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

STATISTICA – 15.02.18 - PROVA GENERALE

TRACCIA DI SOLUZIONE **- Modalità A**

L'ESERCIZIO 8 E' RISERVATO AGLI STUDENTI CHE SCELGONO LA MODALITA' NON FREQUENTANTE SELEZIONANDO LA CASELLA RELATIVA: 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………………..

ESERCIZIO 1 (4 punti)

Il numero di interruzioni nella fornitura di energia elettrica ad un'abitazione civile segue la distribuzione di Poisson. In una settimana si hanno, in media, 2 interruzioni.
a) Si calcoli la probabilità che, in una settimana, si abbiano almeno 3 interruzioni.
b) Si calcoli la probabilità che, in 2 settimane, si abbiano 5 interruzioni.
c) Si calcolino valore atteso e scarto quadratico medio del numero di interruzioni che si hanno in 80 settimane.

a) Sia X="numero interruzioni in una settimana"; X ha distribuzione di Poisson con media 2. Usando le tavole della distribuzione si ottiene:

$$P(X \geq 3) = 1 - P(X \leq 2) = 1 - 0.677 = 0.323.$$

b) Sia Y="numero interruzioni in due settimane"; Y ha distribuzione di Poisson con media 4. Usando le tavole della distribuzione si ottiene:

$$P(Y = 5) = P(Y \leq 5) - P(Y \leq 4) = 0.785 - 0.629 = 0.156.$$

c) Sia Z="numero interruzioni in 80 settimane". Z ha distribuzione di Poisson di media 160. Quindi:

$$E(Z) = 160; \quad \sigma(Z) = \sqrt{160} = 12.649.$$

ESERCIZIO 2 (4 punti)

Una variabile X ha distribuzione binomiale di parametri 2 e 0.8. Una variabile Y ha distribuzione di Poisson con media 2. X e Y sono indipendenti.
a) Si calcoli la probabilità che X sia maggiore o uguale a 1.
b) Sia T=X-3Y-2. Si calcoli il valore atteso di T.
c) Si calcoli la varianza di T.
d) Si calcoli P(X=Y), ovvero la probabilità che X sia uguale a Y.

a)
$$P(X \geq 1) = P(X = 1) + P(X = 2) = 2 \cdot 0.8 \cdot 0.2 + 0.8^2 = 0.96$$
o, equivalentemente,
$$P(X \geq 1) = 1 - P(X = 0) = 1 - 0.2^2 = 0.96$$

b) Si ha:
$$E(T) = E(X) - 3E(Y) - 2 = 2 \cdot 0.8 - 3 \cdot 2 - 2 = -6.4.$$

c) Si ha:

$$Var(T) = Var(X) + 9Var(Y) = 2 \cdot 0.8 \cdot 0.2 + 9 \cdot 2 = 18.32.$$

d) Si ha:
$$P(X = Y) = P(X = 0, Y = 0) + P(X = 1, Y = 1) + P(X = 2, Y = 2) =$$
$$= P(X = 0)P(Y = 0) + P(X = 1)P(Y = 1) + P(X = 2)P(Y = 2) =$$
$$= 0.04 \cdot 0.135 + 0.32 \cdot 0.271 + 0.64 \cdot 0.271 = 0.265.$$

ESERCIZIO 3 (4.5 punti)

Il rendimento annuo di un portafoglio è T=0.4X+0.6Y, dove X e Y sono i rendimenti annui di due titoli A e B, entrambi con distribuzione normale, il primo di media 0.01 e deviazione standard (o scarto quadratico medio) 0.01, il secondo di media 0.02 e deviazione standard 0.02. X e Y sono indipendenti.
a) Si calcoli la probabilità che il rendimento annuo del titolo A sia positivo.
b) Si determini il quantile di ordine 0.5 del rendimento annuo del titolo A.
c) Si determini il rendimento annuo medio del portafoglio.
d) Si determini lo scarto quadratico medio del rendimento annuo del portafoglio.

a) $P(X>0)=P\left(Z>\frac{0-0.01}{0.01}\right)=P(Z>-1)=0.841$.

b) Il quantile di ordine 0.5 è la mediana, che coincide con il valore atteso essendo la distribuzione normale simmetrica; quindi il quantile cercato è 0.01.

c) E(T)=0.4E(X)+0.6E(Y)=0.4*0.01+0.6*0.02=0.016.

d) $\sigma(T)=\sqrt{Var(T)}=\sqrt{0.4^2\cdot 0.01^2+0.6^2\cdot 0.02^2}=0.0126$

<u>ESERCIZIO 4 (5 punti)</u>

Si vuole stimare la proporzione di clienti di un'azienda di vendite online che hanno un'opinione positiva sul servizio ricevuto. A questo scopo, si considera un campione di 200 clienti, dei quali 130 esprimono opinione positiva.

a) Si fornisca un intervallo di confidenza al 99% per la proporzione di clienti dell'azienda che hanno opinione positiva.

b) Si può affermare, a livello 0.05, che la proporzione di clienti dell'azienda che hanno opinione positiva è inferiore a 0.7? Si riportino le ipotesi nulla ed alternativa, la regione di rifiuto del test e, ovviamente, la conclusione raggiunta (con la risposta esplicita alla domanda posta).

a) La stima puntuale è $\hat{p}=\bar{x}=\frac{130}{200}=0.65$, mentre lo standard error è SE= $\sqrt{\frac{\bar{x}\,(1-\bar{x})}{n}}=\sqrt{\frac{0.65*0.35}{200}}=0.0337$.

L'intervallo è quindi

$$(\bar{x}-2.576*SE,\bar{x}+2.576*SE)=(0.65-2.576*0.0337,0.65+2.576*0.0337)=(0.5632,0.7368).$$

b) Le ipotesi sono: $H_0: p=0.7$ e $H_1: p<0.7$. La regione di rifiuto del test è:

$$R: Z=\frac{\bar{X}-0.7}{\sqrt{0.7\cdot 0.3/200}}<-1.645$$

Poiché

$$Z_{oss}=\frac{0.65-0.7}{\sqrt{\frac{0.7\cdot 0.3}{200}}}=-1.543>-1.645,$$

non si rifiuta l'ipotesi nulla; quindi non c'è evidenza per affermare, a livello 0.05, che la proporzione di clienti dell'azienda che hanno opinione positiva è inferiore a 0.7.

<u>ESERCIZIO 5 (5 punti)</u>

Si vogliono confrontare i tempi medi (in ore) di consegna di pacchi di due diversi corrieri, A e B. Si estrae quindi un campione di 100 consegne effettuate dal corriere A, che hanno avuto un tempo medio pari a 58 ore, con una deviazione standard pari a 5 ore. Allo stesso modo, si estrae un campione di 100 consegne del corriere B, per le quali il tempo medio è stato pari a 60 ore, con una deviazione standard pari a 6 ore. Si assumono varianze uguali per i tempi di consegna dei due corrieri.

a) Si scrivano ipotesi nulla ed alternativa del test per stabilire se sono diversi i tempi medi di consegna dei due corrieri, precisando cosa rappresentano simboli e notazioni introdotte.

b) Si determini una stima della varianza (comune) dei tempi di consegna dei due corrieri.

c) Si determini il p-value relativo al test in esame.

d) Sulla base del p-value calcolato al punto precedente, si traggano le conclusioni dell'analisi, a livello 0.05.

Siano $\mu_X, \mu_Y\ e\ \sigma^2$, rispettivamente, il tempo medio di consegna del corriere A, il tempo medio di consegna del corriere B e la varianza (comune) dei tempi di consegna di A e di B.

a) Le ipotesi sono: $H_0: \mu_X=\mu_Y$ e $H_1: \mu_X\neq\mu_Y$.

b) La stima della varianza comune è

$$s_p^2=\frac{5^2*99+6^2*99}{100+100-2}=30.5$$

c) Il p-value è

$$p-value=2P(Z>|Z_{oss}|)=2P\left(Z>\left|\frac{58-60}{\sqrt{\frac{30.5}{100}+\frac{30.5}{100}}}\right|\right)=2P(Z>2.561)=0.0104;$$

è stata usata la distribuzione normale in quanto entrambi i campioni sono sufficientemente grandi.

d) Poiché il p-value è inferiore al livello fissato 0.05, si rifiuta l'ipotesi nulla a questo livello; c'è evidenza che i tempi medi di consegna dei due corrieri siano diversi.

ESERCIZIO 6 (3 punti)

Attraverso un'indagine effettuata su un campione di 232 pazienti che hanno effettuato una degenza in ospedali pubblici si vuole studiare la dipendenza della spesa sostenuta dall'ospedale per la degenza da alcune possibili variabili. Sul campione vengono rilevate:

- **Spesa (Y)……………………......**spesa complessiva, in euro, sostenuta dall'ospedale per la degenza del paziente
- **Età (X_1)…………………..…….**età del paziente
- **Giorni (X_2)………..………......**numero di giorni di degenza del paziente

Sulla base dei dati rilevati nel campione, si ottiene la seguente equazione del modello di regressione stimato:

$$\hat{y} = 138 + 4x_1 + 227x_2$$

Inoltre, sempre sulla base dei dati rilevati nel campione, sono state ottenute le due seguenti quantità:

DEVIANZA SPIEGATA=SSR=12586316
DEVIANZA RESIDUA=SSE=4677765

a) Si descriva in dettaglio l'effetto del numero di giorni di degenza sulla spesa complessiva sostenuta dall'ospedale.
b) Si preveda la spesa complessiva sostenuta dall'ospedale per la degenza di 4 giorni di un paziente di 60 anni di età.
c) Si calcoli il coefficiente di determinazione R^2 e si fornisca un'interpretazione del valore ottenuto.

a) La stima del coefficiente della variabile Giorni è 227; quindi si stima che, a parità di età del paziente, ad ogni giorno in più di degenza è associato un incremento della spesa media sostenuta dall'ospedale pari a 227 euro.

b) La previsione è: 138+4*60+227*4=1286.

c) Il coefficiente di determinazione è R^2=SSR/SST=SSR/(SSR+SSE)=12586316/(12586316+4677765)=0.729.
Si può affermare che le due variabili Età e Giorni spiegano il 73% circa della variabilità complessiva della spesa sostenuta dall'ospedale per la degenza di un paziente.

ESERCIZIO 7 (1.5 punti)

Il Data Base delle fatture attive dell'azienda MEDIASTOK S.p.a. contiene, tra le altre, la colonna, o variabile T, "importo fattura" e la colonna, o variabile Z, "scadenza" in mesi della fattura. Inoltre il Data Base contiene 3000 righe, o "record". Per fini di analisi gestionale si sono considerate tali variabili ottenendo la seguente tabella a doppia entrata:

Z\T	100	500	800	1000
1	40,00%			
2		10,00%	20,00%	
3			10,00%	20,00%

a) Si specifichi l'importo fattura mediano, ed il numero di fatture con importo fattura non superiore a quello mediano.
b) Si calcoli la varianza σ^2(T) dell'importo fattura (l'importo fattura medio M(T) è pari a 530)
c) Si calcoli la frequenza relativa % delle fatture con importo fattura compresso nell'intervallo di estremi (inclusi) M(T) $\mp$ k σ(T), con k=1.

a) Frequenze relative % cumulate di importo fattura T: fr(100) + fr(500) = 40% + 10% = 50% → importo mediano = 500;
numero fatture con importo non superiore a quello mediano = 3000 x 50% = 1500
b) $\sigma^2(T) = 100^2\ 0{,}4 + 500^2\ 0{,}1 + 800^2\ 0{,}3 + 1000^2\ 0{,}2 - 530^2 = 140100$
c) σ(T) = 374,299; estremo inferiore intervallo = 155, 701; estremo superiore intervallo = 904,299;
frequenza relativa % delle fatture con importo nell'intervallo = fr(500) + fr(800) = 10% + 30% = 40%.

ESERCIZIO 8 (3 punti) SOLO PER STUDENTI NON FREQUENTANTI

a) Si fornisca un enunciato (o descrizione) del teorema centrale del limite.
b) Si fornisca un enunciato (o descrizione) della proprietà di correttezza (non distorsione) della media campionaria.

a) Considerando n variabili aleatorie indipendenti ed identicamente distribuite con valore atteso μ e varianza σ^2, con n sufficientemente grande, la media campionaria (cioè la media delle n variabili) ha, approssimativamente, distribuzione normale con valore atteso μ e varianza σ^2/n.

b) La media campionaria è uno stimatore corretto della media μ della popolazione, nel senso che il suo valore atteso è uguale a μ; in simboli, $E(\bar{X}) = \mu$.