PROVA SCRITTA DI STATISTICA (COD 4038-5047-371-377)
26 Marzo 2004

MODALITÀ A

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (9 punti)

a)

Di seguito sono riportati alcuni dati relativi ai clienti di un'azienda che opera nel settore delle macchine utensili.
Per ciascuna azienda cliente sono riportati:

AZIENDA — Codice azienda
FATTURATO — Fatturato in milioni di euro
AREA — Area Geografica (NO=Nord Ovest, NE=Nord Est, C=Centro, S=Sud e Isole)
NUMERO DIPENDENTI IN CLASSI
ANZIANITA' COME CLIENTI (in anni)
SETTORE DI APPARTENENZA (quattro modalità distinte S1, S2, S3, S4)

Azienda	FATTURATO	AREA	NUMERO DIPENDENTI IN CLASSI	ANZIANITA' COME CLIENTE (anni)	SETTORE	FATTURATO^2	ANZIANITA' COME CLIENTE^2
A	15	NO	[1-5)	5	S1	225	25
B	12	NE	[5-10)	6	S2	144	36
C	11	NO	[10-50]	12	S4	121	144
D	13	C	[10-50]	10	S1	169	100
E	9	NO	[5-10)	11	S2	81	121
F	11	C	[5-10)	9	S3	121	81
G	10	C	[5-10)	7	S4	100	49
H	7	S	[1-5)	11	S3	49	121
I	8	S	[1-5)	6	S1	64	36
L	20	S	[1-5)	12	S1	400	144
SOMMA	116			89		1474	857

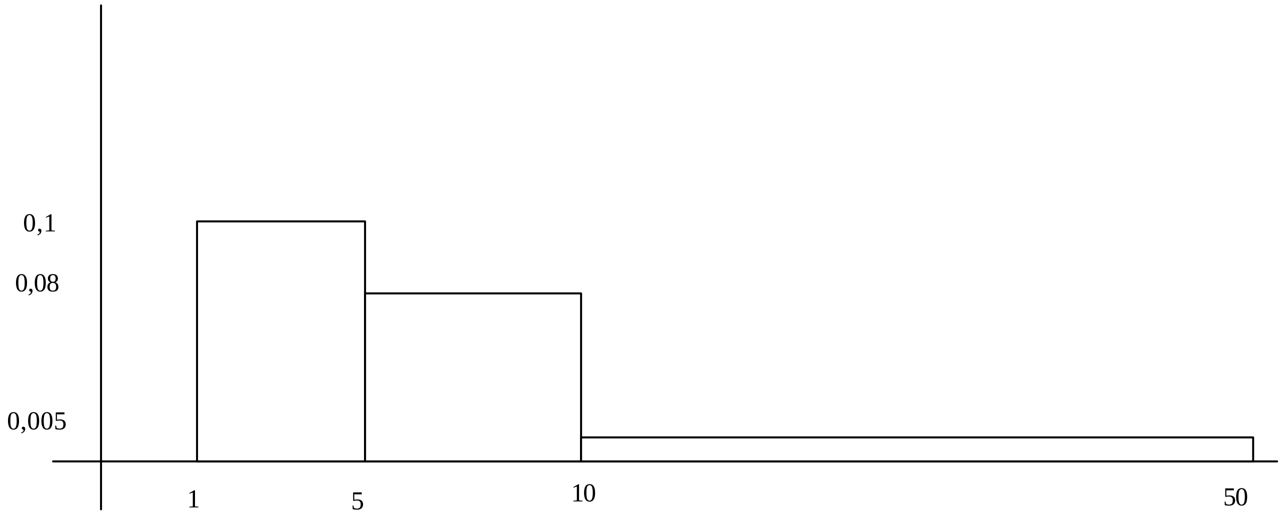
a1) *Fornire un'opportuna rappresentazione grafica del carattere NUMERO DIPENDENTI IN CLASSI*.

Trattandosi di variabile statistica continua per intervallo, si richiede la costruzione di un'istogramma.

La tabella per il calcolo della densità di frequenza è la seguente:

classe	Frequenza assoluta	Frequenza relativa	Densità di frequenza
[1-5)	4	0,4	0,4/4=0,1
[5-10)	4	0,4	0,4/5=0,08
[10-50]	2	0,2	0,2/40=0,005

E il grafico conseguente:



a2) *Calcolare media e mediana del FATTURATO.*

La media aritmetica è pari a 116/10=11,6

Per calcolare la mediana occorre ordinare i dati:

7 8 9 10 11 11 12 13 15 20

Pertanto la mediana, essendo la media aritmetica dei valori di posizione 5 e 6, è pari a 11.

a3) *La distribuzione del FATTURATO è simmetrica? Giustificare la risposta.*

La distribuzione non è sicuramente simmetrica in quanto la mediana è diversa dalla media aritmetica

a4) *Il FATTURATO è regressivamente indipendente dal SETTORE? Giustificare la risposta.*

Si verifica con facilità che non vi è indipendenza regressiva considerando le distribuzioni subordinate a SETTORE di FATTURATO e quindi le medie delle distribuzioni stesse (si veda p.es. S2 ed S3):

S1: 8, 13, 15, 20
S2: 9, 12
S3: 9, 11
S4: 7, 12

b)

Si consideri la tabella seguente, che riporta la distribuzione bidimensionale di due variabili quantitative X e Y. Si dispongano le frequenze congiunte relative in modo che vi sia perfetta connessione ma non perfetta dipendenza lineare:

$X\backslash Y$	6	8	10	
0				
1				
2				
				1

Un esempio possibile è il seguente:

$X\backslash Y$	6	8	10	
0		0,2		
1	0,3			
2			0,5	
				1

ESERCIZIO 2 (6 punti)

a) *Considerando 30 titoli azionari, solo 5 di questi hanno attualmente una quotazione superiore a 10 anni fa. Si estraggono a caso con reimmissione 6 titoli.*

Qual è la probabilità che

a1) Nessun titolo estratto abbia una quotazione superiore a 10 anni fa

La distribuzione da utilizzare è la binomiale, con parametri n=6 e probabilità di successo p=5/30=1/6=0,1667.

Pertanto $$p_X(x) = \binom{6}{x}\left(\frac{1}{6}\right)^x\left(\frac{5}{6}\right)^{6-x} \qquad x = 0,1,2,\dots,6$$

In questo caso siamo interessati a

$$p_X(0) = \binom{6}{0}\left(\frac{1}{6}\right)^0\left(\frac{5}{6}\right)^{6-0} = 0{,}3349 = 33{,}49\%$$

a2) *Almeno 2 titoli estratti abbiano una quotazione inferiore a 10 anni fa.*

In questo caso la distribuzione da utilizzare è sempre la binomiale, con parametri n=6 e probabilità di successo p=25/30=5/6=0,8333.

Pertanto $$p_X(x) = \binom{6}{x}\left(\frac{5}{6}\right)^x \left(\frac{1}{6}\right)^{6-x} \qquad x = 0,1,2,...,6$$

e siamo interessati in questo caso a:

$$\Pr\{X \geq 2\} = 1 - \Pr\{X = 0\} - \Pr\{X = 1\} = 1 - \binom{6}{0}\left(\frac{5}{6}\right)^0 \left(\frac{1}{6}\right)^{6-0} - \binom{6}{1}\left(\frac{5}{6}\right)^1 \left(\frac{1}{6}\right)^{6-1} = 1 - 0{,}000021 - 0{,}00064 \cong 0{,}9994$$

b) Si consideri una variabile aleatoria X con varianza V(X)=3. Si calcoli la varianza della variabile aleatoria Y=-3X + 4.

Vale il seguente risultato per una trasformazione lineare:

$$V(Y)=V(a+bX)=b^2V(X)$$

Pertanto nel caso specifico V(Y)=9 V(X)=27.

c) *Si enunci il teorema di Bayes.*

Si consideri uno spazio campionario Ω e un insieme di eventi A_1, A_2,…, A_n che soddisfano le relazioni $\Omega = \bigcup_{j=1}^{n} A_j$ e $A_i \bigcap A_j = \varnothing$ $i^1 j$ $P(A_j)>0$ per $j=1,2,...,n$; preso un qualunque altro evento B con P(B)>0, si ha che

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{n} P(B|A_j)P(A_j)}$$

ESERCIZIO 3 (12 punti)

a) Si consideri una popolazione con distribuzione uniforme continua tra 2 e 6 e un campione bernoulliano di ampiezza 100 estratto da tale popolazione. Si calcoli la varianza dello stimatore media campionaria.

La varianza della media campionaria è pari a $\frac{s^2}{n}$, dove s^2 è pari alla varianza della popolazione. Nel caso di un'uniforme continua tra a e b, la varianza è pari a $(b-a)^2/12$. Pertanto, la varianza della media campionaria è pari a =4/300.

b)

b1) *Si consideri una popolazione normale X con media m e con varianza nota e pari a 4. Si fornisca l'espressione dell'intervallo di confidenza per m con livello di confidenza fissato ad a=5%.*

L'intervallo di confidenza richiesto si ricava dalla seguente espressione:

$$\overline{X} - z_{0,975}\frac{\sqrt{4}}{\sqrt{n}} \;;\; \overline{X} + z_{0,975}\frac{\sqrt{4}}{\sqrt{n}}$$

dove $\overline{X}$ è pari alla media campionaria, n è l'ampiezza campionaria e $z_{0,975}$=1,96 è il valore corrispondente a una probabilità cumulata del 97,5% nella distribuzione normale standardizzata.

b2) *Si calcoli l'intervallo di confidenza corrispondente alla realizzazione campionaria (3,4 ; 4,5 ; 4,8 ; 3,8 ; 3,1).*

In questo caso, n=5, $\overline{x} = 3{,}92$ e pertanto l'intervallo cercato è

$$3{,}92 \pm 1{,}96\frac{\sqrt{4}}{\sqrt{5}} \;;\; 3{,}92 \pm 1{,}7530$$

$$(2{,}167 \;;\; 5{,}673)$$

b3) *Quale deve essere la numerosità campionaria minima per garantire che la lunghezza dell'intervallo proposto al punto (1) non superi in ampiezza 1,5?*

In questo caso la lunghezza dell'intervallo è data dalla seguente espressione

$$2 \cdot z_{0,975}\frac{s}{\sqrt{n}}$$

Pertanto, la determinazione dell'ampiezza campionaria minima richiesta si basa sull'espressione:

$$2 \cdot 1{,}96\frac{\sqrt{4}}{\sqrt{n}} \leq 1{,}5$$

da cui si ricava

$$\sqrt{n} \geq \frac{7{,}84}{1{,}5}$$

$$\sqrt{n} \geq 5{,}2267$$

$$n \geq 27{,}32 \; da\, cui$$

$$n = 28$$

c)

c1) *Si consideri una popolazione bernoulliana X di parametro p. Ipotizzando n sufficientemente grande per utilizzare i risultati asintotici, si fornisca l'espressione della regione di rifiuto per provare:*

$$H_0 : p \geq 0{,}75 \quad contro \quad H_1 : p < 0{,}75$$

con una dimensione del test fissata al livello α=5%.

Definito $\hat{P} = \overline{X}$ lo stimatore della percentuale p e tenendo conto del teorema centrale del limite, si può utilizzare la seguente regione di rifiuto:

$$R \equiv \left\{ \underline{x} : \hat{p} < p_0 - z_{0,95} \sqrt{\frac{p_0(1-p_0)}{n}} \right\}$$

dove n è l'ampiezza campionaria e $z_{0,95}$=1,645 è il valore corrispondente a una probabilità cumulata del 95% nella distribuzione normale standardizzata. Nel caso specifico si ha pertanto:

$$R \equiv \left\{ \underline{x} : \hat{p} < 0{,}75 - 1{,}645 \sqrt{\frac{0{,}75 \cdot 0{,}25}{n}} \right\}$$

c2) *Avendo estratto un campione bernoulliano di ampiezza n=200 e avendo riscontrato un numero di successi pari a 120, accettereste o rifiutereste l'ipotesi* H_0*?*

In questo caso pertanto si ha $\hat{p} = \frac{120}{200} = 0{,}6$ ed n=200. La regione di rifiuto è pertanto:

$$R \equiv \left\{ \underline{x} : \hat{p} < 0{,}75 - 1{,}645 \sqrt{\frac{0{,}75 \cdot 0{,}25}{200}} \right\}$$

$$R \equiv \left\{ \underline{x} : \hat{p} < 0{,}75 - 1{,}645 \sqrt{0{,}0009} \right\}$$

$$R \equiv \left\{ \underline{x} : \hat{p} < 0{,}75 - 0{,}0493 \right\}$$

$$R \equiv \left\{ \underline{x} : \hat{p} < 0{,}7007 \right\}$$

Considerando la realizzazione campionaria si rifiuta l'ipotesi nulla.

c3) *Si calcoli la probabilità di errore di seconda specie corrispondente a p=0,65*

Se p=0,65, lo stimatore $\hat{P}$ sarà caratterizzato asintoticamente da una distribuzione normale di valore atteso 0,65 e varianza $\frac{0{,}65 \cdot 0{,}35}{200}$=0,0011.
La probabilità di errore di seconda specie corrispondente a p=0,65 si ricava pertanto dalla seguente espressione (con Z caratterizzata da una distribuzione normale standardizzata):

$$b(0{,}65) = \Pr\{A | p = 0{,}65\} = \Pr\{\hat{p} > 0{,}7007 | p = 0{,}65\} = \Pr\left\{ Z > \frac{0{,}7007 - 0{,}65}{\sqrt{0{,}0011}} \right\} =$$

$$= \Pr\left\{ Z > \frac{0{,}0507}{0{,}0332} \right\} = \Pr\{Z > 1{,}53\} = 1 - 0{,}9370 = 0{,}063$$

d) *Si definisca la proprietà di consistenza di uno stimatore in senso forte.*

Si consideri una successione di stimatori T_n di un parametro $q \in \Theta$. T_n è consistente in senso forte per q se

$$\lim_{n\to\infty} E\left[(T_n - q)^2\right] = 0 \qquad \forall q \in \Theta$$

PROVA SCRITTA DI STATISTICA (COD 4038-5047-371-377)
26 Marzo 2004

MODALITÀ B

APPROSSIMARE TUTTI I CALCOLI ALLA QUARTA CIFRA DECIMALE

ESERCIZIO 1 (9 punti)

a)
Di seguito sono riportati alcuni dati relativi ai clienti di un'azienda che opera nel settore delle macchine utensili.
Per ciascuna azienda cliente sono riportati:

AZIENDA	Codice azienda
FATTURATO	Fatturato in milioni di euro
AREA	Area Geografica (NO=Nord Ovest, NE=Nord Est, C=Centro, S=Sud e Isole)

ANZIANITA' COME CLIENTI (in anni)
SETTORE DI APPARTENENZA (quattro modalità distinte S1, S2, S3, S4)

Azienda	FATTURATO	AREA	ANZIANITA' COME CLIENTE (anni)	SETTORE	FATTURATO^2	ANZIANITA' COME CLIENTE^2	FATTURATO x ANZIANITA' COME CLIENTE
A	15	NO	5	S1	225	25	75
B	12	NE	6	S2	144	36	72
C	11	NO	12	S4	121	144	132
D	13	C	10	S1	169	100	130
E	11	NO	11	S2	121	121	121
F	9	C	9	S3	81	81	81
G	10	C	7	S4	100	49	70
H	7	S	11	S3	49	121	77
I	8	S	6	S1	64	36	48
L	20	S	12	S1	400	144	240
SOMMA	116		89		1474	857	1046

a1) *Costruire la distribuzione di frequenza, calcolando le frequenze relative, della variabile SETTORE*.

La distribuzione di frequenza richiesta è la seguente:

Settore	Frequenza assoluta	Frequenza relativa
S1	4	40%
S2	2	20%
S3	2	20%
S4	2	20%
Totale	10	100%

a2) *Calcolare mediana e scarto quadratico medio di ANZIANITA' COME CLIENTE.*

Per calcolare la mediana, occorre ordinare in modo non decrescente i dati di *ANZIANITA' COME CLIENTE*:

5 6 6 7 9 10 11 11 12 12

Pertanto la mediana, essendo la media aritmetica dei valori di posizione 5 e 6, è pari a 9,5.

Lo scarto quadratico medio può essere calcolato facilmente utilizzando i dati forniti nella tabella:

$$s = \sqrt{857/10-(89/10)^2} = \sqrt{85{,}7-79{,}21} = \sqrt{6{,}49} = 2{,}5475$$

a3) *La distribuzione di ANZIANITA' COME CLIENTE è obliqua destra? Giustificare la risposta.*

Per essere obliqua destra occorrerebbe che la mediana risultasse inferiore alla media aritmetica. In questo caso invece la mediana, essendo pari a 9,5, è superiore alla media aritmetica, uguale a 8,9. La risposta è pertanto negativa.

a4) *Le variabili FATTURATO e ANZIANITA' COME CLIENTE sono linearmente indipendenti? Giustificare la risposta.*

Per rispondere, è sufficiente calcolare la covarianza e verificare se è nulla o meno.
In questo caso, utilizzando i risultati forniti in tabella:

$$Cov(F,A) = 1046/10 - 116/10 \cdot 89/10 = 104{,}6 - 103{,}24 = 1{,}36$$

da cui si ricava che le due variabili non sono linearmente indipendenti.

b**)**

Si consideri la tabella seguente, che riporta la distribuzione bidimensionale di due variabili quantitative X e Y. Si dispongano le <u>*frequenze congiunte relative*</u> *in modo che Y risulti regressivamente indipendente da X, ma che le due variabili non siano statisticamente indipendenti.*

X\Y	*6*	*8*	*10*	
0				
1				
2				
				1

Un esempio potrebbe essere il seguente:

X\Y	*6*	*8*	*10*	
0		0,25		
1	0,3		0,3	
2		0,15		
				1

ESERCIZIO 2 (6 punti)

a) Considerando 20 titoli azionari, solo 4 di questi hanno attualmente una quotazione superiore a 10 anni fa. Si estraggono a caso con reimmissione 5 titoli.

Qual è la probabilità che:

a1) Tutti i titoli estratti abbiano una quotazione superiore a 10 anni fa

La distribuzione da utilizzare è la binomiale, con parametri n=5 e probabilità di successo p=4/20=1/5=0,2.

Pertanto $$p_X(x) = \binom{5}{x}(0{,}2)^x(0{,}8)^{5-x} \qquad x = 0,1,2,...,5$$

In questo caso siamo interessati a

$$p_X(5) = \binom{5}{5}(0{,}2)^5(0{,}8)^0 = 0{,}00032 = 0{,}032\%$$

a2) Almeno 2 titoli estratti abbiano una quotazione inferiore a 10 anni fa

In questo caso la distribuzione da utilizzare è sempre la binomiale, con parametri n=5 e probabilità di successo p=16/20=4/5=0,8.

Pertanto $$p_X(x) = \binom{5}{x}(0{,}8)^x(0{,}2)^{5-x} \qquad x = 0,1,2,...,5$$

e siamo interessati in questo caso a:

$$\Pr\{X \geq 2\} = 1 - \Pr\{X = 0\} - \Pr\{X = 1\} = 1 - \binom{5}{0}(0{,}8)^0(0{,}2)^{5-0} - \binom{5}{1}(0{,}8)^1(0{,}2)^{5-1} = 1 - 0{,}00032 - 0{,}0064 \cong 0{,}9933$$

b) *Si consideri una variabile aleatoria X con valore atteso E(X)=4 e una variabile aleatoria Y con valore atteso E(Y)=6. Si calcoli il valore atteso della variabile aleatoria Z=4X – 6Y.*

Per le proprietà dell'operatore valore atteso:

$$E(Z) = E(aX + bY) = aE(X) + bE(Y)$$

In questo caso pertanto:

$$E(Z) = E(4X - 6Y) = 4E(X) - 6E(Y) = 4 \cdot 4 - 6 \cdot 6 = -20$$

c) *Si enunci il teorema delle probabilità totali.*

Si consideri uno spazio campionario Ω e un insieme di eventi A_1, A_2,…, A_n che soddisfano le relazioni $\Omega = \bigcup_{j=1}^{n} A_j$ $A_i \bigcap A_j = \phi$ i^1j $P(A_j)>0$ per $j=1,2,...,n$; preso un qualunque altro evento B si ha che

$$P(B) = \sum_{j=1}^{n} P(B|A_j)P(A_j)$$

ESERCIZIO 3 (12 punti)

a) *Si consideri una popolazione con distribuzione bernoulliana di parametro p=0,6. e un campione bernoulliano di ampiezza 100 estratto da tale popolazione. Si calcoli la probabilità che lo stimatore media campionaria assuma valori superiori a 0,4.*

Per il teorema centrale del limite, asintoticamente lo stimatore $\hat{P} = \overline{X}$ assume una distribuzione normale di media p e di varianza $\frac{p(1-p)}{n}$. Nel caso specifico, si ha pertanto che la distribuzione asintotica è normale di media 0,6 e varianza 0,0024. Pertanto la probabilità cercata è:

$$P\{\hat{P} > 0,4\} = P\left\{Z > \frac{0,4 - 0,6}{\sqrt{0,0024}}\right\} = P\{Z > -4,082\} \cong 1$$

b)

b1) *Si consideri una popolazione normale X con media m e con varianza non nota. Si fornisca l'espressione dell'intervallo di confidenza per m con livello di confidenza fissato ad a=1%.*

L'intervallo richiesto è dato dalla seguente espressione:

$$\overline{X} - t_{0,995}^{n-1} \frac{S_c}{\sqrt{n}} \ ; \ \overline{X} + t_{0,995}^{n-1} \frac{S_c}{\sqrt{n}}$$

dove $\overline{X}$ è pari alla media campionaria, S_c è la radice della varianza campionaria corretta, n è l'ampiezza campionaria e $t_{0,995}^{n-1}$ è il valore corrispondente a una probabilità cumulata del 99,5% nella distribuzione T di Student con *n-1* gradi di libertà.

b2) *Si calcoli l'intervallo corrispondente alla realizzazione campionaria (5 ; 4,5 ; 5,8 ; 6,4 ; 3,3 ; 6,6).*

In questo caso $\overline{x} = 5,2667, s_c = 1,2547, n = 6$ e pertanto l'intervallo cercato è

$$5,2667 \pm 4,032\frac{1,2549}{\sqrt{6}} \; ; \; 5,2667 \pm 2,0656$$
$$(3,204 \; ; \; 7,3323)$$

b3) Che cosa si intende per *livello di confidenza*?

La *confidenza* è la frequenza con cui un elevato numero di intervalli, costruiti come indicato nei punti precedenti ripetendo più volte il campionamento, conterrebbe il vero valore del parametro incognito.

c)

c1) *Si considerino due popolazioni normali X_1 e X_2 rispettivamente con media m_1 e m_2 e con varianze non note, ma uguali. Si fornisca l'espressione della regione di rifiuto per provare:*

$$H_0 : m_1 - m_2 = 0 \quad contro \quad H_1 : m_1 - m_2 \neq 0$$

con una dimensione del test fissata al livello α=10%

Definite rispettivamente $\overline{X}_1$ e $\overline{X}_2$ le medie campionarie ottenute a partire dai campioni estratti dalle popolazioni X_1 e X_2, di ampiezza n_1 e n_2 , nonché $S_p^2 = \frac{(n_1 - 1)S_{1c}^2 + (n_2 - 1)S_{2c}^2}{n_1 + n_2 - 2}$, dove S_{1c}^2 e S_{2c}^2 sono le varianze campionarie corrette, si può utilizzare la seguente regione di rifiuto:

$$R \equiv \left\{ \underline{x}_1 ; \underline{x}_2 : \left| \overline{x}_1 - \overline{x}_2 \right| > {}_{n_1+n_2-2}t_{0,95} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right\}$$

dove ${}_{n_1+n_2-2}t_{0,95}$ è il valore corrispondente a una probabilità cumulata del 95% nella distribuzione *T di* Student con n_1+n_2-2 gradi di libertà.

c2) *Supponendo di aver estratto da X_1 e X_2 due campioni rispettivamente di ampiezza n_1=10 e n_2=15, con media pari a 20,1 e 24,6 e varianze campionarie corrette S_{1c}^2=30 e S_{2c}^2=35, si dica, giustificando la risposta, se accettereste o rifiutereste l'ipotesi nulla.*

Nel caso specifico, si ha $s_p^2 = \frac{9 \cdot 30 + 14 \cdot 35}{23} = \frac{760}{23} = 33,0435$ e pertanto:

$$R \equiv \left\{ \underline{x}_1 ; \underline{x}_2 : \left| \overline{x}_1 - \overline{x}_2 \right| > 1,714 \cdot 5,7483\sqrt{\frac{1}{10} + \frac{1}{15}} \right\}$$
$$R \equiv \left\{ \underline{x}_1 ; \underline{x}_2 : \left| \overline{x}_1 - \overline{x}_2 \right| > 1,714 \cdot 5,7483 \cdot \sqrt{0,1667} \right\}$$

$$R \equiv \left\{ \underline{x}_1 ; \underline{x}_2 : \left| \bar{x}_1 - \bar{x}_2 \right| > 4{,}0228 \right\}$$

Essendo la differenza tra le due medie pari a 4,5, si rifiuta l'ipotesi nulla.

d) *Si enunci il teorema centrale del limite.*

Si consideri un campione bernoulliano di ampiezza n estratto da una popolazione con media m e varianza s^2 finite; sia inoltre $\overline{X}_n$ la media campionaria calcolata utilizzando tale campione.
La variabile aleatoria

$$Z_n = \frac{\overline{X}_n - m}{s / \sqrt{n}}$$

per n che tende a infinito assume una distribuzione normale standardizzata.