Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2018-2019

STATISTICA – 17.01.19 - PROVA GENERALE

Traccia soluzioni Modalità A

FREQUENTANTE□.. NON FREQUENTANTE..

(A) ai fini della valutazione verranno considerate **solo** le risposte riportate dallo studente **negli appositi riquadri** bianchi. **(B) nello svolgimento del compito si utilizzino almeno tre cifre decimali.**

COGNOME…………………………NOME…..……………………..MATR………………..

L'ESERCIZIO 9 E' RISERVATO AGLI STUDENTI NON FREQUENTANTI (SENZA PUNTI NELL'ASSIGNMENT)

ESERCIZIO 1 (punti 4) Si considerino due variabili aleatorie X e Y indipendenti; X ha funzione di probabilità

$$p_X(x) = \begin{cases} 0.3 & x = -1, 0 \\ 0.4 & x = 1 \\ 0 & altrove \end{cases}$$

mentre Y ha distribuzione bernoulliana di parametro 0.2.

a) Si calcoli P(X=1,Y=1).

b) Si determini la varianza di W=Y-X.

c) Si determini la distribuzione della variabile aleatoria T=2-X^2.

a) $Y \sim Be(0.2);\ p_Y(1) = 0.2;\ P(X=1, Y=1) = p_X(1)\, p_Y(1) = 0.08$

b) $E(Y) = 0.2;\ E(X) = -0.3 + 0.4 = 0.1;\ E(W) = E(Y) - E(X) = 0.1$

$V(Y) = 0.2 \times (1-0.2) = 0.16;\ V(X) = (-1)^2\, 0.3 + 0.4 - 0.1^2 = 0.7 - 0.01 = 0.69$

$V(W) = V(Y) + (-1)^2 V(X) = 0.16 + 0.69 = 0.85$

c)

x	$t = 2 - x^2$	$p_T(t)$
-1	1	$p_T(1) = p_X(-1) + p_X(1) = 0.7$
0	2	$p_T(2) = p_X(0) = 0.3$
1	1	

$$\Rightarrow\ p_T(t) = \begin{cases} 0.7, & t = 1 \\ 0.3, & t = 2 \\ 0, & altrove \end{cases}$$

ESERCIZIO 2 (punti 6). Un prodotto viene venduto da un rivenditore ad un prezzo variabile X. Per il rivenditore, il prodotto ha un costo complessivo costituito da una parte variabile Y e da una parte fissa, pari a 5. Il prezzo X ha distribuzione normale con media 27 e scarto quadratico medio 2, il costo variabile Y ha distribuzione normale con media 15 e scarto quadratico medio 1.5. Il coefficiente di correlazione lineare tra prezzo X e costo variabile Y è 0.3.

a) Si calcoli la probabilità che il prezzo di vendita sia compreso tra 25 e 30.

b) Qual è il prezzo di vendita che viene superato con probabilità 0.3?

c) Si calcoli lo scarto quadratico medio del ricavo del rivenditore sul prodotto (differenza tra prezzo di vendita e costo complessivo).

a) prezzo $X \sim N(27; 2^2)$;

$$P(25 < X < 30) = P\left(\frac{25-27}{2} < Z < \frac{30-27}{2}\right) = P(-1 < Z < 1.5) = P(Z < 1.5) - [1 - P(Z < 1)] =$$

$$= 0.9332 - 1 + 0.8413 = 0.7745$$

b) $P(X > x) = 0.3 \Rightarrow 1 - P(X < x) = 0.3 \Rightarrow P(X < x) = 0.7 \Rightarrow z_{0.7} = \dfrac{x-27}{2} = 0.52$ (oppure 0.53)

$x = 0.52 \times 2 + 27 = 28.04$ (oppure 28.06)

c) ricavo $R = X - (Y+5)$; $Y \sim N(15; 1.5^2)$; $\rho_{XY} = 0.3$; $Cov(X,Y) = \sigma_X \sigma_Y 0.3 = 2 \times 1.5 \times 0.3 = 0.9$

$$\sqrt{V(R)} = \sqrt{V(X) + V(Y) - 2Cov(X,Y)} = \sqrt{2^2 + 1.5^2 - 1.8} = \sqrt{4.45} = 2.1095$$

ESERCIZIO 3 (punti 2) Il 30% delle famiglie di un Paese possiede un impianto satellitare per la televisione. Si estrae un campione di 10 famiglie del Paese; si calcoli la probabilità che almeno 2 di esse abbiano un impianto satellitare.

$Y \sim Bi(10;0.3); P(X \geq 2) = 1 - P(X \leq 1) = 0.850692$

$P(X \leq 1) = p_X(0) + p_X(1) = 0.149308$

$p_X(0) = \binom{10}{0} 0.3^0 0.7^{10} = 0.028248; p_X(1) = \binom{10}{1} 0.3^1 0.7^9 = 0.121061$

ESERCIZIO 4 (punti 1.5)
Il Data Base delle fatture attive dell'azienda AZ S.p.a. contiene, tra le altre, la colonna, o variabile X, "importo fattura" e la colonna, o variabile Y, "scadenza" in mesi della fattura (cioè 1, 2 o 4 mesi). Inoltre il Data Base contiene 1000 righe, o "record". Per fini di analisi gestionale si sono considerate tali variabili ottenendo la seguente tabella a doppia entrata:

X\Y	1	2	4
100	50,00%		
400		25,00%	5,00%
700		5,00%	10,00%
900			5,00%

a) Si specifichi la moda della variabile Y ed il numero delle fatture attive con scadenza pari a tale moda.
b) Si specifichi la scadenza media delle fatture attive.
c) Si completi la seguente tabella delle frequenze per la variabile X nei due intervalli qui sotto specificati:

X importo fattura	frequenza assoluta	frequenza relativa%	freq. ass. cumulata	freq.rel.% cumulata
[100, 450)	800	80%	800	80%
[450, 900]	200	20%	1000	100%

a) mod(Y) = 1 mese; n. fatture con scadenza un mese = 0.5 x 1000 = 500
b) M(Y) = 0.5 x 1 + 0.3 x 2 + 0.2 x 4 = 1.9
c) Vedasi tabella.

ESERCIZIO 5 (punti 3) Su un campione di 200 clienti di un istituto bancario 74 usufruiscono di servizi di home-banking.
a) Si determini una stima puntuale della proporzione di clienti dell'istituto bancario che usufruiscono di servizi di home-banking.
b) Si determini un intervallo di confidenza al 95% per la proporzione di clienti dell'istituto bancario che usufruiscono di servizi di home-banking.

a) $X \sim Be(p)$, $n = 200$, n. successi = 74, stima puntuale $\hat{p} = \bar{x}_{200} = \frac{74}{200} = 0.37$

b) $n = 200 > 25$, $1 - \alpha = 0.95$, $z_{0.975} = 1.96$,

SE = $\sqrt{\frac{\hat{\sigma}_X^2}{n}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{200}} = \sqrt{\frac{0.37 \times 0.63}{200}} = \sqrt{0.001166} = 0.03413$; ME = $1.96 \times \text{SE} = 0.066913$

$IC_{0.95}(p) = (0.37 - 0.066913,\ 0.37 + 0.066913) = (0.303087,\ 0.436916)$

ESERCIZIO 6 (punti 6). Su un campione di 120 giorni viene rilevato il numero di email classificate come SPAM che arrivano giornalmente ad una casella di posta elettronica; i dati ottenuti forniscono un numero medio giornaliero di SPAM pari a 42 ed una deviazione standard pari a 4.
a) Si determini una stima puntuale del numero medio giornaliero di SPAM che arrivano alla casella di posta elettronica. Si descrivano le proprietà dello stimatore che ha fornito la stima.
b) Si determini una stima puntuale della varianza dello stimatore di cui al punto precedente.
c) Si stabilisca, a livello 0.1, se il numero medio di SPAM che arrivano alla casella di posta elettronica giornalmente è diverso da 40.

a) stima puntuale $\hat{\mu} = \bar{x}_{120} = 42$; proprietà stimatore $\bar{X}_{120}$ (media campionaria): non distorsione: $E(\bar{X}_n) = \mu_X$,

consistenza: $\lim_{n\to\infty} V(\bar{X}_n) = \lim_{n\to\infty} \frac{\sigma_X^2}{n} = 0$; inoltre con $n = 120 > 25$ per T.C.L. si ha $\bar{X}_n \sim N\left(\mu_X; \frac{\sigma_X^2}{n}\right)$.

b) con $\hat{\sigma}_X = 4$ si ha la stima puntuale: $\hat{\sigma}_{\bar{X}}^2 = \frac{\hat{\sigma}_X^2}{n} = \frac{4^2}{120} = 0.1\bar{3}$

c) $H_1: \mu_X \neq 40; H_0: \mu_X = 40; \mu_0 = 40; \alpha = 0.1, 1 - \alpha/2 = 0.95, z_{0.95} = 1.645$

$$z_{oss} = \frac{\overline{x}_{120} - \mu_0}{SE} = 5.477226$$

$RR = \left\{ \left| z_{oss} \right| \geq z_{0.95} \right\}$; la condizione è verificata, si rifiuta H_0 al livello $\alpha = 0.1$.

Oppure lo stesso si ha con il p-value: $p_v = 2P\left(Z \geq \left| z_{oss} \right|\right) = 2P\left(Z \geq 5.477226\right) = 2 \times 0 = 0 < \alpha = 0.1$.

ESERCIZIO 7 (punti 2.5). Una recente indagine, condotta su un campione di 40 studenti (16 femmine e 24 maschi) che hanno conseguito la laurea triennale in materie scientifiche in un'università, ha fornito i seguenti dati in relazione al voto di laurea:

	Voto medio	Scarto quadratico medio
FEMMINE	103.1	5.2
MASCHI	101.1	5.1

Si assumano indipendenti e con distribuzioni normali con uguale varianza i voti di laurea triennale dei maschi e delle femmine.
Si stabilisca, a livello 0.05, se i voti medi di laurea triennale in materie scientifiche di maschi e femmine nell'università in esame sono diversi.

Siano X il voto di laurea delle femmine e Y quello dei maschi. X e Y sono indipendenti, con uguale varianza σ^2, con medie rispettivamente $\mu_X \; e \; \mu_Y$.
Le ipotesi sono $H_0: \mu_X = \mu_Y \; e \; H_1: \mu_X \neq \mu_Y$. La regione di rifiuto del test di livello 0.05 è:

$$R: |T| = \left| \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2 \left(\frac{1}{16} + \frac{1}{24} \right)}} \right| \geq t_{0.025}^{38} \simeq t_{0.025}^{40} = 2.021.$$

Poichè

$$|t_{oss}| = \frac{2}{\sqrt{\frac{26.4166}{16} + \frac{26.4166}{24}}} = 1.2057 < 2.021,$$

non si rifiuta l'ipotesi nulla a livello 0.05; non c'è quindi evidenza che i voti di maschi e femmine siano diversi. In t_{oss}, il valore di S^2_p è stato ottenuto come s^2_p =(5.2^2*15+5.1^2*23)/38=26.4166.

ESERCIZIO 8 (punti 2). Su un campione di alberghi in località turistiche si rilevano le seguenti variabili:

- **CLIENTI (Y)……………** Numero di clienti che hanno alloggiato nell'ultimo anno
- **SUP (X)..…………..........** Superficie complessiva dell'albergo (in mq)

Si consideri il modello di regressione che studia la dipendenza del numero di clienti che hanno alloggiato in un albergo nell'ultimo anno dalla superficie dell'albergo. I dati rilevati nel campione hanno fornito le seguenti stime (e standard error corrispondenti) per i parametri del modello:

$$b_0 = intercetta = 1420 \; (con \; standard \; error = 423); b_1 = coefficiente \; angolare = 7 \; (con \; standard \; error = 0.8)$$

a) Si stabilisca, a livello 0.01, se la superficie dell'albergo è significativa per la spiegazione del numero di clienti. Si giustifichi la risposta.
b) Se la risposta alla domanda precedente è positiva, si descriva l'effetto della superficie dell'albergo sul numero dei suoi clienti.

a) Occorre effettuare un test relativo alle ipotesi

$$H_0: \beta_1 = 0, H_1: \beta_1 \neq 0$$

La regione di rifiuto del test di livello 0.01 è

$$R: |Z| = \frac{|B_1|}{S_{b_1}} \geq 2.576.$$

Poichè il valore osservato della statistica test è 7/0.8=8.75>2.576, si rifiuta l'ipotesi nulla e si conclude quindi che la superficie è significativa. La stessa conclusione può essere raggiunta calcolando il p-value o determinando l'intervallo di confidenza per β_1.

b) La stima del coefficiente della superficie è 7; quindi, ad un incremento di superficie dell'albergo di 1 metro quadrato è associato un incremento del numero medio di clienti pari a 7.

ESERCIZIO 9 (punti 3). Si consideri una popolazione normale con media μ e varianza σ^2 incognite, da cui si estrae un campione di ampiezza 10 che fornisce una media pari a 18 ed uno scarto quadratico medio pari a 3. Si consideri il problema di verifica di ipotesi con

$$H_0: \mu = 20; \qquad H_1: \mu < 20$$

Si stabilisca, usando il p-value, se l'ipotesi nulla è da rifiutare o meno, a livello 0.05. Si giustifichi la risposta.

Il valore osservato della statistica test è, in questo caso,

$$t_{oss} = \frac{\bar{x} - 20}{s/\sqrt{10}} = \frac{18 - 20}{3/\sqrt{10}} = -2.1082.$$

Il p-value è, quindi,

$$p - value = P(T \leq t_{oss}) = P(T \leq -2.1082) = 0.0321,$$

dove T ha distribuzione t di student con 9 gradi di libertà. Essendo il p-value minore del livello 0.05, si rifiuta l'ipotesi nulla a questo livello.