Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2018-2019
STATISTICA – 17.01.19 - II PROVA PARZIALE
Traccia soluzioni Modalità A

(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 (punti 8) 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 lo standard error della stima precedente.
c) Si determini un intervallo di confidenza al 95% per la proporzione di clienti dell'istituto bancario che usufruiscono di servizi di home-banking.
d) Si verifichi, a livello 0.01, se la proporzione di clienti dell'istituto bancario che usufruiscono di servizi di home-banking è aumentata rispetto all'anno precedente, in cui era pari a 0.35.

a) $X \sim Be(p)$, $n = 200 > 25$, n. successi = 74, stima puntuale $\hat{p} = \bar{x}_{200} = \frac{74}{200} = 0.37$

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

c) $1-\alpha = 0.95$, $z_{0.975} = 1.96$, ME = $1.96 \times \text{SE} = 0.066913$

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

d) $H_1 : p > 0.35; H_0 : p = 0.35; p_0 = 0.35, \alpha = 0.01, 1-\alpha = 0.99, z_{0.99} = 2.326$

$$z_{oss} = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{0.37 - 0.35}{0.033727} = 0.592999$$

$RR = \{z_{oss} \geq z_{0.99}\}$; la condizione non è verificata, non si rifiuta H_0 al livello $\alpha = 0.01$

ESERCIZIO 2 (punti 9). Su un campione di 120 giorni viene rilevato il numero di email classificate come SPAM che arrivano ad una casella di posta elettronica; i dato 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 determini un intervallo di confidenza al 90% per il numero medio di SPAM che arrivano giornalmente alla casella di posta elettronica.
d) 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) $1-\alpha = 0.9$, $z_{0.95} = 1.645$; SE = $\sqrt{\hat{\sigma}_{\bar{X}}^2} = \sqrt{0.1\bar{3}} = 0.365148$; ME = $1.645 \times \text{SE} = 0.600669$

$IC_{0.9}(\mu) = (42 - 0.600669,\ 42 + 0.600669) = (41.39933, 42.60067)$

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

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

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

Oppure: con $\alpha = 0.1$, $1-\alpha = 0.9$ ed il test bilaterale, per **c)** sopra, si ha $\mu_0 = 40 \notin IC_{0.9}(\mu)$ e dunque si rifiuta H_0 al livello $\alpha = 0.1$.

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

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

<u>ESERCIZIO 3</u> (punti 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.
a) Si determini una stima puntuale della differenza tra il voto medio di laurea triennale in materie scientifiche dei maschi e quello delle femmine, nell'università in esame. Si determini la distribuzione dello stimatore che ha fornito la stima richiesta.
b) 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 $\mu_X\ e\ \mu_Y$.
a) La stima puntuale della differenza $\mu_Y - \mu_X$ è $\bar{y} - \bar{x} = 101.1 - 103.1 = -2$. Lo stimatore che ha fornito la stima è la differenza delle medie campionarie e la sua distribuzione è:

$$\bar{Y} - \bar{X} \sim N(\mu_Y - \mu_X, \sigma^2\left(\frac{1}{16} + \frac{1}{24}\right))$$

b) 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_p^2 è stato ottenuto come $s_p^2 = (5.2^2*15+5.1^2*23)/38=26.4166$.

<u>ESERCIZIO</u> 4 (punti 5).
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_1)..………….............** Superficie complessiva dell'albergo (in mq)
- **PREZZO (X_2)......................**Costo tipico di una notte in camera doppia (solo pernottamento), in euro

a) Si consideri il modello di regressione semplice 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 in esame:

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

a1) Si stabilisca, a livello 0.01, se la superficie dell'albergo è significativa per la spiegazione del numero di clienti. Si giustifichi la risposta.
a2) Se la risposta alla domanda precedente è positiva, si descriva l'effetto della superficie dell'albergo sul numero dei suoi clienti.
b) Si consideri ora il modello di regressione multipla che studia la dipendenza del numero di clienti che hanno alloggiato in un albergo nell'ultimo anno dalla superficie dell'albergo e dal costo tipico di una notte.
I dati rilevati nel campione hanno fornito le seguenti stime per i parametri del modello in esame:

$$b_0 = intercetta = 720\ ; b_1 = coefficiente\ di\ SUP = 6; b_2 = coefficiente\ di\ PREZZO = 22.$$

b1) Si preveda, sulla base di questo modello, il numero annuo di clienti di un albergo con superficie di 200 mq e con costo per una notte pari a 50 euro.

a1) 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.

a2) 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.

b1) La previsione è: 720+6*200+22*50=3020.