto precedente per indici adattamento. Si osserva che il modello I) è una retta con coefficiente angolare nullo; non sussiste, quindi, correlazione lineare tra le variabili in gioco. Il modello II) è una parabola con asse di simmetria in $x = 3$ e passa esattamente per le tre medie condizionate. L'indice di adattamento del secondo modello coincide, quindi, con il rapporto di correlazione $\eta_{Y|X}^2$.

4. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 4$.
 $Y^*(4) = 13.6667 + 6.3333(4 - 3)^2 = 20$.

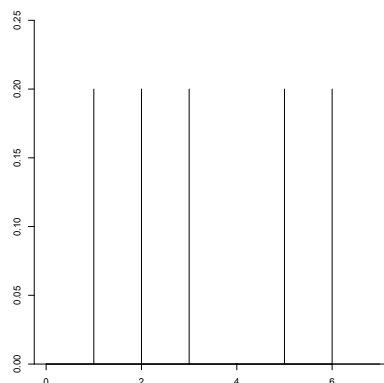
ESERCIZIO 4.

Un dado Δ viene truccato in modo che la faccia 4 non esca mai:

1. Descrivere (valori e probabilità) e rappresentare graficamente la v.c. $X =$ 'risultato del lancio del dado Δ '.

$$\Omega = \boxed{1, 2, 3, 5, 6}$$

x_i	$P(X = x_i)$
1	0.2
2	0.2
3	0.2
4	0
5	0.2
6	0.2



Nel prospetto si è riportato il valore 4, anche se caratterizzato da probabilità pari a 0, in quanto valore di interesse, dal momento che la corrispondente faccia è fisicamente presente sul dado.

2. Calcolare la probabilità che, su 120 lanci del dado Δ , almeno la metà dia risultato pari.

Indicata con X la variabile casuale 'numero di successi' nelle 120 prove indipendenti risulta $X \sim \text{Bin}(n = 120, p = 0.4)$. $M(X) = 48$; $\text{Var}(X) = 28.8$. Dal momento che $np = 48 > 5$ ed $n(1 - p) = 72 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 48, \sigma^2 = 28.8)$.

Si ottiene $P(X > 60) = P\left(\frac{X - \mu}{\sigma} > \frac{60 - \mu}{\sigma}\right) = P\left(Z > \frac{60 - 48}{\sqrt{28.8}}\right) = P(Z > 2.24) = 1 - P(Z \leq 2.24) = 1 - 0.9875 = 0.0125$.

3. Calcolare la probabilità che in una generica estrazione del LOTTO (5 numeri estratti da 90, senza reimmissione) i primi 3 siano dispari e gli altri siano pari.

$$\begin{aligned} P(d_1 \cap d_2 \cap d_3 \cap p_4 \cap p_5) &= P(d_1) \cdot P(d_2|d_1) \cdot P(d_3|d_1 \cap d_2) \cdot P(p_4|d_1 \cap d_2 \cap d_3) \cdot P(p_5|d_1 \cap d_2 \cap d_3 \cap p_4) = \\ &= \frac{45}{90} \cdot \frac{44}{89} \cdot \frac{43}{88} \cdot \frac{45}{87} \cdot \frac{44}{86} = 0.03196 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 28.01.2016 (Tema 287)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

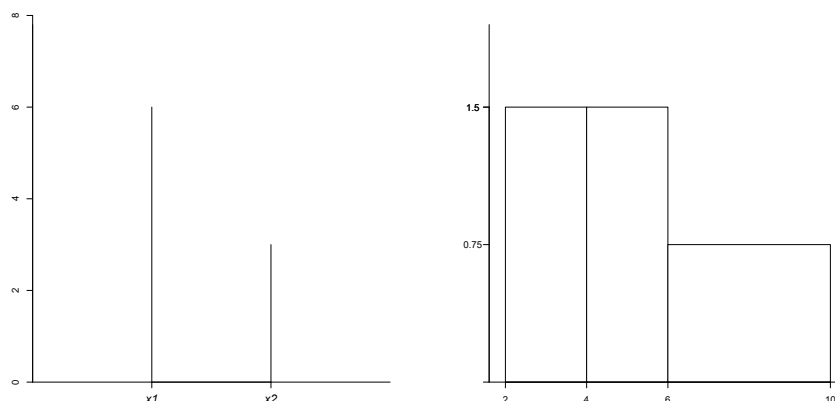
ESERCIZIO 1.

La seguente tabella riporta alcune informazioni sulla distribuzione congiunta della variabile statistica doppia (X, Y) , con X carattere qualitativo.

$X \setminus Y$	y_1	y_2	y_3	
x_1	2	2	2	6
x_2	1	1	1	3
	3	3	3	9

1. Si completino le frequenze nella tabella in modo tale che siano verificate contemporaneamente le tre seguenti condizioni:
a) la distribuzione $Y|x_1$ sia caratterizzata da massima eterogeneità; b) $\chi^2_{\max} = 9$; c) sussista indipendenza stocastica tra le variabili X e Y .

2. Si sostituiscano alle modalità y_1, y_2 e y_3 di Y le classi $2 \rightarrow 4, 4 \rightarrow 6$ e $6 \rightarrow 10$, e si rappresentino graficamente le distribuzioni marginali X e Y .



3. Posto ora $n_{11} = n_{13} = 2$ e $n_{21} = n_{22} = n_{23} = 11$, si calcolino le medie e le varianze delle distribuzioni condizionate $Y|x_1$ e $Y|x_2$ e si confronti la variabilità delle due distribuzioni condizionate.

$$M(Y|x_1) = 5.3333 \quad Var(Y|x_1) = 4.2222 \quad \text{Indice per confronto variabilità } CV(Y|x_1) = 0.3853$$

$$M(Y|x_2) = 5.3333 \quad Var(Y|x_2) = 4.2222 \quad \text{Indice per confronto variabilità } CV(Y|x_2) = 0.3853$$

Dal momento che sussiste indipendenza stocastica le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono somiglianti e, quindi, caratterizzate da eguali valori di media aritmetica, varianza e coefficiente di variazione.

4. Si ricostruisca la varianza della marginale Y a partire dai valori delle medie e delle varianze condizionate determinati al punto precedente.

$Var(Y) = 4.2222$. Dal momento che sussiste indipendenza stocastica le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono somiglianti alla marginale Y e, quindi, caratterizzate dalla stessa varianza.

Con riferimento al risultato di scomposizione della varianza, $Var(Y) = Var_X\{M(Y|X)\} + M_X\{Var(Y|X)\}$, le due distribuzioni condizionate hanno la stessa media e, quindi, la varianza spiegata assume valore nullo; le due distribuzioni condizionate hanno la stessa varianza e, quindi, la varianza residua assume lo stesso valore delle varianze condizionate.

ESERCIZIO 2.

Si consideri la seguente serie storica:

t	x_t	$NIBF$	$NIBM$
2008	500	2	—
2009	250	1	0.5
2010	125	0.5	0.5
2011	250	1	2
2012	500	2	2

1. Si ricostruiscono le serie dei numeri indici a base fissa, con base 2009, e a base mobile.

2. Si calcolino e si commentino i seguenti numeri indici ${}_{2010}I_{2008}$ e ${}_{2010}I_{2012}$.

$${}_{2010}I_{2008} = 4. \quad x_{2008} \text{ è superiore del } 300\% \text{ a } x_{2010}.$$

$${}_{2010}I_{2012} = 4. \quad \text{Tra il 2010 e il 2012 si è registrato un incremento del } 300\%.$$

3. Si calcoli il tasso medio di variazione dal 2008 al 2011 e si commenti il risultato ottenuto.

$$\sqrt[3]{{}_{2008}I_{2011}} = \sqrt[3]{\frac{250}{500}} = \sqrt[3]{0.5} = 0.7937, \text{ numero indice a base mobile medio, a cui corrisponde una variazione relativa media del } -20.63\%.$$

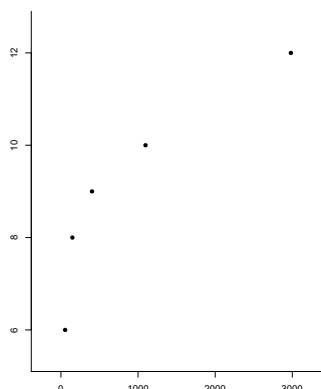
La grandezza esaminata ha registrato un decremento medio annuo del 20.63% tra l'anno 2008 e l'anno 2011.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	54.6	148.41	403.43	1096.63	2980.96
y_i	6	8	9	10	12

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot \ln(X) \quad II) Y^* = c$$

$$\hat{a} = 0.6, \hat{b} = 1.4, \text{ indice adattamento } \rho^2 = 0.98 \quad \hat{c} = 9, \text{ indice adattamento } R^2 = 0$$

Il modello migliore è $Y^* = 0.6 + 1.4 \cdot \ln(X)$.

3. Si indichino i valori assunti da $\eta_{Y|X}^2$ ed $\eta_{X|Y}^2$.

$\eta_{Y|X}^2 = 1$ ed $\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale bijectiva.

ESERCIZIO 4.

Per il seguente gioco è prevista una posta di 4 Euro: si lancia un dado regolare; se esce la faccia 3 oppure la 4 si vince 1 Euro, se esce la faccia 5 oppure la 6 si vincono 8 Euro, altrimenti si perde.

1. Costruita la variabile casuale Y = 'guadagno del giocatore' si stabilisca se il gioco è equo.

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
<i>posta</i>	-4	0.3333	-1.3333
<i>vincita se esce 3 o 4 - posta</i>	-3	0.3333	-1
<i>vincita se esce 5 o 6 - posta</i>	4	0.3333	1.3333
		1	-1

Il valore atteso della variabile casuale guadagno è negativo, quindi il gioco non è equo.

2. Nell'ipotesi di ripetere il gioco 3 volte si calcoli la probabilità di vincere almeno 2 volte.

Sia X la variabile casuale 'n. di vincite nelle 3 prove indipendenti'.

$$X \sim \text{Bin}(n = 3, p = 0.6667).$$

$$\begin{aligned} P(X \geq 2) &= P(X = 2) + P(X = 3) = \\ &= \binom{3}{2} 0.6667^2 \cdot 0.3333^1 + \binom{3}{3} 0.6667^3 \cdot 0.3333^0 = \\ &= 0.4444 + 0.2963 = 0.7407. \end{aligned}$$

3. Nell'ipotesi di ripetere il gioco 100 volte si calcoli la probabilità di vincere almeno 59 volte.

Indicata con X la variabile casuale 'numero di successi' nelle 100 prove indipendenti risulta $X \sim \text{Bin}(n = 100, p = 0.6667)$. $M(X) = 66.67$; $Var(X) = 22.22$. Dal momento che $np = 66.67 > 5$ ed $n(1 - p) = 33.33 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 66.67, \sigma^2 = 22.22)$. Si ottiene

$$\begin{aligned} P(X \geq 59) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{59 - \mu}{\sigma}\right) = P\left(Z \geq \frac{59 - 66.67}{4.7148}\right) = \\ &= P(Z \geq -1.63) = P(Z \leq 1.63) = 0.9484. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 11.02.2016 (Tema 288)

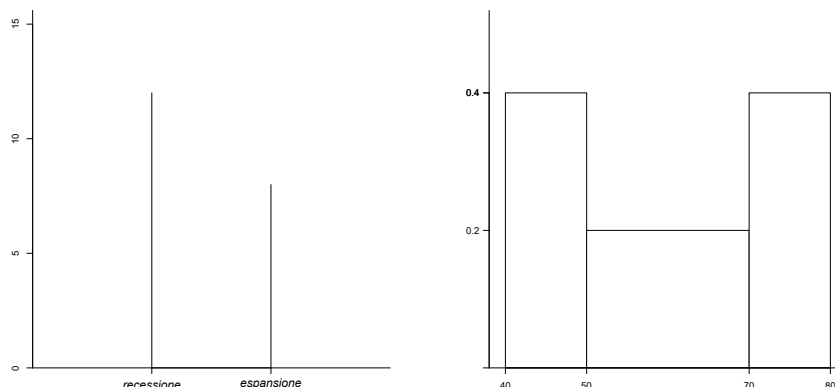
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Al fine di studiare la relazione esistente tra il prezzo di un determinato prodotto e l'andamento economico, vengono rilevate le variabili X = 'andamento economico' e Y = 'prezzo del prodotto'. I dati sono raccolti nella seguente tabella:

$X \setminus Y$	40 - 50	50 - 70	70 - 80
<i>recessione</i>	4	4	4
<i>espansione</i>	4	4	0

1. Fornire la rappresentazione grafica della distribuzione *marginale* di X e della *condizionata* $Y|X = recessione$.



2. Si calcoli un opportuno indice di posizione per la variabile X e la media geometrica di Y .

indice di posizione per X : $\text{Moda}(X) = x_{0.5} = \text{recessione}$

media geometrica di Y : formula $\mu^{(0)} = \prod_{j=1}^3 y_j^{f \cdot j} = 55.9188$

3. Si calcoli con un opportuno indice la mutabilità di X , commentandone il risultato.

indice di mutabilità per X : $F_N = 0.8$. Sussiste un livello elevato di eterogeneità.

4. Misurare la connessione tra le variabili X e Y e commentare il risultato.

$\chi_N = 0.40825$. Sussiste un livello di connessione medio/basso.

ESERCIZIO 2.

La seguente tabella riporta la serie dei NIBF di un titolo azionario quotato sul mercato finanziario:

t	$NIBF$	$NIBM$
2011	0.95	—
2012	1.05	1.1053
2013	1	0.9524
2014	0.95	0.9500
2015	0.9	0.9474

1. Si identifichi l'anno base e si interpreti il significato del numero indice a base fissa riferito al tempo 2015.

Tra l'anno base 2013 e l'anno 2015 si osserva una riduzione del 10% del prezzo del titolo.

Abbiamo infatti $100 \cdot (0.9 - 1)\% = -10\%$

2. Calcolare, mostrandone il procedimento, la serie dei NIBM.

3. Sapendo che il prezzo nel 2015 era 100, calcolare il prezzo del titolo nel 2011.

$x_{2011} = 105.5556$. Ricordando che la serie dei NIBF è proporzionale a quella dei dati abbiamo

$$x_{2015} : 0.9 = x_{2011} : 0.95 \quad \text{ovvero} \quad \frac{x_{2015}}{x_{2011}} = \frac{0.9}{0.95}$$

da cui si ottiene

$$x_{2011} = x_{2015} \cdot \frac{0.95}{0.9} = 105.5556.$$

4. Calcolare il tasso medio di variazione annuo nell'intervallo 2012-2015, commentandone il risultato.

$\sqrt[3]{2012 I_{2015}} = \sqrt[3]{\frac{0.9}{1.05}} = \sqrt[3]{0.8571} = 0.9499$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del -5.01% .

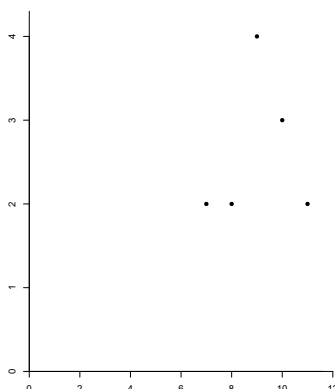
Il prezzo medio del prodotto B ha registrato un decremento medio annuo del 25.99% tra l'anno 2012 e l'anno 2015.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	7	8	9	10	11
y_i	2	2	4	3	2

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = \beta \cdot X$$

$\hat{a} = 1.7$, $\hat{b} = 0.1$, indice adattamento $M(\hat{E}^2) = 0.62$ $\hat{\beta} = 0.2843$, indice adattamento $M(\hat{E}^2) = 0.6896$

3. Si indichi, motivando, quale modello presenta il migliore adattamento.

Il modello migliore è $Y^* = 1.7 + 0.1 \cdot X$ in quanto caratterizzato da residuo quadratico medio (varianza residua) inferiore.

ESERCIZIO 4.

Nell'urna U_1 vengono inserite 2 palline blu e 2 palline rosse e si esegue la seguente procedura:

- si estrae una prima pallina;
- la pallina estratta viene reinserita nell'urna, insieme a 2 ulteriori palline del suo stesso colore;
- si estrae una seconda pallina dall'urna così modificata.

1. Calcolare la probabilità che la prima pallina estratta sia blu.

$$P(\text{prima pallina estratta blu}) = \frac{n. \text{ palline blu}}{n. \text{ palline}} = \frac{2}{4} = 0.5.$$

2. Usando la legge delle probabilità totali, calcolare la probabilità che la seconda pallina estratta sia blu.

Indicati con B_i e R_i gli eventi pallina blu alla i -esima estrazione e pallina rossa alla i -esima estrazione ($i = 1, 2$) abbiamo

$$P(B_2) = P(B_1) \cdot P(B_2|B_1) + P(R_1) \cdot P(B_2|R_1) = \frac{2}{4} \cdot \frac{4}{6} + \frac{2}{4} \cdot \frac{2}{6} = 0.3333 + 0.1667 = 0.5.$$

3. Gli eventi $B_1 = \{\text{prima pallina estratta è blu}\}$ e $B_2 = \{\text{seconda pallina estratta è blu}\}$ sono indipendenti?

Gli eventi B_1 e B_2 non sono indipendenti, abbiamo infatti

$$P(B_2) = 0.5 \neq P(B_2|B_1) = \frac{4}{6}$$

vale inoltre

$$P(B_1 \cap B_2) = 0.3333 \neq P(B_1) \cdot P(B_2) = 0.5 \cdot 0.5 = 0.25.$$

4. Trovare il percentile di ordine 0.025 di X Gaussiana con media 1 e varianza 4, cioè x_p tale che $P(X < x_p) = 0.025$.

$$\begin{aligned} P(X \leq x_p) &= 0.025 \\ P\left(\frac{X - \mu}{\sigma} \leq \frac{x_p - \mu}{\sigma}\right) &= 0.025 \\ P\left(Z \leq \frac{x_p - 1}{2}\right) &= P(Z \leq z_p) = 0.025 \\ \frac{x_p - 1}{2} &= z_p = -1.96 \\ x_p &= 1 - 1.96 \cdot 2 = -2.92. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 09.06.2016 (Tema 289-1)

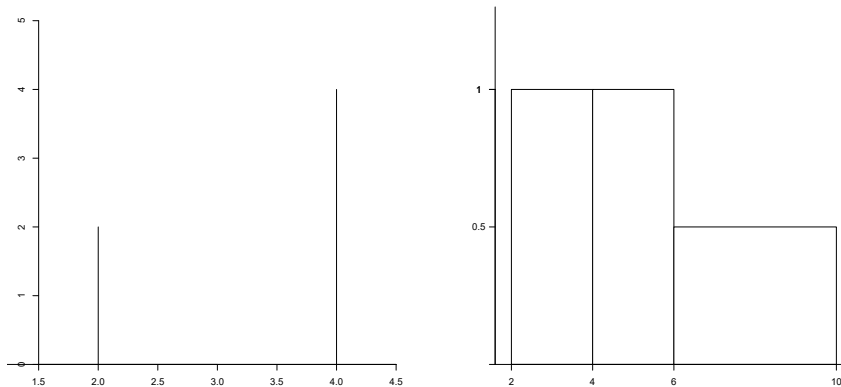
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La seguente tabella riporta alcune informazioni sulla distribuzione congiunta della variabile statistica doppia (X, Y) .

$X \setminus Y$	y_1	y_2	y_3	
2	0	2	0	2
4	2	0	2	4
	2	2	2	6

- Si completino le frequenze nella tabella in modo tale che siano verificate contemporaneamente le due seguenti condizioni:
a) la distribuzione $Y|x_1$ sia caratterizzata da assenza di variabilità; b) $n_{21} = n_{23}$; c) $\chi^2 = \chi_{\max}^2 = 6$.
- Si sostituiscano alle modalità y_1, y_2 e y_3 di Y le classi $2 \rightarrow 4$, $4 \rightarrow 6$ e $6 \rightarrow 10$, e si rappresentino graficamente le distribuzioni marginali X e Y .



- Si riassumano le distribuzioni marginali X e Y con opportuni indici di posizione e variabilità.

X Indice di posizione: $M(X) = 3.3333$ Indice di variabilità: $CV(X) = 0.2828$

Y Indice di posizione: $M(Y) = 5.3333$ Indice di variabilità: $CV(Y) = 0.3853$

La variabile Y è caratterizzata da un livello medio e da una variabilità superiori a X .

- Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

Sussiste dipendenza funzionale di X da Y , quindi $\chi_N = 1$.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	x_t	$NIBF(2010 = 1)$	$NIBM$
2008	200	4	—
2009	100	2	0.5
2010	50	1	0.5
2011	100	2	2
2012	200	4	2

- Si completi il prospetto con i dati mancanti.
- Si calcoli e si commenti il numero indice ${}_{2008}I_{2010}$.

${}_{2008}I_{2010} = \frac{1}{4} = \frac{50}{200} = 0.25$ Il valore della grandezza nel 2010 è inferiore del 75% a quello del 2008.

- Si calcoli il tasso medio di variazione dal 2008 al 2012 e si commenti il risultato ottenuto.

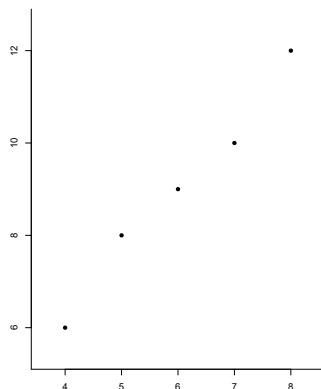
La grandezza esaminata non ha registrato variazione tra l'anno 2008 e l'anno 2012.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	4	5	6	7	8
y_i	6	8	9	10	12

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 0.6$, $\hat{b} = 1.4$, indice adattamento $\rho^2 = 0.98$ $\hat{c} = 9$, indice adattamento $R^2 = 0$

3. Si giustifichi se vale la relazione $\eta_{Y|X}^2 < \eta_{X|Y}^2$.

Nel caso in esame sussiste dipendenza funzionale bijectiva tra X e Y , quindi vale $\eta_{Y|X}^2 = \eta_{X|Y}^2 = 1$.

ESERCIZIO 4.

Nell'urna U_1 vengono inserite 2 palline blu e 2 palline rosse e si esegue la seguente procedura:

- si estrae una prima pallina;
- la pallina estratta viene reinserita nell'urna, insieme a 4 ulteriori palline: 2 del suo stesso colore e 2 dell'altro colore;
- si estrae una seconda pallina dall'urna così modificata.

1. Calcolare la probabilità che la prima pallina estratta sia blu.

$$P(\text{prima pallina estratta blu}) = \frac{n. \text{ palline blu}}{n. \text{ palline}} = \frac{2}{4} = 0.5.$$

2. Usando la legge delle probabilità totali, calcolare la probabilità che la seconda pallina estratta sia blu.

Indicati con B_i e R_i gli eventi pallina blu alla i -esima estrazione e pallina rossa alla i -esima estrazione ($i = 1, 2$) abbiamo

$$P(B_2) = P(B_1) \cdot P(B_2|B_1) + P(R_1) \cdot P(B_2|R_1) = \frac{2}{4} \cdot \frac{4}{8} + \frac{2}{4} \cdot \frac{4}{8} = 0.25 + 0.25 = 0.5.$$

3. Gli eventi $A_1 = \{\text{prima pallina estratta è blu}\}$ e $A_2 = \{\text{seconda pallina estratta è blu}\}$ sono indipendenti?

Gli eventi B_1 e B_2 sono indipendenti, abbiamo infatti

$$P(B_2) = 0.5 = P(B_2|B_1) = \frac{4}{8}$$

vale inoltre

$$P(B_1 \cap B_2) = 0.25 = P(B_1) \cdot P(B_2) = 0.5 \cdot 0.5.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 23.06.2016 (Tema 289)

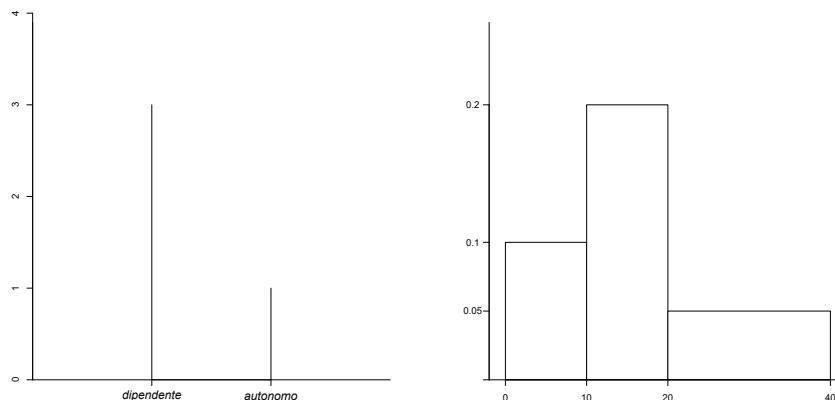
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

Data la seguente tabella a doppia entrata della distribuzione congiunta dei caratteri X = 'tipo di contratto di lavoro' e Y = 'salario lordo orario (in euro)' in una società industriale:

$X \setminus Y$	0 - 10	10 - 20	20 - 40
<i>dipendente</i>	1	2	0
<i>autonomo</i>	0	0	1

1. Fornire la rappresentazione grafica delle distribuzioni marginali.



2. Confrontare con un opportuno indice la variabilità delle distribuzioni condizionate ($Y|X$) del salario rispetto al tipo di contratto. Commentare i risultati.

Indice per confronto variabilità: Coefficiente di variazione

$$CV(Y|x_1) = 0.4041 \quad CV(Y|x_2) = 0$$

La distribuzione del salario dei lavoratori dipendenti presenta variabilità superiore a quella dei lavoratori autonomi

3. Calcolare moda, media e mediana per la marginale Y .

$$\text{Moda}(Y) = 10 - 20, \quad M(Y) = 16.25, \quad y_{0.5} = 15$$

4. Misurare la dipendenza tra le due variabili con un opportuno indice di connessione.

Sussiste dipendenza funzionale di X da Y , quindi $\chi_N = 1$.

ESERCIZIO 2.

La seguente tabella riporta la serie dei NIBM di un titolo azionario quotato in borsa:

t	$NIBM$	$NIBF(2010)$
2010	-	1
2011	1.25	1.25
2012	0.64	0.8
2013	1.25	1
2014	0.75	0.75

1. Si interpreti il significato di ${}_{2010}I_{2011}$, numero indice riferito al tempo 2011 con base 2010.

${}_{2010}I_{2011} = 1.25$. Tra il 2010 e il 2011 si è registrato un incremento del 25% del prezzo del titolo azionario.

2. Calcolare, mostrandone il procedimento, la serie dei NIBF con base 2010.

3. Sapendo che il prezzo del titolo era pari a 200 nel 2013, calcolare il prezzo nel 2011.

$$x_{2011} = 250 \quad \text{Infatti } {}_{2011}I_{2013} = {}_{2011}I_{2012} \cdot {}_{2012}I_{2013} = 0.64 \cdot 1.25 = 0.80 = \frac{x_{2013}}{x_{2011}} = \frac{200}{x_{2011}}, \quad \text{quindi } x_{2011} = \frac{200}{0.80} = 250$$

4. Calcolare il tasso medio di variazione annuo nell'intervallo 2011-2014, commentandone il risultato.

$$\sqrt[3]{{}_{2011}I_{2014}} = \sqrt[3]{{}_{2011}I_{2012} \cdot {}_{2012}I_{2013} \cdot {}_{2013}I_{2014}} = \sqrt[3]{0.64 \cdot 1.25 \cdot 1.75} = \sqrt[3]{0.60} = 0.8434, \quad \text{numero indice a base mobile medio, a cui corrisponde una variazione relativa media del } -15.66\%.$$

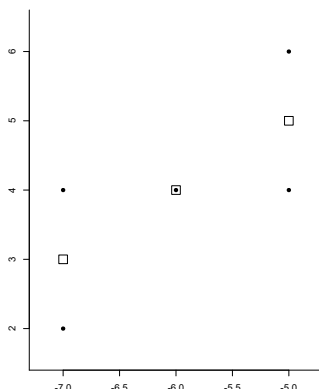
Il prezzo hanno subito un decremento medio annuo del 15.66% tra il 2011 e il 2014.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolte le seguenti 5 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 5$

x_i	-7	-7	-6	-5	-5
y_i	2	4	4	4	6

1. Fornire la rappresentazione grafica delle coppie di punti, insieme alla funzione di regressione di Y dato X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot (X - 3)$$

$\hat{a} = 10$, $\hat{b} = 1$, indice adattamento $M(\hat{E}^2) = 0.8$ $\hat{c} = -0.4303$, indice adattamento $M(\hat{E}^2) = 2.4528$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello retta presenta sicuramente un adattamento superiore al secondo modello in quanto coincide con il modello funzione di regressione, essendo le medie condizionate allineate.

4. Senza svolgere i conti, cosa è possibile dire sul valore di $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = \rho^2 = 0.5$ in quanto le medie condizionate sono allineate.

5. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 10$.

$$\hat{Y}(10) = \hat{a} + \hat{b} \cdot 10 = 10 + 1 \cdot 10 = 20.$$

ESERCIZIO 4.

La durata (in mesi) delle batterie prodotte da un'azienda segue una distribuzione normale di media 21 e scarto quadratico medio 9. Sono considerate difettose le batterie che durano meno di 17 mesi.

1. Qual è la probabilità che una batteria estratta a caso dal flusso produttivo sia difettosa?

Indicata con X la variabile casuale 'durata delle batterie' abbiamo $X \sim N(\mu = 21, \sigma^2 = 81)$.

$$\begin{aligned} P(X \leq 17) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{17 - \mu}{\sigma}\right) \\ &= P\left(Z \leq \frac{17 - 21}{9}\right) \\ &= P(Z \leq -0.44) = P(Z \geq 0.44) \\ &= 1 - P(Z \leq 0.44) = 1 - 0.67 = 0.33. \end{aligned}$$

2. Estraeendo con reimmissione un campione di 10 batterie, quale è la probabilità che meno di due siano difettose? Sia X la variabile casuale 'n. di difettose nelle 10 prove indipendenti'.

$$X \sim \text{Bin}(n = 10, p = 0.33).$$

$$P(X < 2) = P(X = 0) + P(X = 1) = \binom{10}{0} 0.33^0 0.67^{10} + \binom{10}{1} 0.33^1 0.67^9 = 0.01823 + 0.08978 = 0.10801.$$

3. Considerato un campione di 4400 batterie estratte con reimmissione, indicare qual è la probabilità che non più di 400 siano difettose.

Indicata con X la variabile casuale 'numero di difettose' nelle 4400 prove indipendenti risulta $X \sim \text{Bin}(n = 4400, p = 0.33)$. $M(X) = 1452$; $\text{Var}(X) = 972.84$. Dal momento che $np = 1452 > 5$ ed $n(1 - p) = 2948 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 1452, \sigma^2 = 972.84)$.

Si ottiene $P(X \leq 400) = P\left(\frac{X - \mu}{\sigma} \leq \frac{400 - \mu}{\sigma}\right) = P\left(Z \leq \frac{400 - 1452}{31.1904}\right) = P(Z \leq -33.73) = 0$.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 14.07.2016 (Tema 289-2)

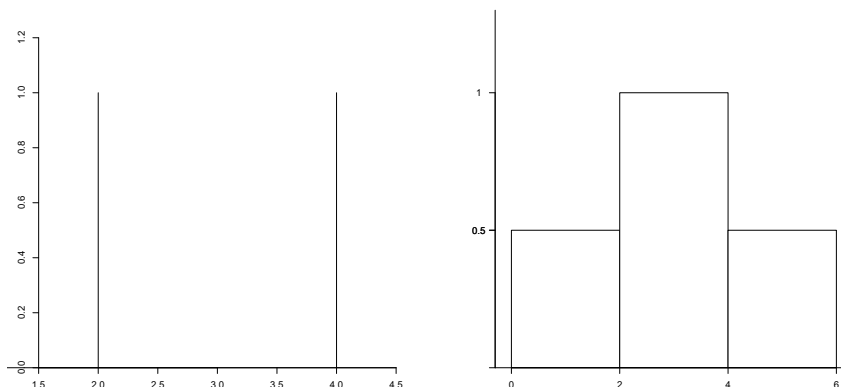
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La seguente tabella riporta la distribuzione congiunta dei caratteri quantitativi X e Y .

$X \setminus Y$	0 + 2	2 + 4	4 + 6
2	1	2	1
4	1	0	1

2. Si rappresentino graficamente le distribuzioni condizionate $X|y_1$ e $Y|x_1$.



3. Si riassumano le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 3$ Indice di variabilità: $CV(Y|x_1) = 0.4714$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 3$ Indice di variabilità: $CV(Y|x_2) = 0.6667$

Le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono caratterizzate dallo stesso livello medio; $Y|x_2$ presenta variabilità superiore a $Y|x_1$.

3. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

$\chi_N = 0.5$. Sussiste un livello di connessione medio.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	x_t	$NIBF(2010 = 1)$	$NIBM$
2008	200	4	—
2009	100	2	0.5
2010	50	1	0.5
2011	100	2	2
2012	200	4	2

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli e si commenti il numero indice ${}_{2012}I_{2010}$.

${}_{2012}I_{2010} = \frac{50}{200} = 0.25$ Il valore della grandezza nel 2010 era inferiore del 75% a quello del 2012.

3. Si calcoli il tasso medio di variazione dal 2008 al 2012 e si commenti il risultato ottenuto.

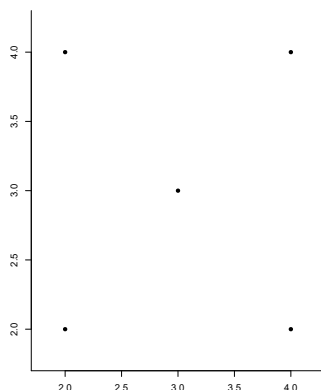
La grandezza esaminata non ha registrato variazione tra l'anno 2008 e l'anno 2012.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	2	3	4	4
y_i	2	4	3	2	4

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$\hat{a} = 3$, $\hat{b} = 0$, indice adattamento $M(\hat{E}^2) = 0.8$ $\hat{c} = 0.9184$, indice adattamento $M(\hat{E}^2) = 1.5347$

3. Si giustifichi se vale la relazione $\eta_{Y|X}^2 \geq \eta_{X|Y}^2$.

Dalla rappresentazione grafica si evince che nel caso in esame sussistono contemporaneamente indipendenza in media di Y da X e di X da Y ; quindi vale $\eta_{Y|X}^2 = \eta_{X|Y}^2 = 0$.

ESERCIZIO 4.

In un gioco si possono vincere, con un biglietto del costo di 1 Euro, fino a due premi: il premio a di 1 Euro con probabilità 0.11 e il premio b di 2 Euro con probabilità 0.15. La probabilità complessiva di vincere o uno o l'altro dei due premi è pari a 0.20.

1. Si calcoli la probabilità di vincere entrambi i premi.

$$P(a \cap b) = 0.06$$

Vale infatti

$$\begin{aligned} P(a \cup b) &= P(a) + P(b) - P(a \cap b) \\ 0.20 &= 0.11 + 0.15 - P(a \cap b) \\ P(a \cap b) &= 0.11 + 0.15 - 0.20 = 0.06 \end{aligned}$$

1. Si stabilisca se il gioco è equo.

La variabile casuale guadagno risulta

x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
-1	$1 - P(a \cup b) = 1 - .20 = 0.80$	-0.80
0	$P(a - b) = P(a) - P(a \cap b) = 0.11 - 0.06 = 0.05$	0.00
1	$P(b - a) = P(b) - P(a \cap b) = 0.15 - 0.06 = 0.09$	0.09
2	$P(a \cap b) = 0.06$	0.12
	1	-0.59

Il valore atteso della variabile casuale guadagno è negativo, quindi il gioco è sfavorevole al giocatore.

2. Nell'ipotesi di ripetere il gioco 3 volte si calcoli la probabilità di vincere il premio a oppure il premio b almeno 2 volte. Sia X la variabile casuale 'n. di vittorie nelle 3 prove indipendenti'.

$$X \sim \text{Bin}(n = 3, p = 0.20).$$

$$P(X \geq 2) = P(X = 2) + P(X = 3) = \binom{3}{2} 0.20^2 0.80^1 + \binom{3}{3} 0.20^3 0.80^0 = 0.096 + 0.008 = 0.104.$$

3. Nell'ipotesi di ripetere il gioco 100 volte si calcoli la probabilità di vincere il premio a oppure il premio b almeno 12 volte.

Indicata con X la variabile casuale 'numero di successi' nelle 100 prove indipendenti risulta $X \sim \text{Bin}(n = 100, p = 0.2)$. $M(X) = 20$; $\text{Var}(X) = 16$. Dal momento che $np = 20 > 5$ ed $n(1 - p) = 80 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 20, \sigma^2 = 16)$.

Si ottiene $P(X \geq 12) = P\left(\frac{X - \mu}{\sigma} \geq \frac{12 - \mu}{\sigma}\right) = P\left(Z \geq \frac{12 - 20}{4}\right) = P(Z \geq -2) = P(Z \leq 2) = 0.9772$.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 23.08.2016 (Tema 290-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La seguente tabella riporta alcune informazioni sulla distribuzione congiunta della variabile statistica doppia (X, Y) , con X carattere qualitativo sconnesso.

$X \setminus Y$	y_1	y_2	y_3	y_4	
x_1	2	2	2	2	8
x_2	1	1	1	1	4
	3	3	3	3	12

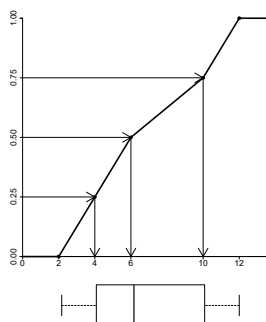
1. Si completino le frequenze nella tabella in modo tale che siano verificate contemporaneamente le tre seguenti condizioni:
a) la distribuzione $Y|x_1$ sia caratterizzata da massima eterogeneità; b) $\chi^2_{\max} = 12$; c) sussista indipendenza stocastica tra le variabili X e Y .

2. Si sostituiscano ora alle modalità y_1, y_2, y_3 , e y_4 di Y le classi $2 \rightarrow 4$, $4 \rightarrow 6$, $6 \rightarrow 10$ e $10 \rightarrow 12$ e le frequenze congiunte $n_{11} = n_{13} = n_{14} = 2$ e $n_{21} = n_{22} = n_{23} = n_{24} = 11$. Si rappresentino graficamente le funzioni di ripartizione delle distribuzioni marginali X e Y .

$X \setminus Y$	y_1	y_2	y_3	y_4	
x_1	2	2	2	2	8
x_2	11	11	11	11	44
	13	13	13	13	52

Dal momento che il carattere qualitativo X è sconnesso, le frequenze cumulate e la funzione di ripartizione non sono definite. Si veda il punto successivo per la funzione di ripartizione di Y .

3. Si costruisca il grafico Box & Whiskers plot della marginale Y .



$x_{0.25} = 4, x_{0.50} = 6, x_{0.75} = 10$, lung. teorica baffo = $1.5 \cdot (10 - 4) = 9$, baffo inferiore = 2, baffo superiore = 12

4. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

Sussiste indipendenza stocastica, quindi $\chi^2 = \chi^2_N = \chi_N = 0$

ESERCIZIO 2.

Si consideri la seguente serie storica:

t	x_t	$NIBF$	$NIBM$
2011	300	2	
2012	150	1	0.5
2013	75	0.5	0.5
2014	150	1	2
2015	300	2	2

1. Si ricostruiscano le serie dei numeri indici a base fissa, con base 2012, e a base mobile.

2. Si calcolino e si commentino i seguenti numeri indici ${}_{2011}I_{2013}$ e ${}_{2013}I_{2015}$.

${}_{2011}I_{2013} = \frac{75}{300} = 0.25$. Tra il 2011 e il 2013 si è verificato un decremento del 75% della grandezza oggetto di studio.

${}_{2013}I_{2015} = \frac{300}{75} = 4$. Tra il 2013 e il 2015 si è verificato un incremento del 300% della grandezza oggetto di studio.

3. Si calcoli il tasso medio di variazione dal 2012 al 2015 e si commenti il risultato ottenuto.

$\sqrt[3]{{}_{2012}I_{2015}} = \sqrt[3]{\frac{300}{150}} = \sqrt[3]{2} = 1.2599$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media del 25.99%.

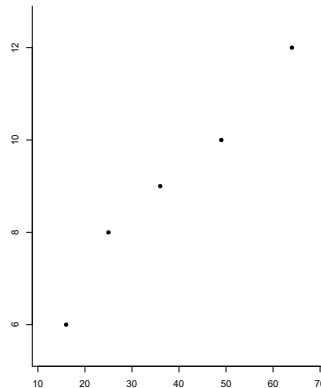
Le vendite hanno subito un incremento medio annuo del 25.99% tra il 2012 e il 2015.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	16	25	36	49	64
y_i	6	8	9	10	12

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot \sqrt{X} \quad II) Y^* = c$$

$\hat{a} = 0.6$, $\hat{b} = 1.4$, indice adattamento $\rho^2 = 0.98$ $\hat{c} = 9$, indice adattamento $R^2 = 0$

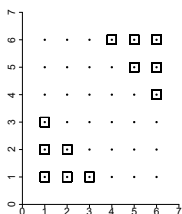
3. Si giustifichi se vale la relazione $\eta_{Y|X}^2 > \eta_{X|Y}^2$.

La relazione non vale; infatti, $\eta_{Y|X}^2 = 1$ ed $\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale bijectiva.

ESERCIZIO 4.

Per il seguente gioco è prevista una posta di 1 Euro: si lanciano due dadi; se la somma dei punteggi ottenuti con i due dadi è inferiore a 5 si vincono 10 Euro, se invece è superiore a 9 si vincono 15 Euro.

1. Si stabilisca se il gioco è equo.



$$P(\text{vincere 10 Euro}) = \frac{6}{36}$$

$$P(\text{vincere 15 Euro}) = \frac{6}{36}$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
posta	-1	$\frac{24}{36}$	$-\frac{24}{36}$
vincita se somma < 5	9	$\frac{6}{36}$	$\frac{54}{36}$
vincita se somma > 9	14	$\frac{6}{36}$	$\frac{84}{36}$
		1	$\frac{114}{36}$

Il valore atteso della variabile casuale guadagno è positivo, quindi il gioco non è equo, ma favorevole al giocatore.

2. Nell'ipotesi di ripetere il gioco 10 volte si calcoli la probabilità di vincere almeno 2 volte.

Sia X la variabile casuale 'n. di vincite nelle 10 prove indipendenti'. Abbiamo $X \sim \text{Bin}(n = 10, p = \frac{12}{36} = 0.3333)$.

$$\begin{aligned}
 P(X \geq 2) &= 1 - P(X < 2) = 1 - \{P(X = 0) + P(X = 1)\} = \\
 &= 1 - \left\{ \binom{10}{0} 0.3333^0 \cdot 0.6667^{10} + \binom{10}{1} 0.3333^1 \cdot 0.6667^9 \right\} = \\
 &= 1 - (0.0173 + 0.0867) = 1 - 0.1040 = 0.8960.
 \end{aligned}$$

3. Nell'ipotesi di ripetere il gioco 100 volte si calcoli la probabilità di vincere almeno 40 volte.

Indicata con X la variabile casuale 'numero di successi' nelle 100 prove indipendenti risulta $X \sim \text{Bin}(n = 100, p = 0.3333)$. $M(X) = 33.33$; $\text{Var}(X) = 22.22$. Dal momento che $np = 33.33 > 5$ ed $n(1 - p) = 66.67 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 33.33, \sigma^2 = 22.22)$.

Si ottiene

$$\begin{aligned}
 P(X \geq 40) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{40 - \mu}{\sigma}\right) = P\left(Z \geq \frac{40 - 33.33}{4.7148}\right) = \\
 &= P(Z \geq 1.41) = 1 - P(Z \leq 1.41) = 1 - 0.9207 = 0.0793.
 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano
FACOLTÀ DI ECONOMIA

Prova scritta di STATISTICA del 08.09.2016 (Tema 290)

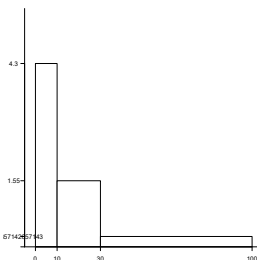
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La tabella seguente raccoglie i dati relativi a 91 nazioni, sul reddito pro-capite X (in migliaia di dollari) e tasso di mortalità infantile Y (n. decessi ogni 1000 nati vivi):

$X \setminus Y$	0 - 10	10 - 40	40 - 100
0 - 10	1	16	26
10 - 30	10	18	3
30 - 100	15	2	0

1. Rappresentare graficamente la marginale X .



2. Confrontare la variabilità delle marginali X e Y .

Indice per confronto variabilità: coefficiente di variazione

valore X 1.0308 valore Y 0.7784

Il reddito pro-capite presenta un livello di variabilità superiore a quello del tasso di mortalità infantile.

3. Valutare la connessione tra X e Y con un opportuno indice.

$\chi_N = 0.5694$. Sussiste un livello di connessione medio.

ESERCIZIO 2.

A partire dalla serie storica delle vendite, in n. pezzi, $Y(i)$, $i = 1, 2, \dots, 5$, di un certo prodotto in 5 successivi anni, si è costruita la seguente tabella, dove $\Delta_i = Y(i) - Y(i-1)$, mentre ${}_1I_i$ sono i numeri indici con base $i = 1$:

i	Δ_i	${}_1I_i$	$NIBM$
1	-	1	-
2	697	1.043	1.043
3	696	1.086	1.0412
4	211	1.099	1.0120
5	600	1.136	1.0337

1. Fornire il significato di ${}_1I_2$, numero indice riferito al tempo 2 con base 1.

${}_1I_2 = 1.043$ Tra il tempo 1 e il tempo 2 si è registrato un incremento del 4.3% corrispondente a una variazione assoluta di 697.

2. Calcolare, mostrandone il procedimento, la serie dei NIBM.

3. Calcolare il valore di $Y(1)$.

$Y(1) = 16209.3023$ Vale infatti: $0.043 = \frac{Y_2}{Y_1} - 1 = \frac{Y_2 - Y_1}{Y_1} = \frac{697}{Y_1}$; da cui segue $Y_1 = \frac{697}{0.043}$.

4. Calcolare il tasso medio di variazione annuo tra il secondo ($i = 2$) e il quinto anno ($i = 5$), interpretando il risultato.

$\sqrt[3]{{}_2I_5} = \sqrt[3]{\frac{1.136}{1.043}} = \sqrt[3]{1.0892} = 1.0289$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del 2.89%.

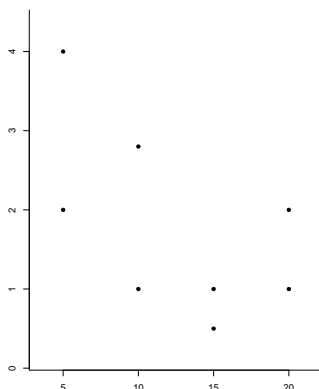
La grandezza esaminata ha registrato un incremento medio annuo del 2.89% tra il tempo 2 e il tempo 5.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolte le seguenti 8 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 8$

x_i	5	5	10	10.0	15	15.0	20	20
y_i	2	4	1	2.8	1	0.5	1	2

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d/X$$

$\hat{a} = 3.2$, $\hat{b} = -0.1130$, indice adattamento $\rho^2 = 0.3350$ $\hat{c} = 0.4782$, $\hat{d} = 12.5692$, indice adattamento $R^2 = 0.4490$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 0.4782 + 12.5692/X$.

4. In base al modello migliore fornire il valore teorico di Y per $X = 30$.

$$Y^*(30) = 0.4782 + 12.5692/30 = 0.8972$$

ESERCIZIO 4.

Si lanciano 2 dadi equilibrati, il primo regolare (con simboli 1, 2, 3, 4, 5, 6) e il secondo con simboli di valore 1, 2, 0, 0, 0, 0.

1. Descrivere (specificando valori e probabilità) le v.c. $X_1 =$ 'risultato del dado 1' e $X_2 =$ 'risultato del dado 2'.

x_{1i}	$P(X_1 = x_{1i})$	x_{2i}	$P(X_2 = x_{2i})$
1	1/6	0	4/6
2	1/6	1	1/6
3	1/6	2	1/6
4	1/6		
5	1/6		
6	1/6		
	1		1

2. Calcolare $M(X_1 + X_2)$.

$$M(X_1 + X_2) = M(X_1) + M(X_2) = 3.5 + 0.5 = 4$$

3. Calcolare la probabilità che, su 6 lanci, il secondo dado presenti 3 volte valore 0.

Sia X la variabile casuale 'n. di volte che si presenta 0 nelle 6 prove indipendenti'. Abbiamo $X \sim \text{Bin}(n = 6, p = \frac{4}{6} = 0.6667)$.

$$P(X = 3) = \binom{6}{3} 0.6667^3 \cdot 0.3333^3 = 0.2195.$$

4. Calcolare la probabilità che su $n = 200$ lanci l'evento 'secondo dado di valore 0' si presenti non più di 127 volte.

Indicata con X la variabile casuale 'numero di volte che si presenta 0' nelle 200 prove indipendenti risulta $X \sim \text{Bin}(n = 200, p = 0.6667)$. $M(X) = 133.33$; $\text{Var}(X) = 44.44$. Dal momento che $np = 133.33 > 5$ ed $n(1-p) = 66.67 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 133.33, \sigma^2 = 44.44)$.

Si ottiene

$$\begin{aligned} P(X \leq 127) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{127 - \mu}{\sigma}\right) = P\left(Z \leq \frac{127 - 133.33}{6.6667}\right) = \\ &= P(Z \leq -0.95) = P(Z \geq 0.95) = 1 - P(Z < 0.95) = 1 - P(Z \leq 0.95) = 1 - 0.8289 = 0.1711. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 20.01.2017 (Tema 291)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

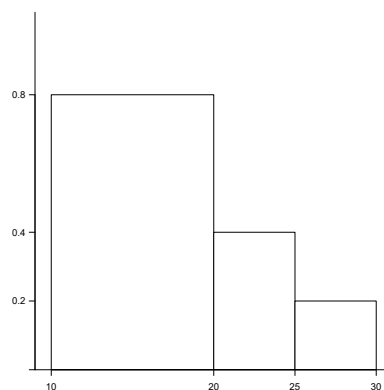
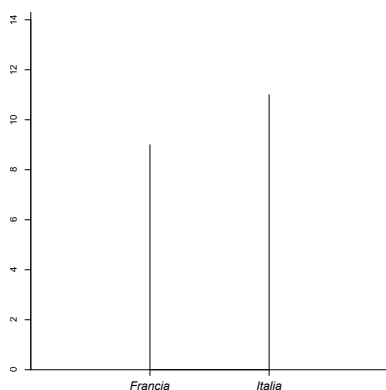
Studente _____ matricola _____

ESERCIZIO 1.

Al fine di analizzare la relazione tra la variabile X = 'Nazione' e Y = 'Salario annuo (in migliaia di euro)' rilevate su 20 unità statistiche, si sono raccolti i seguenti dati:

$X \setminus Y$	10 - 20	20 - 25	25 - 30
Francia	2	7	0
Italia	8	2	1

1. Si rappresentino graficamente la distribuzione *marginale* di X e la *condizionata* di $Y|X = Italia$.



2. Si calcoli un opportuno indice di posizione per la variabile X . Si calcolino inoltre la media aritmetica e la mediana di $Y|X = Francia$.

indice posizione X : $\text{Moda}(X) = Italia$

$$M(Y|X = Francia) = 20.8333$$

$$y_{0.5}|X = Francia = 21.7857$$

3. Si calcoli l'indice di mutabilità normalizzato di X , commentandone il risultato.

Indice $F_N = 0.9$. Sussiste un livello elevato di eterogeneità.

4. Misurare e commentare la connessione tra le variabili X e Y .

$\chi_N = 0.6021$. Sussiste un livello di connessione medio.

ESERCIZIO 2. Si consideri il seguente prospetto:

t	x_t	$NIBF(t = 2010)$	$NIBM$
2008	2400	16	
2009	600	4	0.25
2010	150	1	0.25
2011	150	1	1
2012	75	0.5	0.5

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli e si commenti ${}_{2008}I_{2010}$, numero indice riferito al tempo 2010 con base 2008.

${}_{2008}I_{2010} = 0.0625$. Tra il 2008 e il 2010 si è registrata una riduzione del 93.75%.

3. Si calcoli il tasso medio di variazione dal 2008 al 2010 e si commenti il risultato ottenuto.

$\sqrt{{}_{2008}I_{2010}} = \sqrt{0.0625} = 0.25$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del -75%.

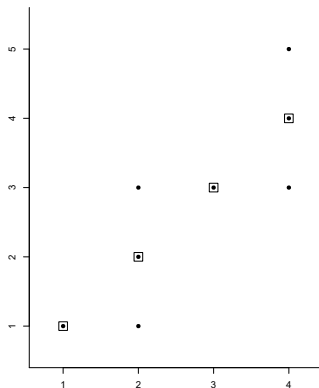
La grandezza esaminata ha registrato un decremento medio annuo del 75% tra l'anno 2008 e l'anno 2010.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 capofamiglia con riferimento a due caratteri quantitativi X = numero dei componenti della famiglia e Y = spesa per un vacanza di due giorni (in centinaia di euro).

x_i	1	2	2	2	3	4	4	4
y_i	1	1	2	3	3	3	4	5

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando la spesa in funzione del numero dei componenti della famiglia.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 0, \hat{b} = 1, \text{ indice adattamento } M(\hat{E}^2) = 0.5 \quad \hat{c} = 1, \text{ indice adattamento } M(\hat{E}^2) = 0.5$$

3. Confrontare gli indici di adattamento, commentando i risultati.
I due modelli sono equivalenti.

4. Si determini il valore di $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = \rho^2 = 0.7037$ in quanto le medie condizionate sono allineate.

5. Si dica, senza effettuare calcoli e motivando la risposta, se $\eta_{X|Y}^2 = 1$.

$\eta_{X|Y}^2 < 1$ in quanto le distribuzioni $X|y_j$ non sono degeneri.

ESERCIZIO 4.

Una indagine di una compagnia telefonica ha stabilito che la durata (in secondi) delle chiamate dei propri utenti è distribuita come una Normale di media μ e varianza σ^2 .

1. Sapendo che il 50% degli utenti effettua chiamate inferiori a 190 secondi, stabilire il valore della media μ .

$\mu = 190$ in quanto la funzione di densità della variabile casuale Normale è simmetrica, quindi la media aritmetica coincide con la mediana 190.

2. Supponendo che $\mu = 290$ e $\sigma = 80$, si calcoli la probabilità che una telefonata duri più di 370 secondi.
Sia X la variabile casuale 'durata di una telefonata'.

$$X \sim N(\mu = 290, \sigma^2 = 80^2 = 6400).$$

$$P(X > 370) = P\left(\frac{X-\mu}{\sigma} > \frac{370-\mu}{\sigma}\right) = P\left(Z > \frac{370-290}{80}\right) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587.$$

3. Si calcoli la probabilità che su 10 telefonate almeno una duri più di 370 secondi.

Sia Y la variabile casuale 'n. di telefonate con durata superiore a 370 secondi nelle 10 prove indipendenti'.

$$Y \sim \text{Bin}(n = 10, p = 0.1587).$$

$$P(Y \geq 1) = 1 - P(Y = 0) = 1 - \binom{10}{0} 0.1587^0 0.8413^{10} = 1 - 0.8413^{10} = 1 - 0.1776 = 0.8224.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 06.02.2017 (Tema 292)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

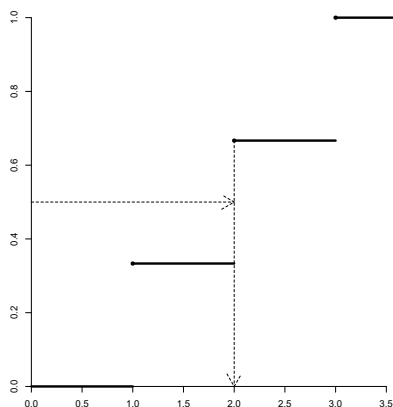
Il prospetto seguente riporta le distribuzioni relative del carattere quantitativo Y condizionate rispetto alle categorie del carattere qualitativo X .

$Y x_i$	1	2	3	
x_1	1	0	0	1
x_2	0	0.5	0.5	1

1. Sapendo che le frequenze marginali di X sono $n_{1\bullet} = 10$ e $n_{2\bullet} = 20$ si ricostruiscano le frequenze congiunte della variabile doppia (X, Y) .

$X \setminus Y$	1	2	3	$n_{i\bullet}$
x_1	10	0	0	10
x_2	0	10	10	20
$n_{\bullet j}$	10	10	10	30

2. Si rappresentino graficamente le frequenze cumulate della marginale Y e se ne determini il valore della mediana.



$y_{0.5} = \frac{1}{2} (y_{(15)} + y_{(16)}) = 2$. Si osserva che la distribuzione è simmetrica.

3. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ mediante opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 1$ Indice di variabilità: $CV(Y|x_1) = 0$.

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 2.5$ Indice di variabilità: $CV(Y|x_2) = 0.2$.

4. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

$\chi_N = 1$ dal momento che sussiste dipendenza funzionale di X da Y (le distribuzioni condizionate $X|y_j$ sono degeneri).

ESERCIZIO 2.

La seguente tabella riporta la serie dei NIBF del fatturato annuo di un'ipotetica azienda:

t	NIBF	NIBM
2012	1	
2013	1.1	1.1
2014	1.045	0.95
2015	1.254	1.2
2016	1.254	1

1. Calcolare, mostrandone il procedimento, la serie dei NIBM e interpretare il valore di ${}_{2012}I_{2013}$, numero indice al tempo 2013 con base 2012.

${}_{2012}I_{2013} = 1.1$. Tra il 2012 e il 2013 il fatturato ha registrato un incremento del 10%.

2. Sapendo che il fatturato nel 2016 era 100000 Euro, calcolare il fatturato nel 2013.

$x_{2013} = 87719.30$. Da ${}_{2012}I_{2013} = \frac{x_{2013}}{x_{2012}} = 1.1$ e ${}_{2012}I_{2016} = \frac{x_{2016}}{x_{2012}} = 1.254$ si ottiene $x_{2013} = x_{2016} \cdot \frac{{}_{2012}I_{2013}}{{}_{2012}I_{2016}} = 100000 \cdot \frac{1.1}{1.254}$.

3. Si calcoli il tasso medio di variazione dal 2013 al 2016 e si commenti il risultato ottenuto.

$\sqrt[3]{{}_{2013}I_{2016}} = \sqrt[3]{1.14} = 1.0446$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del 4.46%.

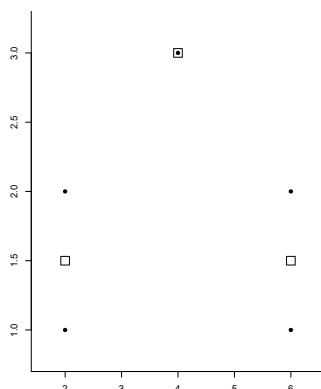
La grandezza esaminata ha registrato un incremento medio annuo del 4.46% tra l'anno 2013 e l'anno 2016.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	2	4	6	6
y_i	1	2	3	1	2

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$$\hat{a} = 1.8, \hat{b} = 0, \text{ indice adattamento } M(\hat{E}^2) = 0.56 \quad \hat{c} = 1.8, \text{ indice adattamento } M(\hat{E}^2) = 0.56$$

3. Confrontare gli indici di adattamento, commentando i risultati.

I due modelli sono equivalenti.

4. Si motivi perché $\eta_{X|Y}^2 < \eta_{Y|X}^2$.

$\eta_{X|Y}^2 < \eta_{Y|X}^2$ perché $0 = \eta_{X|Y}^2 < \eta_{Y|X}^2 < 1$.
Abbiamo, infatti,

$$M(X|y=1) = M(X|y=2) = M(X|y=3) = 4,$$

mentre vale

$$M(Y|x=2) = M(Y|x=6) = 1.5 \neq M(Y|x=4) = 3.$$

ESERCIZIO 4.

L'urna $U1$ contiene 3 palline nere e 1 pallina rossa. L'urna $U2$ contiene 9 palline nere e 10 palline rosse.

1. Si calcoli la probabilità che estraendo una pallina dall'urna $U1$ e una pallina dall'urna $U2$ solo una delle due palline sia nera (si ottengano una pallina nera e una pallina rossa).

$$P(1 \text{ rossa e } 1 \text{ nera}) = 0.5132.$$

$$P(1 \text{ rossa e } 1 \text{ nera}) = P\{(N_1 \cap R_2) \cup (R_1 \cap N_2)\} = P(N_1 \cap R_2) + P(R_1 \cap N_2) = 0.3947 + 0.1184 = 0.5132.$$

2. Si calcoli la probabilità che estraendo senza reimmissione 4 palline dall'urna $U2$ ce ne siano almeno 3 nere.

$$P(\text{almeno } 3N) = \binom{4}{3} \frac{9}{19} \frac{8}{18} \frac{7}{17} \frac{10}{16} + \frac{9}{19} \frac{8}{18} \frac{7}{17} \frac{6}{16} = 0.2167 + 0.0325 = 0.2492$$

o mediante la v.c. ipergeometrica

$$P(\text{almeno } 3N) = \frac{\binom{9}{3} \binom{10}{1}}{\binom{19}{4}} + \frac{\binom{9}{4} \binom{10}{0}}{\binom{19}{4}} = 0.2167 + 0.0325 = 0.2492.$$

3. Si estraiga una pallina dall'urna $U1$ e la si metta nell'urna $U2$. Si calcoli la probabilità che estraendo con reimmissione 2 palline dall'urna $U2$ si ottengano 2 palline nere.

In base alla legge delle probabilità totali abbiamo

$$\begin{aligned} P(2N) &= P((R_1 \cap 2N) \cup (N_1 \cap 2N)) = P(R_1 \cap 2N) + P(N_1 \cap 2N) = P(R_1)P(2N|R_1) + P(N_1)P(2N|N_1) = \\ &= \frac{1}{4} \left(\frac{9}{20} \right)^2 + \frac{3}{4} \left(\frac{10}{20} \right)^2 = \\ &= 0.25 \cdot 0.2025 + 0.75 \cdot 0.25 = 0.050625 + 0.1875 = 0.238125 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 21.02.2017 (Tema 293)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

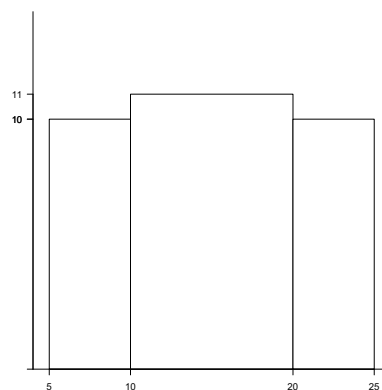
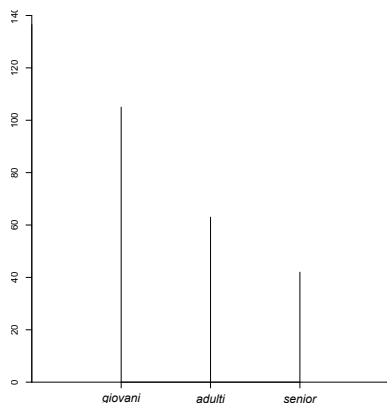
Studente _____ matricola _____

ESERCIZIO 1.

Una indagine su 210 cittadini riporta la categoria di età X e la spesa media mensile per l'attività sportiva Y . Le distribuzioni delle frequenze relative di X e delle densità di frequenza di Y sono le seguenti:

x_i	f_i	n_i	$h_{j-1} \rightarrow h_j$	d_j	n_j
<i>giovani</i>	0.50	105	5 → 10	10	50
<i>adulti</i>	0.30	63	10 → 20	11	110
<i>senior</i>	0.20	42	20 → 25	10	50

1. Si costruiscano le distribuzioni di frequenza assolute delle due variabili.
2. Si rappresentino graficamente le due variabili e si riassumano le stesse con opportuni indici di posizione e di variabilità.



X Indice di posizione: $\text{Moda}(X) = \text{giovani}$ Indice di variabilità: $F_N = 0.7354$.

Y Indice di posizione: $M(Y) = 15$ Indice di variabilità: $CV(Y) = 0.3450$.

3. Si calcoli l'indice di asimmetria della variabile Y .

La distribuzione è simmetrica, quindi gli indici di asimmetria assumono valore nullo

4. Si costruisca la tabella delle frequenze congiunte delle due variabili nel caso di indipendenza stocastica e si dica invece quanto varrebbe l'indice di connessione χ^2 nel caso di massima dipendenza.

$X \setminus Y$	5 → 10	10 → 20	20 → 25	$n_{i\bullet}$
<i>giovani</i>	25	55	25	105
<i>adulti</i>	15	33	15	63
<i>senior</i>	10	22	10	42
$n_{\bullet j}$	50	110	50	210

in caso di massima dipendenza $\chi^2 = n \cdot \min(h - 1, k - 1) = 420$

ESERCIZIO 2.

Con riferimento alla seguente serie di Numeri Indici a Base Mobile:

t	$NIBM$	x_t
2008		104.1190
2009	0.92	95.7895
2010	0.95	91.0000
2011	1.05	95.5500
2012	1.08	103.1940

1. Si ricostruisca, mostrandone il procedimento, la serie storica dei dati x_t sapendo che $x_{2010} = 91$.

2. Si calcoli il numero indice al tempo 2012 con base 2010 e si commenti il risultato ottenuto.

${}_{2010}I_{2012} = 1.134$. La grandezza in oggetto ha subito tra il 2010 e il 2012 un incremento del 13.4%.

3. Si calcoli il tasso medio di variazione nel periodo 2008-2010 interpretando il risultato ottenuto.

$\sqrt{{}_{2008}I_{2010}} = \sqrt{0.92 \cdot 0.95} = \sqrt{0.874} = 0.9349$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del -6.51%.

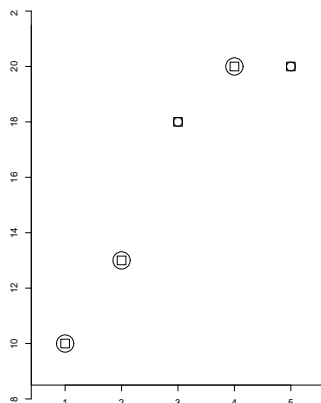
La grandezza esaminata ha registrato un decremento medio annuo del 6.51% tra l'anno 2008 e l'anno 2010.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	2	2	3	4	4	5
y_i	10	10	13	13	18	20	20	20

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = \exp(cX)$$

$$\hat{a} = 7.5161, \hat{b} = 2.9032, \text{ indice adattamento } M(\hat{E}^2) = 1.1694 \quad \hat{c} = 0.8221, \text{ indice adattamento } M(\hat{E}^2) = 256.5069$$

3. Si individui il modello migliore.

Il modello migliore è $Y^* = 7.5161 + 2.9032 \cdot X$ in quanto caratterizzato da residuo quadratico medio inferiore.

4. Si dica, motivando la risposta ma senza fare calcoli, se $\eta_{Y|X}^2 \neq \eta_{X|Y}^2$.

La relazione risulta verificata; abbiamo, infatti, $\eta_{Y|X}^2 = 1$ in quanto le distribuzioni $Y|x_i$ sono degeneri, mentre $0 < \eta_{X|Y}^2 < 1$ in quanto le distribuzioni condizionate $X|y_j$ non sono degeneri e le medie condizionate $M(Y|x_i)$ non sono tutte uguali.

ESERCIZIO 4.

Si considerano due urne contenenti palline numerate. L'urna U_1 contiene palline coi numeri 1, 2 e 3, in eguale numerosità, mentre l'urna U_2 è così composta: $\{4, 4, 4, 4, 5, 6, 7, 8, 8, 8\}$.

1. Si costruisca la distribuzione di probabilità delle variabili casuali X = 'risultato di una estrazione dall'urna U_1 ' e Y = 'risultato di una estrazione dall'urna U_2 '.

x	$P(X = x)$	y	$P(Y = y)$
1	1/3	4	0.4
2	1/3	5	0.1
3	1/3	6	0.1
		7	0.1
		8	0.3

2. Si estrae una pallina da ciascuna urna. Si calcoli la probabilità che la somma dei numeri estratti sia minore o uguale a 6.

$P(\text{somma} \leq 6) = 0.3$. Abbiamo

$$\begin{aligned} P(\text{somma} \leq 6) &= P((U_1 = 1) \cap (U_2 = 4) \cup (U_1 = 1) \cap (U_2 = 5) \cup (U_1 = 2) \cap (U_2 = 4)) = \\ &= P((U_1 = 1) \cap (U_2 = 4)) + P((U_1 = 1) \cap (U_2 = 5)) + P((U_1 = 2) \cap (U_2 = 4)) = \\ &= P(U_1 = 1) \cdot P(U_2 = 4) + P(U_1 = 1) \cdot P(U_2 = 5) + P(U_1 = 2) \cdot P(U_2 = 4) = \\ &= \frac{1}{3} \cdot 0.4 + \frac{1}{3} \cdot 0.1 + \frac{1}{3} \cdot 0.4 = \frac{1}{3} \cdot 0.9 = 0.3 \end{aligned}$$

3. Considerando solo l'urna U_2 , si calcoli la probabilità che su 200 estrazioni con reimmissione il numero 8 esca almeno 50 volte.

Indicata con X la variabile casuale 'numero di volte che si presenta il numero 8' nelle 200 estrazioni con reimmissione risulta $X \sim \text{Bin}(n = 200, p = 0.3)$. $M(X) = 60$; $\text{Var}(X) = 42$. Dal momento che $np = 60 > 5$ ed $n(1 - p) = 140 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 60, \sigma^2 = 42)$. Si ottiene

$$\begin{aligned} P(X \geq 50) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{50 - \mu}{\sigma}\right) = P\left(Z \geq \frac{50 - 60}{\sqrt{42}}\right) = \\ &= P(Z \geq -1.54) = P(Z \leq 1.54) = 0.9382. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 08.06.2017 (Tema 294-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

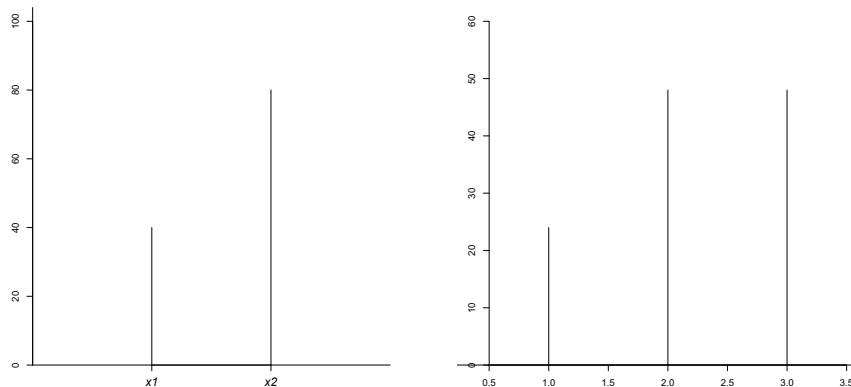
Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta la distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	1	2	3
x_1	8	16	16
x_2	16	32	32

1. Si rappresentino graficamente le distribuzioni di frequenza delle marginali X e Y .



2. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ mediante opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = M(Y|x_2) = 2.2$ Indice di variabilità: $CV(Y|x_1) = CV(Y|x_2) = 0.3401$

Le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono caratterizzate dallo stesso livello medio e dalla stessa variabilità in quanto sussiste indipendenza stocastica.

3. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

L'indice di connessione χ_N^2 assume valore nullo in quanto sussiste indipendenza stocastica.

ESERCIZIO 2.

La seguente tabella riporta la serie storica del fatturato annuo di un'azienda:

t	x_t	$NIBM$
2012	50	
2013	55.00	1.10
2014	52.25	0.95
2015	62.70	1.20
2016	62.70	1.00

1. Si costruisca la serie dei numeri indici a base mobile e si calcoli, commentando il risultato, il valore di ${}_{2013}I_{2015}$, numero indice al tempo 2015 con base 2013.

${}_{2013}I_{2015} = \frac{x_{2015}}{x_{2013}} = \frac{62.7}{55} = 1.14$. Tra il tempo 2013 e il tempo 2015 si è registrato un incremento complessivo del 14% della grandezza oggetto di studio.

2. Si calcoli il tasso medio di variazione dal 2013 al 2015 e si commenti il risultato ottenuto.

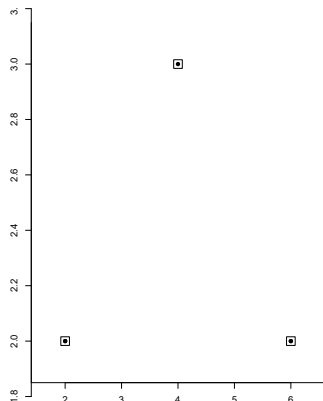
$\alpha = \sqrt[2]{{}_{2013}I_{2015}} = \sqrt[2]{{}_{2013}I_{2014} \cdot {}_{2014}I_{2015}} = 1.0677$. Da cui segue una variazione relativa media pari a 0.0677, corrispondente a un incremento medio annuo tra il 2013 e il 2015 del 6.77%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	2	4	6	6
y_i	2	2	3	2	2

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 2.2, \hat{b} = 0, \text{ indice adattamento } M(\hat{E}^2) = 0.16 \quad \hat{c} = 0.4583, \text{ indice adattamento } M(\hat{E}^2) = 0.9667$$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 2.2$ in quanto caratterizzato da residuo quadratico medio inferiore.

ESERCIZIO 4.

L'urna $U1$ contiene 3 palline nere e 1 pallina rossa. L'urna $U2$ contiene 9 palline nere e 10 palline rosse.

1. Si calcoli la probabilità che estraendo una pallina dall'urna $U1$ e una pallina dall'urna $U2$ si ottengano due palline dello stesso colore (si ottengano due palline nere oppure due palline rosse).

$$\begin{aligned} P(2 \text{ nere o } 2 \text{ rosse}) &= P\{(N_1 \cap N_2) \cup (R_1 \cap R_2)\} = \\ &= P(N_1 \cap N_2) + P(R_1 \cap R_2) = \\ &= P(N_1) \cdot P(N_2) + P(R_1) \cdot P(R_2) = \\ &= \frac{3}{4} \cdot \frac{9}{19} + \frac{1}{4} \cdot \frac{10}{19} = \\ &= 0.3553 + 0.1316 = 0.4868. \end{aligned}$$

2. Si calcoli la probabilità che estraendo con reimmissione 4 palline dall'urna $U2$ ce ne siano almeno 3 nere.

$$\begin{aligned} P(\text{almeno } 3N) &= \binom{4}{3} \left(\frac{9}{19}\right)^3 \left(\frac{10}{19}\right) + \left(\frac{9}{19}\right)^4 = \\ &= 0.2238 + 0.0503 = 0.2741 \end{aligned}$$

3. Si estraiano contemporaneamente due palline dall'urna $U1$ e le si mettano nell'urna $U2$. Si calcoli la probabilità che estraendo con reimmissione 2 palline dall'urna $U2$ (modificata) si ottengano 2 palline nere.

In base alla legge delle probabilità totali abbiamo

$$\begin{aligned} P(2N) &= P(((2R)_1 \cap 2N) \cup ((2N)_1 \cap 2N) \cup ((1N1R)_1 \cap 2N)) = \\ &= P((2R)_1 \cap 2N) + P((2N)_1 \cap 2N) + P((1N1R)_1 \cap 2N) = \\ &= P((2R)_1)P(2N|(2R)_1) + P((2N)_1)P(2N|(2N)_1) + P((1N1R)_1)P(2N|(1N1R)_1) = \\ &= 0 + \left(\frac{3}{4} \cdot \frac{2}{3}\right) \left(\frac{11}{21}\right)^2 + \left(2 \cdot \frac{3}{4} \cdot \frac{1}{3}\right) \left(\frac{10}{21}\right)^2 = \\ &= 0.5 \cdot 0.2744 + 0.5 \cdot 0.2268 = \\ &= 0.1372 + 0.1134 = 0.2506 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 22.06.2017 (Tema 294)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

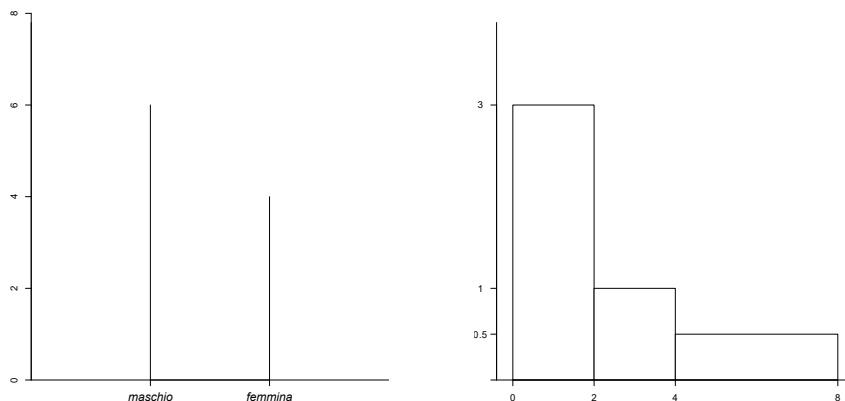
Studente _____ matricola _____

ESERCIZIO 1.

Data la seguente tabella a doppia entrata che riporta la distribuzione congiunta dei caratteri X = 'genere' e Y = 'tempo dedicato settimanalmente ai lavori domestici (in ore)' per un collettivo di 10 persone:

$X \setminus Y$	0 ÷ 2	2 ÷ 4	4 ÷ 8
maschio	6	0	0
femmina	0	2	2

1. Fornire una opportuna rappresentazione grafica delle distribuzioni marginali di X e Y .



2. Calcolare moda, media e mediana per la marginale Y .

$$\text{Moda}(Y) = 0 \div 2 \quad M(Y) = 2.4 \quad y_{0.5} = 1.6667$$

3. Si calcoli un opportuno indice di eterogeneità per la variabile X e se ne commenti il valore.

$$F_N = 0.8. \text{ Sussiste un livello alto di eterogeneità.}$$

4. Si calcolino la varianza within e la varianza between del tempo dedicato ai lavori domestici considerando come gruppi quelli definiti dal genere e si commentino i risultati ottenuti.

$$\sigma_{\text{Between}}^2 = 2.94 \quad \sigma_{\text{Within}}^2 = 0.9 \text{ da cui abbiamo } \eta_{Y|X}^2 = \frac{2.94}{2.94+0.9} = 0.7656.$$

Sussiste, quindi, un livello di dipendenza medio alto.

5. Senza fare calcoli, si dica che valore assume l'indice χ^2 normalizzato.

$\chi_N^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y essendo le distribuzioni condizionate $x|y_j$ degeneri.

ESERCIZIO 2.

Il prospetto seguente riporta la serie storica dei prezzi di due beni A e B fra il 2012 e il 2016:

t	Ap_t	Bp_t	$NIBFA(2012)$
2012	200	400	1.00
2013	210	410	1.05
2014	220	390	1.10
2015	240	390	1.20
2016	260	410	1.30

1. Costruire la serie dei numeri indice a base fissa con base nel 2012 per il bene A .

2. Sapendo che le quantità $Aq_{2012} = 100$ e $Bq_{2012} = 200$, calcolare l'indice di Laspeyres dei prezzi al 2013 con base al 2012.

$$\frac{p}{2012} I_{2013}^L = \frac{1}{1e+05} 103000 = 1.03.$$

La variazione relativa nel livello generale dei prezzi, secondo l'Indice di Laspeyres, è del 3%.

3. Si calcoli il tasso medio di variazione del prezzo del prodotto A nel periodo 2014-2016 interpretando il risultato ottenuto.

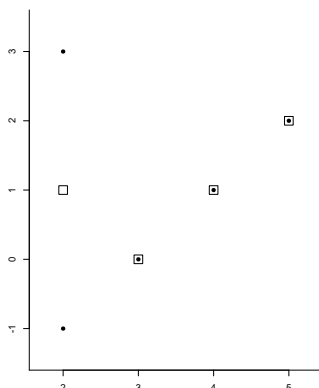
$$\alpha = \sqrt[3]{\frac{I_{2016}}{I_{2014}}} = 1.0871. \text{ Da cui segue una variazione relativa media pari a } 0.0871, \text{ corrispondente a un incremento medio annuo tra il 2014 e il 2016 del } 8.71\%.$$

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolti i dati relativi alle seguenti 5 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 5$

x_i	2	3	4	5	2
y_i	-1	0	1	2	3

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 0.0588, \hat{b} = 0.2941, \text{ indice adattamento } M(\hat{E}^2) = 1.8824 \quad \hat{c} = 0.3103, \text{ indice adattamento } M(\hat{E}^2) = 1.8828$$

3. Confrontare gli indici di adattamento, commentando le eventuali differenze.

I due modelli sono quasi equivalenti, è marginalmente migliore il modello $Y^* = 0.0588 + 0.2941 \cdot X$ in quanto caratterizzato da residuo quadratico medio leggermente inferiore.

4. Senza svolgere i conti, cosa è possibile dire sui valori di $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

Vale la relazione $\rho^2 (= 0.0588) < \eta_{Y|X}^2 < \eta_{X|Y}^2 = 1$. Sussiste, infatti, dipendenza funzionale di X da Y in quanto a ogni y_j corrisponde una sola x_i ; $0 < \eta_{Y|X}^2 < 1$ in quanto la distribuzione condizionata $Y|X = 2$ non è degenera.

5. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 5$.

$$Y^*(5) = 0.0588 + 0.2941 \cdot 5 = 1.5294$$

ESERCIZIO 4.

Nel corso della sperimentazione dei metodi di indagine per l'intolleranza al glutine viene messo a punto un nuovo test. I possibili esiti che il test può dare sono due: POSITIVO se il test dichiara l'individuo come intollerante o NEGATIVO se nell'esame non emerge una intolleranza al glutine.

Tuttavia il metodo di screening non è infallibile e la positività al test non predice perfettamente se l'individuo è davvero intollerante (e quindi MALATO) oppure se è SANO.

Sapendo che la probabilità che un individuo sia malato nella popolazione di riferimento è pari a 0.2. È inoltre noto che la diagnosi attraverso il test clinico ha le seguenti caratteristiche: la probabilità che un soggetto risulti positivo dato che nella realtà è malato è pari a 0.85 mentre la probabilità che un individuo risulti positivo dato che è sano è pari 0.1.

Si risponda alle seguenti domande:

1. Quale è la probabilità che un generico individuo risulti positivo al test?

Indicati con M ed S gli eventi 'soggetto malato' e 'soggetto sano' e con $+$ e $-$ gli eventi 'positivo al test' e 'negativo al test' sono disponibili le seguenti informazioni:

$$P(M) = 0.2, P(+|M) = 0.85, P(+|S) = 0.1.$$

Applicando la legge delle probabilità totali abbiamo:

$$\begin{aligned} P(+) &= P(+ \cap (M \cup S)) = P((+ \cap M) \cup (+ \cap S)) = P(+ \cap M) + P(+ \cap S) = P(M) \cdot P(+|M) + P(S) \cdot P(+|S) = \\ &= P(M) \cdot P(+|M) + (1 - P(M)) \cdot P(+|S) = 0.2 \cdot 0.85 + 0.8 \cdot 0.1 = 0.17 + 0.08 = 0.25 \end{aligned}$$

2. Quale è la probabilità che un individuo sia davvero malato dato che è risultato positivo al test?

$$P(M|+) = \frac{P(M \cap +)}{P(+)} = \frac{0.17}{0.25} = 0.68$$

3. Estratti con reinserimento 10 individui dalla popolazione, dire quale è la probabilità che almeno uno di essi sia malato. Sia X la variabile casuale 'n. di individui malati tra i 10 estratti'. Supponendo che la popolazione sia composta da un numero elevato di soggetti le estrazioni 'effettuate in maniera casuale' possono ritenersi indipendenti. Abbiamo

$$X \sim \text{Bin}(n = 10, p = 0.2)$$

e

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0} 0.2^0 \cdot 0.8^{10} = 1 - 0.1074 = 0.8926.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 13.07.2017 (Tema 294-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

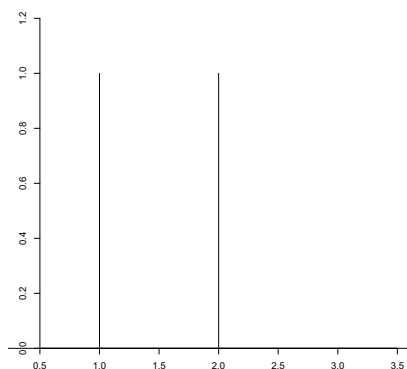
Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta la distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	1	2	3
x_1	2	2	0
x_2	0	2	2

1. Si rappresenti graficamente la distribuzione di frequenza della condizionata $Y|x_1$ e si misuri l'eterogeneità delle distribuzioni $X|y_2$ e $X|y_3$.



$$G_N(X|y_2) = F_N(X|y_2) = 1 \text{ (max. eterogeneità)} \quad G_N(X|y_3) = F_N(X|y_3) = 0 \text{ (assenza di eterogeneità)}$$

2. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ mediante opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 1.5$ Indice di variabilità: $CV(Y|x_1) = 0.3333$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 2.5$ Indice di variabilità: $CV(Y|x_2) = 0.25$

La distribuzione condizionata $Y|x_2$ presenta un livello medio superiore, ma una variabilità inferiore a $Y|x_1$.

3. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

$\chi_N = 0.7071$. Sussiste un livello di connessione medio alto.

ESERCIZIO 2.

La seguente tabella riporta la serie storica del fatturato annuo di un'azienda:

t	x_t	$NIBF(t = 2013)$
2012	50.00	0.9091
2013	55.00	1.0000
2014	52.25	0.9500
2015	62.70	1.1400
2016	62.70	1.1400

1. Si costruisca la serie dei numeri indici a base fissa con base 2013 e si calcoli, commentando il risultato, il valore di ${}_{2015}I_{2012}$, numero indice al tempo 2012 con base 2015.

${}_{2015}I_{2012} = 0.7974$. La grandezza al tempo 2012 era inferiore del 20.26% rispetto alla grandezza al tempo 2015; Ovvero il valore della grandezza nel 2012 era il 79.74% del valore registrato nel 2015.

2. Si calcoli il tasso medio di variazione dal 2012 al 2016 e si commenti il risultato ottenuto.

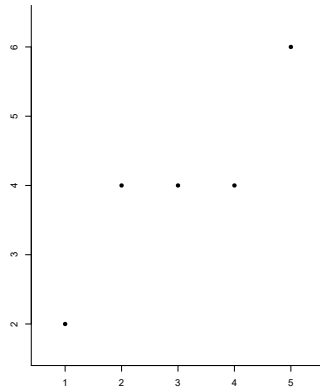
$\alpha = \sqrt[4]{{}_{2012}I_{2016}} = 1.0582$. Da cui segue una variazione relativa media pari a 0.0582, corrispondente a un incremento medio annuo tra il 2012 e il 2016 del 5.82%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	3	4	5
y_i	2	4	4	4	6

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 1.6$, $\hat{b} = 0.8$, indice adattamento $\rho^2 = 0.8$ $\hat{c} = 4$, indice adattamento $R^2 = 0$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 1.6 + 0.8 \cdot X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Senza svolgere i conti, cosa è possibile dire con riferimento ai dati in esame riguardo alla relazione tra ρ^2 ed $\eta_{Y|X}^2$.

Vale la relazione $\rho^2 (= 0.8) < \eta_{Y|X}^2 = 1$. Sussiste, infatti, dipendenza funzionale di Y da X in quanto le distribuzioni condizionate $Y|X_i$ sono degeneri.

5. Si motivi perché $\eta_{X|Y}^2 < \eta_{Y|X}^2$.

Vale la relazione $\eta_{X|Y}^2 = \rho^2 (= 0.8) < \eta_{Y|X}^2 = 1$ in quanto le medie condizionate $M(X|y_j)$ sono allineate.

ESERCIZIO 4.

L'urna $U1$ contiene 6 palline nere e 2 palline rosse. L'urna $U2$ contiene 18 palline nere e 1 pallina rossa.

1. Si calcoli la probabilità che estraendo prima una pallina dall'urna $U1$ e poi una pallina dall'urna $U2$ la prima sia rossa e la seconda sia nera.

$$P(R_1 \cap N_2) = P(R_1) \cdot P(N_2) = \frac{2}{8} \cdot \frac{18}{19} = 0.2368.$$

2. Si calcoli la probabilità che estraendo con reimmissione 40 palline dall'urna $U2$ almeno una sia rossa.

Sia X la variabile casuale 'n. di palline rosse tra le 40 estratte con reimmissione'.

Abbiamo

$$X \sim \text{Bin}(n = 40, p = \frac{1}{19} = 0.0526)$$

e

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{40}{0} 0.0526^0 \cdot 0.9474^{40} = 1 - 0.1150 = 0.8850.$$

Si osserva che non sussistono le condizioni per l'approssimazione della variabile casuale binomiale con la normale in quanto $np = 2.1053 < 5$, pur essendo $n(1-p) = 37.8947 > 5$.

3. Si estraiano, con reimmissione, due palline dall'urna $U1$ e le si mettano nell'urna $U2$. Si calcoli la probabilità che estraendo con reimmissione 2 palline dall'urna $U2$ (modificata) si ottengano 2 palline nere.

Ricordando che la frazione di palline nere nell'urna U_1 è pari a 0.75, in base alla legge delle probabilità totali abbiamo

$$\begin{aligned} P(2N) &= P(((2R)_1 \cap 2N) \cup ((2N)_1 \cap 2N) \cup ((1N1R)_1 \cap 2N)) = \\ &= P((2R)_1 \cap 2N) + P((2N)_1 \cap 2N) + P((1N1R)_1 \cap 2N) = \\ &= P((2R)_1)P(2N|(2R)_1) + P((2N)_1)P(2N|(2N)_1) + P((1N1R)_1)P(2N|(1N1R)_1) = \\ &= \left\{ \binom{2}{0} 0.75^0 \cdot 0.25^2 \right\} \cdot \left\{ \binom{2}{2} \frac{18^2}{21^2} \cdot \frac{3}{21} \right\} + \left\{ \binom{2}{2} 0.75^2 \cdot 0.25^0 \right\} \cdot \left\{ \binom{2}{2} \frac{20^2}{21^2} \cdot \frac{1}{21} \right\} + \\ &\quad + \left\{ \binom{2}{1} 0.75^1 \cdot 0.25^1 \right\} \cdot \left\{ \binom{2}{2} \frac{19^2}{21^2} \cdot \frac{2}{21} \right\} \\ &= \left\{ \binom{2}{0} 0.75^0 \cdot 0.25^2 \right\} \cdot \frac{18^2}{21^2} + \left\{ \binom{2}{2} 0.75^2 \cdot 0.25^0 \right\} \cdot \frac{20^2}{21^2} + \left\{ \binom{2}{1} 0.75^1 \cdot 0.25^1 \right\} \cdot \frac{19^2}{21^2} = \\ &= 0.0625 \cdot 0.7347 + 0.5625 \cdot 0.9070 + 0.375 \cdot 0.8186 = \\ &= 0.0459 + 0.5102 + 0.3070 = 0.8631 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 21.08.2017 (Tema 295-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta alcune informazioni sulla distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	2	3	4	
x_1	4	4	4	12
x_2	2	2	2	6
	6	6	6	18

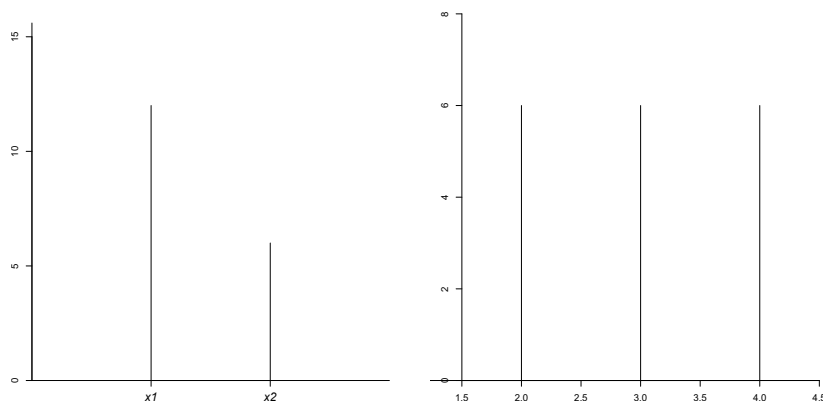
1. Sapendo che la varianza between di Y rispetto ai gruppi definiti dalle categorie della variabile X è nulla, si costruisca la distribuzione delle frequenze congiunte delle due variabili.

Si osserva in primo luogo che $n_{1\bullet} = 12$ e $n_{21} = 2$.

Vale $M(Y|x_2) = 3$ dal momento che la distribuzione $Y|x_2$ risulta simmetrica.

Affinché la varianza between di Y rispetto ai gruppi definiti dalle categorie della variabile X sia nulla le due medie di gruppo devono essere identiche, $M(Y|x_1) = M(Y|x_2) = 3$ e, quindi, anche la distribuzione $Y|x_1$ deve essere simmetrica, con $n_{21} = n_{23} = 4$.

2. Si rappresentino graficamente le distribuzioni marginali e si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e di variabilità.



$Y|x_1$ Indice di posizione: $M(Y|x_1) = 3$ Indice di variabilità: $CV(Y|x_1) = 0.2722$ $Y|x_2$ Indice di posizione: $M(Y|x_2) = 3$ Indice di variabilità: $CV(Y|x_2) = 0.2722$

Le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ sono somiglianti e, quindi, caratterizzate da identici valori degli indici di posizione e di variabilità.

ESERCIZIO 2.

Della serie storica X_t sono noti i valori riportati nel seguente prospetto :

t	x_t	$NIBF$
2010	20	1
2011	22	1.1
2012	24	1.2
2013	26	1.3
2014	28	1.4

1. Sapendo che nel periodo considerato la grandezza X_t ha subito variazioni assolute costanti, si ricostruiscano, mostrando il procedimento, la serie storica dei dati x_t e la serie dei numeri indici a base fissa con base 2010.

Dal momento che $x_t - x_{t-1} = cost$, $t = 2011, \dots, 2014$ e che $x_{2012} - x_{2010} = 24 - 20 = 4$, abbiamo $cost = 2$.

2. Si calcoli il numero indice al tempo 2013 con base 2011 e si commenti il risultato ottenuto.

${}_{2011}I_{2013} = 1.1818$. Tra il tempo 2011 e il tempo 2013 si è registrato un incremento complessivo del 18.18% della grandezza oggetto di studio.

3. Si calcoli il tasso medio di variazione tra il 2012 e il 2014, interpretando il risultato ottenuto.

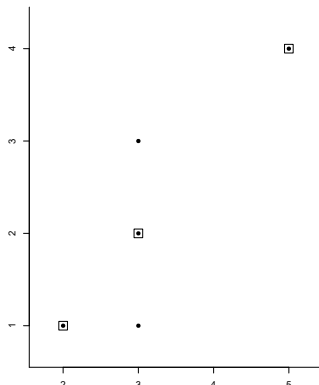
$\sqrt[2]{{}_{2012}I_{2014}} = 1.0801$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua dell'8.01%. La grandezza oggetto di studio ha subito un incremento medio annuo tra il 2012 e il 2014 dell'8.01%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	3	3	3	5
y_i	1	1	2	3	4

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = -1$, $\hat{b} = 1$, indice adattamento $\rho^2 = 0.7059$ $\hat{c} = 2.2$, indice adattamento $R^2 = 0$

3. Si individui il modello migliore.

Il modello migliore è $Y^* = -1 + X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Si indichi, motivando la risposta ma senza fare calcoli, l'intervallo dei valori che può assumere $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = \rho^2 = 0.7059$ essendo le medie condizionate allineate.

5. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 10$.

$$\hat{Y}(x = 10) = -1 + 10 = 9.$$

ESERCIZIO 4.

Si consideri il seguente gioco. Si lancia in primo luogo una moneta. Se esce testa si puntano 2 Euro sul risultato del lancio di un dado e si vincono 3 Euro se esce un numero pari. Se, invece, esce croce si puntano 4 Euro sul risultato del lancio di un dado e si vincono 9 Euro se esce un numero dispari.

1. Si stabilisca se il gioco è equo.

$$P(T \cap \text{Pari}) = \frac{1}{2} \cdot \frac{3}{6} = 0.25 \quad P(T \cap \text{Dispari}) = \frac{1}{2} \cdot \frac{3}{6} = 0.25 \quad P(C \cap \text{Pari}) = \frac{1}{2} \cdot \frac{3}{6} = 0.25 \quad P(C \cap \text{Dispari}) = \frac{1}{2} \cdot \frac{3}{6} = 0.25$$

	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
posta se T	-2	0.25	-0.5
posta se C	-4	0.25	-1.0
vincita se $T \cap \text{Pari} - \text{posta}$	1	0.25	0.25
vincita se $C \cap \text{Dispari} - \text{posta}$	5	0.25	1.25
		1	0

La variabile casuale guadagno ha valore atteso nullo e, quindi, il gioco è equo.

2. Si lanciano 100 volte 2 dadi. Calcolare la probabilità che almeno 2 volte entrambi diano risultato 1.

Sia X la variabile casuale 'n. di volte che entrambi i dadi danno risultato 1 nelle 100 prove indipendenti'. Abbiamo $X \sim \text{Bin}(n = 100, p = (\frac{1}{6})^2 = 0.02778)$.

$$\begin{aligned} P(X \geq 2) &= 1 - P(X < 2) = 1 - \{P(X = 0) + P(X = 1)\} = \\ &= 1 - \left\{ \binom{100}{0} 0.02778^0 \cdot 0.97222^{100} + \binom{100}{1} 0.02778^1 \cdot 0.97222^{99} \right\} = \\ &= 1 - (0.05978 + 0.1708) = 1 - 0.23058 = 0.76942. \end{aligned}$$

3. Si lanciano 100 volte 2 dadi. Calcolare la probabilità che almeno 17 volte entrambi diano un risultato dispari.

Indicata con X la variabile casuale 'n. di volte che entrambi i dadi danno risultato dispari nelle 100 prove indipendenti' risulta $X \sim \text{Bin}(n = 100, p = (\frac{1}{2})^2 = 0.25)$. $M(X) = 25$; $\text{Var}(X) = 18.75$. Dal momento che $np = 25 > 5$ ed $n(1 - p) = 75 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 25, \sigma^2 = 18.75)$.

Si ottiene

$$\begin{aligned} P(X \geq 17) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{17 - \mu}{\sigma}\right) = P\left(Z \geq \frac{17 - 25}{4.3301}\right) = \\ &= P(Z \geq -1.85) = P(Z \leq 1.85) = 0.9678. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 07.09.2017 (Tema 295)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta alcune informazioni sulla distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	2	3	4	
x_1	0		0	
x_2		2		6

$X \setminus Y$	2	3	4	
x_1	0	6	0	6
x_2	2	2	2	6
	2	8	2	12

1. Sapendo che la distribuzione marginale di X è caratterizzata da massima eterogeneità e che $M(Y|x_1) = M(Y|x_2)$ si costruisca la distribuzione delle frequenze congiunte.

Quando la distribuzione marginale di X è caratterizzata da massima eterogeneità, abbiamo una situazione di equidistribuzione delle frequenze, quindi $n_{1\bullet} = n_{2\bullet} = 6$.

La distribuzione condizionata $Y|x_1$ è degenera dal momento che solo $n_{22} = 6 \neq 0$, e quindi $M(Y|x_1) = 3$.

Affinché anche $M(Y|x_2) = 3$ la distribuzione $Y|x_2$ deve essere simmetrica, e quindi $n_{21} = n_{23} = 2$

2. Si calcolino la media e la varianza della Y .

$$M(Y) = 3 \quad Var(Y) = 0.3333$$

3. Considerati i due gruppi di osservazioni relativi alle condizionate $Y|x_1$ e $Y|x_2$ si indichi il valore della varianza within. varianza within = 0.3333

Dal momento che $M(Y|x_1) = M(Y|x_2)$ (le medie di gruppo sono identiche) abbiamo che $\sigma_{Between}^2 = 0$ e $\sigma_{Within}^2 = \sigma_Y^2$.

4. Senza effettuare calcoli, indicare se, e perché, le marginali X e Y sono indipendenti.

Le variabili X e Y non sono stocasticamente indipendenti in quanto le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ non sono somiglianti.

ESERCIZIO 2.

Della serie storica X_t sono noti i valori riportati nel seguente prospetto:

t	x_t	$NIBM$
2010	20	
2011	22	1.1
2012	24.2	1.1
2013	26.62	1.1
2014	29.282	1.1

1. Sapendo che nel periodo considerato la grandezza X_t ha subito variazioni relative costanti, si ricostruiscano, mostrandone il procedimento, la serie storica dei dati x_t e la serie dei numeri indici a base mobile.

Dal momento che le variazioni relative $\frac{x_t - x_{t-1}}{x_{t-1}} = cost$, $t = 2011, \dots, 2014$, abbiamo $\frac{x_t}{x_{t-1}} = 1 + cost$, $t = 2011, \dots, 2014$.

Essendo $\frac{x_{2012}}{x_{2010}} = \frac{x_{2011}}{x_{2010}} \cdot \frac{x_{2012}}{x_{2011}} = (1 + cost)^2 = 1.21$, abbiamo $(1 + cost) = 1.1$ e $cost = 0.1$.

Nel periodo considerato la grandezza oggetto di studio ha subito un incremento del 10% tra un anno e il successivo e, quindi

- $x_t = 1.1 \cdot x_{t-1}, \quad t = 2011, \dots, 2014.$
- $\frac{x_t}{x_{t-1}} = 1.1, \quad t = 2011, \dots, 2014.$

2. Si calcoli il numero indice al tempo 2013 con base 2011 e si commenti il risultato ottenuto.

${}_{2011}I_{2013} = 1.21$ Tra il tempo 2011 e il tempo 2013 si è registrato un incremento complessivo del 21% della grandezza oggetto di studio.

3. Si calcoli il tasso medio di variazione tra il 2012 e il 2014, interpretando il risultato ottenuto.

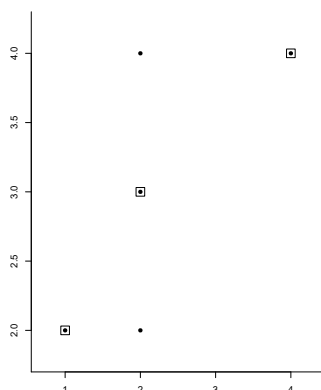
Dal momento che la serie ha subito variazioni relative costanti nel periodo, abbiamo che la variazione relativa media coincide con $cost = 0.1$. Tra ogni anno e il successivo la serie ha subito un incremento del 10%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	2	2	4
y_i	2	2	3	4	4

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d \cdot \sqrt{X}$$

$$\hat{a} = 1.625, \hat{b} = 0.625, \text{ indice adattamento } \rho^2 = 0.4688 \quad \hat{c} = 0.1532, \hat{d} = 1.9653, \text{ indice adattamento } R^2 = 0.4913$$

3. Si individui il modello migliore.

Il modello migliore è $Y^* = 0.1532 + 1.9653 \cdot \sqrt{X}$ in quanto caratterizzato da indice di adattamento più elevato.

4. Si indichi, **motivando** la risposta ma senza fare calcoli, quale relazione è vera:

- $\eta_{X|Y}^2 > \rho^2$
- $\eta_{X|Y}^2 = \rho^2$
- $\eta_{X|Y}^2 < \rho^2$

Dal momento che le medie condizionate non sono allineate abbiamo che vale $\eta_{X|Y}^2 > \rho^2 = 0.4688$

ESERCIZIO 4.

Da un mazzo di carte da bridge (52 carte) vengono tolte le 13 carte di picche.

1. Considerando l'esperimento di estrazione casuale, con reimmissione, di 3 carte costruire la variabile casuale X = 'n. di carte con seme di colore nero' (valori e probabilità).

Una volta tolte le 13 carte di picche, il mazzo di carte da bridge risulta composto da 39 carte, 13 cuori, 13 quadri e 13 fiori, e quindi da 26 carte rosse (R) e 13 carte nere (N).

La variabile casuale X = 'n. di carte con seme di colore nero' tra le 3 carte estratte con reimmissione risulta una variabile casuale Binomiale $X \sim \text{Bin}(n = 3, p = \frac{13}{39} = \frac{1}{3})$ con funzione di probabilità

$$P(X = x) = \binom{3}{x} \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{3-x}, \quad x = 0, 1, 2, 3.$$

2. Calcolare $M(X)$.

La media di una variabile casuale $X \sim (n, p)$ risulta $M(X) = n \cdot p$; nel caso in esame abbiamo $M(X) = 1$.

3. Considerando ora $n = 30$ successive estrazioni (sempre con reimmissione) calcolare la probabilità che almeno 15 siano di colore nero.

Indicata con X la variabile casuale 'n. di carte con seme nero tra le 30 estratte con reimmissione' (prove indipendenti) risulta $X \sim \text{Bin}(n = 30, p = \frac{1}{3})$. $M(X) = 10$; $\text{Var}(X) = 6.6667$. Dal momento che $np = 10 > 5$ ed $n(1 - p) = 20 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 10, \sigma^2 = 6.6667)$.

Si ottiene

$$\begin{aligned} P(X \geq 15) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{15 - \mu}{\sigma}\right) = P\left(Z \geq \frac{15 - 10}{2.582}\right) = \\ &= P(Z \geq 1.94) = 1 - P(Z \leq 1.94) = 1 - 0.9738 = 0.0262. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 17.01.2018 (Tema 296)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta le distribuzioni relative condizionate della variabile $Y = \text{Numero di ingressi al museo nell'ultimo mese}$ rispetto alla variabile $X = \text{Zona di provenienza del visitatore}$, ricavata a seguito di una indagine campionaria:

$Y X$	0 + 2	2 + 7	7 + 10	
<i>NORD</i>	0.3	0.4	0.3	1
<i>CENTRO</i>	0.2	0.6	0.2	1
<i>SUD</i>	0.3	0.4	0.3	1

1. Senza effettuare calcoli, si dica se è possibile concludere che esiste indipendenza stocastica tra X e Y . Motivare la risposta.

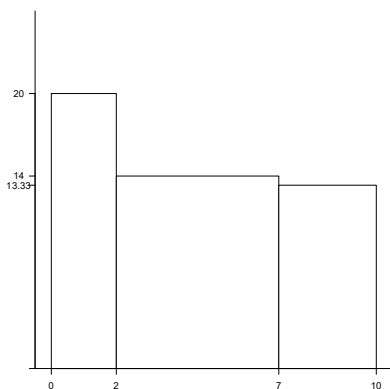
Non sussiste indipendenza stocastica tra le variabili X e Y in quanto le distribuzioni condizionate $Y|x_i$ non sono tra loro somiglianti. Abbiamo, infatti, $Y|x_1 \sim Y|x_3 \not\sim Y|x_2$.

2. Costruire la tabella delle frequenze congiunte sapendo che 50 dei visitatori intervistati provengono dal *NORD*, 50 dal *CENTRO* e 50 dal *SUD*.

Ricordando che per la generica frequenza condizionata relativa vale $f(y_j|x_i) = \frac{n_{ij}}{n_{i\bullet}}$ abbiamo che $n_{ij} = f(y_j|x_i) \cdot n_{i\bullet}$ e

$X \setminus Y$	0 + 2	2 + 7	7 + 10	$n_{i\bullet}$
<i>NORD</i>	15	20	15	50
<i>CENTRO</i>	10	30	10	50
<i>SUD</i>	15	20	15	50
$n_{\bullet j}$	40	70	40	150

3. Si rappresenti graficamente la variabile Y e se ne calcoli moda e varianza.



$$\text{Moda}(Y) = 0 + 2 \quad \text{Var}(Y) = 7.5156$$

4. Si calcoli l'indice di mutabilità normalizzato di X e se ne commenti il risultato.

$F_N = 1$. Sussiste massima eterogeneità.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	$NIBF(\text{base} = 2014)$	$NIBM$
2012	1.25	
2013	1	0.8
2014	1	1
2015	1	1
2016	0.6	0.6

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli il numero indice al tempo 2016 con base 2013 e si commenti il risultato ottenuto.

${}_{2013}I_{2016} = 0.6$ La grandezza oggetto di studio ha subito un decremento del 40% tra il 2013 e il 2016.

3. Si calcoli il tasso medio di variazione dal 2013 al 2016 e si commenti il risultato ottenuto.

$\sqrt[3]{{}_{2013}I_{2016}} = 0.8434$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del -15.66%.

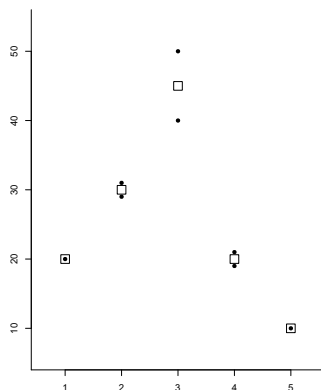
La grandezza oggetto di studio ha subito un decremento medio annuo tra il 2013 e il 2016 del 15.66%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 capofamiglia con riferimento a due caratteri quantitativi X = numero dei componenti della famiglia e Y = spesa mensile per musei e mostre (in euro).

x_i	1	2	2	3	3	4	4	5
y_i	20	29	31	40	50	21	19	10

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando la spesa in funzione del numero dei componenti della famiglia.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot (X - 3)^2$$

$$\hat{a} = 37.5, \hat{b} = -3.3333, \text{ indice adattamento } M(\hat{E}^2) = 127.583 \quad \hat{c} = 6.1111, \text{ indice adattamento } M(\hat{E}^2) = 732.444$$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 37.5 - 3.3333 \cdot X$ in quanto caratterizzato da residuo quadratico medio più basso.

4. Senza effettuare calcoli e motivando la risposta, si stabilisca: a) l'intervallo di possibili valori assumibili da $\eta_{Y|X}^2$; b) se $\eta_{X|Y}^2 < 1$.

$\rho^2 < \eta_{Y|X}^2 < 1$ in quanto le medie condizionate non sono allineate e non sussiste dipendenza funzionale di Y da X dal momento che le distribuzioni condizionate $Y|x, x \in \{2, 3, 4\}$ non sono degeneri.

$\eta_{X|Y}^2 = 1$ in quanto le distribuzioni condizionate $X|y$ sono tutte degeneri.

ESERCIZIO 4.

La retribuzione media mensile lorda di un laureato in Economia è distribuita come una Normale di media $\mu = 1600$ euro e varianza $\sigma^2 = 10000$ euro.

1. Quale percentuale di laureati guadagna più di 1700 euro al mese?

Indicata con $X \sim N(\mu = 1600, \sigma^2 = 10000)$ la variabile casuale che descrive la retribuzione media mensile lorda, occorre, in primo luogo, determinare la seguente probabilità: $P(X > 1700)$. Abbiamo

$$\begin{aligned} P(X > 1700) &= P\left(\frac{X - \mu}{\sigma} > \frac{1700 - \mu}{\sigma}\right) = P\left(Z > \frac{1700 - 1600}{100}\right) = \\ &= P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587. \end{aligned}$$

La percentuale di laureati che guadagna più di 1700 euro al mese risulta 15.87%.

2. Quale è il reddito limite mensile che definisce il 10% superiore della retribuzione dei laureati in Economia?

Il reddito limite mensile che definisce il 10% superiore della retribuzione dei laureati in Economia corrisponde al percentile di ordine 0.90 della distribuzione di X .

$$\{x : P(X \leq x) = 0.90\} = \left\{x : P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = 0.90\right\} = \left\{x : P\left(Z \leq \frac{x - 1600}{100}\right) = 0.90\right\}.$$

Dalla lettura inversa della Tavola abbiamo $z = \frac{x - 1600}{100} = 1.28$ da cui segue $x = 1728$.

3. Si calcoli la probabilità che su 10 laureati selezionati casualmente almeno uno guadagni più di 1700 euro al mese.

Sia X la variabile casuale 'n. di laureati con guadagno superiore a 1700 euro tra i 10 selezionati'. Dal momento che la popolazione dei laureati ha dimensione molto elevata, abbiamo $X \sim \text{Bin}(n = 10, p = 0.1587)$.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = \\ &= 1 - \binom{10}{0} 0.1587^0 \cdot 0.8413^{10} = \\ &= 1 - 0.1777 = 0.8223. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 31.01.2018 (Tema 297)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

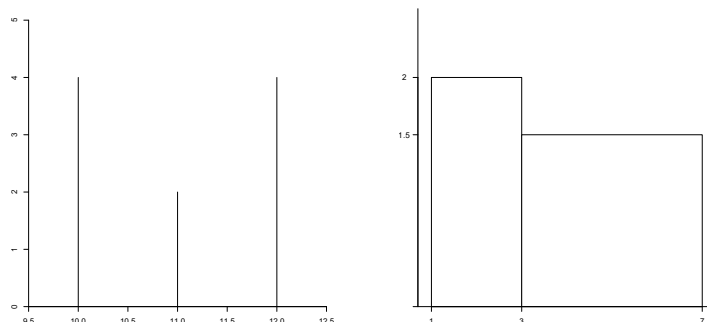
Studente _____ matricola _____

ESERCIZIO 1.

Con riferimento a 10 famiglie sono disponibili le distribuzioni marginali dei due caratteri X = 'numero di visite effettuate a musei nello scorso anno' e Y = 'durata massima della visita'.

x_i	n_i	y_j	n_j
10	4	1 + 3	4
11	2	3 + 7	6
12	4		

1. Si rappresentino graficamente le due variabili X e Y .



2. Si confrontino le due distribuzioni mediante opportuni indici di posizione e di variabilità.

X Indice di posizione: $M(X) = 11$ Indice di variabilità: $CV(X) = 0.0813$

Y Indice di posizione: $M(Y) = 3.8$ Indice di variabilità: $CV(Y) = 0.3868$

Non ha senso il confronto mediante indici di posizione tra le due variabili.

La durata media della visita presenta un livello di variabilità superiore rispetto al numero medio delle visite.

3. Si costruisca la distribuzione congiunta di X e Y nell'ipotesi che tra i due caratteri sussista dipendenza funzionale.

Occorre costruire una tabella nella quale le distribuzioni condizionate $Y|x_i$ sono degeneri. Si riportano due possibili soluzioni.

$X \ Y$	1 + 3	3 + 7	
10	4	0	4
11	0	2	2
12	0	4	4
	4	6	10

$X \ Y$	1 + 3	3 + 7	
10	0	4	4
11	0	2	2
12	4	0	4
	4	6	10

4. Si costruisca la distribuzione congiunta di X e Y nell'ipotesi di indipendenza stocastica.

Occorre costruire una tabella con distribuzioni condizionate somiglianti: $Y|x_i \sim Y$ e $X|y_j \sim X$.

$X \ Y$	1 + 3	3 + 7	
10	1.6	2.4	4
11	0.8	1.2	2
12	1.6	2.4	4
	4	6	10

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	x_t	$NIBF(base = 2013)$	$NIBM$
2012	48	2	—
2013	24	1	0.5
2014	24	1	1
2015	12	0.5	0.5
2016	6	0.25	0.5

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli il numero indice al tempo 2013 con base 2015 e si commenti il risultato ottenuto.

${}_{2015}I_{2013} = \frac{x_{2013}}{x_{2015}} = \frac{24}{12} = 2$ La grandezza nel 2013 assumeva valore doppio rispetto al 2015.

3. Si calcoli il tasso medio di variazione dal 2013 al 2015 e si commenti il risultato ottenuto.

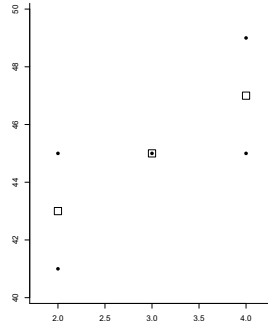
$\sqrt[2]{{}_{2013}I_{2015}} = 0.7071$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del -29.29% . La grandezza oggetto di studio ha subito un decremento medio annuo tra il 2013 e il 2015 del 29.29% .

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 capifamiglia con riferimento ai tre caratteri X = 'numero dei componenti della famiglia', Y = 'spesa mensile per musei e mostre (in euro)' e W = 'livello di gradimento registrato nelle visite dello scorso anno' con categorie b = 'basso' e a = 'alto'.

x_i	2	2	3	4	4
y_i	41	45	45	45	49
w_i	a	a	b	b	a

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando la spesa in funzione del numero dei componenti della famiglia.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d \cdot V$$

con V variabile dummy che ricodifica le categorie di W nel modo seguente $v_i = \begin{cases} 0 & \text{se } w_i = b \\ 1 & \text{se } w_i = a \end{cases}$

$\hat{a} = 39$, $\hat{b} = 2$, indice adattamento $\rho^2 = 0.5$ $\hat{c} = 45$, $\hat{d} = 0$, indice adattamento $\rho^2 = 0$

3. Si indichi, motivando la risposta, quale modello presenta il migliore adattamento.

Il modello migliore è $Y^* = 39 + 2 \cdot X$ in quanto caratterizzato da indice di adattamento ρ^2 più elevato.

4. Si indichi l'intervallo dei valori possibili per $\eta_{Y|X}^2$ e $\eta_{Y|W}^2$.

$\eta_{Y|X}^2 = \rho_{XY}^2 = 0.5$ in quanto le medie condizionate $M(Y|x_i)$ sono allineate. $\eta_{Y|W}^2 = \rho_{XV}^2 = 0$ in quanto il modello II) include solo variabili indicatrici tra i regressori, e risulta quindi equivalente alla funzione di regressione.

ESERCIZIO 4.

L'urna A contiene 4 elementi contrassegnati con i simboli 1, 2, 3, 4, l'urna B 3 elementi contrassegnati con 2, 4, 6.

1. Si paga una posta di 3€ per partecipare al seguente gioco. Si lancia una moneta con $P(testa) = 0.1$ e $P(croce) = 0.9$. Se esce testa si estrae un elemento dall'urna A altrimenti uno dall'urna B . Si vince una somma in € pari al simbolo che contrassegna l'elemento estratto. Si costruisca la variabile casuale guadagno del giocatore e si stabilisca se il gioco è equo. Lo spazio probabilistico è composto dai seguenti eventi elementari: $\Omega = \{T1, T2, T3, T4, C2, C4, C6\}$, a cui corrispondono le probabilità $\{0.1 \cdot 0.25 = 0.025, 0.025, 0.025, 0.025, 0.9 \cdot 0.3333 = 0.3, 0.3, 0.3\}$. La variabile casuale guadagno

g	$P(G = g)$	$g \cdot P(G = g)$
-2	0.025	-0.050
-1	0.325	-0.325
0	0.025	0.000
1	0.325	0.325
3	0.300	0.900
1.000		0.850

ha valore atteso pari a $0.85 \neq 0$. Il gioco non è equo ma favorevole al giocatore.

2. Si consideri il seguente esperimento. Si estrae un elemento dall'urna B e lo si inserisce nell'urna A ; si estrae poi un elemento dall'urna A (modificata). Si calcoli la probabilità che l'elemento estratto dall'urna A rechi un simbolo dispari. Indicati con D_B e D_A gli eventi estrazione di un elemento con simbolo dispari dall'urna B e A modificata, occorre calcolare la seguente probabilità

$$P(D_B) \cdot P(D_A|D_B) + P(\bar{D}_B) \cdot P(D_A|\bar{D}_B) = 0 \cdot \frac{2}{4} + 1 \cdot \frac{2}{5} = 0.4.$$

3. Si calcoli la probabilità che in 576 estrazioni con reimmissione dall'urna B ve ne siano almeno 181 con il simbolo 2.

Sia X la variabile casuale 'n. elementi con il simbolo 2 nelle 576 estrazioni con reimmissione dall'urna B '. Abbiamo $X \sim \text{Bin}(n = 576, p = \frac{1}{3})$ con $E(X) = n \cdot p = 576 \cdot \frac{1}{3} = 192$ e $\text{Var}(X) = n \cdot p \cdot (1 - p) = 576 \cdot \frac{1}{3} \cdot \frac{2}{3} = 128$.

Dal momento che $np = 192 > 5$ ed $n(1 - p) = 384 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 192, \sigma^2 = 128)$.

Si ottiene

$$\begin{aligned} P(X \geq 181) &= P\left(\frac{X - \mu}{\sigma} \geq \frac{181 - \mu}{\sigma}\right) = P\left(Z \geq \frac{181 - 192}{11.3137}\right) = \\ &= P(Z \geq -0.97) = P(Z \leq 0.97) = 0.8340. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 20.02.2018 (Tema 298)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

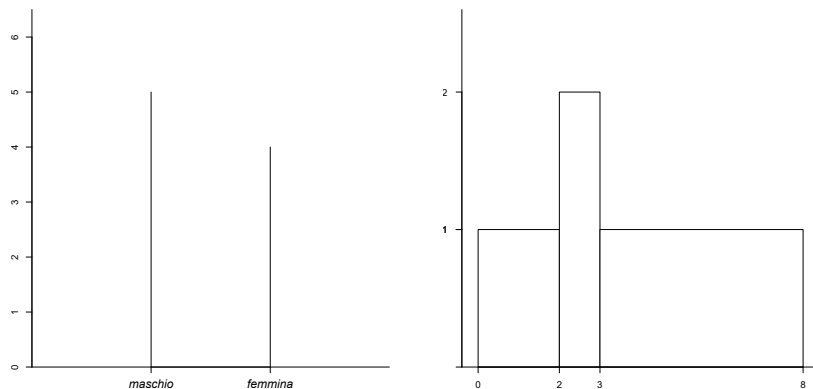
Studente _____ matricola _____

ESERCIZIO 1.

Data la seguente tabella a doppia entrata che riporta la distribuzione congiunta dei caratteri X = 'genere' e Y = 'numero di visite mediche specialistiche fatte in un anno' di 9 persone:

$X \setminus Y$	0 + 2	2 + 3	3 + 8
maschio	0	0	5
femmina	2	2	0

1. Rappresentare graficamente le distribuzioni marginali di X e di Y .



2. Calcolare media aritmetica e media geometrica della marginale Y .

$$M(Y) = \mu^{(1)} = 3.8333, \quad \mu^{(0)} = 3.1604$$

3. Calcolare la varianza within e la varianza between di Y considerando come gruppi quelli definiti dal genere (X) e commentare i risultati ottenuti.

$$\sigma_{within}^2 = 0.25, \quad \sigma_{between}^2 = 3.4722$$

4. Senza fare calcoli e giustificando la risposta, si indichi il valore dell'indice χ^2 normalizzato.

$$\chi_N^2 = 1 \text{ in quanto sussiste dipendenza funzionale di } X \text{ da } Y$$

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi medi annui di 2 beni di uso quotidiano nel periodo 2012-2016:

t	bene A	bene B	$NIBF_A(2013)$	$NIBM_B$
2012	9	77	0.8182	
2013	11	65	1	0.8442
2014	12	47	1.0909	0.7231
2015	14	35	1.2727	0.7447
2016	15	33	1.3636	0.9429

1. Costruire la serie dei numeri indici a base fissa dei prezzi con base 2013 per il bene A e la serie dei numeri indici a base mobile dei prezzi per il bene B.

2. Calcolare il tasso medio di variazione per il prezzo del bene B tra il 2013 e il 2015, interpretando il risultato.

$$\sqrt[2]{\frac{I_{2015}}{I_{2013}}} = 0.7338, \text{ numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del } -26.62\%.$$

La grandezza oggetto di studio ha subito un decremento medio annuo tra il 2013 e il 2015 del 26.62%.

3. Note le quantità $A_{q2013} = 150$ e $B_{q2013} = 850$, calcolare l'indice di Laspeyres dei prezzi al 2016 con base 2013.

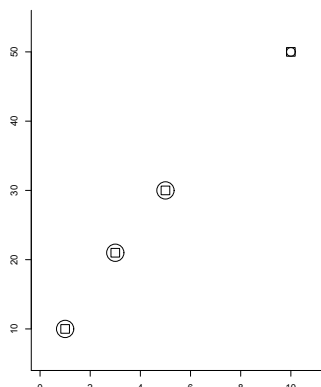
$$\frac{p}{2013} I_{2016}^L = 0.5325. \text{ I prezzi dei due prodotti hanno subito un decremento medio del } 46.75\% \text{ tra il 2013 e il 2016.}$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 7 unità statistiche con riferimento a due caratteri quantitativi X e Y per i quali si vuole studiare la dipendenza di Y da X :

x_i	1	1	3	3	5	5	10
y_i	10	10	21	21	30	30	50

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione di Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot \ln(X)$$

$\hat{a} = 6.7783$, $\hat{b} = 4.4483$, indice adattamento $M(\hat{E}^2) = 1.1513$ $\hat{c} = 19.9933$, indice adattamento $M(\hat{E}^2) = 32.433$

3. Si indichi quale modello presenta il migliore adattamento ai dati.

Il modello migliore è $Y^* = 6.7783 + 4.4483 \cdot X$ in quanto caratterizzato da residuo quadratico medio più basso.

4. Si dica, motivando teoricamente la risposta, se $\eta_{Y|X}^2 \neq \eta_{X|Y}^2$.

$\eta_{Y|X}^2 = \eta_{X|Y}^2$ in quanto sussiste dipendenza funzionale bijectiva.

ESERCIZIO 4.

Un'urna è composta da 4 gettoni gialli e 16 gettoni blu. Si consideri l'esperimento casuale 'estrazione con reimmissione di 3 gettoni dall'urna'.

1. Si costruisca lo spazio probabilistico Ω associato all'esperimento.

Indicati con G_i e B_i gli eventi estrazione di gettone giallo e blu nella i -esima estrazione ($i = 1, 2, 3$), lo spazio probabilistico risulta composto dai seguenti eventi elementari:

$$\Omega = \{G_1G_2G_3, G_1G_2B_3, G_1B_2G_3, B_1G_2G_3, G_1B_2B_3, B_1G_2B_3, B_1B_2G_3, B_1B_2B_3\},$$

a cui corrispondono le probabilità

$$\{0.2^3 = 0.008, 0.2^2 \cdot 0.8 = 0.032, 0.032, 0.032, 0.2 \cdot 0.8^2 = 0.128, 0.128, 0.128, 0.8^3 = 0.512\}$$

tenendo conto che le estrazioni sono con reimmissione (prove indipendenti) e che $P(G_i) = P(G) = 0.2$, $P(B_i) = P(B) = 0.8$, $i = 1, 2, 3$.

2. Sapendo che ogni gettone giallo reca impresso il numero 2, ogni gettone blu il numero 3, si costruisca la variabile casuale X = 'somma dei valori impressi sui gettoni estratti' e si calcoli la probabilità che X assuma valori maggiori o uguali a 8.

Gli eventi elementari di Ω hanno le seguenti immagini in $\mathbb{R}$:

$$\begin{aligned} X(G_1G_2G_3) &= 6, & X(G_1G_2B_3) &= 7, & X(G_1B_2G_3) &= 7, & X(B_1G_2G_3) &= 7 \\ X(G_1B_2B_3) &= 8, & X(B_1G_2B_3) &= 8, & X(B_1B_2G_3) &= 8, & X(B_1B_2B_3) &= 9. \end{aligned}$$

La variabile casuale X è caratterizzata dalla seguente distribuzione di probabilità

x	$P(X = x)$
6	0.008
7	0.096
8	0.384
9	0.512
	1.000

Abbiamo, infine, $P(X \geq 8) = P(X = 8) + P(X = 9) = 0.384 + 0.512 = 0.896$.

3. Facendo ora 4 estrazioni **senza reimmissione**, si calcoli la probabilità di ottenere almeno 1 gettone giallo.

$P(\text{almeno 1 gettone giallo}) = 1 - P(\text{nessun gettone giallo}) = 1 - P(B_1 \cap B_2 \cap B_3 \cap B_4) = 1 - P(B_1) \cdot P(B_2|B_1) \cdot P(B_3|B_1 \cap B_2) \cdot P(B_4|B_1 \cap B_2 \cap B_3) = 1 - \frac{16}{20} \cdot \frac{15}{19} \cdot \frac{14}{18} \cdot \frac{13}{17} = 1 - 0.3756 = 0.6244$.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 04.06.2018 (Tema 299-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

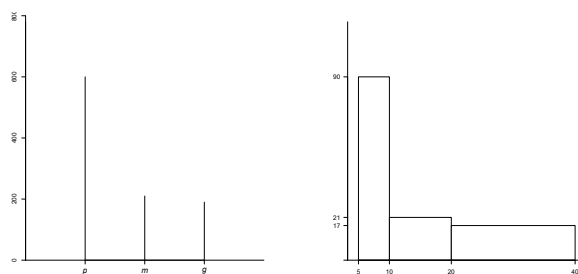
Si sono rilevati la dimensione X (p = piccola; m = media; g = grande) e il fatturato annuo in centinaia di migliaia di euro Y di 1000 imprese. Si riportano la distribuzione delle frequenze relative di X e le densità di frequenza di Y .

x_i	f_i	y_i	d_i
p	0.60	5 - 10	90
m	0.21	10 - 20	21
g	0.19	20 - 40	17

1. Si costruiscano le distribuzioni delle frequenze assolute di X e di Y .

x_i	n_i	y_i	n_i
p	600	5 - 10	450
m	210	10 - 20	210
g	190	20 - 40	340

2. Si rappresentino graficamente le due variabili e si riassumano le stesse con opportuni indici di posizione e di variabilità.

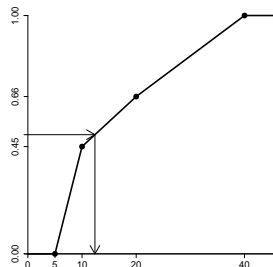


X Indice di posizione: $\text{Moda}(X) = p$ Indice di variabilità: $F_N = 0.5996$

Y Indice di posizione: $M(Y) = 16.725$ Indice di variabilità: $CV(Y) = 0.5944$

3. Si calcoli la media geometrica e si individui su un opportuno grafico la mediana della variabile Y .

$$\mu_{(0)} = 13.8988$$



ESERCIZIO 2.

Il prospetto seguente contiene le serie storiche dei prezzi dei prodotti A , B e C :

t	Ap_t	Bp_t	Cp_t
2011	200	400	600
2012	220	445	650
2013	210	420	Cp_{2013}
2014	230	460	Cp_{2014}
2015	250	490	700

1. Sapendo che nel 2011 sono state scambiate le seguenti quantità: $Aq_{2011} = 100$, $Bq_{2011} = 200$ e $Cq_{2011} = 100$ si determini, giustificando la risposta, il prezzo del prodotto C nel 2013 tale che la variazione del livello generale dei prezzi dei 3 prodotti al tempo 2013 con base 2011, determinata secondo l'indice di Laspeyres, risulti pari al 5%.

Dal momento che i prezzi relativi riferiti ai prodotti A e B risultano

$$\frac{Ap_{2013}}{Ap_{2011}} = \frac{210}{200} = 1.05 \quad \text{e} \quad \frac{Bp_{2013}}{Bp_{2011}} = \frac{420}{400} = 1.05$$

al fine di ottenere $\frac{p}{2011} I_{2013}^L = 1.05$ deve valere anche $\frac{Cp_{2013}}{Cp_{2011}} = 1.05$ da cui segue $Cp_{2013} = 630$.

2. Si calcoli il numero indice del prezzo del prodotto A al tempo 2013 con base 2011 e si commenti il risultato ottenuto. $2011 I_{2013} = 1.05$. Il prezzo del prodotto A ha subito un incremento del 5% tra il 2011 e il 2013.

3. Si calcoli il tasso medio di variazione del prezzo del prodotto A tra il 2012 e il 2014, interpretando il risultato ottenuto.

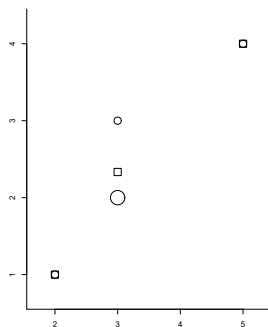
$\sqrt[2]{2012 I_{2014}} = \sqrt[2]{\frac{230}{220}} = \sqrt[2]{1.0455} = 1.0225$, numero indice a base mobile medio, a cui corrisponde un incremento medio annuo del 2.25%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	2	3	3	3	5
y_i	1	2	2	3	4

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = -0.6667, \hat{b} = 0.9583, \text{ indice adattamento } M(\hat{E}^2) = 0.1583 \quad \hat{c} = 0.7679, \text{ indice adattamento } M(\hat{E}^2) = 0.1964$$

3. Si individui il modello migliore.

Il modello migliore è $Y^* = -0.6667 + 0.9583 \cdot X$ in quanto caratterizzato da residuo quadratico medio più basso.

4. Si indichi, motivando la risposta ma senza fare calcoli, l'intervallo dei valori che può assumere $\eta_{X|Y}^2$.

$\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale di X da Y dal momento che le distribuzioni condizionate $X|y_j$ sono degeneri.

ESERCIZIO 4.

Si consideri l'esperimento di estrazione di 4 palline da un'urna contenente 21 palline rosse e 29 palline blu.

1. Si calcoli la probabilità di ottenere almeno 3 palline blu nell'ipotesi che le 4 estrazioni siano effettuate senza reimmissione.

$$P(\text{almeno } 3 B) = P(3B) + P(4B) = \binom{4}{3} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{21}{47} + \binom{4}{4} \frac{29}{50} \frac{28}{49} \frac{27}{48} \frac{26}{47} = 0.3332 + 0.1031 = 0.4363$$

o mediante la v.c. ipergeometrica

$$P(X = 3) + P(X = 4) = \frac{\binom{29}{3} \binom{21}{1}}{\binom{50}{4}} + \frac{\binom{29}{4} \binom{21}{0}}{\binom{50}{4}} = 0.4363$$

2. Si calcoli la probabilità di ottenere almeno 3 palline blu nell'ipotesi che le 4 estrazioni siano effettuate con reimmissione.

Sia X la variabile casuale 'n. di palline blu nelle 4 estrazioni (con reimmissione) (prove indipendenti)'. $X \sim \text{Bin}(n = 4, p = 0.58)$.

$$P(X \geq 3) = \binom{4}{3} 0.58^3 0.42^1 + \binom{4}{4} 0.58^4 0.42^0 = 0.3278 + 0.1132 = 0.4410.$$

3. Si calcoli la probabilità di ottenere almeno 36 palline blu nell'ipotesi che siano effettuate 60 estrazioni con reimmissione.

Indicata con X la variabile casuale 'numero di successi' nelle 60 prove indipendenti risulta $X \sim \text{Bin}(n = 60, p = 0.58)$. $M(X) = 34.8$; $\text{Var}(X) = 14.6160$. Dal momento che $np = 34.8 > 5$ ed $n(1 - p) = 25.2 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 34.8, \sigma^2 = 14.6160).$$

Si ottiene

$$P(X > 36) = P\left(\frac{X - \mu}{\sigma} > \frac{36 - \mu}{\sigma}\right) = P\left(Z > \frac{36 - 34.8}{3.8231}\right) = P(Z > 0.31) = 1 - P(Z \leq 0.31) = 1 - 0.6217 = 0.3783.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 18.06.2018 (Tema 299)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

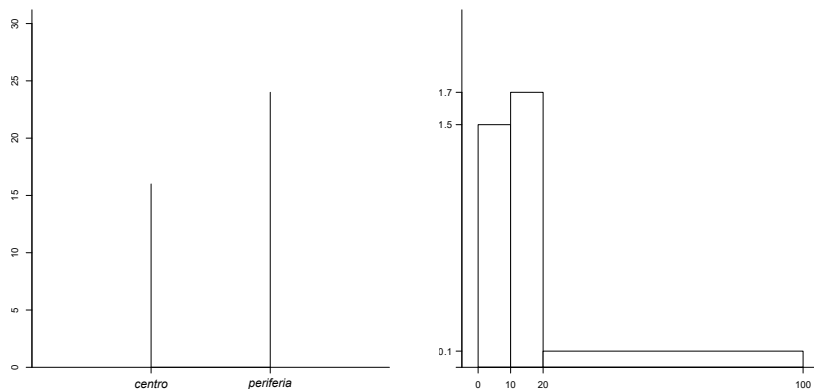
Studente _____ matricola _____

ESERCIZIO 1.

Un'azienda che produce cibo e prodotti per animali ha condotto una indagine campionaria su 40 individui residenti in una città e proprietari di animali domestici al fine di valutare il luogo ideale per aprire un nuovo punto vendita. Nella tabella sottostante sono state rilevate le frequenze congiunte di due variabili X = 'Luogo di abitazione dell'intervistato' e Y = 'Spesa settimanale in cibo e prodotti per animali' (in euro):

$X \setminus Y$	0 - 10	10 - 20	20 - 100
centro	7	9	0
periferia	8	8	8

1. Fornire una opportuna rappresentazione grafica delle distribuzioni marginali di X e di Y .



2. Calcolare moda, media e mediana per la marginale Y .

moda = 10 - 20 media aritmetica $M(Y) = 20.25$ $y_{0.5} = 12.9412$

3. Calcolare un opportuno indice di eterogeneità per la variabile X e commentarne il valore.

$F_N = 0.8$. Sussiste un elevato livello di eterogeneità.

4. Valutare il grado di connessione fra i caratteri X e Y utilizzando un opportuno indice normalizzato.

$\chi_N = 0.4122$. Sussiste un livello di connessione medio/basso.

ESERCIZIO 2.

Il prospetto seguente riporta la serie storica dei prezzi di due beni A e B fra il 2014 e il 2017:

anno	p_A	p_B	$NIBFA(2014)$
2014	100	200	1
2015	105	205	1.05
2016	110	220	1.1
2017	120	190	1.2

1. Costruire la serie dei numeri indici a base fissa dei prezzi con base 2014 per il bene A .

2. Calcolare il tasso medio di variazione per il prezzo del bene A nell'intervallo 2014-2017 e commentare il risultato ottenuto.

$\sqrt[3]{I_{2014}^F} = 1.0627$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del 6.27%. La grandezza oggetto di studio ha subito un incremento medio annuo tra il 2014 e il 2017 del 6.27%.

3. Sapendo che le quantità ${}_{2014}q_A = 100$ e ${}_{2014}q_B = 200$ e che ${}_{2016}q_A = 150$ e ${}_{2016}q_B = 180$, calcolare l'indice di Fisher dei prezzi al 2016 con base al 2014.

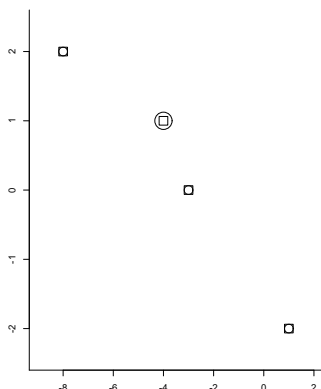
${}_{2014}^p I_{2016}^F = 1.10$. La variazione relativa nel livello generale dei prezzi, secondo l'Indice di Fisher, è del 10%. Si osserva che i prezzi di entrambi i beni hanno subito un incremento del 10% nel periodo considerato.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolte le seguenti 5 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 5$

x_i	-8	-4	-3	-4	1
y_i	2	1	0	1	-2

1. Fornire la rappresentazione grafica delle coppie di punti e della funzione di regressione di Y dato X .



2. Calcolare, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X^2$$

$$\hat{a} = -1.2427, \hat{b} = -0.4563, \text{ indice adattamento } M(\hat{E}^2) = 0.1243 \quad \hat{c} = 0.03369, \text{ indice adattamento } M(\hat{E}^2) = 0.9354$$

3. Confrontare gli indici di adattamento, commentando le eventuali differenze.

Il modello migliore è $Y^* = -1.2427 - 0.4563 \cdot X$ in quanto caratterizzato da residuo quadratico medio più basso.

4. Sulla base dei risultati precedenti e senza svolgere ulteriori conti, cosa è possibile dire sul valore di $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X .

5. Sulla base del miglior modello individuato al punto 3. prevedere il valore di Y in corrispondenza di $X = 5$.

$$Y^*(5) = -1.2427 - 0.4563 \cdot 5 = -3.5243$$

ESERCIZIO 4.

Con riferimento a un processo produttivo è noto che la proporzione di componenti difettose nella popolazione è pari a 0.12.

1. Dopo avere estratto con reimmissione un campione di 4 componenti, quale è la probabilità che non più di una componente risulti difettosa?

Sia X la variabile casuale 'n. di difettosi nelle 4 estrazioni con reimmissione'. Abbiamo $X \sim \text{Bin}(n = 4, p = 0.12)$.

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) = \\ &= \binom{4}{0} 0.12^0 \cdot 0.88^4 + \binom{4}{1} 0.12^1 \cdot 0.88^3 = \\ &= 0.5997 + 0.3271 = 0.9268. \end{aligned}$$

2. Calcolare il valore atteso e la varianza del numero di pezzi difettosi per un campione di 100 componenti.

Sia X la variabile casuale 'n. di difettosi nelle 100 estrazioni con reimmissione'. Abbiamo $X \sim \text{Bin}(n = 100, p = 0.12)$.
 $E(X) = n \cdot p = 100 \cdot 0.12 = 12$ $Var(X) = n \cdot p \cdot (1 - p) = 100 \cdot 0.12 \cdot 0.88 = 10.56$

3. Con riferimento al campione di 100 componenti del punto precedente, calcolare la probabilità che la quota di pezzi difettosi sia inferiore al 19%.

Dal momento che $np = 12 > 5$ ed $n(1 - p) = 88 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 12, \sigma^2 = 10.56)$.

Si ottiene

$$\begin{aligned} P(X \leq 19) &= P\left(\frac{X - \mu}{\sigma} \leq \frac{19 - \mu}{\sigma}\right) = P\left(Z \leq \frac{19 - 12}{3.2496}\right) = \\ &= P(Z \leq 2.15) = 0.9842. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 10.07.2018 (Tema 299-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

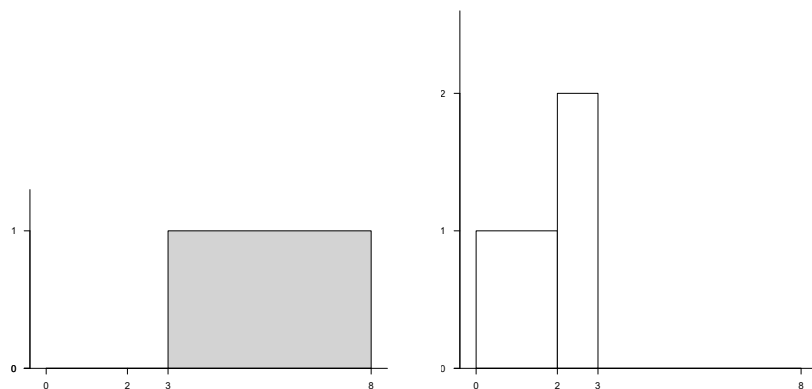
Studente _____ matricola _____

ESERCIZIO 1.

Data la seguente tabella a doppia entrata che riporta la distribuzione congiunta dei caratteri X = 'genere' e Y = 'numero di visite mediche specialistiche fatte in un anno' riferite a 9 persone:

$X \setminus Y$	0 + 2	2 + 3	3 + 8
maschio	0	0	5
femmina	2	2	0

1. Rappresentare graficamente le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.



2. Si riassume la marginale X con opportuni indici di posizione e di variabilità.

$Moda(X) = maschio$; $F_N(X) = 0.8889$. Sussiste un elevato livello di eterogeneità. La moda non è rappresentativa della distribuzione.

3. Si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 5.5$ Indice di variabilità: $CV(Y|x_1) = 0$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 1.75$ Indice di variabilità: $CV(Y|x_2) = 0.4286$

Nel campione considerato si osserva un numero medio di visite specialistiche per i maschi superiore a quello delle femmine. La distribuzione del numero medio di visite specialistiche per le femmine presenta un livello di variabilità superiore a quello dei maschi.

4. Senza fare calcoli e giustificando la risposta, si indichi il valore dell'indice χ^2 (assoluto) e dell'indice χ_N normalizzato.

$\chi^2 = \chi_{\max}^2 = 9$ e $\chi_N = 1$. Sussiste dipendenza funzionale di X da Y essendo le distribuzioni condizionate $X|y_j$ degeneri.

ESERCIZIO 2.

Il prospetto seguente riporta i prezzi medi annui di 2 beni di uso quotidiano nel periodo 2012-2016:

t	bene A	bene B	$NIBF_A(2013)$	$NIBM_B$
2012	9	27	0.9	
2013	10	20	1	0.7407
2014	12	27	1.2	1.3500
2015	14	35	1.4	1.2963
2016	15	30	1.5	0.8571

1. Costruire la serie dei numeri indici a base fissa dei prezzi con base 2013 per il bene A e la serie dei numeri indici a base mobile dei prezzi per il bene B.

2. Calcolare il tasso medio di variazione per il prezzo del bene A tra il 2013 e il 2016, interpretando il risultato.

$\sqrt[3]{\frac{15}{10}} = \sqrt[3]{1.5} = 1.1447$, numero indice a base mobile medio, a cui corrisponde un incremento medio annuo del 14.47%.

3. Nota la quantità $Aq_{2013} = 150$, si determini la quantità Bq_{2013} in modo che l'indice dei prezzi di Laspeyres al tempo 2016 con base 2013 risulti pari a 1.5.

Dal momento che i prezzi relativi riferiti ai prodotti A e B risultano

$$\frac{AP_{2016}}{AP_{2013}} = \frac{15}{10} = 1.5 \quad \text{e} \quad \frac{BP_{2016}}{BP_{2013}} = \frac{30}{20} = 1.5$$

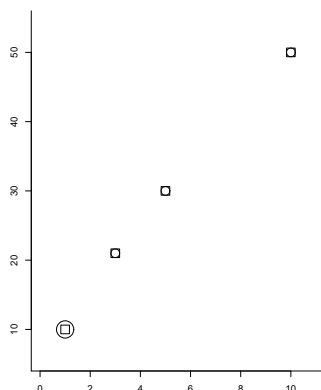
la corrispondente variabile statistica risulta degenera e, quindi, $\frac{p}{2013} I_{2016}^L = 1.5$ per qualsiasi quantità $Bq_{2013} \geq 0$.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y per i quali si vuole studiare la dipendenza di Y da X :

x_i	1	1	3	5	10
y_i	10	10	21	30	50

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione di Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 6.4143, \hat{b} = 4.4464, \text{ indice adattamento } M(\hat{E}^2) = 1.1279 \quad \hat{c} = 5.3897, \text{ indice adattamento } M(\hat{E}^2) = 18.0691$$

3. Si indichi quale modello presenta il migliore adattamento ai dati.

Il modello migliore è $Y^* = 6.4143 + 4.4464 \cdot X$ in quanto caratterizzato da residuo quadratico medio più basso.

4. Si dica, motivando teoricamente la risposta, se $\eta_{Y|X}^2 \neq \eta_{X|Y}^2$.

$\eta_{Y|X}^2 = \eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale bijectiva.

ESERCIZIO 4.

Un'urna è composta da 6 gettoni gialli e 4 gettoni blu. Si consideri l'esperimento casuale 'estrazione con reimmissione di 2 gettoni dall'urna'.

1. Si costruisca lo spazio probabilistico Ω associato all'esperimento.

Indicati con G_i e B_i gli eventi estrazione di gettone giallo e blu nella i -esima estrazione ($i = 1, 2$), lo spazio probabilistico risulta composto dai seguenti eventi elementari:

$$\Omega = \{G_1G_2, G_1B_2, B_1G_2, B_1B_2\},$$

a cui corrispondono le probabilità

$$\{0.6^2 = 0.36, 0.6 \cdot 0.4 = 0.24, 0.24, 0.4^2 = 0.16\}$$

tenendo conto che le estrazioni sono con reimmissione (prove indipendenti) e che $P(G_i) = P(G) = 0.6$, $P(B_i) = P(B) = 0.4$, $i = 1, 2$.

2. Un gioco, per cui è prevista una posta di 1 euro, consiste nell'estrazione con reimmissione di 2 gettoni dall'urna e comporta l'obbligo di pagare altri 4 euro per ogni gettone giallo che si presenta, mentre si vincono 8 euro per ogni gettone blu che si presenta. Si costruisca la variabile casuale G = 'guadagno del giocatore' e si stabilisca se il gioco è equo.

La variabile casuale guadagno

evento	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
$2G$	-9	0.36	-3.24
$1G1B$	3	0.48	1.44
$2B$	15	0.16	2.40
		1.00	0.60

ha valore atteso positivo pari a 0.60 e, quindi, il gioco è favorevole al giocatore.

3. Si calcoli la probabilità di ottenere almeno 52 gettoni gialli in 100 estrazioni con reimmissione dall'urna.

Indicata con X la variabile casuale 'numero di gettoni gialli' nelle 100 estrazioni con reimmissione dall'urna risulta $X \sim \text{Bin}(n = 100, p = 0.6)$. $M(X) = 60$; $\text{Var}(X) = 24$. Dal momento che $np = 60 > 5$ ed $n(1 - p) = 40 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 60, \sigma^2 = 24).$$

Si ottiene

$$P(X > 52) = P\left(\frac{X - \mu}{\sigma} > \frac{52 - \mu}{\sigma}\right) = P\left(Z > \frac{52 - 60}{4.899}\right) = P(Z > -1.63) = P(Z < 1.63) = P(Z \leq 1.63) = 0.9484.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 23.08.2018 (Tema 300-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta le distribuzioni relative condizionate della variabile $Y = \text{Numero di ingressi al museo nell'ultimo mese}$ rispetto alla variabile $X = \text{Zona di provenienza del visitatore}$, ricavata a seguito di una indagine campionaria:

$Y X$	0 ÷ 2	2 ÷ 7	7 ÷ 10	
$NORD$	0.2	0.6	0.2	1
$CENTRO$	0.2	0.6	0.2	1
SUD	0.2	0.6	0.2	1

1. Senza effettuare calcoli, si dica se è possibile concludere che esiste indipendenza stocastica tra X e Y . Motivare la risposta.

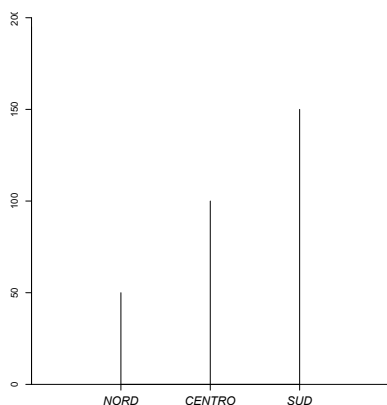
Tra le variabili X e Y sussiste indipendenza stocastica in quanto le distribuzioni condizionate $Y|x_i$ sono tra loro somiglianti. Abbiamo, infatti, $Y|x_1 \sim Y|x_2 \sim Y|x_3$.

2. Costruire la tabella delle frequenze congiunte sapendo che 50 dei visitatori intervistati provengono dal $NORD$, 100 dal $CENTRO$ e 150 dal SUD .

Ricordando che per la generica frequenza condizionata relativa vale $f(y_j|x_i) = \frac{n_{ij}}{n_{i\bullet}}$ abbiamo che $n_{ij} = f(y_j|x_i) \cdot n_{i\bullet}$ e

$X \setminus Y$	0 ÷ 2	2 ÷ 7	7 ÷ 10	$n_{i\bullet}$
$NORD$	10	30	10	50
$CENTRO$	20	60	20	100
SUD	30	90	30	150
$n_{\bullet j}$	60	180	60	300

3. Si rappresenti graficamente la variabile X e se ne calcoli moda ed eterogeneità.



$$\text{Moda}(X) = SUD \quad F_N = 0.7113 \text{ (livello medio alto di eterogeneità)}$$

4. Si confrontino le distribuzioni condizionate $Y|X = NORD$ e $Y|X = SUD$ con opportuni indici di posizione e di variabilità.

$Y|X = NORD$ Indice di posizione: $M(Y|X = NORD) = 4.6$ Indice di variabilità: $CV(Y|X = NORD) = 0.5163$

$Y|X = SUD$ Indice di posizione: $M(Y|X = SUD) = 4.6$ Indice di variabilità: $CV(Y|X = SUD) = 0.5163$

Dal momento che sussiste indipendenza stocastica le distribuzioni sono caratterizzate dal medesimo valore medio e dallo stesso livello di variabilità.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	$NIBF(\text{base} = 2014)$	$NIBM$
2012	0.89	
2013	0.89	1
2014	1	1.1236
2015	0.5	0.5
2016	0.5	1

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli il numero indice al tempo 2013 con base 2016 e si commenti il risultato ottenuto.

${}_{2016}I_{2013} = 1.78$ La grandezza oggetto di studio nel 2013 era superiore del 78% rispetto al 2016.

3. Si calcoli il tasso medio di variazione dal 2012 al 2016 e si commenti il risultato ottenuto.

$\sqrt[4]{{}_{2012}I_{2016}} = 0.8658$, numero indice a base mobile medio, a cui corrisponde una variazione relativa media annua del -13.42%.

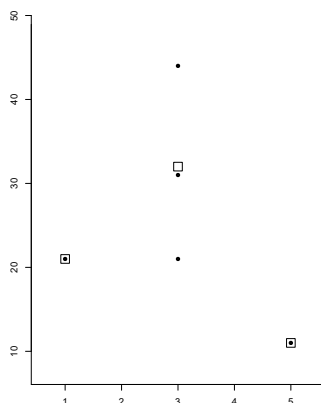
La grandezza oggetto di studio ha subito un decremento medio annuo tra il 2012 e il 2016 del 13.42%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 capofamiglia con riferimento a due caratteri quantitativi X = numero dei componenti della famiglia e Y = spesa mensile per musei e mostre (in euro).

x_i	1	3	3	3	5
y_i	21	31	44	21	11

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando la spesa in funzione del numero dei componenti della famiglia.



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$$\hat{a} = 33.1, \hat{b} = -2.5, \text{ indice adattamento } \rho^2 = 0.0802 \quad \hat{c} = 25.6, \text{ indice adattamento } R^2 = 0$$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 33.1 - 2.5 \cdot X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Senza effettuare calcoli e motivando la risposta, si stabilisca: a) l'intervallo di possibili valori assumibili da $\eta_{Y|X}^2$; b) se $\eta_{X|Y}^2 < 1$.

$\rho^2 < \eta_{Y|X}^2 < 1$ in quanto le medie condizionate non sono allineate e non sussiste dipendenza funzionale di Y da X dal momento che la distribuzione condizionata $Y|x = 3$ non è degenere.

$\rho^2 < \eta_{X|Y}^2 < 1$ in quanto le medie condizionate non sono allineate e non sussiste dipendenza funzionale di X da Y dal momento che le distribuzioni condizionate $X|y = 21$ non è degenere.

ESERCIZIO 4.

La retribuzione media mensile lorda di un laureato in Economia è distribuita come una Normale di media $\mu = 2300$ euro e varianza $\sigma^2 = 10000$ euro.

1. Quale percentuale di laureati guadagna meno di 2200 euro al mese?

Indicata con $X \sim N(\mu = 2300, \sigma^2 = 10000)$ la variabile casuale che descrive la retribuzione media mensile lorda, occorre, in primo luogo, determinare la seguente probabilità: $P(X < 2200)$. Abbiamo

$$\begin{aligned} P(X < 2200) &= P\left(\frac{X - \mu}{\sigma} < \frac{2200 - \mu}{\sigma}\right) = P\left(Z < \frac{2200 - 2300}{100}\right) = \\ &= P(Z < -1) = P(Z > 1) = 1 - P(Z \leq 1) = 1 - 0.8413 = 0.1587. \end{aligned}$$

La percentuale di laureati che guadagna meno di 2200 euro al mese risulta 15.87%.

2. Quale è il reddito limite mensile che definisce il 10% inferiore della retribuzione dei laureati in Economia?

Il reddito limite mensile che definisce il 10% inferiore della retribuzione dei laureati in Economia corrisponde al percentile di ordine 0.10 della distribuzione di X .

$$\{x : P(X \leq x) = 0.10\} = \left\{x : P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = 0.10\right\} = \left\{x : P\left(Z \leq \frac{x - 2300}{100}\right) = 0.10\right\}.$$

Dalla lettura inversa della Tavola possiamo ricavare il percentile di ordine 0.90, che risulta 1.28 (infatti $P(Z \leq 1.28) = 0.90$). Tenendo ora conto che la distribuzione della variabile casuale Normale è simmetrica, abbiamo, con riferimento al percentile di ordine 0.10, $z = \frac{x - 2300}{100} = -1.28$ da cui segue $x = 2172$.

3. Si calcoli la probabilità che su 10 laureati selezionati casualmente almeno uno guadagni meno di 2200 euro al mese.

Sia X la variabile casuale 'n. di laureati con guadagno inferiore a 2200 euro tra i 10 selezionati'. Dal momento che la popolazione dei laureati ha dimensione molto elevata, abbiamo $X \sim \text{Bin}(n = 10, p = 0.1587)$.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = \\ &= 1 - \binom{10}{0} 0.1587^0 \cdot 0.8413^{10} = \\ &= 1 - 0.1777 = 0.8223. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 06.09.2018 (Tema 300)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

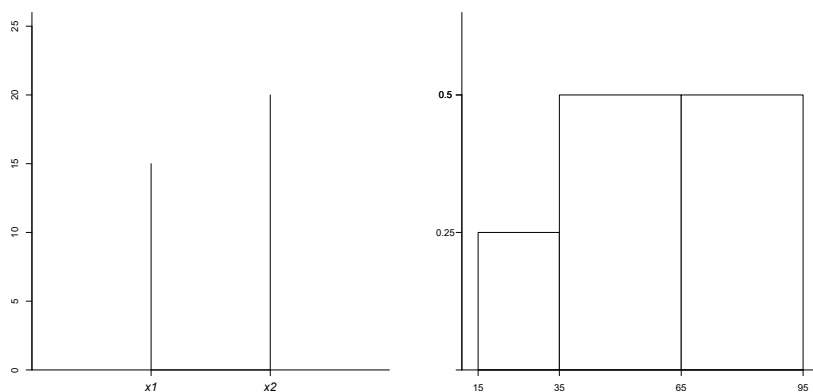
Studente _____ matricola _____

ESERCIZIO 1.

Con riferimento a due teatri italiani (x_1 = Teatro alla Scala, x_2 = Teatro Regio di Torino) si riporta nella seguente tabella la distribuzione della variabile Y 'età dello spettatore':

$X \setminus Y$	15 - 35	35 - 65	65 - 95
x_1	0	10	5
x_2	5	5	10

1. Si rappresentino graficamente le distribuzioni marginali di X e di Y .



2. Si confronti la distribuzione delle età degli spettatori nei due teatri attraverso opportuni indici di posizione e di variabilità, commentando i risultati ottenuti.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 60$ Indice di variabilità: $CV(Y|x_1) = 0.2357$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 58.75$ Indice di variabilità: $CV(Y|x_2) = 0.3917$

L'età degli spettatori del Teatro alla Scala risulta in media leggermente più elevata e meno variabile di quella degli spettatori del Teatro Regio di Torino.

3. Si determini il valore dell'indice χ^2 normalizzato e si commenti il risultato ottenuto.

$\chi_N^2 = 0.2222$; $\chi_N = 0.4714$. Sussiste un livello medio di connessione tra le due variabili.

ESERCIZIO 2.

Il prospetto seguente riporta la serie storica delle quotazioni dei titoli azionari x_t di una società calcistica italiana in cinque giornate del mese di luglio 2018:

t	x_t	$NIBF(12)$	$NIBM$
12	65	1	
13	70	1.0769	1.0769
14	80	1.2308	1.1429
15	85	1.3077	1.0625
16	85	1.3077	1

1. Si costruiscano la serie dei numeri indici a base fissa con base $t = 12$ e la serie dei numeri indici a base mobile.

2. Si calcoli il tasso medio di variazione delle quotazioni tra $t = 12$ e $t = 16$ e si interpreti il valore ottenuto.

$\sqrt[6]{12I_{16}} = \sqrt[6]{\frac{85}{65}} = \sqrt[6]{1.3077} = 1.0694$, numero indice a base mobile medio, a cui corrisponde un incremento medio annuo del 6.94%.

3. Sapendo che il numero indice a base mobile per il tempo $t = 17$ è risultato pari a 0.95, si determini il valore del titolo x_{17} .

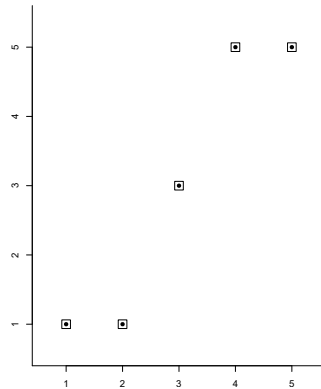
$$x_{17} = x_{16} \cdot {}_{16}I_{17} = 85 \cdot 0.95 = 80.75.$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	3	4	5
y_i	1	1	3	5	5

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione di Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot \ln(X) \quad II) Y^* = c$$

$\hat{a} = 0.2705$, $\hat{b} = 2.8506$, indice adattamento $\rho^2 = 0.8205$ $\hat{c} = 3$, indice adattamento $R^2 = 0$

3. Si indichi quale modello presenta il migliore adattamento ai dati.

Il modello migliore è $Y^* = 0.2705 + 2.8506 \cdot \ln X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Si stabilisca, senza effettuare calcoli e giustificando teoricamente la risposta, il valore assunto dagli indici $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

$\eta_{Y|X}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X .

$0 < \eta_{X|Y}^2 < 1$ in quanto non sussistono né indipendenza in media né dipendenza funzionale di X da Y .

Si osserva come non sia possibile stabilire la relazione esistente tra $\eta_{X|Y}^2$ e $\rho_{\ln(X), Y}^2$ non essendo $\ln(X)$ trasformazione lineare di X .

ESERCIZIO 4.

Una pasticceria produce Sachertorte il cui peso in grammi X si distribuisce secondo una variabile aleatoria Normale con media $\mu = 500$ e scarto quadratico medio $\sigma = 16$.

1. Si calcoli la probabilità che una torta pesi meno di 460 grammi.

Indicata con $X \sim N(\mu = 500, \sigma^2 = 256)$ la variabile casuale che descrive il peso in grammi delle Sachertorte, occorre determinare la seguente probabilità: $P(X < 460)$. Abbiamo

$$\begin{aligned} P(X < 460) &= P\left(\frac{X - \mu}{\sigma} < \frac{460 - \mu}{\sigma}\right) = P\left(Z < \frac{460 - 500}{16}\right) = \\ &= P(Z < -2.5) = P(Z > 2.5) = 1 - P(Z \leq 2.5) = 1 - 0.9938 = 0.0062. \end{aligned}$$

2. Sapendo che vengono esclusi dalla vendita i prodotti con peso inferiore a 460 grammi e che in una giornata sono state prodotte 10 torte, si calcoli la probabilità che tutte vengano messe in vendita, descrivendo la variabile aleatoria utilizzata.

La probabilità che una singola Sachertorte venga messa in vendita risulta pari a

$$p = 1 - P(X < 460) = 1 - 0.0062 = 0.9938$$

Ipotizzando, l'indipendenza nella produzione, il numero di Sachertorte messe in vendita risulta distribuito secondo una variabile casuale Binomiale con parametri $n = 10$ e $p = 0.9938$. Abbiamo, quindi

$$P(X = 10) = 0.9938^{10} = 0.9397$$

3. Il guadagno sul prezzo di vendita di ciascuna torta è di 10 euro. Si calcoli il guadagno medio atteso per la giornata (si ipotizzi che tutte le Sachertorte messe in vendita vengano acquistate).

Dal momento che il numero Y di Sachertorte messe in vendita risulta distribuito secondo una variabile casuale Binomiale $Y \sim \text{Bin}(n = 10, p = 0.9938)$, abbiamo $E(Y) = n \cdot p = 9.938$, numero atteso di Sachertorte vendute a cui corrisponde un guadagno G atteso pari a $E(G) = 10 \cdot 9.938 = 99.38$, tenendo conto della relazione $G = 10 \cdot Y$ tra guadagno e numero di torte vendute e che la media (valore atteso) è un operatore lineare.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 16.01.2019 (Tema 301)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

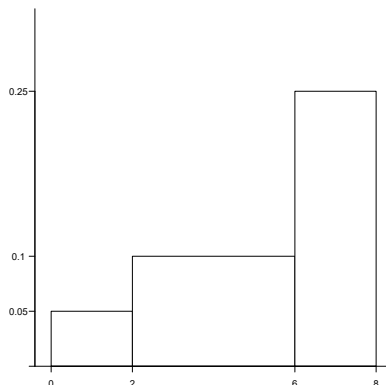
Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta le distribuzioni relative condizionate - ricavate a seguito di una indagine campionaria condotta su $n = 20$ lavoratori autonomi - della variabile $Y = \text{Reddito mensile}$, espressa in migliaia di Euro, rispetto alla variabile $X = \text{Genere}$:

$Y X$	0 - 2	2 - 6	6 - 8	
m	0.1	0.4	0.5	1
f	0.5	0.4	0.1	1

1. Dopo avere rappresentato graficamente la variabile $Y|X = m$, si dica se è possibile concludere che la distribuzione è simmetrica.



La distribuzione non è simmetrica perché le densità di frequenza della prima e terza classe non sono uguali tra loro.

2. Costruire la tabella delle frequenze congiunte sapendo che $n/2$ dei lavoratori intervistati sono *maschi* e $n/2$ sono *femmine*.

$X \setminus Y$	0 - 2	2 - 6	6 - 8	
m	1	4	5	10
f	5	4	1	10
	6	8	6	20

3. Confrontare, con opportuni indici di posizione e di variabilità, la distribuzione della variabile reddito nei *maschi* (m) e nelle *femmine* (f). Commentare il risultato ottenuto.

$Y|X = m$ Indice di posizione: $M(Y|X = m)$ valore 5.2 Indice di variabilità: $CV(Y|X = m)$ valore 0.3827

$Y|X = f$ Indice di posizione: $M(Y|X = f)$ valore 2.8 Indice di variabilità: $CV(Y|X = f)$ valore 0.7107

Il reddito medio mensile dei maschi è superiore a quello delle femmine. Il reddito delle femmine presenta un livello di variabilità superiore a quello dei maschi.

4. Calcolare l'indice di mutabilità normalizzato di X e commentare il risultato ottenuto.

indice $G_N(X) = F_N(X) = 1$. Situazione di massima eterogeneità in quanto sussiste equidistribuzione delle frequenze.

5. Senza effettuare calcoli, dire se è possibile concludere che esiste indipendenza stocastica tra X e Y .

Le variabili non sono stocasticamente indipendenti in quanto le distribuzioni condizionate $Y|X = m$ e $Y|X = f$ non sono somiglianti.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	$NIBM$	$NIBF(base = 2014)$
2014	—	1
2015	1	1
2016	1	1
2017	0.9	0.9
2018	0.2668	0.2401

1. Sapendo che la grandezza oggetto di studio ha subito un decremento medio annuo del 30% dal 2014 al 2018, trovare il valore del $NIBM$ del 2018.

A un decremento medio annuo del 30% corrisponde una media geometrica dei numeri indici a base mobile pari a 0.7.

Occorre, quindi, risolvere l'equazione $0.7 = \sqrt[4]{1 \cdot 1 \cdot 0.9 \cdot {}_{2017}I_{2018}}$ da cui segue ${}_{2017}I_{2018} = \frac{0.7^4}{0.9} = 0.2668$.

2. Ricostruire la serie dei $NIBF$ con base 2014.

3. Calcolare il valore del bene nel 2015, sapendo che nel 2017 è risultato pari a 33.

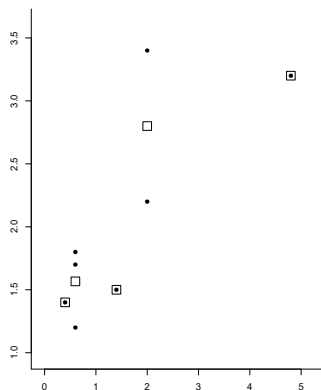
Dal momento che ${}_{2015}I_{2017} = \frac{x_{2017}}{x_{2015}} = 0.9$ segue che $x_{2015} = \frac{33}{0.9} = 36.6667$.

ESERCIZIO 3.

Su un insieme di 8 negozi vengono rilevate le seguenti variabili: Y = fatturato medio giornaliero (in migliaia di Euro), X = spesa annua di riscaldamento (in migliaia di Euro) e L = localizzazione del negozio (c = centrale, p = periferico). I dati sono riportati di seguito:

x_i	0.6	0.6	2	1.4	4.8	0.4	0.6	2
y_i	1.2	1.8	2.2	1.5	3.2	1.4	1.7	3.4
l_i	c	c	p	c	p	c	c	p

1. Rappresentare graficamente le coppie dei dati (x_i, y_i) e la funzione di regressione ipotizzando il fatturato in funzione della spesa di riscaldamento.



2. Calcolare, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d \cdot W$$

con W variabile dummy, indicatrice della localizzazione periferica del negozio.

$\hat{a} = 1.3668$, $\hat{b} = 0.4407$, indice adattamento $\rho^2 = 0.6079$ $\hat{c} = 1.52$, $\hat{d} = 1.4133$, indice adattamento $\rho^2 = 0.7803$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 1.52 + 1.4133 \cdot W$ in quanto caratterizzato da indice di adattamento più elevato.

4. Fornire una interpretazione delle stime dei coefficienti del modello II).

$\hat{c} = M(Y|W = c) = 1.52$, $\hat{d} = M(Y|W = p) - M(Y|W = c) = 1.4133$

5. Senza effettuare calcoli e motivando la risposta, stabilire l'intervallo di possibili valori assumibili da $\eta_{Y|X}^2$.

$0.6079 < \eta_{Y|X}^2 < 1$ in quanto le medie condizionate non sono allineate e le distribuzioni condizionate $Y|X$ non sono degeneri.

ESERCIZIO 4.

Il valore di un portafoglio titoli, costituito da un paniere di azioni, si distribuisce come una variabile aleatoria X con distribuzione Normale, di media $\mu = 100$ euro e varianza $\sigma^2 = 100$ euro². Il portafoglio titoli è considerato *svantaggioso* quando il suo valore è inferiore a 82 euro.

1. Calcolare la probabilità che il valore del portafoglio diventi *svantaggioso*.

Indicata con $X \sim N(\mu = 100, \sigma^2 = 100)$ la variabile casuale che descrive il valore del portafoglio titoli, occorre determinare la seguente probabilità: $P(X < 82)$. Abbiamo

$$\begin{aligned} P(X < 82) &= P\left(\frac{X - \mu}{\sigma} < \frac{82 - \mu}{\sigma}\right) = P\left(Z < \frac{82 - 100}{10}\right) = \\ &= P(Z < -1.8) = P(Z > 1.8) = 1 - P(Z \leq 1.8) = 1 - 0.9641 = 0.0359. \end{aligned}$$

2. Indicare quale è il valore limite che definisce il 2% superiore dei valori del portafoglio titoli.

Il reddito limite mensile che definisce il 2% superiore dei valori del portafoglio titoli corrisponde al percentile di ordine 0.98 della distribuzione di X .

$$\{x : P(X \leq x) = 0.98\} = \left\{x : P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = 0.98\right\} = \left\{x : P\left(Z \leq \frac{x - 100}{10}\right) = 0.98\right\}.$$

Dalla lettura inversa della Tavola possiamo ricavare il percentile di ordine 0.98, che risulta 2.05 (infatti $P(Z \leq 2.05) = 0.98$). Abbiamo, quindi, $z = \frac{x - 100}{10} = 2.05$ da cui segue $x = 120.5$.

3. Assumendo di avere investito in 5 diversi portafogli titoli, tra loro indipendenti ma con le stesse caratteristiche del portafoglio titoli definito sopra, calcolare la probabilità che al più uno sia *svantaggioso*.

Sia X la variabile casuale 'n. di portafogli svantaggiosi tra i 5 in cui si è investito'. Abbiamo $X \sim \text{Bin}(n = 5, p = 0.0359)$.

$$\begin{aligned} P(X \leq 1) &= P(X = 0) + P(X = 1) = \\ &= \binom{5}{0} 0.0359^0 \cdot 0.9641^5 + \binom{5}{1} 0.0359^1 \cdot 0.9641^4 = \\ &= 0.8328 + 0.1552 = 0.9880. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 30.01.2019 (Tema 302)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

La seguente tabella riporta i dati di 9 località sciistiche italiane per le variabili X : tipologia di neve (con modalità a = artificiale e f = fresca) e Y : % di impianti aperti, nella settimana di Natale.

x_i	a	a	a	a	a	f	f	f	f
y_i	99	89	94	84	94	100	81	89	91

1. Si confronti la variabilità della variabile Y nei gruppi definiti dalla variabile X .

indice variabilità $Y|X = a$ CV valore 0.0554

indice variabilità $Y|X = f$ CV valore 0.0749

La percentuale di impianti aperti risulta caratterizzata da una variabilità più elevata nelle località con neve fresca.

2. Si calcolino la varianza between e la varianza within di Y nei gruppi definiti dalla variabile X . Si commentino i risultati ottenuti.

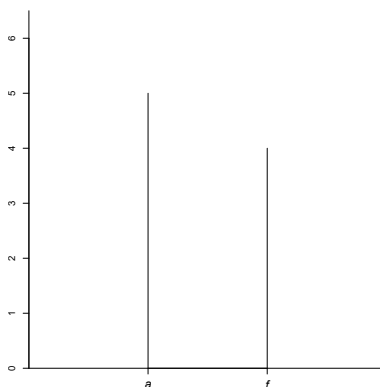
$$\sigma_{Between}^2 = \frac{1}{9} \left\{ (\mu_a - \mu)^2 \cdot 5 + (\mu_f - \mu)^2 \cdot 4 \right\} = \frac{1}{9} \left(\mu_a^2 \cdot 5 + \mu_f^2 \cdot 4 \right) - \mu^2 = \frac{1}{9} (92^2 \cdot 5 + 90.25^2 \cdot 4) - 91.2222^2 = 0.7562$$

$$\sigma_{Within}^2 = \frac{1}{9} \left(\sigma_a^2 \cdot 5 + \sigma_f^2 \cdot 4 \right) = \frac{1}{9} (26 \cdot 5 + 45.6875 \cdot 4) = 34.75.$$

Il rapporto di correlazione $\eta_{Y|tipologia\ neve}^2$ risulta pari a 0.0213.

Sussiste, quindi, un livello di dipendenza molto basso del numero di impianti aperti rispetto alla tipologia della neve.

3. Si rappresenti graficamente la distribuzione di frequenze della variabile X e si calcolino un indice di posizione e uno di variabilità.



X Indice di posizione: Moda valore *artificiale* Indice di variabilità: F_N valore 0.8889.

La tipologia della neve risulta caratterizzata da un elevato livello di eterogeneità.

4. Costruita la tabella a doppia entrata (X, Y) , con Y riclassificata in 2 classi di uguale ampiezza (pari a 10), a partire da 80, si indichi se le variabili possono ritenersi stocasticamente indipendenti.

$X \backslash Y$	80 - 90	90 - 100
a	2	3
f	2	2

Le due variabili non possono ritenersi stocasticamente indipendenti dal momento che le distribuzioni condizionate non sono tra loro somiglianti.

ESERCIZIO 2.

Si consideri la seguente serie dei numeri indici a base fissa, con base 2009:

<i>anno</i>	<i>NIBF</i>	<i>NIBM</i>
2009	1	—
2010	0.90	0.9000
2011	0.93	1.0333
2012	1.02	1.0968
2013	1.09	1.0686

1. Si costruisca la corrispondente serie dei numeri indici a base mobile.

2. Sapendo che $x_{2011} = 26$, si calcoli x_{2013} .

$$x_{2013} = x_{2011} \cdot {}_{2011}I_{2013} = 26 \cdot \frac{1.09}{0.93} = 30.4731$$

3. Si calcoli il tasso medio di variazione dal 2011 al 2013 e si commenti il risultato ottenuto.

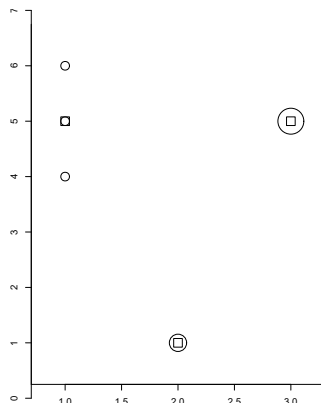
tasso medio di variazione 0.08261. Tra il 2009 e il 2013 si è osservato un incremento medio annuo del 8.26%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati su 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	3	1	2	3	2	3	1
y_i	6	5	5	1	5	1	5	4

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione, individuando il modello migliore:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d \cdot X^2$$

$\hat{a} = 4$, $\hat{b} = 0$, indice adattamento $\rho^2 = 0$ $\hat{c} = 3.7077$, $\hat{d} = 0.0615$, indice adattamento $\rho^2 = 0.0142$

Il modello migliore è $Y^* = 3.7077 + 0.0615X^2$.

3. Si interpretino i coefficienti del modello I).

In base al primo modello il valore atteso della Y non subisce variazioni al variare del livello della X e risulta pari a 4.

4. Si determinino i possibili valori del rapporto di correlazione $\eta_{Y|X}^2$ motivando la risposta.

$0.0142 < \eta_{Y|X}^2 < 1$ in quanto le medie condizionate non sono una parabola avente come asse l'asse delle ordinate e le distribuzioni condizionate $Y|X$ non sono degeneri.

5. Per il modello migliore, si fornisca la previsione in corrispondenza di $X = 5$.

$$Y^*(5) = 3.7077 + 0.0615 \cdot 5^2 = 3.7077 + 0.0615 \cdot 25 = 5.2462.$$

ESERCIZIO 4.

Da una ricerca condotta da una rivista automobilistica è risultato che le auto fuoristrada (tipo SUV, Van e Pickup) sono più soggette a ribaltamento se coinvolte in incidenti stradali rispetto alle altre auto. Nel 1997 il 24% degli incidenti ha comportato un ribaltamento di auto; di questi il 15 % ha coinvolto dei fuoristrada, mentre negli incidenti senza ribaltamento la percentuale di fuoristrada coinvolti scende al 5 %.

1. Si calcoli la probabilità che in un incidente, estratto a caso tra i sinistri del 1997, sia coinvolto un fuoristrada.

Indicati con F ed $\bar{F}$ gli eventi 'auto fuoristrada' e 'altre auto' e con R e $\bar{R}$ gli eventi 'incidente con ribaltamento' e 'incidente senza ribaltamento' sono disponibili le seguenti informazioni:

$$P(R) = 0.24, P(F|R) = 0.15, P(F|\bar{R}) = 0.05.$$

Applicando la legge delle probabilità totali abbiamo:

$$\begin{aligned} P(F) &= P(F \cap (R \cup \bar{R})) = P((F \cap R) \cup (F \cap \bar{R})) = P(F \cap R) + P(F \cap \bar{R}) = P(R) \cdot P(F|R) + P(\bar{R}) \cdot P(F|\bar{R}) = \\ &= P(R) \cdot P(F|R) + (1 - P(R)) \cdot P(F|\bar{R}) = 0.24 \cdot 0.15 + 0.76 \cdot 0.05 = 0.036 + 0.038 = 0.074 \end{aligned}$$

2. Si calcoli la probabilità che un incidente, estratto a caso tra i sinistri del 1997, sia con ribaltamento sapendo che in esso è coinvolto un fuoristrada.

$$P(R|F) = \frac{P(R \cap F)}{P(F)} = \frac{0.036}{0.074} = 0.4865$$

3. Si calcoli la probabilità che su 10 incidenti, estratti a caso tra i sinistri del 1997, in al più 2 ci sia un ribaltamento di un fuoristrada.

Sia X la variabile casuale 'n. di incidenti con ribaltamento di un fuoristrada tra i 10 estratti'. Supponendo che la popolazione sia composta da un numero elevato di soggetti le estrazioni 'effettuate in maniera casuale' possono ritenersi indipendenti. Abbiamo

$$X \sim \text{Bin}(n = 10, p = 0.036)$$

e

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) = \\ &= \binom{10}{0} 0.036^0 \cdot 0.964^{10} + \binom{10}{1} 0.036^1 \cdot 0.964^9 + \binom{10}{2} 0.036^2 \cdot 0.964^8 = 0.6931 + 0.2588 + 0.0435 = 0.9954. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 19.02.2019 (Tema 303)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

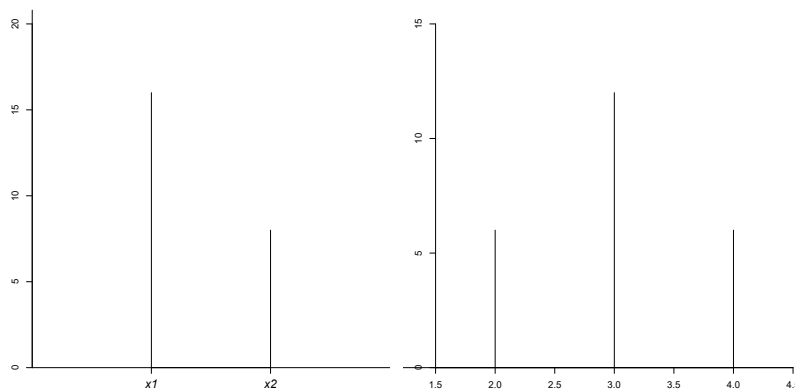
Il prospetto seguente riporta alcune informazioni sulla distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	2	3	4	
x_1	4	8	4	16
x_2	2	4	2	8
	6	12	6	24

1. Sapendo che la varianza within di Y rispetto ai gruppi definiti dalle categorie della variabile X è uguale alla varianza della distribuzione marginale di Y , si costruisca la distribuzione delle frequenze congiunte delle due variabili.

Se la varianza within è uguale alla varianza della distribuzione marginale, allora la varianza between è nulla e le medie condizionate sono tra loro uguali. Dal momento che la frequenza $n_{21} = 2$, $M(Y|x_2) = 3$ (con $Y|x_2$ distribuzione simmetrica). Affinché $M(Y|x_1) = 3$ occorre fissare $n_{11} = n_{13}$.

2. Si rappresentino graficamente le distribuzioni marginali e si confrontino le distribuzioni condizionate $Y|x_1$ e $Y|x_2$ con opportuni indici di posizione e di variabilità.



$Y|X = x_1$ Indice di posizione: $M(Y|X = x_1)$ valore 3 Indice di variabilità: $CV(Y|X = x_1)$ valore 0.2357

$Y|X = x_2$ Indice di posizione: $M(Y|X = x_2)$ valore 3 Indice di variabilità: $CV(Y|X = x_2)$ valore 0.2357

Si osserva che le variabili X e Y sono stocasticamente indipendenti; quindi, le distribuzioni condizionate saranno caratterizzate dallo stesso livello medio e da eguale variabilità.

3. Si determini il grado di connessione tra X e Y mediante il calcolo di un opportuno indice.

Dal momento che sussiste indipendenza stocastica vi è assenza di connessione

ESERCIZIO 2.

La quotazione del Bitcoin ha registrato negli ultimi 5 giorni le seguenti variazioni percentuali:

t	variaz.rel.%	NIBM	NIBF(base = 3)
1	—		1.0122
2	+11%	1.11	1.1236
3	-11%	0.89	1
4	+12%	1.12	1.12
5	-12%	0.88	0.9856

1. Si costruiscano la serie dei numeri indici a base mobile e la serie dei numeri indici a base fissa con base al tempo 3 per la quotazione del Bitcoin.

2. Si calcoli il numero indice al tempo 4 con base 2, interpretando il valore ottenuto.

${}_2I_4 = 0.9968$ Tra il tempo 2 e il tempo 4 si è registrato un decremento complessivo dello 0.32% della quotazione del Bitcoin.

3. Si calcoli la media aritmetica dei 4 numeri indici a base mobile ottenuti al punto 1 e si fornisca un intervallo di possibili valori per il tasso medio di variazione della quotazione del Bitcoin tra il tempo 1 e il tempo 5.

$\frac{1}{4}(1.11 + 0.89 + 1.12 + 0.88) = 1$. In base al criterio di scelta della media secondo Chisini la media obiettivo dei numeri indici a base mobile ai fini del calcolo della tasso medio di variazione è la media geometrica. Abbiamo quindi, dal momento che la variabile statistica costituita dai numeri indici a base mobile non è degenere:

$$I_{min} = 0.88 < \mu^{(0)} < \mu = 1$$

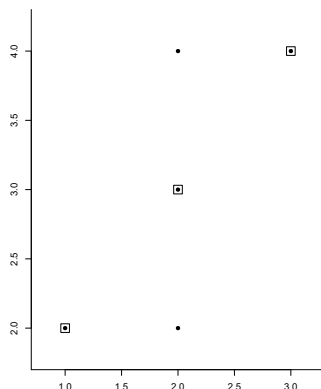
e

$$-12\% < \text{variazione relativa media percentuale} = \text{tasso medio di variazione} < 0\%$$

ESERCIZIO 3. Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	2	2	3
y_i	2	2	3	4	4

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$$\hat{a} = 1, \hat{b} = 1, \text{indice adattamento } \rho^2 = 0.5 \quad \hat{c} = 3, \text{indice adattamento } R^2 = 0$$

3. Si individui il modello migliore.

Il modello migliore è $Y^* = 1 + 1 \cdot X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Si indichino i valori assunti da $\eta_{Y|X}^2$ ed $\eta_{X|Y}^2$.

$\eta_{Y|X}^2 = \eta_{X|Y}^2 = \rho^2 = 0.5$ in quanto sia le medie condizionate $M(Y|X)$ che le medie condizionate $M(X|Y)$ sono allineate.

5. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 4$.

$$Y^*(4) = 1 + 1 \cdot 4 = 5.$$

ESERCIZIO 4.

L'urna U_1 contiene 71 palline bianche e 9 palline rosse; l'urna U_2 contiene 21 palline bianche e 29 palline rosse. Si consideri il seguente esperimento. Si lanciano 3 dadi regolari; se questi mostrano tutti la stessa faccia si effettuano 2 estrazioni senza reimmissione dall'urna U_1 altrimenti le 2 estrazioni vengono effettuate dall'urna U_2 .

1. Si calcoli la probabilità che le due palline estratte siano entrambe bianche.

Si osserva, in primo luogo, che la probabilità che i 3 dadi regolari mostrino tutti la stessa faccia risulta pari a $6 \cdot \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} = \frac{6}{216} = \frac{1}{36}$. Indicando con A l'evento 'i 3 dadi mostrano tutti la stessa faccia', abbiamo:

$$\begin{aligned} P(2B) &= P(A) \cdot P(2B|A) + P(\bar{A}) \cdot P(2B|\bar{A}) = \\ &= \frac{1}{36} \cdot \frac{71}{80} \cdot \frac{70}{79} + \frac{35}{36} \cdot \frac{21}{50} \cdot \frac{20}{49} = 0.02778 \cdot 0.7864 + 0.9722 \cdot 0.1714 = 0.02185 + 0.1666 = 0.1885 \end{aligned}$$

2. Si consideri un gioco per cui si paga una posta di 2Euro e si vincono 6Euro nel caso le 2 palline di cui all'esperimento sopra descritto siano entrambe bianche. Si stabilisca se il gioco è equo.

La variabile casuale guadagno

evento	x_i	$P(X = x_i)$	$x_i \cdot P(X = x_i)$
2B	-2	0.8115	-1.623
2B	4	0.1885	0.754
		1.0000	-0.869

ha valore atteso negativo pari a -0.869 e, quindi, il gioco è sfavorevole al banco.

3. Si calcoli la probabilità che effettuando 200 estrazioni con reimmissione dall'urna U_2 si ottengano almeno 86 palline bianche.

Indicata con X la variabile casuale 'numero palline bianche nelle 200 estrazioni con reimmissione dall'urna U_2 ' risulta $X \sim \text{Bin}(n = 200, p = 0.42)$. $M(X) = 84$; $\text{Var}(X) = 48.72$. Dal momento che $np = 84 > 5$ ed $n(1 - p) = 116 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 84, \sigma^2 = 48.72).$$

Si ottiene

$$P(X > 86) = P\left(\frac{X - \mu}{\sigma} > \frac{86 - \mu}{\sigma}\right) = P\left(Z > \frac{86 - 84}{6.98}\right) = P(Z > 0.29) = 1 - P(Z \leq 0.29) = 1 - 0.6141 = 0.3859.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 03.06.2019 (Tema 304-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

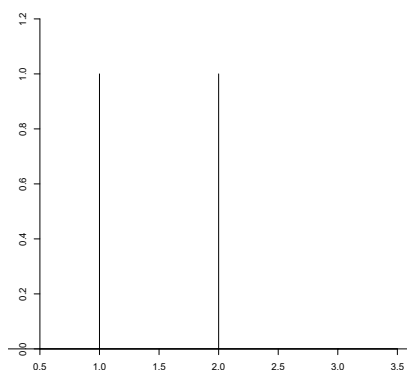
Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta la distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	1	2	3
x_1	1	1	0
x_2	0	1	1

1. Si rappresenti graficamente la distribuzione di frequenza della condizionata $Y|x_1$ e si confronti il livello di eterogeneità della distribuzione $X|y_2$ con quello della marginale X .



eterogeneità $X|y_2$: $G_N(X) = F_N(X) = 1$

eterogeneità X : $G_N(X) = F_N(X) = 1$

Situazione di massima eterogeneità in quanto sussiste equidistribuzione delle frequenze per entrambe le distribuzioni.

2. Si confrontino la distribuzione condizionata $Y|x_1$ e la marginale Y mediante opportuni indici di posizione e di variabilità.

$Y|x_1$ Indice di posizione: $M(Y|x_1)$ valore 1.5 Indice di variabilità: $CV(Y|x_1)$ valore 0.3333

Y Indice di posizione: $M(Y)$ valore 2 Indice di variabilità: $CV(Y)$ valore 0.3536

La distribuzione condizionata $Y|x_1$ è caratterizzata da un livello medio e da variabilità inferiori rispetto alla marginale Y .

3. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

indice di connessione χ_N valore 0.7071

Sussiste un livello elevato di connessione tra le due variabili.

ESERCIZIO 2.

La seguente tabella riporta la serie storica del fatturato annuo di un'azienda:

t	x_t	$NIBM$
2012	50	—
2013	55	1.10
2014	52.25	0.95
2015	62.7	1.20
2016	62.7	1.00

1. Si costruisca la serie dei numeri indici a base mobile e si calcoli, commentando il risultato, il valore di ${}_{2014}I_{2016}$, numero indice al tempo 2016 con base 2014.

${}_{2014}I_{2016} = 1.2$ Il fatturato ha subito un incremento complessivo del 20% tra il 2014 e il 2016.

2. Si calcoli il tasso medio di variazione dal 2012 al 2014 e si commenti il risultato ottenuto.

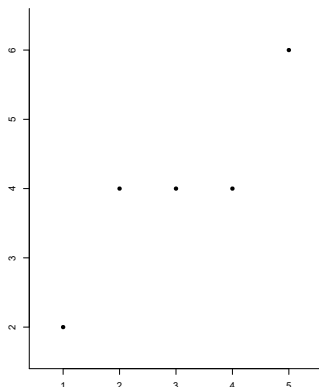
tasso medio di variazione 0.0223. Tra il 2012 e il 2014 si è osservato un incremento medio annuo del 2.23%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	2	3	4	5
y_i	2	4	4	4	6

1. Si rappresentino graficamente le coppie dei dati ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$\hat{a} = 1.6$, $\hat{b} = 0.8$, indice adattamento $M(\hat{E}^2) = 0.32$ $\hat{c} = 1.2364$, indice adattamento $M(\hat{E}^2) = 0.7855$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = 1.6 + 0.8X$.

4. Senza svolgere i conti, cosa è possibile dire con riferimento ai dati in esame riguardo alla relazione tra ρ^2 ed $\eta_{Y|X}^2$.

Vale la relazione $\rho^2 = 0.8 < \eta_{Y|X}^2 = 1$, in quanto sussiste dipendenza funzionale di Y da X .

5. Senza svolgere i conti, cosa è possibile dire con riferimento ai dati in esame riguardo alla relazione tra ρ^2 ed $\eta_{X|Y}^2$.

Vale la relazione $\rho^2 = \eta_{X|Y}^2 = 0.8$, in quanto le medie condizionate $M(X|Y)$ sono allineate.

ESERCIZIO 4.

L'urna $U1$ contiene 3 palline nere e 1 pallina rossa. L'urna $U2$ contiene 18 palline nere e 1 pallina rossa.

1. Si calcoli la probabilità che estraendo con reimmissione 800 palline dall'urna $U1$ almeno 200 siano rosse.

Indicata con X la variabile casuale 'numero palline rosse nelle 800 estrazioni con reimmissione dall'urna $U1$ ' risulta $X \sim \text{Bin}(n = 800, p = 0.25)$. $M(X) = 200$; $\text{Var}(X) = 150$. Dal momento che $np = 200 > 5$ ed $n(1 - p) = 600 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 200, \sigma^2 = 150).$$

Dal momento che la variabile casuale Normale è simmetrica rispetto alla media, si ottiene

$$P(X > 200) = 0.5.$$

2. Si calcoli la probabilità che estraendo con reimmissione 40 palline dall'urna $U2$ almeno una sia rossa.

Sia X la variabile casuale 'numero palline rosse nelle 40 estrazioni con reimmissione dall'urna $U2$ '. Abbiamo $X \sim \text{Bin}(n = 40, p = 0.0526)$.

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{40}{0} 0.0526^0 \cdot 0.9474^{40} = 1 - 0.1150 = 0.8850.$$

3. Si estraiga una pallina dall'urna $U1$ e la si metta nell'urna $U2$. Si calcoli la probabilità che estraendo con reimmissione 2 palline dall'urna $U2$ (modificata) si ottengano 2 palline nere.

Siano N l'evento 'estrazione di una pallina nera dall'urna $U1$ ', $X|N$ e $X|\bar{N}$ le variabile casuale 'n. di palline nere nelle 2 estrazioni con reimmissione dall'urna $U2$ ' condizionate al verificarsi degli eventi N e $\bar{N}$.

$$\begin{aligned} P(2 \text{ nere in 2 estraz. con reimmissione}) &= P(N) \cdot P\{(X|N) = 2\} + P(\bar{N}) \cdot P\{(X|\bar{N}) = 2\} = \\ &= \frac{3}{4} \cdot \left(\frac{19}{20}\right)^2 + \frac{1}{4} \cdot \left(\frac{18}{20}\right)^2 = 0.75 \cdot 0.9025 + 0.25 \cdot 0.81 = 0.8794. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 17.06.2019 (Tema 304)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

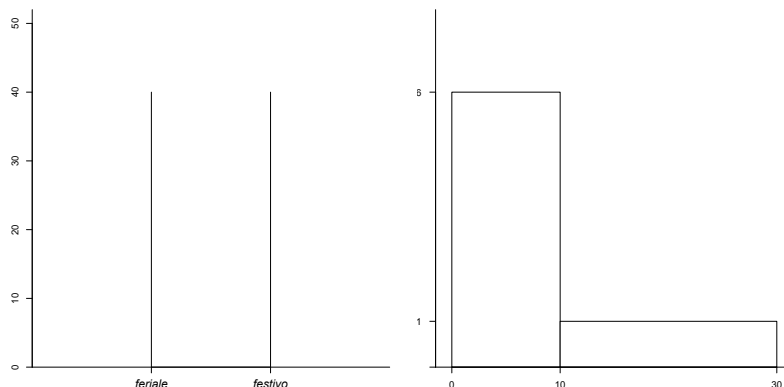
Studente _____ matricola _____

ESERCIZIO 1.

Per valutare la dipendenza fra costo orario (in euro) di una prestazione professionale (Y) e giorno di erogazione (X) è stata svolta una indagine di mercato su un campione di 80 committenze. Nella tabella sottostante sono riportate le frequenze congiunte di (X, Y) .

$X \setminus Y$	0 - 10	10 - 30
<i>feriale</i>	40	0
<i>festivo</i>	20	20

1. Fornire una opportuna rappresentazione grafica delle distribuzioni marginali di X e Y .



2. Calcolare il valore della varianza between e della varianza within per il carattere Y rispetto ai gruppi definiti dal carattere X . Commentare il risultato ottenuto.

$$\sigma_{Between}^2 = \frac{1}{80} \left\{ (\mu_{feriale} - \mu)^2 \cdot 40 + (\mu_{festivo} - \mu)^2 \cdot 40 \right\} = \frac{1}{80} \left(\mu_{feriale}^2 \cdot 40 + \mu_{festivo}^2 \cdot 40 \right) - \mu^2 = \frac{1}{80} (5^2 \cdot 40 + 12.5^2 \cdot 40) - 8.75^2 = 14.0625$$

$$\sigma_{Within}^2 = \frac{1}{80} \left(\sigma_{feriale}^2 \cdot 40 + \sigma_{festivo}^2 \cdot 40 \right) = \frac{1}{80} (0 \cdot 40 + 56.25 \cdot 40) = 28.125.$$

Il rapporto di correlazione $\eta_{Y|giorno\ erogazione}^2$ risulta pari a 0.3333.

Sussiste, quindi, un livello di dipendenza medio basso del costo della prestazione professionale rispetto al giorno di erogazione.

3. Calcolare un opportuno indice di eterogeneità per la variabile X e commentarne il valore.

indice $G_N(X) = F_N(X) = 1$ (situazione di massima eterogeneità in quanto sussiste equidistribuzione delle frequenze)

4. Valutare il grado di connessione fra i caratteri X e Y utilizzando un opportuno indice normalizzato.

indice di connessione χ_N valore 0.5774

Sussiste un livello medio di connessione tra le due variabili.

ESERCIZIO 2.

La tabella seguente riporta la serie dei numeri indice dei prezzi a base fissa (con anno base 2015) per due beni A e B registrati fra il 2015 e il 2018:

t	$NIBF_A$	$NIBF_B$	$NIBM_A$
2015	1	1	—
2016	1.05	1	1.050
2017	1.05	1.1	1.000
2018	1.1	1.1	1.048

1. Costruire la serie dei numeri indice a base mobile per il bene A .

2. Calcolare il tasso medio di variazione annuo del bene A nell'intervallo 2015-2018 e commentare il risultato ottenuto. tasso medio di variazione 0.03228. Tra il 2015 e il 2018 si è osservato un incremento medio annuo del 3.23%.

3. Sapendo che ${}_{2015}q_A = 100$ e ${}_{2015}q_B = 100$ sono le quantità vendute per i beni A e B nel 2015, calcolare l'indice di Laspeyres dei prezzi al 2016 con base al 2015 nell'ipotesi che il prezzo del bene B nell'anno base sia doppio rispetto a quello del bene A .

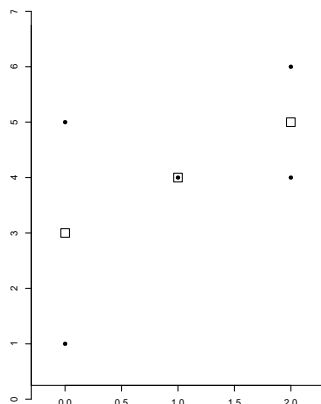
${}_{2015}^p I_{2016}^L = 1.0167$. Nel calcolo dell'indice come media aritmetica dei prezzi relativi, si tenga conto che il prodotto B ha peso doppio rispetto al prodotto A . Tra il 2015 e il 2016 si è registrato un incremento medio dei prezzi dell'1.67%.

ESERCIZIO 3.

Con riferimento allo studio della dipendenza tra due caratteri quantitativi X e Y , si sono raccolti i dati relativi alle seguenti 5 coppie di osservazioni (x_i, y_i) $i = 1, 2, \dots, 5$

x_i	0	0	1	2	2
y_i	1	5	4	4	6

1. Fornire la rappresentazione grafica delle coppie di punti e della funzione di regressione di Y dato X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$\hat{a} = 3$, $\hat{b} = 1$, indice adattamento $M(\hat{E}^2) = 2$ $\hat{c} = 2.6667$, indice adattamento $M(\hat{E}^2) = 6$

3. Confrontare gli indici di adattamento, commentando le eventuali differenze.

Il modello migliore è $Y^* = 3 + X$.

4. Sulla base dei risultati precedenti e senza svolgere ulteriori conti, cosa è possibile dire sui valori di $\eta_{Y|X}^2$?

$\eta_{Y|X}^2 = \rho^2$ in quanto le medie condizionate sono allineate.

5. Sulla base del miglior modello prevedere il valore di Y in corrispondenza di $X = 3$.

$$Y^*(3) = 3 + 1 \cdot 3 = 6.$$

ESERCIZIO 4.

In una comunità la percentuale di pazienti allergici a pollini e graminacee è nota essere pari a 0.2. Sapendo che la probabilità di risultare positivo ad un test se un individuo è allergico è del 0.8 mentre la probabilità di risultare negativo se un individuo non è allergico è pari a 0.6 si risponda alle seguenti domande:

1. Calcolare la probabilità di risultare positivo al test per un individuo selezionato casualmente.

Indicati con A ed $\bar{A}$ gli eventi 'allergico' e 'non allergico' e con $+$ e $-$ gli eventi 'positivo al test' e 'negativo al test' sono disponibili le seguenti informazioni:

$$P(A) = 0.2, P(+|A) = 0.8, P(-|\bar{A}) = 0.6.$$

Applicando la legge delle probabilità totali abbiamo:

$$\begin{aligned} P(+) &= P(+ \cap (A \cup \bar{A})) = P((+ \cap A) \cup (+ \cap \bar{A})) = P(+ \cap A) + P(+ \cap \bar{A}) = P(A) \cdot P(+|A) + P(\bar{A}) \cdot P(+|\bar{A}) = \\ &= P(A) \cdot P(+|A) + (1 - P(A)) \cdot (1 - P(-|\bar{A})) = 0.2 \cdot 0.8 + 0.8 \cdot 0.4 = 0.16 + 0.32 = 0.48 \end{aligned}$$

2. Calcolare la probabilità di essere allergico, dato che un paziente è risultato positivo al test.

$$P(A|+) = \frac{P(A \cap +)}{P(+)} = \frac{0.16}{0.48} = 0.3333$$

3. Estraendo quattro individui con reimmissione a caso dalla comunità, calcolare che almeno uno risulti allergico.

Sia X la variabile casuale 'n. di individui allergici tra i 4 estratti'. Abbiamo

$$X \sim \text{Bin}(n = 4, p = 0.2)$$

e

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{4}{0} 0.2^0 \cdot 0.8^4 = 1 - 0.4096 = 0.5904.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 11.07.2019 (Tema 304-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

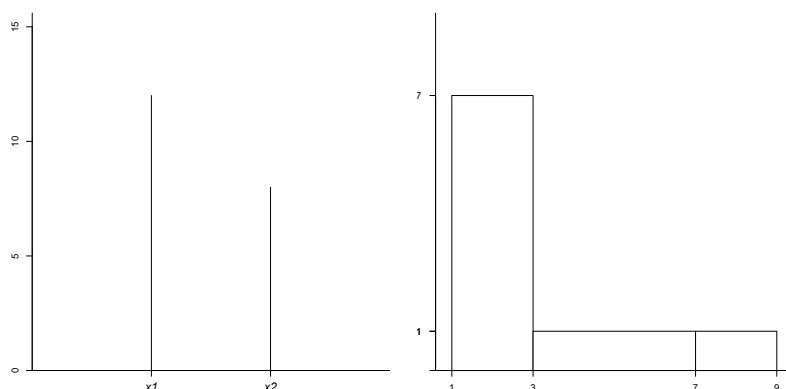
Il prospetto seguente riporta la distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	1 + 3	3 + 7	7 + 9	
x_1	12	0	0	12
x_2	2	4	2	8
	14	4	2	20

1. Si completi la distribuzione delle frequenze congiunte, sapendo che il numero totale di unità statistiche è 20, che $M(Y|x_1) = 2$ e che $M(Y|x_2) = 5$.

$M(Y|x_1) = 2$ implica che la distribuzione condizionata $Y|x_1$ è degenere e, quindi, $n_{12} = n_{13} = 0$. $n_{21} = 2$ realizza $M(Y|x_2) = 5$

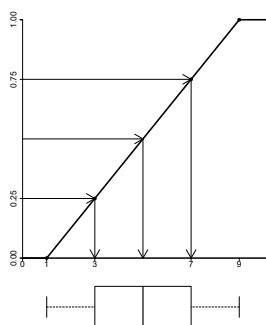
2. Si rappresentino graficamente le distribuzioni marginali X e Y .



2. Si calcolino media, varianza e coefficiente di variazione per la distribuzione marginale Y .

$$M(Y) = 3.2 \quad Var(Y) = 3.96 \quad CV(Y) = 0.6219$$

3. Si costruisca il grafico Box & Whiskers Plot per la distribuzione condizionata $Y|x_2$.



$$x_{0.25} = 3, x_{0.50} = 5, x_{0.75} = 7, \text{lunghezza teorica baffo} = 6.5, \text{baffo inferiore} = 1, \text{baffo superiore} = 9$$

ESERCIZIO 2.

Si riportano i prezzi di un prodotto registrati fra il 2015 e il 2018:

t	x_t	$NIBM$	$NIBF(2015)$
2015	20	—	1.000
2016	22	1.1000	1.100
2017	25	1.1364	1.250
2018	27.1	1.0840	1.355

1. Si costruiscano le serie dei numeri indice a base mobile e a base fissa con base 2015.

2. Si calcoli il numero indice al tempo 2016 con base 2017 interpretando il valore ottenuto.

${}_{2017}I_{2016} = 0.88$ Il prezzo nel 2016 era inferiore del 12% rispetto al prezzo nel 2017

3. Si calcoli il tasso medio di variazione tra il 2016 e il 2018.

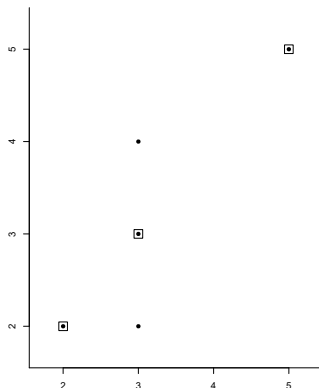
tasso medio di variazione 0.1099. Tra il 2016 e il 2018 si è osservato un incremento medio annuo del 10.99%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y tranne che per l'osservazione y_2

x_i	2	3	3	3	5
y_i	2		3	4	5

1. Si rappresentino graficamente le coppie dei dati, con anche il valore di y_2 tale per cui $\eta_{Y|X}^2 = \rho^2$. Si completi il grafico con la funzione di regressione ipotizzando Y in funzione di X .



$\eta_{Y|X}^2 = \rho^2$ se le medie condizionate $M(Y|X)$ sono allineate e con $y_2 = 2$ abbiamo $M(Y|X = 3) = 3$.

2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 0$, $\hat{b} = 1$, indice adattamento $\rho^2 = 0.7059$ $\hat{c} = 3.2$, indice adattamento $R^2 = 0$

3. Si indichi, motivando, quale modello presenta il migliore adattamento.

Il modello migliore è $Y^* = X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Indicare l'intervallo dei valori possibili per $\eta_{X|Y}^2$.

$\rho^2 = 0.7059 < \eta_{X|Y}^2 < 1$ in quanto le medie condizionate $M(X|Y)$ non sono allineate e la distribuzione condizionata $X|Y = 2$ non è degenere.

5. Sulla base del miglior modello prevedere il valore di Y in corrispondenza di $X = 3$.

$$Y^*(3) = 1 \cdot 3 = 3.$$

ESERCIZIO 4.

In una sala di un museo sono presenti le 4 opere $\{a, b, c, d\}$. Ogni visitatore può visionare le opere secondo un ordine arbitrario.

1. Si calcoli la probabilità che un visitatore visioni le opere secondo la sequenza d, b, a, c oppure secondo la sequenza b, d, a, c .

Le due sequenze hanno la stessa probabilità di verificarsi:

$$P(d \cap b \cap a \cap c) = P(d) \cdot P(b|d) \cdot P(a|d \cap b) \cdot P(c|d \cap b \cap a) = \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{1}{24}.$$

Abbiamo quindi $P(\{d \cap b \cap a \cap c\} \cup \{b \cap d \cap a \cap c\}) = 2 \cdot \frac{1}{24} = \frac{1}{12} = 0.08333$.

2. Si calcoli la probabilità che scelti a caso 3 visitatori, che hanno comportamento indipendente, questi adottino la stessa sequenza di visione delle opere.

$$24 \cdot \frac{1}{24} \cdot \frac{1}{24} \cdot \frac{1}{24} = \frac{24}{24^3} = \frac{1}{24^2} = 0.001736$$

3. In prossimità dell'orario di chiusura sono presenti 10 visitatori. Viene data loro la possibilità di visionare una sola opera. Si calcoli la probabilità che almeno 3 visitatori scelgano l'opera a , assumendo che la scelta dell'opera venga effettuata dai visitatori in maniera casuale e indipendente.

Sia X la variabile casuale 'n. di visitatori che scelgono l'opera a tra i 10 presenti'. Abbiamo

$$X \sim \text{Bin}(n = 10, p = 0.25)$$

e

$$\begin{aligned} P(X \geq 3) &= 1 - \{P(X = 0) + P(X = 1) + P(X = 2)\} = \\ &= 1 - \left\{ \binom{10}{0} 0.25^0 \cdot 0.75^{10} + \binom{10}{1} 0.25^1 \cdot 0.75^9 + \binom{10}{2} 0.25^2 \cdot 0.75^8 \right\} = \\ &= 1 - (0.05631 + 0.18771 + 0.28157) = 1 - 0.52559 = 0.47441. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 29.08.2019 (Tema 305-1)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

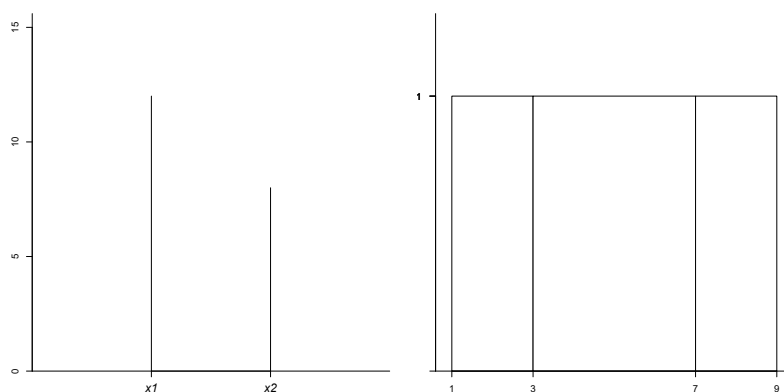
Il prospetto seguente riporta la distribuzione congiunta del carattere qualitativo X e del carattere quantitativo Y .

$X \setminus Y$	1 → 3	3 → 7	7 → 9	
x_1	12	0	0	12
x_2	2	4	2	8
	14	4	2	20

1. Si completi la distribuzione delle frequenze congiunte, sapendo che il numero totale di unità statistiche è 20, che $Var(Y|x_1) = 0$ e che la distribuzione $Y|x_2$ è simmetrica.

$Var(Y|x_1) = 0$ implica che la distribuzione condizionata $Y|x_1$ è degenere e, quindi, $n_{12} = n_{13} = 0$. $n_{21} = 2$ realizza la simmetria della distribuzione $Y|x_2$.

2. Si rappresentino graficamente la distribuzione marginale X e la distribuzione condizionata $Y|x_2$.



3. Si confrontino con opportuni indici di posizione e variabilità le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.

$Y|x_1$ Indice di posizione: $M(Y|x_1)$ valore 2 Indice di variabilità: $CV(Y|x_1)$ valore 0

$Y|x_2$ Indice di posizione: $M(Y|x_2)$ valore 5 Indice di variabilità: $CV(Y|x_2)$ valore 0.4243

La distribuzione condizionata $Y|x_1$ è caratterizzata da un livello medio e da variabilità inferiori rispetto alla condizionata $Y|x_2$.

4. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

indice di connessione χ_N valore 0.8018

Sussiste un livello elevato di connessione tra le due variabili.

ESERCIZIO 2.

Si consideri il seguente prospetto:

t	x_t	$NIBF(base = 2013)$	$NIBM$
2012	48	2	—
2013	24	1	0.5
2014	12	0.5	0.5
2015	6	0.25	0.5
2016	12	0.5	2

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli il numero indice al tempo 2013 con base 2015 e si commenti il risultato ottenuto.

${}_{2015}I_{2013} = 4$ Il valore della grandezza nel 2013 era il quadruplo del valore della grandezza nel 2015.

3. Si calcoli il tasso medio di variazione dal 2013 al 2015 e si commenti il risultato ottenuto.

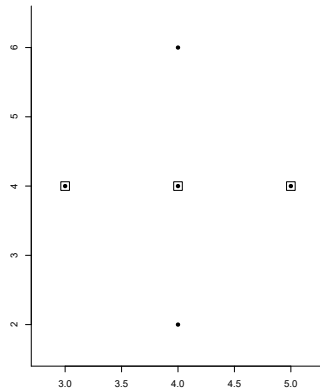
tasso medio di variazione 0.5. Tra il 2013 e il 2015 si è osservato un decremento medio annuo del 50%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y tranne che per l'osservazione y_3

x_i	3	4	4	4	5
y_i	4	2		6	4

1. Si rappresentino graficamente le coppie dei dati, con anche il valore di y_3 tale per cui $\eta_{Y|X}^2 = \eta_{X|Y}^2$. Si completi il grafico con la funzione di regressione ipotizzando Y in funzione di X .



È agevole realizzare la condizione $\eta_{Y|X}^2 = \eta_{X|Y}^2 = 0$ fissando $y_3 = 4$.

2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c$$

$\hat{a} = 4$, $\hat{b} = 0$, indice adattamento $\rho^2 = 0$ $\hat{c} = 4$, indice adattamento $R^2 = 0$

Dal momento che sussiste indipendenza in media di Y da X , essendo $\eta_{Y|X}^2 = 0$, abbiamo anche indipendenza lineare, $\rho^2 = Cov(X, Y) = b = 0$ e i due modelli risultano equivalenti con $a = c = M(Y)$.

ESERCIZIO 4.

1. Si consideri l'allestimento di una sala quadrata (con 4 pareti) con 4 quadri: 2 di Monet e 2 di Chagall.

Si calcoli la probabilità che, disponendo in maniera casuale un quadro su ciascuna delle 4 pareti, i due quadri di Monet figurino su 2 pareti opposte.

Se si considerano le 4 opere distinte:

$$8 \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{8}{24} = \frac{1}{3} = 0.3333$$

Se si considerano 2 opere di Monet e 2 di Chagall:

$$2 \cdot \frac{2}{4} \cdot \frac{2}{3} \cdot \frac{1}{2} \cdot \frac{1}{1} = \frac{2}{\binom{4}{2}} = \frac{2}{6} = \frac{1}{3} = 0.3333$$

2. Si scelgono in maniera casuale 5 opere tra 8 quadri di Manet, 8 di Degas e 4 di Sisley per la composizione di un art visual tour. Si calcoli la probabilità che almeno 3 tra le 5 opere scelte siano di Manet.

$$\begin{aligned} P(\text{almeno } 3 M) &= P(3M) + P(4M) + P(5M) = \\ &= \binom{5}{3} \frac{8}{20} \frac{7}{19} \frac{6}{18} \frac{12}{17} \frac{11}{16} + \binom{5}{4} \frac{8}{20} \frac{7}{19} \frac{6}{18} \frac{5}{17} \frac{12}{16} + \binom{5}{5} \frac{8}{20} \frac{7}{19} \frac{6}{18} \frac{5}{17} \frac{4}{16} = \\ &= 0.23839 + 0.05418 + 0.003612 = 0.2962 \end{aligned}$$

o mediante la v.c. ipergeometrica

$$P(X = 3) + P(X = 4) + P(X = 5) = \frac{\binom{8}{3} \binom{12}{2}}{\binom{20}{5}} + \frac{\binom{8}{4} \binom{12}{1}}{\binom{20}{5}} + \frac{\binom{8}{5} \binom{12}{0}}{\binom{20}{5}} = 0.2962.$$

3. Il visual tour dura 3 minuti. La probabilità che un generico visitatore completi la visione del tour è pari a 0.98. Si calcoli la probabilità che tra 200 visitatori con comportamento indipendente almeno 198 completino la visione del tour.

Sia X la variabile casuale 'n. di visitatori che completano il tour tra i 200 considerati'. Abbiamo

$$X \sim Bin(n = 200, p = 0.98)$$

e

$$\begin{aligned} P(X \geq 198) &= \{P(X = 198) + P(X = 199) + P(X = 200)\} = \\ &= \binom{200}{198} 0.98^{198} \cdot 0.02^2 + \binom{200}{199} 0.98^{199} \cdot 0.02^1 + \binom{200}{200} 0.98^{200} \cdot 0.02^0 = \\ &= 0.14577 + 0.07179 + 0.01759 = 0.2351. \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 05.09.2019 (Tema 305)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

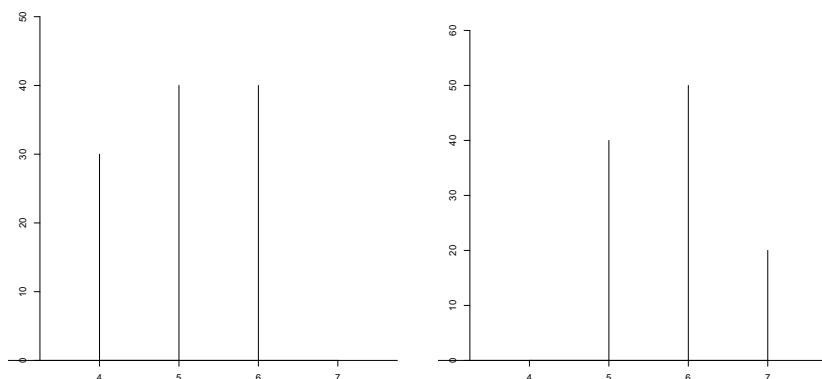
Studente _____ matricola _____

ESERCIZIO 1.

Con riferimento a due classi di studenti ($x_1 = A$, $x_2 = B$) viene riportata nel seguente prospetto la distribuzione del numero di esami Y sostenuti durante il primo anno accademico:

$X \setminus Y$	4	5	6	7
x_1	30	40	40	0
x_2	0	40	50	20

1. Si rappresentino graficamente le distribuzioni condizionate di Y dato X .



2. Si confrontino le distribuzioni condizionate di Y dato X calcolando medie e varianze.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 5.0909$ Indice di variabilità: $CV(Y|x_1) = 0.1557$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 5.8182$ Indice di variabilità: $CV(Y|x_2) = 0.1230$

La distribuzione condizionata $Y|x_1$ è caratterizzata da un livello medio leggermente inferiore a quello della condizionata $Y|x_2$ ma da variabilità superiore. Per il confronto in termini di variabilità non è possibile utilizzare le varianze $Var(Y|x_1) = 0.6281$, $Var(Y|x_2) = 0.5124$, ma occorre calcolare i coefficienti di variazione.

3. Utilizzando i risultati al punto precedente si calcoli un indice di dipendenza in media di Y dato X .

$$\sigma_{Between}^2 = \frac{1}{220} \left\{ (\mu_A \cdot 110 - \mu)^2 + (\mu_B \cdot 110 - \mu)^2 \right\} = \frac{1}{220} (\mu_A^2 \cdot 110 + \mu_B^2 \cdot 110) - \mu^2 = \frac{1}{220} (5.0909^2 \cdot 110 + 5.8182^2 \cdot 110) - 5.4545^2 = 0.1322$$

$$\sigma_{Within}^2 = \frac{1}{220} (\sigma_A^2 \cdot 110 + \sigma_B^2 \cdot 110) = \frac{1}{220} (0.6281 \cdot 110 + 0.5124 \cdot 110) = 0.5702.$$

Il rapporto di correlazione $\eta_{Y|X}^2$ risulta pari a 0.1882.

Sussiste, quindi, un livello di dipendenza basso del numero di esami sostenuti rispetto alla classe.

4. Senza effettuare calcoli, si dica se può sussistere indipendenza stocastica tra le due variabili motivando la risposta.

Non sussiste indipendenza stocastica in quanto le distribuzioni condizionate non sono somiglianti.

ESERCIZIO 2.

Con riferimento alla serie storica di un titolo azionario x_t sono note le informazioni riportate nel seguente prospetto:

t	x_t	$NIBM_t$
1	95.2381	—
2	100	1.05
3	100	1.00
4	95	0.95
5	85.5	0.90

1. Si ricostruisca la serie delle osservazioni x_t a partire dalle informazioni disponibili.

2. Si calcoli il tasso medio di variazione tra $t = 1$ e $t = 5$ e si interpreti il valore ottenuto.

tasso medio di variazione -0.0266. Tra $t = 1$ e $t = 5$ si è osservato un decremento medio annuo del 2.661%.

3. Si determini il valore del titolo al tempo $t = 7$ assumendo tra $t = 5$ e $t = 7$ la stessa variazione (media annua) osservata nei periodi precedenti.

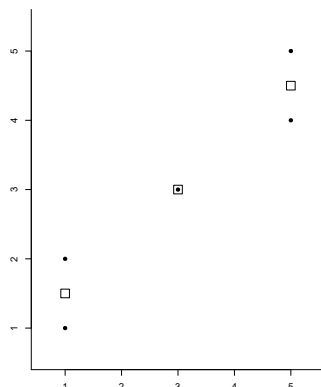
$$x_7 = x_5 \cdot 0.9734 \cdot 0.9734 = 81.0110$$

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y :

x_i	1	1	3	5	5
y_i	1	2	3	4	5

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione di Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 0.75, \hat{b} = 0.75, \text{ indice adattamento } M(\hat{E}^2) = 0.2 \quad \hat{c} = 0.9344, \text{ indice adattamento } M(\hat{E}^2) = 0.3475$$

Il modello migliore è $Y^* = 0.75 + 0.75X$.

3. In corrispondenza del modello migliore, si fornisca la previsione per $x = 7$.

$$Y^*(7) = 0.75 + 0.75 \cdot 7 = 6.$$

4. Si stabilisca, senza effettuare calcoli e giustificando teoricamente la risposta, il valore assunto dagli indici $\eta_{Y|X}^2$ e $\eta_{X|Y}^2$.

$$\eta_{Y|X}^2 = \rho^2 = 0.9 \text{ in quanto le medie condizionate } M(Y|X) \text{ sono allineate}$$

$$\eta_{X|Y}^2 = 1 \text{ in quanto sussiste dipendenza funzionale di } X \text{ da } Y.$$

ESERCIZIO 4.

Un'urna contiene 50 palline, di cui 10 rosse, mentre si sa che le restanti possono essere blu o verdi.

1. Si effettuano 5 estrazioni con reimmissione. Si calcoli la probabilità che nessuna delle palline estratte sia rossa.

Sia X la variabile casuale 'n. palline rosse tra le 5 estratti'. Abbiamo

$$X \sim \text{Bin}(n = 5, p = 0.2)$$

e

$$P(X = 0) = \binom{5}{0} 0.2^0 \cdot 0.8^5 = 0.32768.$$

2. Si ipotizzi ora di effettuare 100 estrazioni con reimmissione. Si calcoli la probabilità di ottenere almeno 20 palline rosse.

Indicata con X la variabile casuale 'numero palline rosse nelle 100 estrazioni con reimmissione dall'urna U_1 ' risulta $X \sim \text{Bin}(n = 100, p = 0.2)$. $M(X) = 20$; $\text{Var}(X) = 16$. Dal momento che $np = 20 > 5$ ed $n(1 - p) = 80 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 20, \sigma^2 = 16).$$

Dal momento che la variabile casuale Normale è simmetrica rispetto alla media, si ottiene

$$P(X > 20) = 0.5.$$

3. Sapendo che la probabilità di estrarre 3 palline blu in 3 estrazioni con reimmissione è pari a 0.125, si determini il numero di palline blu contenute nell'urna.

Indicata con p la frazione di palline blu contenute nell'urna, che coincide con la probabilità di ottenere una pallina blu in una generica estrazione Bernoulliana, la probabilità di ottenere palline blu in 3 estrazioni con reimmissione risulta

$$p^3 = 0.125$$

da cui segue $p = 0.5$. Il numero di palline blu contenute nell'urna risulta, quindi, $50 \cdot p = 50 \cdot 0.5 = 25$.

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 12.09.2019 (Tema 305-2)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

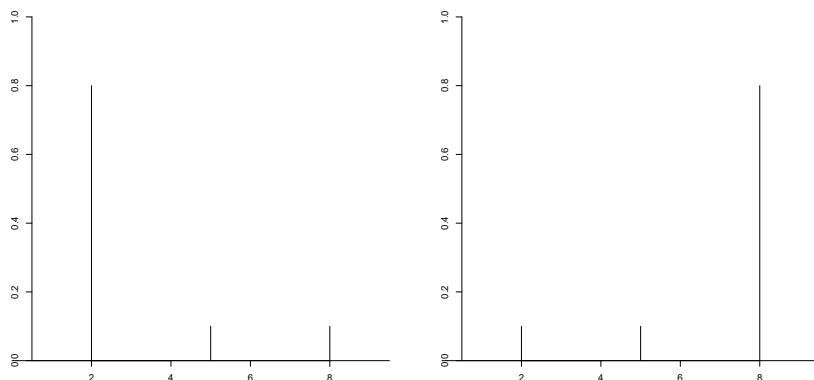
Il prospetto seguente riporta la distribuzione delle frequenze relative del carattere quantitativo Y condizionate rispetto alle categorie del carattere qualitativo X .

$Y X$	2	5	8	
x_1	0.80	0.10	0.10	1
x_2	0.10	0.10	0.80	1

1. Si completi la distribuzione delle frequenze congiunte sapendo che il numero di unità statistiche con categoria x_1 è 20, e che numero di unità statistiche con categoria x_2 è 80.

$X \backslash Y$	2	5	8	
x_1	16	2	2	20
x_2	8	8	64	80
	24	10	66	100

2. Si rappresentino graficamente le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.



3. Si confrontino con opportuni indici di posizione e variabilità le distribuzioni condizionate $Y|x_1$ e $Y|x_2$.

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 2.9$ Indice di variabilità: $CV(Y|x_1) = 0.6624$

$Y|x_2$ Indice di posizione: $M(Y|x_2) = 7.1$ Indice di variabilità: $CV(Y|x_2) = 0.2705$

La distribuzione condizionata $Y|x_1$ è caratterizzata da un livello medio inferiore a quello della condizionata $Y|x_2$ ma da variabilità superiore.

4. Si determini il grado di connessione tra le variabili X e Y mediante il calcolo di un opportuno indice.

indice di connessione $\chi_N = 0.6675$. Sussiste un livello medio alto di connessione.

ESERCIZIO 2.

Il seguente prospetto riporta i valori del fatturato di un'azienda tra il 2012 e il 2016:

t	x_t	$NIBF(base = 2014)$	$NIBM$
2012	10	0.2	—
2013	30	0.6	3
2014	50	1	1.6667
2015	30	0.6	0.6000
2016	10	0.2	0.3333

1. Si costruiscano le serie dei numeri indici a base fissa, con base 2014, e a base mobile.

2. Si calcoli il numero indice al tempo 2014 con base 2013 e si commenti il risultato ottenuto.

${}_{2013}I_{2014} = 1.6667$ Il fatturato ha subito un incremento del 66.67% tra il 2013 e il 2014.

3. Si calcoli il tasso medio di variazione tra il 2012 e il 2015 e si commenti il risultato ottenuto.

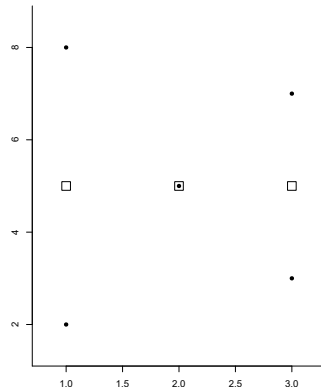
tasso medio di variazione 0.4422. Tra il 2012 e il 2015 il fatturato ha subito un incremento medio annuo del 44.22%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y tranne che per l'osservazione y_3

x_i	1	1	2	3	3
y_i	2	8		3	7

1. Si rappresentino graficamente le coppie dei dati, con anche il valore di y_3 tale per cui $\eta_{Y|X}^2 = 0$ ed $\eta_{X|Y}^2 = 1$. Si completi il grafico con la funzione di regressione ipotizzando Y in funzione di X .



$\eta_{Y|X}^2 = 0$ solo se sussiste indipendenza in media di Y da X , vale a dire se $M(Y|x=1) = M(Y|x=2) = M(Y|x=3) = 5$. Segue che $y_3 = 5$. $\eta_{X|Y}^2 = 1$ in quanto sussiste dipendenza funzionale di Y da X qualsiasi sia il valore assegnato a y_3 .

2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$\hat{a} = 5$, $\hat{b} = 0$, indice adattamento $M(\hat{E}) = 5.2$ $\hat{c} = 2.0833$, indice adattamento $M(\hat{E}) = 9.3667$

Il modello migliore è $Y^* = 5$.

Dal momento che sussiste indipendenza in media di Y da X , essendo $\eta_{Y|X}^2 = 0$, abbiamo anche indipendenza lineare, $\rho^2 = \text{Cov}(X, Y) = b = 0$ e $a = M(Y)$.

3. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 3$.

$$Y^*(3) = 5 + 0 \cdot 3 = 5.$$

ESERCIZIO 4.

Lo sfondo di una pagina web viene ridefinito a ogni visualizzazione estraendo in maniera casuale e indipendente un'opera tra 8 quadri di Kandinskij, 8 di Mondrian e 4 di Delaunay.

1. Si calcoli la probabilità che un soggetto che effettua 5 visualizzazioni trovi come sfondo un quadro di Kandinskij in 3 visualizzazioni consecutive.

$$P(3K \text{ consecutive in 5 visualizzazioni}) = 0.06912.$$

La probabilità della generica sequenza contenente 3 quadri di Kandinskij in 5 visualizzazioni risulta

$$P(K \cap K \cap K \cap \bar{K} \cap \bar{K}) = 0.4 \cdot 0.4 \cdot 0.4 \cdot 0.6 \cdot 0.6 = 0.02304$$

e

$$\begin{aligned} P(3K \text{ consecutive in 5 visualizzazioni}) &= P(K \cap K \cap K \cap \bar{K} \cap \bar{K}) + P(\bar{K} \cap K \cap K \cap K \cap \bar{K}) + P(\bar{K} \cap \bar{K} \cap K \cap K \cap K) = \\ &= 3 \cdot 0.4^3 \cdot 0.6^2 = 3 \cdot 0.02304 = 0.06912. \end{aligned}$$

2. Si calcoli la probabilità che un soggetto che effettua 5 visualizzazioni trovi come sfondo un quadro di Kandinskij in almeno 3 visualizzazioni.

Sia X la variabile casuale 'n. di visualizzazioni con un quadro di Kandinskij nelle 5 prove indipendenti'. Abbiamo $X \sim \text{Bin}(n=5, p=0.4)$ e

$$\begin{aligned} P(X \geq 3) &= \{P(X=3) + P(X=4) + P(X=5)\} = \\ &= \binom{5}{3} 0.4^3 \cdot 0.6^2 + \binom{5}{4} 0.4^4 \cdot 0.6^1 + \binom{5}{5} 0.4^5 \cdot 0.6^0 = \\ &= 0.2304 + 0.0768 + 0.0102 = 0.3174. \end{aligned}$$

3. Si calcoli la probabilità che un quadro di Kandinskij figuri come sfondo in almeno 31 visualizzazioni su 100.

Indicata con X la variabile casuale 'n. di visualizzazioni con un quadro di Kandinskij nelle 100 prove indipendenti' risulta $X \sim \text{Bin}(n=100, p=0.4)$. $M(X) = 40$; $\text{Var}(X) = 24$. Dal momento che $np = 40 > 5$ ed $n(1-p) = 60 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale $X \sim N(\mu = 40, \sigma^2 = 24)$. Si ottiene

$$P(X \geq 31) = P\left(\frac{X - \mu}{\sigma} \geq \frac{31 - \mu}{\sigma}\right) = P\left(Z \geq \frac{31 - 40}{4.899}\right) = P(Z \geq -1.84) = P(Z \leq 1.84) = 0.9671.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 15.01.2020 (Tema 306)

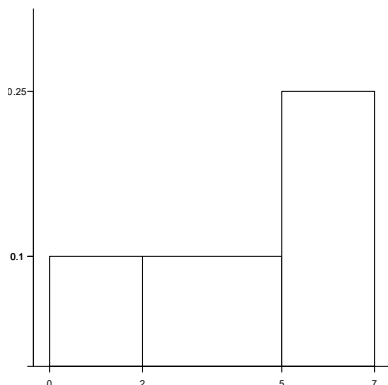
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La seguente tabella riporta le distribuzioni relative condizionate – ricavate a seguito di una indagine campionaria condotta su $n = 150$ automobilisti – della variabile $Y = \text{Ore al giorno trascorse alla guida della macchina}$ rispetto alla variabile $X = \text{Utilizzo delle cinture di sicurezza}$:

$Y X$	0 ÷ 2	2 ÷ 5	5 ÷ 7	
$sì$	0.2	0.3	0.5	1
no	0.6	0.3	0.1	1

1. Dopo avere rappresentato graficamente la variabile $Y|X = sì$, se ne calcoli la moda. È possibile dire che la moda coincide con la mediana di $Y|X = sì$?



La mediana $y_{0.5} = 5$ (ottenuta come estremo superiore della seconda classe, osservando che la somma delle frequenze relative delle prime due classi è pari a 0.5) coincide con l'estremo inferiore della classe modale 5 ÷ 7.

2. Costruire la tabella delle frequenze congiunte sapendo che il 60% dei lavoratori intervistati usa la cintura di sicurezza.

$X \setminus Y$	0 ÷ 2	2 ÷ 5	5 ÷ 7	$n_{i\bullet}$
$sì$	18	27	45	90
no	36	18	6	60
$n_{\bullet j}$	54	45	51	150

3. Confrontare, con opportuni indici di posizione e di variabilità, la distribuzione della variabile Y (ore al giorno trascorse alla guida) nei due gruppi di utilizzatori ($X = sì$) e non utilizzatori ($X = no$) della cintura di sicurezza. Commentare il risultato ottenuto.

$Y|X = sì$ Indice di posizione: $M(Y|X = sì)$ valore 4.25 Indice di variabilità: $CV(Y|X = sì)$ valore 0.4594

$Y|X = no$ Indice di posizione: $M(Y|X = no)$ valore 2.25 Indice di variabilità: $CV(Y|X = no)$ valore 0.7454

Il numero medio di ore al giorno trascorse alla guida di coloro che utilizzano la cintura di sicurezza è superiore a quello di coloro che non la utilizzano. Il numero medio di ore al giorno trascorse alla guida di coloro che non utilizzano la cintura di sicurezza presenta un livello di variabilità superiore a quello di coloro che la utilizzano.

4. Senza effettuare calcoli e motivando la risposta, dire se l'indice di mutabilità normalizzato di X assume un valore più vicino a 0 o a 1.

F_N è più vicino a 1 essendo le frequenze marginali $f_{1\bullet} = \frac{90}{150} = 0.6$ e $f_{2\bullet} = \frac{60}{150} = 0.4$ più prossime alla situazione di equidistribuzione ($f_{1\bullet} = f_{2\bullet} = 0.50$) che alla situazione di variabile statistica degenerare.

5. Calcolare un indice di connessione tra X e Y e commentare il risultato ottenuto.

indice di connessione χ_N valore 0.4686.

Sussiste un livello medio di connessione tra le due variabili.

ESERCIZIO 2.

Si consideri il seguente prospetto che riporta i $NIBF$ (base=2015) di un bene dal 2015 al 2018:

t	$NIBF(\text{base} = 2015)$	$NIBM$
2015	1	-
2016	1	1
2017	1.1	1.1
2018	0.96	0.8727
2019	0.6561	0.6834

1. Ricostruire la serie dei $NIBM$ dal 2015 al 2018.

2. Sapendo che il bene ha subito un decremento medio annuo del 10% dal 2015 al 2019, completare la tabella con il valore dei numeri indici a base fissa e a base mobile al tempo 2019.

Un decremento medio del 10% corrisponde a una misura relativa media dello 0.90; quindi, ${}_{2015}I_{2019} = 0.90^4 = 0.6561$.

3. Calcolare il valore del bene nel 2019, sapendo che nel 2018 è risultato pari a 21.

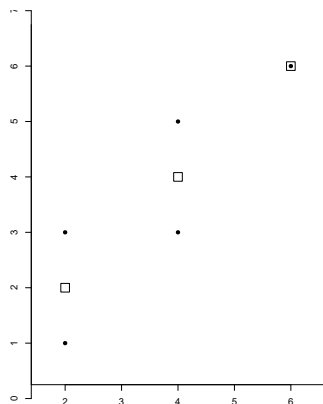
$x_{2019} = 14.3522$. Vale ${}_{2018}I_{2019} = \frac{{}_{2015}I_{2019}}{{}_{2015}I_{2018}} = \frac{0.6561}{0.96} = \frac{x_{2019}}{x_{2018}} = \frac{x_{2019}}{21} = 0.6834$.

ESERCIZIO 3.

Su un insieme di 5 individui vengono rilevate le seguenti variabili: X = numero di regali di Natale, Y = importo speso per i regali (in centinaia di Euro) e G = genere (f = femmina, m = maschio). I dati sono riportati di seguito:

x_i	2	6	4	2	4
y_i	1	6	5	3	3
g_i	f	f	f	m	m

1. Costruire il grafico delle coppie dei dati (x_i, y_i) e della funzione di regressione della spesa in funzione del numero di regali.



2. Calcolare, secondo il metodo dei minimi quadrati, i parametri dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c + d \cdot W$$

con W variabile dummy, indicatrice del genere femmina, ($w_i = 0$ se $g_i = m$, $w_i = 1$ se $g_i = f$).

$\hat{a} = 0$, $\hat{b} = 1$, indice adattamento $\rho^2 = 0.7368$ $\hat{c} = 3$, $\hat{d} = 1$, indice adattamento $\rho^2 = 0.0789$

3. Confrontare gli indici di adattamento, commentando i risultati.

Il modello migliore è $Y^* = X$ in quanto caratterizzato da indice di adattamento più elevato.

4. Fornire una interpretazione delle stime dei parametri del modello II).

$\hat{c} = M(Y|W = m) = 3$, $\hat{d} = M(Y|W = f) - M(Y|W = m) = 1$ ($= 4 - 3$)

5. Senza effettuare calcoli e motivando la risposta, indicare il valore di $\eta_{Y|X}^2$.

$\eta_{Y|X}^2 = 0.7368$ in quanto le medie condizionate $M(Y|x = 2)$, $M(Y|x = 4)$ e $M(Y|x = 6)$ sono allineate.

ESERCIZIO 4.

Una ricerca ha evidenziato che i tempi di gestazione delle donne in gravidanza seguono una distribuzione Normale con media $\mu = 270$ giorni e varianza $\sigma^2 = 900$.

1. Indicare quale è la probabilità che un bambino nasca dopo 274 giorni.

Indicata con $X \sim N(\mu = 270, \sigma^2 = 900)$ la variabile casuale che descrive i tempi di gestazione delle donne in gravidanza, occorre determinare la seguente probabilità: $P(X > 274)$. Abbiamo

$$\begin{aligned} P(X > 274) &= P\left(\frac{X - \mu}{\sigma} > \frac{274 - \mu}{\sigma}\right) = P\left(Z > \frac{274 - 270}{30}\right) = \\ &= P(Z > 0.13) = 1 - P(Z \leq 0.13) = 1 - 0.5517 = 0.4483. \end{aligned}$$

2. Sapendo che la probabilità che un bambino nasca prematuro è del 2%, stabilire quale è il tempo gestazionale minimo (numero di giorni) affinché un bambino non sia considerato prematuro.

Il tempo gestazionale minimo (numero di giorni) affinché un bambino non sia considerato prematuro corrisponde al percentile di ordine 0.02 della distribuzione di X .

$$\{x : P(X \leq x) = 0.02\} = \left\{x : P\left(\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right) = 0.02\right\} = \left\{x : P\left(Z \leq \frac{x - 270}{30}\right) = 0.02\right\}.$$

Dalla lettura inversa della Tavola possiamo ricavare il percentile di ordine 0.02, che risulta -2.05 (infatti, $P(Z \leq -2.05) = 0.02$ dal momento che la distribuzione della variabile casuale Normale è simmetrica e $P(Z \leq 2.05) = 0.98$). Abbiamo, quindi, $z = \frac{x - 270}{30} = -2.05$ da cui segue $x = 208.5 \simeq 209$.

3. Un ginecologo ha in carico 10 donne in gravidanza: quale è la probabilità che almeno una di loro abbia un parto prematuro?

Sia X la variabile casuale 'n. pazienti in carico con parto prematuro'. Dal momento che la popolazione è composta da un numero elevato di soggetti si può ritenere che l'evento parto prematuro si presenti in maniera indipendente in diversi soggetti. Abbiamo

$$X \sim \text{Bin}(n = 10, p = 0.02)$$

e

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0} 0.02^0 \cdot 0.98^{10} = 1 - 0.8171 = 0.1829.$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 29.01.2020 (Tema 307)

Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

Studente _____ matricola _____

ESERCIZIO 1.

Il prospetto seguente riporta la distribuzione delle frequenze congiunte del carattere qualitativo X e del carattere quantitativo Y .

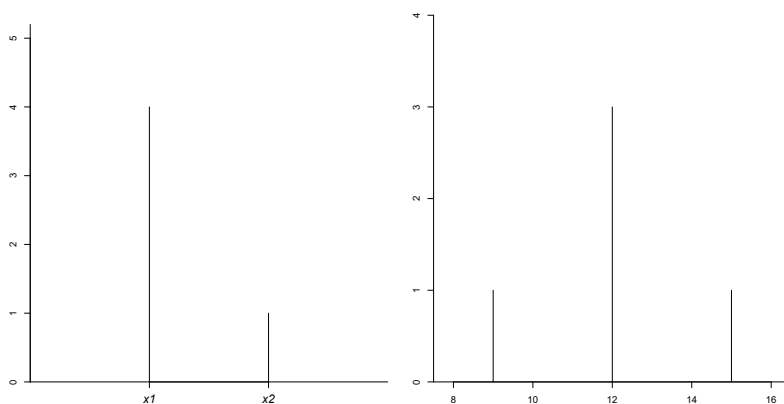
$X \backslash Y$	9	12	15	$n_{i\bullet}$
x_1	1	2	1	4
x_2	0	1	0	1
$n_{\bullet j}$	1	3	1	5

1. Si completi la distribuzione delle frequenze congiunte in modo che la mediana e la media aritmetica della distribuzione condizionata $Y|x_1$ risultino 12 e che il coefficiente di variazione della distribuzione condizionata $Y|x_2$ sia nullo.

Con $n_{13} = 1$ la distribuzione $Y|x_1$ risulta simmetrica e vale $M(Y|x_1) = y_{0.5}|x_1 = 12$.

$CV(Y|x_2) = 0$ implica che la distribuzione condizionata $Y|x_2$ è degenera e, quindi, $n_{21} = n_{23} = 0$.

2. Si rappresentino graficamente le distribuzioni marginali X e Y .



3. Si ricavino i valori della media e della varianza della distribuzione marginale Y a partire dalle medie e varianze delle distribuzioni condizionate $Y|x_1$ e $Y|x_2$ e si indichi il valore del rapporto di correlazione $\eta_{Y|X}^2$.

Da $M(Y|x_1) = M(Y|x_2) = 12$ e $Var(Y|x_1) = 4.5$ e $Var(Y|x_2) = 0$ segue:

$$M(Y) = M_X\{M(Y|X)\} = 12$$

$$Var(Y) = Var_X\{M(Y|X)\} + M_X\{Var(Y|X)\} = 0 + 3.6 = 3.6$$

$$\eta_{Y|X}^2 = \frac{Var_X\{M(Y|X)\}}{Var(Y)} = \frac{0}{3.6} = 0$$

4. Si confrontino con opportuni indici di posizione e variabilità la distribuzione condizionata $Y|x_1$ e la distribuzione marginale Y .

$Y|x_1$ Indice di posizione: $M(Y|x_1) = 12$ Indice di variabilità: $CV(Y|x_1) = 0.1768$

Y Indice di posizione: $M(Y) = 12$ Indice di variabilità: $CV(Y) = 0.1581$

Le due distribuzioni hanno lo stesso livello medio e la distribuzione condizionata $Y|x_1$ è caratterizzata da variabilità superiore rispetto alla distribuzione marginale Y .

ESERCIZIO 2.

Il seguente prospetto riporta alcune informazioni sui numeri indici con base fissa 2012 ($NIBF(base = 2012)$), sui numeri indici a base mobile ($NIBM$) e sulle variazioni percentuali rispetto all'anno precedente ($variaz.relative\%$) riferiti ai valori del fatturato di un'azienda tra il 2012 e il 2016:

t	$NIBF(base = 2012)$	$NIBM$	$variaz.relative\%$
2012	1	—	
2013	1.11	1.11	+11%
2014	1.2321	1.11	+11%
2015	1.3676	1.11	+11%
2016	0.78	0.5703	-42.97%

1. Si completi il prospetto con i dati mancanti.

2. Si calcoli il numero indice al tempo 2013 con base 2016 e si commenti il risultato ottenuto.

${}_{2016}I_{2013} = 1.4231$ Il fatturato nel 2013 era superiore del 42.31% al fatturato nel 2016.

3. Si calcoli il tasso medio di variazione tra il 2013 e il 2016 e si commenti il risultato ottenuto.

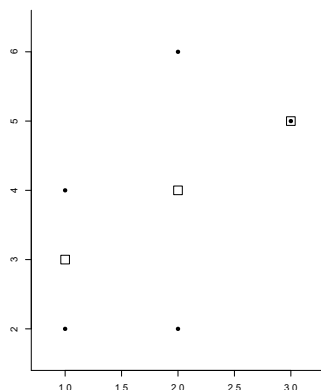
tasso medio di variazione -0.1110 . Tra il 2013 e il 2016 si è osservato un decremento medio annuo del fatturato del 11.10%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 5 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	2	2	3
y_i	2	4	2	6	5

1. Si rappresentino graficamente le coppie dei dati e la funzione di regressione ipotizzando Y in funzione di X .



2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = c \cdot X$$

$$\hat{a} = 2, \hat{b} = 1, \text{ indice adattamento } M(\hat{E}^2) = 2 \quad \hat{c} = 1.9474, \text{ indice adattamento } M(\hat{E}^2) = 2.5895$$

Il modello migliore è $Y^* = 2 + X$.

3. Senza svolgere i conti, cosa è possibile dire con riferimento ai dati in esame riguardo alla relazione tra ρ^2 ed $\eta_{X|Y}^2$?

$\rho^2 = \eta_{X|Y}^2 = 0.2188$ in quanto le medie condizionate $M(Y|X)$ sono allineate.

4. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 3$.

$$Y^*(3) = 2 + 1 \cdot 3 = 5.$$

ESERCIZIO 4.

Un albero di Natale è addobbato con led di diversi colori: 20 bianchi, 15 rossi e 15 verdi. A ogni accensione 3 diversi led si illuminano mentre i restanti 47 restano spenti.

1. Si dica quale v.c. descrive il numero di led bianchi che si illuminano in una accensione e si calcoli il suo valore atteso.

L'esperimento corrisponde alla scelta, senza reimmissione, di 3 led tra i 50 presenti sull'albero. Il numero di led bianchi tra i 3 led che si illuminano in una accensione può essere descritto mediante una variabile casuale ipergeometrica

$$P(X = x) = \frac{\binom{20}{x} \binom{30}{3-x}}{\binom{50}{3}}, \quad x = 0, 1, 2, 3$$

Il valore atteso risulta pari a $3 \cdot \frac{20}{50} = 1.2$

2. Si calcoli la probabilità che in una accensione sia illuminato almeno 1 led bianco.

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - \frac{\binom{20}{0} \binom{30}{3}}{\binom{50}{3}} = \\ &= 1 - P(\bar{B}_1 \cap \bar{B}_2 \cap \bar{B}_3) = 1 - P(\bar{B}_1) \cdot P(\bar{B}_2 | \bar{B}_1) \cdot P(\bar{B}_3 | \bar{B}_1 \cap \bar{B}_2) = \\ &= 1 - \frac{30}{50} \frac{29}{49} \frac{28}{48} = 1 - 0.2071 = 0.7929. \end{aligned}$$

3. Si calcoli la probabilità che in una accensione siano illuminati 3 led tutti dello stesso colore.

$$\begin{aligned} P((B_1 \cap B_2 \cap B_3) \cup (R_1 \cap R_2 \cap R_3) \cup (V_1 \cap V_2 \cap V_3)) &= \\ &= P(B_1) \cdot P(B_2 | B_1) \cdot P(B_3 | B_1 \cap B_2) + P(R_1) \cdot P(R_2 | R_1) \cdot P(R_3 | R_1 \cap R_2) + P(V_1) \cdot P(V_2 | V_1) \cdot P(V_3 | V_1 \cap V_2) = \\ &= \frac{20}{50} \frac{19}{49} \frac{18}{48} + \frac{15}{50} \frac{14}{49} \frac{13}{48} + \frac{15}{50} \frac{14}{49} \frac{13}{48} = \\ &= 0.05816 + 0.02321 + 0.02321 = 0.10458 \end{aligned}$$

Università Cattolica del Sacro Cuore - Milano

FACOLTÀ DI ECONOMIA - LETTERE E FILOSOFIA

Prova scritta di STATISTICA del 19.02.2020 (Tema 308)

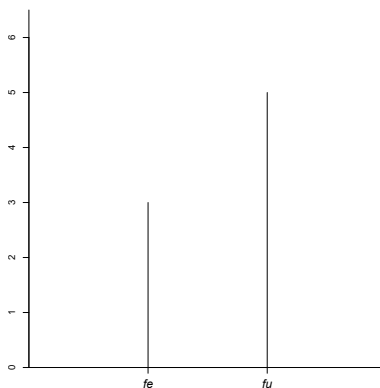
Svolgere per esteso la prova sui fogli quadrettati, indicando formule, calcoli, risultati e commenti.

ESERCIZIO 1.

La seguente tabella riporta le informazioni sul peso, variabile X (misurato in Kg), e sul tipo di sospensione, variabile Y (con modalità fu =integrare e fe =frontale), di 8 tipi di mountain bike:

x_i	7	12	12	12	12	12	13	14
y_i	fu	fe	fe	fu	fu	fe	fu	fu

1. Si rappresenti graficamente la distribuzione di frequenza della variabile Y e si calcolino opportuni indici di posizione e di variabilità. Si commentino i risultati.



Y Indice di posizione: Moda valore fu =integrare Indice di variabilità: F_N valore 0.75.

La tipologia di sospensione delle mountain bike risulta caratterizzata da un livello medio/alto di eterogeneità.

2. Si confronti la variabilità di X nei gruppi definiti dalla variabile Y .

indice variabilità $X|Y = fe$ CV valore 0

indice variabilità $X|Y = fu$ CV valore 0.2083

Il peso delle mountain bike con sospensione frontale presenza assenza di variabilità.

3. Si calcolino la varianza between e la varianza within di X nei gruppi definiti dalla variabile Y . Si commentino i risultati ottenuti.

$$\sigma_{Between}^2 = \frac{1}{8} \left\{ (\mu_{fe} - \mu)^2 \cdot 3 + (\mu_{fu} - \mu)^2 \cdot 5 \right\} = \frac{1}{8} \left(\mu_{fe}^2 \cdot 3 + \mu_{fu}^2 \cdot 5 \right) - \mu^2 = \frac{1}{8} (12^2 \cdot 3 + 11.6^2 \cdot 5) - 11.75^2 = 0.0375$$

$$\sigma_{Within}^2 = \frac{1}{8} \left(\sigma_{fe}^2 \cdot 3 + \sigma_{fu}^2 \cdot 5 \right) = \frac{1}{8} (0 \cdot 3 + 5.84 \cdot 5) = 3.65.$$

Il rapporto di correlazione $\eta_{X|tipologia\ sospensione}^2$ risulta pari a 0.0102.

Sussiste, quindi, un livello di dipendenza molto basso del peso delle mountain bike rispetto alla tipologia di sospensione.

4. Si costruisca la tabella a doppia entrata (W, Y), con W variabile qualitativa con categorie: $w_1 = leggera$ se il peso X è minore o UGUALE a 12 e $w_2 = pesante$ se il peso X è maggiore di 12. Si indichi se le variabili possono ritenersi stocasticamente indipendenti.

$W \backslash Y$	fe	fu
$leggera$	3	3
$pesante$	0	2

Le variabili W e Y non sono stocasticamente indipendenti in quanto le distribuzioni condizionate non sono somiglianti.

ESERCIZIO 2.

Per i prezzi di due beni A e B sono state calcolate le seguenti serie dei numeri indici negli anni 2012-2016, con base fissa 2012 ($NIBF_A$) per il bene A e con base mobile per il bene B ($NIBM_B$):

t	$NIBF_A$	$NIBM_A$	$NIBF_B$	$NIBM_B$
2012	1	-	1	-
2013	1.01	1.0100	1.0100	1.01
2014	1.05	1.0396	1.1110	1.1
2015	1.16	1.1048	1.3332	1.2
2016	1.15	0.9914	1.7332	1.3

1. Costruire le serie dei $NIBM$ per il bene A e dei $NIBF$ con base 2012 per il bene B .

2. Dire, giustificando la risposta, quale dei due beni ha avuto maggior variazione media annua negli anni considerati.

Tra il 2012 e il 2016 i beni A e B hanno registrato incrementi medi annui del 3.56% e 14.74%, rispettivamente.

3. Sapendo che le quantità vendute dei due beni nel 2012 sono rispettivamente $q_A = 1.2$ e $q_B = 0.1$ calcolare l'indice di Laspeyres dei prezzi al 2013, con base 2012 (ipotizzando, se necessario, che i prezzi dei due beni nell'anno base siano uguali ai corrispondenti $NIBF$ dell'anno 2016 con base 2012).

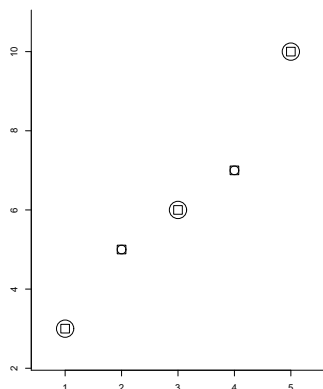
$\frac{P_{2012} I_{2013}^L}{P_{2012} I_{2013}^L} = 1.01$ dal momento che $\frac{P_{2012} I_{2013}^{bene A}}{P_{2012} I_{2013}^{bene B}} = 1.01$. La variazione relativa nel livello generale dei prezzi, secondo l'Indice di Laspeyres, è dell'1%.

ESERCIZIO 3.

Il seguente prospetto riporta i valori rilevati in corrispondenza a 8 unità statistiche con riferimento a due caratteri quantitativi X e Y

x_i	1	1	2	3	3	4	5	5
y_i	3	3	5	6	6	7	10	10

1. Si rappresentino graficamente le coppie dei dati e le medie condizionate di Y in funzione di X e si indichi il grado del polinomio mediante il quale è possibile rappresentare la funzione di regressione.



È possibile rappresentare la funzione di regressione mediante un polinomio di grado al massimo pari a 4.

2. Si calcolino, secondo il metodo dei minimi quadrati, i parametri e l'indice di adattamento dei seguenti modelli di regressione:

$$I) Y^* = a + b \cdot X \quad II) Y^* = e^{cX}$$

$\hat{a} = 1.25$, $\hat{b} = 1.6667$, indice adattamento $M(\hat{E}^2) = 0.1875$ $\hat{c} = 0.5220$, indice adattamento $M(\hat{E}^2) = 4.7636$

Il modello migliore è $Y^* = 1.25 + 1.6667 \cdot X$.

3. Senza svolgere i calcoli ma motivando la risposta si dica quali valori possono assumere $\eta_{X|Y}^2$ ed $\eta_{Y|X}^2$.

$\eta_{X|Y}^2 = \eta_{Y|X}^2 = 1$ in quanto le distribuzioni condizionate $Y|X$ e $X|Y$ sono degeneri.

4. Sulla base del miglior modello si preveda il valore di Y in corrispondenza di $X = 6$.

$Y^*(6) = 1.25 + 1.6667 \cdot 6 = 11.25$.

ESERCIZIO 4.

Un giocatore professionista di basket in un certo momento di una partita deve tirare due tiri liberi in sequenza. Si supponga di sapere che: la probabilità che metta a segno il primo tiro sia 0.1; la probabilità che realizzi il secondo tiro, dato che il primo sia andato a segno, sia pari a 0.2 e che la probabilità che il giocatore realizzi il secondo tiro, dato che il primo NON sia andato a segno sia 0.45.

1. Calcolare la probabilità che realizzi il canestro in entrambi i tiri.

Indicati con S_1 e S_2 gli eventi 'il primo tiro a segno' e 'il secondo tiro a segno' sono disponibili le seguenti informazioni:

$$P(S_1) = 0.1, P(S_2|S_1) = 0.2, P(S_2|\bar{S}_1) = 0.45.$$

Applicando la probabilità composta abbiamo:

$$P(S_1 \cap S_2) = P(S_1) \cdot P(S_2|S_1) = 0.1 \cdot 0.2 = 0.02$$

2. Dire, motivando la risposta, se i risultati dei due tiri sono indipendenti.

Dal momento che $P(S_2|S_1) = 0.2 \neq P(S_2|\bar{S}_1) = 0.45$ i risultati dei due tiri non sono indipendenti.

3. Calcolare la probabilità che realizzi il canestro in uno solo dei due tiri (considerare tutte le possibili coppie di tiri).

$$\begin{aligned} P((S_1 \cap \bar{S}_2) \cup (\bar{S}_1 \cap S_2)) &= P(S_1 \cap \bar{S}_2) + P(\bar{S}_1 \cap S_2) = \\ &= P(S_1) \cdot P(\bar{S}_2|S_1) + P(\bar{S}_1) \cdot P(S_2|\bar{S}_1) = \\ &= P(S_1) \cdot (1 - P(S_2|S_1)) + (1 - P(S_1)) \cdot P(S_2|\bar{S}_1) = \\ &= 0.1 \cdot 0.8 + 0.9 \cdot 0.45 = 0.08 + 0.405 = 0.485 \end{aligned}$$

4. Sapendo che alla fine della stagione il giocatore ha fatto 150 tiri liberi singoli, si calcoli la probabilità che abbia messo a segno almeno 20 tiri, ipotizzando che i risultati dei tiri liberi singoli siano tra loro indipendenti.

Indicata con X la variabile casuale 'n. di tiri a segno tra i 150 tiri liberi singoli indipendenti' risulta $X \sim \text{Bin}(n = 150, p = 0.1)$. $M(X) = 15$; $\text{Var}(X) = 13.5$.

Dal momento che $np = 15 > 5$ ed $n(1 - p) = 135 > 5$ è possibile utilizzare l'approssimazione della variabile casuale Binomiale con la variabile casuale Normale

$$X \sim N(\mu = 15, \sigma^2 = 13.5).$$

Si ottiene

$$P(X \geq 20) = P\left(\frac{X - \mu}{\sigma} \geq \frac{20 - \mu}{\sigma}\right) = P\left(Z \geq \frac{20 - 15}{3.6742}\right) = P(Z \geq 1.36) = 1 - P(Z \leq 1.36) = 1 - 0.9132 = 0.0868.$$