Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2016-2017
STATISTICA – 16.02.17 - PROVA GENERALE
Modalità A

(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 tre cifre decimali.

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

ESERCIZIO 1 (punti 3.5). Una variabile X ha distribuzione binomiale di parametri 16 e 0.5; una variabile Y ha distribuzione di Poisson con media 4. X e Y hanno covarianza pari a -3.
a) Si calcoli la probabilità che X sia minore di 15.
b) Si calcoli la probabilità che Y sia maggiore di 5.
c) Si calcoli il coefficiente di correlazione lineare di X e Y.
d) Si calcoli il valore atteso di T=3X-Y-2.

a) $P(X<15) = 1-P(X=15)-P(X=16) = 1-\frac{16!}{15!1!}(0.5)^{15}(0.5)^{1} - \frac{16!}{16!0!}(0.5)^{16}(0.5)^{0} =$
$= 1-0.0002-0.00002 = 0.9998$

b) $P(Y>5) = 1-P(Y\leq 5) = 1-0.785 = 0.215$

c) $\rho(X,Y) = \frac{cov(X,Y)}{\sigma(X)\cdot\sigma(Y)} = \frac{-3}{\sqrt{16\cdot 0.5\cdot 0.5}\sqrt{4}} = -0.75$

d) $E(T) = E(3X-Y-2) = 3E(X)-E(Y)-2 = 3\cdot(16\cdot 0.5)-4-2 =$
$= 18$

ESERCIZIO 2 (punti 3). Si consideri un cliente di un supermercato. La tabella a doppia entrata che segue riporta la distribuzione di probabilità congiunta di X:"numero di articoli A acquistati dal cliente" e Y:"numero di articoli B acquistati dal cliente".

X \ Y	0	1	2	4
0	0.2	0	0	0
1	0	0.3	0.1	0.1
2	0	0	0.1	0.2

a) Esiste un legame tra il numero di articoli A e il numero di articoli B acquistati? Se la risposta è affermativa, si descrivano brevemente la natura e l'intensità del legame.
b) Si calcoli la probabilità che il cliente acquisti lo stesso numero di articoli di tipo A e di tipo B.
c) Si dica, giustificando la risposta, se Y ha distribuzione binomiale.

a) Si possono calcolare covarianza e coeff. di correl. lineare.
$E(X) = 0\cdot(0.2)+1\cdot(0.5)+2\cdot(0.3) = 1.1$
$E(Y) = 0\cdot(0.2)+1\cdot(0.3)+2\cdot(0.2)+4\cdot(0.3) = 1.9$
$Var(X) = 0^2\cdot(0.2)+1^2\cdot(0.5)+2^2\cdot(0.3)-(1.1)^2 = 0.49$
$Var(Y) = 0^2\cdot(0.2)+1^2\cdot(0.3)+2^2\cdot(0.2)+4^2\cdot(0.3)-(1.9)^2 = 2.29$
$cov(X,Y) = 1\cdot1\cdot(0.3)+1\cdot2\cdot(0.1)+2\cdot2\cdot(0.1)+1\cdot4\cdot(0.1)+2\cdot4\cdot(0.2)-(1.1)\cdot(1.9) =$
$= 0.81$

$\rho(X,Y) = \frac{0.81}{\sqrt{0.49}\sqrt{2.29}} = 0.765$

Essendo diversi da 0 covarianza e coefficiente di correlazione lineare esiste un legame. Essendo i due indici positivi, il legame è piuttosto diretto (o "concordante"). Essendo $\rho(X,Y)$ vicino a 1, c'è una forte associazione lineare (diretta) tra X e Y.

b) $P(X=Y) = P(X=0, Y=0) + P(X=1, Y=1) +$
$+ P(X=2, Y=2) = 0.2 + 0.3 + 0.1 = 0.6$

c) Y non può avere distribuzione binomiale in quanto il suo supporto non coincide con quello di una distribuzione binomiale (dovrebbe includere 3). Inoltre anche le probabilità non sono quelle di una binomiale.

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2016-2017

ESERCIZIO 3 (punti 6). Il rendimento di un portafogli è determinato per il 20% da un titolo A, per il 50% da un titolo B e per il resto da un titolo C. I rendimenti dei 3 titoli hanno tutti distribuzione normale; il rendimento di A ha media 0.01 e scarto quadratico medio 0.01, il rendimento di B ha media 0.02 e scarto quadratico medio 0.02, il rendimento di C ha media 0.03 e scarto quadratico medio 0.03. Inoltre, i rendimenti di A e B sono indipendenti, i rendimenti di A e C hanno covarianza pari a 0.005, i rendimenti di B e C hanno covarianza pari a 0.003.

a) Si calcoli la probabilità che il rendimento di A sia positivo.
b) Si determini il quantile di ordine 0.3 del rendimento di A.
c) Si determini il valore atteso del rendimento del portafogli.
d) Si determini lo scarto quadratico medio del rendimento del portafogli.

X = rend. A; Y = rend. B; W = rend. C. $X \sim N(0.01, 0.01^2)$; $Y \sim N(0.02, 0.02^2)$; $W \sim N(0.03, 0.$

X, Y indip.; $cov(X,W) = 0.005$; $cov(Y,W) = 0.003$.

a) $P(X>0) = P\left(Z > \frac{0-0.01}{0.01}\right) = P(Z>-1) = P(X<1) = 0,8413$

b) $P(X < q_{0.3}) = 0.3$, quindi $P\left(Z < \frac{q_{0.3}-0.01}{0.01}\right) = 0.3$, quindi $\frac{q_{0.3}-0.01}{0.01} = z_{0.3}$

dove $z_{0.3}$ è il quantile di ordine 0.3 di $N(0,1)$.

Poichè $z_{0.3} = -z_{0.7}$, e $z_{0.7} \cong 0.52$, si ha $z_{0.3} = -0.52$, per cui $q_{0.3} = (-0.52)\cdot(0.01) + 0.01 = 0.0048$

c) Rend. portaf. $T = 0.2 \cdot X + 0.5 \cdot Y + 0.3 \cdot W$; $E(T) = (0.2)(0.01) + (0.5)(0.02) + (0.3)(0.03) = 0.021$

d) $Var(T) = (0.2)^2 \cdot (0.01)^2 + (0.5)^2 \cdot (0.02)^2 + (0.3)^2 \cdot (0.03)^2 + 2 \cdot (0.2)(0.3) \cdot (0.005) + 2 \cdot (0.5)(0.3)(0.003)$
$= 0.0017$

$\sigma(T) = 0.041$

ESERCIZIO 4 (punti 6). I laterizi di un certo tipo prodotti da una ditta hanno peso ditribuito in accordo ad una distribuzione normale. Si considera un campione di 18 laterizi; il loro peso medio è pari a 2222 grammi, la deviazione standard (o scarto quadratico medio) del loro peso è 16 grammi.

a) Si determini una stima puntuale del peso medio dei laterizi del tipo considerato prodotti dalla ditta.
b) Si scriva l'espressione dello stimatore usato al punto precedente e si enuncino le sue proprietà.
c) Si determini un intervallo di confidenza al 95% per il peso medio dei laterizi del tipo considerato prodotti dalla ditta.
d) Si stabilisca, a livello 0.05, se il peso medio dei laterizi del tipo considerato prodotti dalla ditta è diverso da 2200.

X = peso laterizi; $X \sim N(\mu, \sigma^2)$. Campione: $n = 18$, $\bar{x} = 2222$, $s = 16$

a) $\bar{x} = 2222$ stima puntuale

b) $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. NON DISTORSIONE: $E(\bar{X}) = \mu$; CONSISTENZA: $\lim_{n\to+\infty} Var(\bar{X}) = \sigma^2/n = 0$ (questo implica la consistenza essendo $\bar{X}$ non distorto)

c) $\left(\bar{x} - t_{17,0.975}\cdot\frac{s}{\sqrt{n}}, \bar{x} + t_{17,0.975}\cdot\frac{s}{\sqrt{n}}\right) = \left(2222 \pm 2.11 \cdot \frac{16}{\sqrt{18}}\right) =$
$= (2222 \pm 7.957) = (2214.043, 2229.957)$

d) $\begin{cases} H_0: \mu = 2200 \\ H_1: \mu \neq 2200 \end{cases}$ Poichè 2200 non appartiene all'intervallo di confidenza (al 95%) costruito sopra, si rifiuta H_0 a livello 0.05; c'è evidenza che il peso medio sia diverso da 2200.

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2016-2017

ESERCIZIO 5 (punti 4). Si considera un campione di 400 clienti di un supermercato e, su ciascuno di essi, si rileva se ha comparto o meno il prodotto P. La rilevazione indica che, dei 400 clienti, 175 hanno comprato il prodotto P.

a) Si determini un intervallo di confidenza al 99% per la proporzione di clienti del supermercato che comprano il prodotto P.

b) Si stabilisca, a livello 0.01, se è diminuita la proporzione di clienti che comprano il prodotto P rispetto all'anno precedente, in cui era pari a 0.48.

$X_1, X_2, \ldots, X_{400}$ iid $\sim Bern(p)$. $\bar{x} = 175/400 = 0.438$

a) $\left(\bar{x} - z_{0.995}\cdot\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{0.995}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right) = \left(0.438 \pm 2.576\sqrt{\frac{0.438\cdot 0.562}{400}}\right) =$
$= (0.438 \pm 0.064) = (0.374, 0.502)$

b) $\begin{cases} H_0: p = 0.48 \\ H_1: p < 0.48 \end{cases}$ $R_{0.01}$: $z = \frac{\bar{x} - 0.48}{\sqrt{\frac{0.48\cdot 0.52}{400}}} \leq -z_{0.99} = -2.326$

$z_{oss} = \frac{0.438 - 0.48}{\sqrt{0.48\cdot 0.52/400}} = -1.681 \not\leq -2.326 \Rightarrow$ NON SI RIFIUTA H_0 (non c'è evidenza che p sia diminuita)

ESERCIZIO 6 (punti 3.5)

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2016-2017

ESERCIZIO 6 (punti 3)

Si vuole analizzare, mediante un modello di regressione lineare, la dipendenza della concentrazione di una certa sostanza inquinante (variabile Y) dal numero di automobili che mediamente transitano nella zona in un'ora (variabile X_1) e dalla densità di popolazione nella zona (abitanti per Kmq, variabile X_2). Sulla base di un campione di 70 zone, in ciascuna della quali si rilevano le tre variabili, si ottiene il seguente output excel:

ANOVA

	gdl	*SQ*	*MQ*	*F*	*Sig. F*
Regressione	4	1909.0237	477.2559	16.1217	3.2175E-09
Residuo	65	1924.2222	29.6034		
Total	69	3833.2459			

	Coefficienti	*Errore Standard*	*Stat t*	*Sig. P-value*	*Inferiore 95%*	*Superiore 95%*
Intercetta	66.449	12.044	5.5172	3.2102E-07	42.361	90.537
Numero automobili	0.1315	0.0281	4.6797	7.5191E-06	0.0753	0.1877
Densità popolazione	0.0678	0.0131	5.1756	1.1885E-02	0.0416	0.0940

a) E' significativo, in questo modello, l'effetto del numero di automobili in transito sulla concentrazione dell'inquinante? Si giustifichi la risposta precisando gli elementi fondamentali della procedura adottata.
b) Fissato il numero di automobili, come varia la concentrazione di inquinante al variare della densità di popolazione?
c) Si dica se il modello è, a livello 0.05, globalmente significativo. A questo proposito, si scrivano ipotesi nulla ed alternativa, regione di rifiuto, valore della statistica test e p-value relativi al test necessario per rispondere.

a) Sì, in quanto il test relativo alle ipotesi

$$\begin{cases} H_0: \beta_{NUM\_AUTO} = 0 \\ H_1: \beta_{NUM\_AUTO} \neq 0 \end{cases}$$

rifiuta H_0 a livello α qualunque, essendo il p-value $7.519 \cdot 10^{-6} \simeq 0 < \alpha$.

b) Fissato il numero di automobili, ad un incremento di 1 della densità di popolazione corrisponde un incremento pari a 0.0678 della concentrazione media di inquinante.

c)

$$\begin{cases} H_0: \beta_{DENSITÀ} = \beta_{NUM\_AUTO} = 0 \\ H_1: \text{almeno un coeff.} \neq 0 \end{cases}$$

La regione di rifiuto è $R: F = \frac{SSR/4}{SSE/65} > F_{4,65,1-\alpha}$

La statistica test ha valore 16.1217.
Il p-value è $3.2175 \cdot 10^{-9}$. Essendo il p-value $< \alpha = 0.05$ si rifiuta H_0, per cui si valuta il modello globalmente significativo.

ESERCIZIO 7 (punti 1.5)
Per i valori z della variabile Z (riportati nella prima colonna a sinistra nella tabella qui sotto) ed i valori t della variabile T (riportati nella prima riga in alto della stessa tabella) si sono rilevate le frequenze assolute congiunte riportate nella stessa tabella.

Z \ T	4	5	6	7
0	20			5
1		5		5
2		5		10
3	5		15	30

Mostrando i calcoli principali, si risponda alle seguenti domande:
a) si determinino il primo quartile ed il quinto decile della variabile T.
b) si determini la frequenza relativa delle coppie (z, t) di valori delle due variabili con z compreso fra 1 e 2 (inclusi) e con t compreso fra 6 e 7 (inclusi).
c) si determinino media e varianza della variabile T.
d) si determini il coefficiente di correlazione lineare delle variabili Z e T.

	T				
Z	4	5	6	7	
0	0,2			0,05	0,25
1		0,05		0,05	0,1
2		0,05		0,1	0,15
3	0,05		0,15	0,3	0,5
	0,25	0,1	0,15	0,5	1

V dec. T	I quart. T	fr.relativa congiunta 1<=Z<=2 6<=T<=7	media T	varianza T	s.q.m. T	cov(Z,T)	rho (Z,T)
6	4	0,15	5,9	1,59	1,260952	0,89	0,5597484

inoltre per calcolare Cov e Rho:

media Z	varianza Z	s.q.m. Z	E(ZT)
1,9	1,59	1,260952	12,10