Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2017-2018

STATISTICA — 02.11.17 - I PROVA PARZIALE

Modalità B/D

(A) ai fini della valutazione verranno considerate **solo** le risposte riportate dallo studente **negli appositi riquadri** bianchi: in caso di necessità **si può anche andare fuori dai margini che delimitano i riquadri**.
(B) nello svolgimento del compito si utilizzino almeno tre cifre decimali.

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

ESERCIZIO 1 (punti 6) Si consideri una variabile aleatoria X con distribuzione binomiale di parametri 3 e 0.6 ed una variabile aleatoria Y con distribuzione binomiale di parametri 2 e 0.8. La distribuzione congiunta di X e Y è riportata nella seguente tabella a doppia entrata che segue:

X \ Y	0	1	2
0	0	0	0.064
1	0	0	0.288
2	0	0.144	0.288
3	0.040	0.176	0

a) Si calcoli il coefficiente di correlazione lineare di X e Y. Che cosa si può affermare in relazione al valore trovato?
b) Si calcoli la probabilità che X sia maggiore di 1 e Y sia maggiore o uguale a 1.
c) Si determini la funzione di probabilità di U=3Y-2.
d) Si determini lo scarto quadratico medio (o deviazione standard) di T=2Y-2X.

a) $E(X)=np=3\cdot0.6=1.8$; $E(Y)=2\cdot0.8=1.6$; $Var(X)=np(1-p)=0.72$; $Var(Y)=0.32$

$cov(X,Y)=1\cdot2\cdot(0.288)+2\cdot1\cdot(0.144)+2\cdot2\cdot(0.288)+3\cdot1\cdot(0.176)-(1.8)\cdot(1.6)=$
$=2.544-2.88=-0.336$

$$\rho(X,Y)=\frac{cov(X,Y)}{\sigma(X)\cdot\sigma(Y)}=\frac{-0.336}{\sqrt{0.72}\cdot\sqrt{0.32}}=-0.7$$

È negativo e abbastanza vicino a 1: c'è una associazione lineare negativa medio-alta tra le due variabili

b) $P(X>1, Y\geq1)=0.144+0.288+0.176=0.608$

c) $S_U=\{-2,1,4\}$

$$P_U(u)=\begin{cases}0.04 & u=-2\\ 0.32 & u=1\\ 0.64 & u=4\\ 0 & \text{altrove}\end{cases}$$

d) $Var(T)=4Var(X)+4Var(Y)-8cov(X,Y)=$
$=4\cdot0.72+4\cdot0.32-8\cdot(-0.336)=6.848$

$\sigma(T)=\sqrt{Var(T)}=\sqrt{6.848}=2.6169$

ESERCIZIO 2 (punti 4). Il numero di clienti che si recano ad un certo ristorante ha distribuzione di Poisson; in un'ora si recano al ristorante mediamente 12 persone.
a) Si calcoli la probabilità che in un quarto d'ora si rechino al ristorante 2 persone.
b) Si calcoli la probabilità che in 20 minuti si rechino al ristorante più di 3 persone.

a) X = numero clienti che si recano al ristorante in 15 minuti
$X \sim Poisson(3)$
$P(X=2) = P(X \leq 2) - P(X \leq 1) = 0.423 - 0.199 = 0.224$

b) Y = numero clienti che si recano al ristorante in 20 minuti
$X \sim Poisson(4)$
$P(X > 3) = 1 - P(X \leq 3) = 1 - 0.433 = 0.567$

ESERCIZIO 3 (punti 4). In un Paese la probabilità che una generica lettera venga smarrita è pari a 0.12.
a) Si considerino 6 lettere. Si calcoli la probabilità che non più di 2 di esse vengano smarrite.
b) Si considerino 300 lettere. Si calcoli la probabilità che più di 30 di queste lettere vengano smarrite.

a) X = numero lettere smarrite su 6; $X \sim Bin(6, 0.12)$
$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) =$
$= \frac{6!}{0!6!}(0.12)^0(0.88)^6 + \frac{6!}{1!5!}(0.12)^1(0.88)^5 + \frac{6!}{2!4!}(0.12)^2(0.88)^4 =$
$= 0.4644 + 0.3800 + 0.1295 = 0.9739$

b) Y = numero lettere smarrite su 300
$Y \sim Bin(300, 0.12)$ ma anche, essendo n grande, $Y \approx N(36, 31.68)$
$P(Y > 30) = P\left(Z > \frac{30-36}{\sqrt{31.68}}\right) = P(Z > -1.066) =$
$= P(Z < 1.066) = 0.8577$

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2017-2018

ESERCIZIO 4 (punti 10). Il voto riportato dagli studenti in un esame universitario è determinato per il 40% da una prova scritta e per il 60% da una prova orale. Il voto della prova scritta ha distribuzione normale con media 22 e varianza 11, il voto della prova orale ha media 25 e scarto quadratico medio (ovvero deviazione standard) pari a 2. I voti della prova scritta e della prova orale hanno coefficiente di correlazione lineare pari a 0.35.

a) Si calcoli la probabilità che il voto della prova scritta sia compreso tra 21 e 24.
b) Si determini il quantile di ordine 0.72 del voto della prova scritta.
c) Si determini il valore atteso del voto finale riportato dallo studente (determinato dai due voti come descritto in precedenza).
d) Si determini la varianza del voto finale.
e) Si dica, giustificando la risposta, se i due voti sono indipendenti.
f) Si considerino ora 4 studenti che sostengono l'esame. Si calcoli la probabilità che almeno 3 dei 4 studenti riportino un voto **nella prova scritta** insufficiente (cioè minore di 18).

X = voto prova scritta; Y = voto prova orale

$X \sim N(22, 11)$; $E(Y) = 25$, $\sigma(Y) = 2$; $\rho(X,Y) = 0.35$

a) $P(21 < X < 24) = P\left(\frac{21-22}{\sqrt{11}} < Z < \frac{24-22}{\sqrt{11}}\right) = P(-0.30 < Z < 0.60) =$

$= P(Z < 0.60) - P(Z < -0.30) = 0.7257 - (1 - 0.6179) = 0.3436$

b) $P(X < k) = 0.72$, ovvero $P\left(Z < \frac{k-22}{\sqrt{11}}\right) = 0.72 \Rightarrow \frac{k-22}{\sqrt{11}} = 0.58$

quindi $k = 0.58 \cdot \sqrt{11} + 22 = 23.9236$

c) T = voto finale; $T = 0.4 \cdot X + 0.6 \cdot Y$

$E(T) = (0.4) \cdot E(X) + (0.6) \cdot E(Y) = (0.4) \cdot 22 + (0.6) \cdot 25 = 23.8$

d) $Var(T) = (0.4)^2 \cdot Var(X) + (0.6)^2 \cdot Var(Y) + 2 \cdot (0.4) \cdot (0.6)\, cov(X,Y) =$

$= (0.16) \cdot 11 + (0.36) \cdot 4 + 2 \cdot (0.4) \cdot (0.6) [(0.35)\sqrt{11} \cdot 2] = 4.3144$

e) Non sono indipendenti; se lo fossero, si avrebbe $\rho(X,Y) = 0$

f) U = numero studenti, sui 4, che riportano voto insuff.

$P(X < 18) = P\left(Z < \frac{18-22}{\sqrt{11}}\right) = P(Z < -1.21) = 1 - 0.8869 = 0.1131$

$U \sim Bin(4, 0.1131)$

$P(U \geq 3) = P(U=3) + P(U=4) = \frac{4!}{3!\,1!}(0.1131)^3(0.8869)^1 + \frac{4!}{4!\,0!}(0.1131)^4(0.8869)^0 =$

$= 0.0513 + 0.0002 = 0.0515$

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2017-2018

ESERCIZIO 5 (punti 3). Il Data Base delle fatture attive dell'azienda AZ.PROD S.p.a. (che produce tre prodotti) contiene, tra le altre, la colonna, o variabile X, "importo fattura" e la colonna, o variabile Y, "numero di tipologie di prodotto fatturate" in ciascuna fattura (cioè 1, 2 o 3 tipi di prodotto). Per certi fini di analisi gestionale di dette fatture attive (200 fatture in totale nel Data Base) si sono considerate tali variabili ottenendo la tabella a doppia entrata qui sotto:

X\Y	1	2	3
100	40,00%		
500		30,00%	
800		18,00%	2,00%
1000			10,00%

a) Si specifichi l'unità statistica (o unità osservativa) cui si riferiscono i valori delle variabili X e Y registrati nel Data Base.
b) Si specifichino le frequenze marginali della variabile Y, la moda della stessa variabile, e quante sono le fatture (delle 200) con un numero di tipologie di prodotto fatturate pari a tale moda.
c) Si calcoli media e varianza della variabile Y.
d) Si specifichi cosa indica (o cosa significa) la frequenza % all'incrocio di X=800 e Y=3 con riferimento alle unità statistiche registrate nel Data Base.
e) Infine sapendo che per la variabile X la media è M(X) = 350 e lo scarto quadratico medio è σ(X) = 429,535, si determini la frequenza relativa % dei valori della variabile X nell'intervallo M(X) ∓ k σ(X) con k=1,5.

a) L'unità statistica è la fattura attiva

b)

valori di Y	1	2	3
freq. marginali	40%	48%	12%

moda(Y) = 2

numero fatture richiesto = 0.48·200 = 96

c) media di Y = 1.72 ; varianza di Y = 0.4416

d) La frequenza specificata indica che il 2% delle 200 fatture attive ha un importo pari a 800 e 3 tipi di prodotti fatturati

e) $M(X) - 1.5 \cdot \sigma(X) = 350 - 1.5 \cdot 428.535 = -284.3025$
$M(X) + 1.5 \cdot \sigma(X) = 350 + 1.5 \cdot 428.535 = 984.3025$
I valori di X nell'intervallo (−284.3025, 994.3025) sono 100, 500 e 800. La frequenza cercata è quindi 40% + 30% + 20% = 90%