1

Università C. Cattaneo, Corso di Laurea in Economia Aziendale, A.A. 2016-2017
STATISTICA — 26.01.17 - II PROVA PARZIALE
Modalità B

(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 9).

a) Si vogliono confrontare i tempi di attesa in coda a due impianti sciistici, A e B. A tale scopo si estrae un campione di 70 sciatori in coda all'impianto A e si osserva un tempo medio di attesa pari a 6 minuti e una deviazione standard pari a 2.5. Si estrae un campione di 80 sciatori in coda all'impianto B e si osserva un tempo medio di attesa pari a 8 minuti e una deviazione standard pari a 2. Ipotizzando che i tempi di attesa in coda ai due impianti abbiano la medesima varianza e distribuzioni normali, si può concludere al livello di significatività del 5% che i due tempi medi di attesa sono diversi?

b) L'impianto A è appena stato rinnovato. Il tempo medio di attesa lo scorso anno era di 6.5 minuti. Si verifichi al livello di significatività del 10% se il rinnovo ha portato ad una riduzione del tempo medio di attesa rispetto allo scorso anno. Si effettui il test usando il p-value, del quale si riporti il valore.

c) Si costruisca un intervallo di confidenza di livello 99% per il tempo medio di attesa all'impianto B. In base all'intervallo costruito si può concludere che il tempo medio di attesa all'impianto B è diverso da 7?

a) $\begin{cases} H_0: \mu_X = \mu_Y \\ H_1: \mu_X \neq \mu_Y \end{cases}$ μ_X = attesa media impianto A, $X \sim N(\mu_X, \sigma^2)$; μ_Y = attesa media impianto B, $Y \sim N(\mu_Y, \sigma^2)$

$\bar{x} = 6$, $s_X = 2.5$, $n_X = 70$; $\bar{y} = 8$, $s_Y = 2$, $n_Y = 80$ $\alpha = 0.05$

$R_{0.05}$: $|Z| = \left| \dfrac{\bar{x} - \bar{y}}{\sqrt{\frac{s_p^2}{n_X} + \frac{s_p^2}{n_Y}}} \right| > z_{1-\alpha/2} = z_{0.975} = 1.96$

$s_p^2 = \dfrac{2.5^2 \cdot 69 + 2^2 \cdot 79}{148} = 5.0490$ $|z_{oss}| = \dfrac{|6-8|}{\sqrt{\frac{5.049}{70} + \frac{5.049}{80}}} = 5.4385$

Poiché $5.4385 > 1.96$, si rifiuta H_0, si conclude che i due tempi medi sono diversi.

b) $\begin{cases} H_0: \mu_X = 6.5 \\ H_1: \mu_X < 6.5 \end{cases}$ p-value $= P\left(Z < \dfrac{6-6.5}{2.5/\sqrt{70}}\right) = P(Z < -1.6733) = 0.0475$

Poiché il p-value è minore di 0.1, si rifiuta H_0 e si conclude che il tempo medio è diminuito.

c) $\left(\bar{y} - 2.576 \cdot \frac{s_Y}{\sqrt{n}}, \bar{y} - 2.576 \cdot \frac{s_Y}{\sqrt{n}}\right) = (\bar{y} - 0.576, \bar{y} + 0.576) = (7.424, 8.576)$

$\begin{cases} H_0: \mu_Y = 7 \\ H_1: \mu_Y \neq 7 \end{cases}$ Poiché $7 \notin$ intervallo, si conclude che H_0 è da rifiutare a livello 0.01, per cui c'è evidenza che $\mu_Y \neq 7$.

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

ESERCIZIO 2 (punti 6). Su un campione di 300 clienti di un supermercato si rilevano: l'entità della spesa effettuata, classificata nelle due modalità BASSA (se non superiore a 40 €) o ALTA (se superiore a 40 €) e la forma di pagamento (CONTANTE o ALTRO). I dati rilevati indicano che 104 clienti del campione hanno usato mezzo diverso dal contante per il pagamento e che 120 clienti del campione hanno effettuato una spesa ALTA.

a) Si determini una stima puntuale della proporzione di clienti del supermercato che effettuano il pagamento con mezzo diverso dal contante.

b) Cosa si può dire della distribuzione dello stimatore da cui proviene la stima ottenuta al punto precedente?

c) Si stabilisca a livello 0.01 se è aumentata la proporzione di clienti che effettuano una spesa ALTA rispetto all'anno precedente, in cui era pari a 0.3.

a) $\bar{x} = \hat{p} = 104/300 = 0.3467$

b) Lo stimatore è la media campionaria (o proporzione campionaria), con distribuzione approssimativamente $N\left(p, \frac{p(1-p)}{n}\right)$ (in quanto n "grande")

c) $\begin{cases} H_0: p_{ALTA} = 0.3 \\ H_1: p_{ALTA} > 0.3 \end{cases}$ $\bar{x}_{ALTA} = \hat{p}_{ALTA} = \frac{120}{300} = 0.4$

$R_{0.01}: \quad Z = \frac{\hat{p}_{ALTA} - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{300}}} > z_{0.99} = 2.326$

$z_{oss} = \frac{0.4 - 0.3}{\sqrt{\frac{0.3 \cdot 0.7}{300}}} = 3.7796 > 2.326$

Si rifiuta H_0 a livello 0.01, per cui si può concludere che la proporzione p_{ALTA} è aumentata.

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

<u>ESERCIZIO</u> **3** (punti 6).
Si considera un campione di 12 mesi, in ciascuno dei quali si rileva il rendimento di un fondo; il rendimento mensile medio nei 12 mesi è stato di 0.023, lo scarto quadratico medio dei rendimenti dei 12 mesi è stato 0.003.
a) Si stabilisca, a livello 0.05, se il rendimento medio mensile del fondo è maggiore di 0.02.
b) Si determini un intervallo di confidenza al 95% per il rendimento mensile medio del fondo.

È necessario assumere che il rendimento abbia distribuzione normale: $X \sim N(\mu, \sigma^2)$.

a) $\begin{cases} H_0: \mu = 0.02 \\ H_1: \mu > 0.02 \end{cases}$ $R_{0.05}: T = \dfrac{\bar{X} - 0.02}{S/\sqrt{n}} > t_{11, 0.95} = 1.796$

$T_{oss} = \dfrac{0.023 - 0.02}{0.003/\sqrt{12}} = 3.4641 > 1.796$, per cui si rifiuta H_0 e si conclude che il rendim. medio è aumentato (a livello 0.05).

b) $\left(\bar{x} - 2.201 \cdot \frac{S}{\sqrt{n}}, \bar{x} + 2.201 \cdot \frac{S}{\sqrt{n}}\right) =$

$= \left(0.023 - 2.201 \cdot \frac{0.003}{\sqrt{12}}, 0.023 + 2.201 \cdot \frac{0.003}{\sqrt{12}}\right) =$

$= (0.023 - 0.0019, 0.023 + 0.0019) =$

$= (0.0211, 0.0249)$

4

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

ESERCIZIO 4 (punti 6).

Si vuole studiare la dipendenza del consumo mensile di acqua (in Mc) delle famiglie italiane dal numero di componenti familiari. A questo scopo si stima un modello di regressione lineare, sulla base di un campione di 120 famiglie; i dati rilevati sul campione hanno fornito un output Excel riportato (in parte) di seguito.

	Coefficienti	*Errore Standard*	*Stat t*	*Sig. P-value*	*Inferiore 95%*	*Superiore 95%*
Intercetta	8.149824921	0.103139800	79.01727285	1.02E-99		
num_comp_fam	2.262517656	0.281574923	8.03522400	4.440892E-16		

a) Si stabilisca, attraverso un'opportuna procedura condotta a livello 0.05, se il numero di componenti familiari ha un effetto significativo sul consumo di acqua.
b) Si preveda, sulla base del modello, la media del consumo di acqua delle famiglie con 3 componenti.
c) Si considera ora il modello che, oltre alla variabile esplicativa X: "numero di componenti familiari", ha la variabile esplicativa W: "superficie dell'appartamento, in mq, in cui vive la famiglia"; l'equazione stimata del modello è:

$$\hat{y} = 7.134 + 2.111\,x + 0.003w$$

Si descriva l'effetto che ha la superficie dell'appartamento sul consumo di acqua di una famiglia.

a) $\begin{cases} H_0: \beta_{NUM\_COMP.FAM} = 0 \\ H_1: \beta_{NUM-COMP-FAM} \neq 0 \end{cases}$

p-value $= 4.44 \cdot 10^{-16} \simeq 0 < 0.05$, per cui si rifiuta H_0; il num. di comp. fam. ha un effetto significativo.

b) PREVISIONE $= 8.1498 + 2.2625 \cdot 3 = 14.9373$

c) La stima del coeff. della variabile W è 0.003; quindi, fissato il numero di componenti familiari, ad un incremento di un mq di superficie è associato un incremento medio di consumo di acqua pari a 0.003 Mc.