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

STATISTICA ─ 24.01.18 - PROVA GENERALE

Modalità B – *TRACCIA DI SOLUZIONE*

(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)

La probabilità che una richiesta di fornitura ad un magazzino resti inevasa per mancanza della merce è pari a 0.2.
a) Vengono effettuate in una giornata 7 richieste di fornitura. Si calcoli la probabilità che almeno due di queste restino inevase.
b) In un mese vengono effettuate 150 richieste di fornitura. Si calcoli la probabilità che più di 50 di tali richieste restino inevase.

a) X=numero richieste, sulle 7, inevase; X ha distribuzione binomiale di parametri 7 e 0.2.

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - 0.210 - 0.367 = 0.423.$$

b) Y= numero richieste, sulle 150, inevase; Y ha distribuzione binomiale di parametri 150 e 0.2 ma, anche, Y ha distribuzione approssimata N(150*0.2,150*0.2*0.8)=N(30, 24) (per il teorema centrale del limite).

$$P(Y > 50) = P\left(Z > \frac{50 - 30}{\sqrt{24}}\right) = P(Z > 4.08) \cong 0.$$

ESERCIZIO 2 (5 punti)

Una variabile X ha distribuzione bernoulliana di parametro 0.6. Una variabile Y ha distribuzione binomiale di parametri 3 e 0.4. X e Y sono indipendenti.
a) Si calcoli la probabilità congiunta che X sia uguale a 1 e Y maggiore di 1.
b) Sia T=3X-2Y+1. Si calcoli il valore atteso di T.
c) Si calcoli lo scarto quadratico medio di T.
d) Si determini il valore del coefficiente di correlazione lineare di X e Y.

a) P(X=1,Y>1)=P(X=1)*P(Y>1)=0.6*(0.288+0.064)=0.211.

b) E(T)=E(3X-2Y+1)=3*E(X)-2*E(Y)+1=3*0.6-2*(0.4*3)+1=0.4.

c) $\sigma(T) = \sqrt{Var(T)} = \sqrt{9 * Var(X) + 4 * Var(Y)} = \sqrt{9 * (0.6 * 0.4) + 4 * (3 * 0.4 * 0.6)} = 2.245.$

d) Il coefficiente di correlazione lineare tra X e Y è uguale a 0 in quanto le due variabili sono indipendenti.

ESERCIZIO 3 (3.5 punti)

Il rendimento annuo di un titolo ha distribuzione gaussiana di media 0.02 e deviazione standard 0.04.
a) Si calcoli la probabilità che il rendimento sia negativo.
b) Si determini il quantile di ordine 0.4 del rendimento del titolo.

X=rendimento annuo del titolo; $X \sim N(0.02, 0.04^2)$.

a) $P(X < 0) = P\left(Z < \frac{0-0.02}{0.04}\right) = P(Z < -0.5) = 1 - 0.6915 = 0.3085.$

b) $P(X < k) = 0.4, ovvero\ P\left(Z < \frac{k-0.02}{0.04}\right) = 0.4, ovvero\ P\left(Z < -\frac{k-0.02}{0.04}\right) = 0.6, da\ cui\ -\frac{k-0.02}{0.04} \cong 0.25, cio\text{è}\ k = 0.01.$

ESERCIZIO 4 (3 punti)

Su un campione di 22 notebook tra quelli in commercio si rileva il peso in Kg. I 22 pesi rilevati hanno una media pari a 1.91 ed una varianza pari a 0.09. Si assume distribuzione normale per il peso dei notebook. Si stabilisca a livello 0.05 se il peso medio dei notebook in commercio è inferiore a 2 Kg, usando il metodo del p-value.

Le ipotesi sono: $H_0: \mu = 2$ e $H_1: \mu < 2$. Il p-value del test è:

$$p - value = P(T \leq t_{oss}) = P\left(T \leq \frac{1.91 - 2}{\sqrt{\frac{0.09}{22}}}\right) = P(T \leq -1.407) = 1 - P(T \leq 1.407),$$

con la probabilità calcolata supponendo vera l'ipotesi nulla, nel qual caso la statistica test *T* ha distribuzione T di Student con 21 gradi di libertà. Dalle tavole si rileva che il valore della probabilità sopra indicata (cioè il p-value) è compreso tra 0.05 e 0.1, per cui è superiore al livello fissato 0.05. Quindi non si rifiuta l'ipotesi nulla a livello 0.05; non c'è dunque evidenza per stabilire, a questo livello, che il peso medio dei notebook in commercio è inferiore a 2 Kg.

ESERCIZIO 5 (3 punti)

Il gestore di una sala cinematografica vuole stabilire se è opportuno introdurre degli sconti sui biglietti nei giorni feriali per incrementare gli incassi. Su un primo campione di 20 giorni feriali in cui non vengono praticati gli sconti rileva un incasso medio giornaliero pari a 2800 euro ed una deviazione standard pari a 410 euro. Su un secondo campione di 10 giorni feriali, in cui vengono praticati gli sconti, l'incasso medio giornaliero è pari a 3000 euro e la deviazione standard corrispondente 405 euro. Si assumono distribuzioni normali con varianze uguali per gli incassi giornalieri nei giorni feriali in cui sono praticati gli sconti e per quelli nei giorni feriali in cui non lo sono. Si stabilisca, a livello 0.01, se l'introduzione degli sconti sui biglietti incrementa gli incassi medi giornalieri dei giorni feriali. Si giustifichi la risposta, riportando ipotesi nulla ed alternativa e regione di rifiuto del test.

Le ipotesi sono: $H_0: \mu_X = \mu_Y$ e $H_1: \mu_X < \mu_Y$, dove $\mu_X e\ \mu_Y$ sono gli incassi medi giornalieri quando non vengono effettuati gli sconti e quando vengono effettuati rispettivamente. La regione di rifiuto del test è

$$R_{0.05}: T = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}} < -t_{0.01}^{28} = -2.467.$$

Usando i dati campionari si ottiene

$$s_p^2 = \frac{410^2 * 19 + 405^2 * 9}{20 + 10 - 2} = 166790.2$$

e quindi

$$T_{oss} = \frac{2800 - 3000}{\sqrt{\frac{166790.2}{20} + \frac{166790.2}{10}}} = -1.264.$$

Essendo il valore osservato della statistica test maggiore di -2.467, non si rifiuta l'ipotesi nulla a livello 0.01, per cui non c'è evidenza, a questo livello, del fatto che l'introduzione degli sconti incrementi gli incassi medi.

ESERCIZIO 6 (3 punti)

Si vuole stimare la proporzione di consegne effettuate da un corriere che arrivano a destinazione in ritardo rispetto alla data stabilita. Si considera un campione di 150 consegne; di queste, 20 arrivano a destinazione in ritardo. Si determini un intervallo di confidenza al 99% per la proporzione di consegne che arrivano a destinazione in ritardo.

La stima puntuale è $\hat{p} = \bar{x} = \frac{20}{150} = 0.133$ e lo standard error è SE= $\sqrt{\frac{\bar{x}\,(1-\bar{x})}{n}} = \sqrt{\frac{0.133*0.867}{150}} = 0.028$.

L'intervallo al 99% è quindi

$$(\bar{x} - 2.576 * SE, \bar{x} + 2.576 * SE) = (0.133 - 2.576 * 0.028, 0.133 + 2.576 * 0.028) = (0.061, 0.205).$$

ESERCIZIO 7 (4 punti)

Si vuole analizzare, mediante un modello di regressione lineare, la dipendenza del tasso di riempimento dell'aereo sui voli effettuati su una certa tratta dal prezzo medio al quale è venduto il biglietto in classe economica e dalla congiuntura economica nel periodo in cui è effettuato il volo, misurata attraverso un indicatore numerico da 0 (che indica congiuntura economica molto negativa) fino a 100 (congiuntura economica molto positiva). Sulla base di un campione di voli, in ciascuno dei quali si rilevano le tre variabili (tasso di riempimento, prezzo medio del biglietto e congiuntura economica), si ottiene il seguente output excel:

ANOVA

	gdl	*SQ*	*MQ*	*F*	*Sig. F*
Regressione	2	1708.0237	854.0119	28.2929	3.2175E-34
Residuo	77	2324.2245	30.1847		
Total	79	4032.2482			

	Coefficienti	*Errore Standard*	*Stat t*	*Sig. P-value*
Intercetta	66.449	12.044	5.5172	3.2102E-07
Prezzo medio biglietto	-0.0523	0.0184	-2.8424	0.0047232
Congiuntura economica	0.0181	0.0021	8.6190	3.4536E-34

a) Si descriva l'effetto della congiuntura economica sul tasso di riempimento del volo.

b) Si preveda il tasso di riempimento di un aereo per un volo con prezzo medio del biglietto pari a 200 effettuato in un periodo di congiuntura economica con valore dell'indicatore pari a 60.

a) La stima del coefficiente $\beta_{congiuntura\ economica}$ è 0.0181; quindi, ad un incremento unitario del livello di congiuntura economica, corrisponde un incremento del tasso medio di riempimento pari a 0.0181, fissato il prezzo medio del biglietto.

b) La previsione è: 66.449-0.0523*200+0.0181*60=57.075.

ESERCIZIO 8 (1.5 punti)

Il Data Base delle fatture passive dell'azienda PROD.AZ S.p.a. contiene, tra le altre, la colonna, o variabile Y, "importo fattura" e la colonna, o variabile S, "scadenza" in mesi della fattura (cioè 1, 2 o 3 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:

Y\S	1	2	3
100	50,00%		
400		25,00%	5,00%
700		5,00%	10,00%
1000			5,00%

a) Si specifichi l'unità statistica (o unità osservativa) cui si riferiscono i valori delle variabili Y e S registrati nel Data Base.
b) Si specifichi la moda della variabile S, il numero delle fatture passive con scadenza pari a tale moda, e la scadenza media delle fatture passive.
c) Si completi la seguente tabella delle frequenze per la variabile Y nei due intervalli specificati:

Y importo fattura	frequenza assoluta	frequenza relativa%	freq. ass. cumulata	freq.rel.% cumulata
[100, 500)				
[500, 1000]				

a) unità statistica = fattura passiva
b) scadenza media = 1,7
scadenza moda = 1 (mese), numero fatture con tale scadenza = 50% x 1000 = 500
c)

Y importo fattura	frequenza assoluta	frequenza relativa%	freq. ass. cumulata	freq.rel.% cumulata
[100, 500)	800	80,00%	800	80,00%
[500, 1000]	200	20,00%	1000	100,00%