CORSO DI STATISTICA DI BASE (Prof. GIORGIO ALLEVA)

Anno Accademico 2019-2020

Prova scritta del 4 febbraio 2020

PROVA A

ESERCIZI

Esercizio 1 (9 punti). Si disponga della seguente distribuzione di frequenza delle variabili X e Y.

X\Y	*1-3*	*3-9*	*9-19*	*Tot*
0	-	4	15	19
5	-	9	2	11
7	10	-	-	10
Tot	10	13	17	40

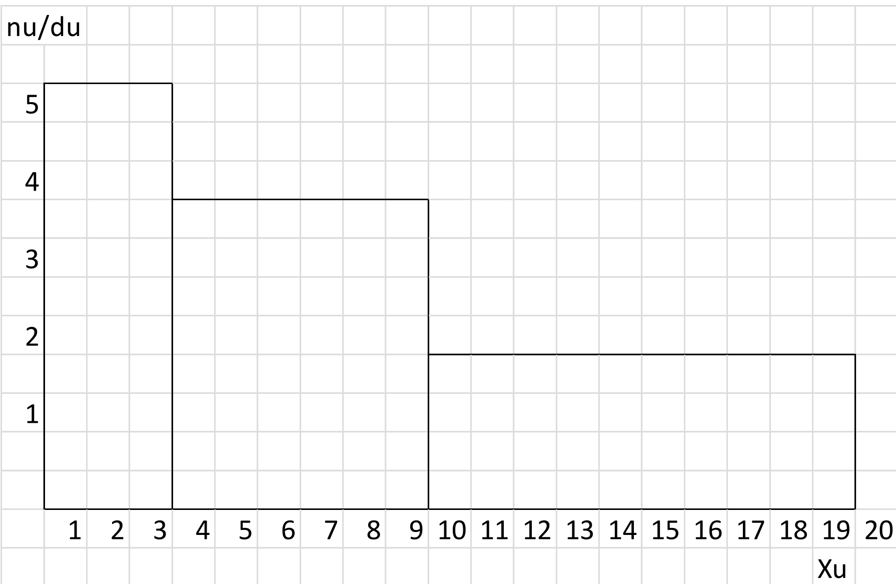
a) si costruisca l'istogramma della distribuzione del carattere continuo Y e si determini la moda.

Svolgimento

Yu	*Frequenza (nu)*	*Ampiezza (du)*	*densità di frequenza (nu/du)*
1-3	*10*	*2*	*5*
3-9	*13*	*4*	*3,25*
9-19	*17*	*10*	*1,7*

La classe modale è 1-3, la classe che presenta la massima densità di frequenza. La moda può essere stimata attraverso il valore centrale di tale classe, pari a 2.

ISTOGRAMMA



b) si misuri la dipendenza assoluta tra i due caratteri, valutandone l'intensità rispetto al massimo che può assumere.

Svolgimento. Come misura di dipendenza assoluta si può considerare l'indice Chi quadrato e per valutare l'intensità della dipendenza l'indice Chi quadrato relativo o l'indice di Cramer.

Per il calcolo di chi quadro si può calcolare la somma dei rapporti tra i quadrati delle contingenze e le frequenze teoriche.

Chi quadrato =

$(22,56^2/4,75)+(4,73^2/6,18)+(47,96^2/8,08)+(7,56^2/2,75)+(29,43^2/3,58)+(7,16^2/4,68)+(56,25^2/2,5)+$

$+ (10,56^2/3,25)+(18,06^2/4,25) = 53,97$

Chi quadro relativo 53,97/80 = 0,6746   *Indice di Cramer = rad(0,6746) = 0,8213*

c) si calcoli la covarianza tra X e Y; si indichino, in generale, quali sono gli estremi dell'intervallo in cui è compresa (indicare l'espressione senza fare calcoli).

Svolgimento. $\sigma_{XY} = Cov(X,Y) = \mu_{XY} - \mu_X \mu_Y = 13,75 - 3,125x8,4 = -12,5$

Infatti, $\mu_X = (0x19+5x11+7x10)/40 = 3,125$ *e, considerando i valori centrali delle classi di Y*

$\mu_Y = (2x10+6x13+14x17)/40 = 8,4$

$\mu_{XY} = (0x2x0+0x6x4+0x14x15+5x2x0+5x6x9+5x14x2+7x2x10+7x6x0+7x14x0)/40 = 550/40 = 13,75$

La covarianza è compresa tra $-\sigma_X\sigma_Y$ *e* $+\sigma_X\sigma_Y$.

d) si calcoli la devianza esterna di Y; si indichino, in generale, quali sono gli estremi dell'intervallo in cui è compresa (indicare l'espressione senza fare calcoli). Quali situazioni configurano tali due estremi?

Svolgimento. La devianza esterna è l

a devianza delle medie condizionate di Y rispetto a X.

Le tre medie condizionate Y, calcolate considerando i valori centrali delle classi, sono:

$\mu_{Y|/X1} = 12,32 \quad \mu_{Y|/X2} = 7,455 \quad \mu_{Y|/X3} = 2 \quad \mu_Y = 8,4$

$Dev\ Est(Y) = \Sigma(\mu_{Y|/XI} - \mu_Y)^2 n_i = (12,32-8,4)^2 19 + (7,455-8,4)^2\ 11 + (2-8,4)^2 10 = 861,27$

La devianza esterna è compresa tra 0 (caso di indipendenza in media di Y da X, in cui $\eta^2_{Y|/X} = 0$*) e Dev Tot(Y) (caso di massima dipendenza in media di Y da X, in cui* $\eta^2_{Y|/X} = 1$*).*

Esercizio 2 (4 punti). La serie storica del numero indice dei prezzi dei trasporti in Italia in base 2015.

	2016	2017	2018	2019
N.I. (2015=100)	98,6	102,0	104,8	105,6

a) si calcoli la variazione percentuale dei prezzi dei trasporti nell'ultimo biennio.
$p_{2019}/p_{2017} = 105,6/104,8 = 1,0076$ *e quindi Var % nell'ultimo biennio è (1,0076-1)x100 = +0,76%.*

b) si calcoli la variazione % media annua dei prezzi dei trasporti tra il 2015 e il 2019.
$p_{2019}/p_{2015} = 105,6/98,6 = 1,8343$ *e quindi Var % media annua è* $(1,8343^{1/4} - 1)\ x100 = +16,38\%$.

Esercizio 3 (8 punti). Sia X una variabile che si distribuisce normalmente con media 15 e varianza 25. Sulla base di un campione estratto con ripetizione di n = 30 unità:

a) si determini la probabilità che la media campionaria sia compresa tra 12 e 16;
Svolgimento. La media campionaria si distribuisce come una normale $(\mu; \sigma^2/n)$ *quindi si tratta di trovare:*
$p\left(\frac{12-15}{\frac{5}{\sqrt{30}}} \le z \le \frac{16-15}{\frac{5}{\sqrt{30}}}\right) = p(-3,289 \le z \le 1,095) = \Phi(1,095) - [(1-\Phi(3,289)] =$

$= 0,8643 - 0,0005 = 0,8638$

b) si indichino quali sono le caratteristiche dello stimatore della media campionaria qualora le n v.c. estrazione siano NIID.
La media campionaria in questo caso è uno stimatore non distorto (il suo valore atteso è μ*) e consistente (al crescere di n la sua varianza tende a 0).*

c) Sulla base di un campione di n=34 unità estratto con ripetizione da una popolazione normale con media e varianza ignota, la devianza campionaria $\sum(X_i - \bar{X})^2$ sia risultata pari a 125. Si verifichi, con $\alpha = 0,05$, se la varianza della popolazione possa essere assunta superiore a 4.
Svolgimento. H_0: $\sigma^2 = 4$; H_1: $\sigma^2 > 4$. *Si cosce la devianza campionaria* $\sum(X_i - \overline{X})^2 = 125$.
Non si conosce μ *quindi la regola decisionale è la seguente:*
Rifiutiamo H_0 *se* $\frac{\sum(X_i-\overline{X})^2}{\sigma_0^2} \ge \chi^2_{\alpha;n-1}$
Poiché $125/4 = 31,25 < \chi^2_{0,05;34} = 47,4$ *non rifiutiamo* H_0

Esercizio 4 (6 punti). La probabilità di fare uso di doping in una certa disciplina sportiva sia ritenuta essere pari a 0,2. I controlli sugli atleti sono effettuati sulla base di una procedura di campionamento indicata nel regolamento internazionale. La procedura, basandosi sull'osservazione delle prestazioni individuali degli atleti nel corso del tempo, ha consentito nel passato che con probabilità pari 0,8 venissero selezionati per i controlli atleti che avevano fatto uso di doping, e con probabilità pari 0,4 che fossero selezionati atleti non ne avevano fatto uso. Sulla base di tali informazioni si determini:

a) la probabilità che un atleta venga selezionato per il controllo antidoping.

   Svolgimento. Indicando con p(D)=0,2 e p(nD)=0,8 la probabilità che un atleta faccia uso e non faccia uso di doping e con p(S|D)=0,8 e p(S|nD)=0,4 la probabilità di selezionare un atleta che abbia fatto o non fatto uso di doping, la probabilità che un atleta venga selezionato per il controllo antidoping è:

   p(S) = p(S|D) p(D) + p(S|nD) p(nD) = 0,8x0,2+0,4x0,8 = 0,48

b) la probabilità che, avendo selezionato un atleta per effettuare controlli, questo risulti aver fatto uso di doping.

   Svolgimento. p(D|S) = p(D∩S)/p(S) = [p(D) p(S|D)] / p(S) = 0,2x0,8/0,48 = 0,333

c) in quale misura la procedura adottata consente risultati migliori nell'individuare atleti che fanno uso di doping rispetto ad un'estrazione a sorte di un atleta da sottoporre a controlli?

   Svolgimento. La procedura fa crescere la probabilità di individuare atleti che fanno uso di doping, infatti P(D|S)=0,33 > p(D)=0,2 . L'aumento è pari a [(0,33/0,2)-1] x 100 = + 65%

QUESITI (barrare la risposta ritenuta esatta)

(PUNTI 2 per risposta corretta, PUNTI -1 per risposta sbagliata, PUNTI 0 per assenza di risposta)

Conoscendo che il coefficiente angolare della retta di regressione di Y su X sia 2, quale potrebbe essere il valore del coefficiente angolare della retta di regressione di X su Y:

- ❒ 0,25
- ❒ 0,5
- ☒ Entrambi

Qualora la correlazione $r_{XZ} = r_{YZ} = 0$, quale espressione assume il coefficiente di correlazione parziale $r_{XY.Z}$?

- ❒ $r_{XY.Z}$ è indeterminato
- ☒ $r_{XY.Z} = r_{XY}$
- ❒ $r_{XY.Z} = 0$

Un rapporto di durata è:

- ❒ un rapporto tra distanze temporali
- ❒ un rapporto tra una variabile di flusso e una variabile di stock
- ☒ un rapporto tra una variabile di stock e una variabile di flusso

Siano Z e W due variabili trasformate lineari di X e Y. In particolare Z = a+bX e W =c+dY.
La covarianza tra Z e W è:

- ❒ $b^2 d^2$ Cov(X,Y)
- ☒ b d Cov(X,Y)
- ❒ |b||d| Cov(X,Y)

Uno stimatore è consistente se:

- ☒ converge in probabilità al parametro della popolazione al crescere della numerosità campionaria
- ❒ la sua varianza converge a 0 al crescere della numerosità campionaria
- ❒ il suo valore atteso converge al parametro della popolazione al crescere della numerosità campionaria

CORSO DI STATISTICA DI BASE (Prof. GIORGIO ALLEVA)

Anno Accademico 2019-2020

Prova scritta del 4 febbraio 2020

PROVA B

ESERCIZI

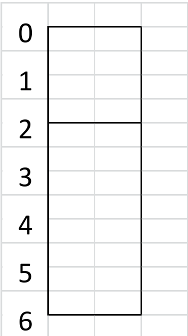
Esercizio 1 (9 punti). Si disponga della seguente distribuzione di frequenza delle variabili X e Y.

X\Y	*1*	*3*	*8*	*Tot*
0	13	-	-	13
2	1	11	3	15
6	-	-	22	22
Tot	14	11	25	50

a) si calcolino i tre quartili di X e si descriva la distribuzione di tale variabile attraverso il box plot.

Svolgimento. Rango Q1: 13; ranghi Q2: 25 e 26; rango Q3: 38. Quindi: Q1=0; Q2=2; Q3=6.

Box plot



b) si misuri la dipendenza assoluta tra i due caratteri, valutandone l'intensità rispetto al massimo che può assumere.

Svolgimento. Come misura di dipendenza assoluta si può considerare l'indice Chi quadrato e per valutare l'intensità della dipendenza l'indice Chi quadrato relativo o l'indice di Cramer.

Per il calcolo di chi quadro si può calcolare la somma dei rapporti tra i quadrati delle contingenze e le frequenze teoriche.

Chi quadrato = $(87,6^2/3,64)+(8,18^2/2,86)+(42,25^2/6,5)+(10,2^2/4,2)+(59,29^2/3,3)+(20,25^2/7,5)+(37,9^2/6,16)+$

$+ (23,43^2/4,84)+(20,25^2/11) = 78,53$

Chi quadro relativo 78,53/100 = 0,7853   Indice di Cramer = rad(0,7853) = 0,8862

c) si calcoli la covarianza tra X e Y; si indichino, in generale, quali sono gli estremi dell'intervallo in cui è compresa (indicare l'espressione senza fare calcoli).

Svolgimento. σ_{XY} = Cov(X,Y) = $\mu_{XY} - \mu_X \, \mu_Y$ = 23,44 – 3,24x4,94 = 7,43

Infatti, μ_X = (0x13+2x15+6x22)/50 = 3,24 e, considerando i valori centrali delle classi di Y

μ_Y = (1x14+3x11+8x25)/50 = 4,94

μ_{XY} = (0x1x13+0x3x0+0x8x0+2x1x2+2x3x11+2x8x3+6x1x0+6x3x0+6x8x22)/50 = 1.172/50 = 23,44

La covarianza è compresa tra $-\sigma_X\sigma_Y$ e $+\sigma_X\sigma_Y$.

d) si calcoli la devianza esterna di Y; si indichino, in generale, quali sono gli estremi dell'intervallo in cui è compresa (indicare l'espressione senza fare calcoli). Quali situazioni configurano tali due estremi?

Svolgimento. La devianza esterna è la devianza delle medie condizionate di Y rispetto a X.

Le tre medie condizionate Y, calcolate considerando i valori centrali delle classi, sono:

$\mu_{Y|X1}$ = 1 $\mu_{Y|X2}$ = 3,87 $\mu_{Y|X3}$ = 8 μ_Y = 4,94

Dev Est(Y) = $\Sigma(\mu_{Y|XI} - \mu_Y)^2 n_i$ = $(1-4,94)^2 13 + (3,87-4,94)^2\, 15 + (8-4,94)^2 22$ = 464,09

La devianza esterna è compresa tra 0 (caso di indipendenza in media di Y da X, in cui $\eta^2_{Y|X}$ =0) e Dev Tot(Y) (caso di massima dipendenza in media di Y da X, in cui $\eta^2_{Y|X}$ =1).

Esercizio 2 (4 Punti). Si conosca la variazione dei prezzi al consumo rispetto all'anno precedente nell'ultimo triennio in Italia (corrispondente al tasso d'inflazione).

	2017	2018	2019
Tasso d'inflazione %	1,3	1,2	0,6

a) si calcoli la variazione percentuale dei prezzi al consumo nell'intero triennio.
p_{2019}/p_{2016} = 1,0013 x 1,012 x 1,006 = 1,031   quindi Var % nel triennio è (1,031-1)x100 = +3,1%

b) si calcoli il tasso di inflazione medio annuo.
Avendo determinato al punto precedente che p_{2019}/p_{2016} = 1,0313 allora il tasso medio annuo è $(1{,}0313^{1/3} - 1)$ x 100 = + 1,03 %

Esercizio 3 (8 punti).

a) Conoscendo che la frazione di elettori che hanno votato il partito G alle scorse elezioni in un certo comune capoluogo è stata pari a al 22,5%, si determini la probabilità che in campione di 150 intervistati, estratto con ripetizione, la percentuale di elettori sia compresa tra il 25 e il 27%.
Svolgimento. Poiché si può assumere che la proporzione campionaria si distribuisca come una normale con media pari a p e varianza pari a p(1-p)/n si tratta di determinare la probabilità che Rn sia compresa tra 0,25 e 0,27. Standardizzando:

$$p\left(\frac{0{,}25-0{,}225}{\sqrt{\frac{0{,}225(1-0{,}225)}{150}}} \leq\ z \leq \frac{0{,}27-0{,}225}{\sqrt{\frac{0{,}225(1-0{,}225)}{150}}}\right) = p(0{,}7332 \leq\ z \leq 1{,}3198) = \Phi(1{,}3198) - \Phi(0{,}7332) =$$

= 0,9066-0,7673 = 0,1393

b) secondo quale teorema, e sotto quali condizioni, la proporzione campionaria converge in distribuzione ad una v.c. normale?
Svolgimento. Per il teorema di De Moivre: date n v.c. bernoulliane indipendenti con lo stesso parametro p, la loro media si distribuisce come una normale con media pari a p e varianza pari a p(1-p)/n. Si tratta del teorema del limite centrale nel caso di v.c. benoulliane anziché qualsiasi.

c) Sulla base di un campione di n = 35 unità estratto con ripetizione da una popolazione normale con media e varianza ignota, la media campionaria sia risultata pari a 3,5 e la devianza campionaria pari a 136. Si verifichi, con a = 0,05, se la media della popolazione possa essere assunta minore di 3,9.

Svolgimento. H_0: μ = 3,9; H_1: μ < 3,9. Si conoscono la media campionaria $\overline{X} = 3{,}5$ e la devianza campionaria $\sum(X_i - \overline{X})^2 = 136$. Non conoscendosi σ^2 si stima $\bar{s}^2 = \frac{\sum(X_i-\overline{X})^2}{n-1} = \frac{136}{34} = 4.$ Quindi $\bar{s} = 2.$
La regola decisionale è la seguente:
Rifiutiamo H_0 se $\overline{X} < \mu_0 - \frac{\bar{s}}{\sqrt{35}} t_{0,05;34} = 3{,}9 - \frac{2}{\sqrt{35}} 1{,}691 = 3{,}9 - 0{,}572 = 2{,}428.$
Poiché 3,5 >2,428 non rifiutiamo H_0.

Esercizio 4 (6 punti). La probabilità di doping in una certa disciplina sportiva sia pari a 0,15. I controlli sugli atleti sono effettuati sulla base di una procedura di campionamento indicata nel regolamento internazionale. La procedura, basandosi sull'osservazione delle prestazioni individuali degli atleti nel corso del tempo, ha consentito nel passato che con probabilità pari 0,75 venissero selezionati per i controlli atleti che avevano fatto uso di doping, e con probabilità pari 0,4 che fossero selezionati atleti non ne avevano fatto uso. Si determini:

a) la probabilità che un atleta venga selezionato per il controllo antidoping.

Svolgimento. Indicando con p(D)=015 e p(nD)=0,85 la probabilità che un atleta faccia uso e non faccia uso di doping e con p(S|D)=0,75 e p(S|nD)=0,4 la probabilità di selezionare un atleta che abbia fatto o non fatto uso di doping, la probabilità che un atleta venga selezionato per il controllo antidoping è:

p(S) = p(S|D) p(D) + p(S|nD) p(nD) = 0,75x0,15+0,4x0,85 = 0,4525

b) la probabilità che, avendo selezionato un atleta per effettuare controlli, questo risulti aver fatto uso di doping

Svolgimento. p(D|S) = p(D∩S)/p(S) = [p(D) p(S|D)]/ p(S) = 0,15x0,75/0,4525 = 0,249

c) in quale misura la procedura adottata consente risultati migliori nell'individuare atleti che fanno uso di doping rispetto ad un'estrazione a sorte di un atleta da sottoporre a controlli?

Svolgimento. La procedura fa crescere la probabilità di individuare atleti che fanno uso di doping, infatti P(D|S)=0,249 > p(D)=0,15 . L'aumento è pari a [(0,249/0,15)-1] x 100 = + 66%

QUESITI (barrare la risposta ritenuta esatta)

(PUNTI 2 per risposta corretta, PUNTI -1 per risposta sbagliata, PUNTI 0 per assenza di risposta)

Conoscendo che la retta di regressione di Y su X sia $\hat{Y} = 8 - 3X$ e che la media di X sia pari a 1, quale valore assume la media Y:

- ☒ 5
- ❒ -1/3
- ❒ Non ci sono elementi per rispondere

Considerando il piano di regressione $\hat{Y} = B_0 + B_1X + B_2Z$, non si ottiene alcun miglioramento nella spiegazione di Y rispetto alla retta di regressione $\hat{Y} = B_0 + B_1X$ qualora:

- ❒ $r_{YZ} = 0$
- ☒ $B_2 = 0$
- ❒ $r_{XY} = r_{YZ}$

Il diagramma a coordinate polari consente di rappresentare:

- ❒ serie temporali
- ❒ serie territoriali
- ☒ serie cicliche

Se il χ^2 tra due variabili X e Y risultato pari a 12. Considerando le due variabili trasformate lineari Z = 2X e W = 3Y il χ^2 tra le due variabili Z e W è uguale a:

- ☒ 12
- ❒ 72
- ❒ Nessuno dei precedenti

L'affermazione "uno stimatore è più efficiente di un altro stimatore se ha una varianza inferiore" è:

- ❒ sempre vera
- ❒ vera se tale stimatore sia anche corretto
- ☒ vera solo se entrambi gli stimatori siano corretti