Nome e cognome: ______________________________ *Laurea a luglio:* ☐

indirizzo email: ______________________________

Statistica – Esercitazione del 18 maggio 2018 – Gruppo B

B1. Per valutare se l'uso del telefono cellulare influisce sui tempi di reazione durante la guida è stato condotto un esperimento.
Due gruppi di studenti, ciascuno dei quali composto da 26 individui selezionati a caso, sono stati sottoposti ad una simulazione di guida. Durante la simulazione, gli individui di uno dei gruppi sentivano musica alla radio mentre quelli dell'altro gruppo erano impegnati in una conversazione al cellulare. Sono stati misurati i tempi di reazione con cui gli individui hanno schiacciato il pedale del freno in seguito all'accensione di una luce rossa. Per il "gruppo radio" il tempo medio di reazione è stato di 533.7 millisecondi con una deviazione standard di 65.3, per il "gruppo cellulare" il tempo medio di reazione è stato 558.2 con deviazione standard di 98.6.
Ipotizzando che i tempi di reazione di ciascun individuo seguano una distribuzione normale la cui varianza non dipende dal gruppo di appartenenza, stabilire se questi dati permettono di affermare che l'uso del cellulare allunghi i tempi di reazione.

(a) Si considerino le ipotesi seguenti:
(i) L'uso del cellulare alla guida non allunga i tempi di reazione;
(ii) L'uso del cellulare alla guida allunga i tempi di reazione.
Stabilire quale delle due debba essere l'ipotesi nulla e quale l'alternativa.

(b) Esprimere il sistema di ipotesi risultante dal punto precedente in termini di uguaglianze e/o disuguaglianze sui parametri delle distribuzioni.

(c) Svolgere il test con un livello di significatività pari al 5%.

(d) Conseguentemente a quanto visto nei punti precedenti, quale conclusione si può trarre?

---

argomentare tutti i passaggi in modo adeguato

(a) Come ipotesi alternativa bisogna considerare quella che vorremmo poter dimostrare, quindi che l'uso del cellulare allunga i tempi di reazione. Di conseguenza:

H_0 : *L'uso del cellulare alla guida non allunga i tempi di reazione*

H_1 : *L'uso del cellulare alla guida allunga i tempi di reazione*

(b) Indicando con μ_R il valore atteso relativo al "gruppo radio" e μ_C quello relativo al "gruppo cellulare", il sistema di ipotesi diventa

$$H_0 : \mu_C - \mu_R \leq 0 \qquad \textit{contro} \qquad H_0 : \mu_C - \mu_R > 0.$$

(c) Si tratta di un test unilaterale sulla differenza tra medie di normali con uguale varianza (incognita). La statistica test è

$$T = \frac{\bar{X}_C - \bar{X}_R}{S_p\sqrt{\frac{1}{n_C} + \frac{1}{n_R}}}$$

che sotto H_0 ha distribuzione T con $n_R + n_C - 2$ gradi di libertà.

Rifiutiamo l'ipotesi nulla per valori alti della statistica test, quindi la regione critica (di rifiuto) è $\{T \geq t_{0.05,50}\}$.

I dati del campione osservato sono:

$$\bar{x}_C = 558.2 \qquad \bar{x}_R = 533.7 \qquad s_C^2 = 98.6^2 = 9721.96 \qquad s_R^2 = 65.3^2 = 4264.09$$

per cui $s_p = \sqrt{\frac{25\cdot 9721.96 + 25\cdot 7276.09}{50}} = 83.62431.$

Il valore osservato della statistica test $\frac{558.2-533.7}{83.62431\sqrt{2/26}} = 1.056344$ *è maggiore di* $t_{0.05,50} = 1.676$ *per cui l'ipotesi nulla NON VA RIFIUTATA.*

(d) I dati non forniscono evidenza sufficiente a stabilire che l'uso del cellulare alla guida allunghi i tempi di reazione.

B2. Un contapassi negli ultimi 7 giorni lavorativi ha riportato i valori seguenti:

6822 5333 7240 7432 6252 7005 6572

Nell'ipotesi che i valori quotidiani del contapassi costituiscano un campione casuale da una popolazione normale, si vuole costruire un intervallo di predizione al 95% per il valore che sarà rilevato domani sul contapassi.
Indichiamo $\bar{X}_7$ e S_7^2 la media e la varianza campionaria relative ad un campione di giorni $X_1, \cdots, X_7$. Sia Y il valore relativo a domani.

(a) Determinare la distribuzione di $Y - \bar{X}_7$.

(b) Dedurre la distribuzione della variabile aleatoria

$$\frac{Y - \bar{X}_7}{S_7\sqrt{1 + \frac{1}{7}}}.$$

(c) Dal risultato in (b) ottenere un intervallo di predizione che contenga il valore Y con un grado di fiducia pari a 95%.

---

argomentare tutti i passaggi in modo adeguato

(a) Dalle ipotesi fatte $X_1, \cdots, X_7, Y$ sono indipendenti, tutte con distribuzione $N(\mu, \sigma^2)$. Segue che

$$Y \sim N(\mu, \sigma^2) \qquad \bar{X}_7 \sim N(\mu, \sigma^2/7) \qquad 6\frac{S_7^2}{\sigma^2} \sim \chi_6^2$$

sono indipendenti l'una dall'altra. Di conseguenza $Y - \bar{X}_7$, essendo differenza di due normali indipendenti, ha distribuzione normale con valore atteso $\mu - \mu = 0$ e varianza $\sigma^2 + \sigma^2/7 = \sigma^2(1 + \frac{1}{7})$.

(b) Le variabili aleatorie $Z = \frac{Y - \bar{X}_7}{\sqrt{\sigma^2(1+\frac{1}{7})}} \sim N(0,1)$ e $6\frac{S_7^2}{\sigma^2} \sim \chi_7^2$ sono indipendenti, per cui

$$\frac{Y - \bar{X}_7}{S_7\sqrt{1 + \frac{1}{7}}} = \frac{Z}{\sqrt{\frac{6\frac{S_7^2}{\sigma^2}}{6}}} \sim T_6.$$

(c) Da quanto visto al punto precedente si ha che

$$\begin{aligned} 1 - \alpha &= P\left(-t_{6,\alpha/2} < \frac{Y - \bar{X}_7}{S_7\sqrt{1 + \frac{1}{6}}} < t_{6,\alpha/2}\right) = \cdots \\ &= P\left(Y \in \left(\bar{X}_7 - t_{6,\alpha/2}S_7\sqrt{1 + \frac{1}{6}}, \bar{X}_7 - t_{6,\alpha/2}S_7\sqrt{1 + \frac{1}{7}}\right)\right). \end{aligned}$$

Nell'intervallo aleatorio appena determinato, inserendo i dati osservati $\bar{x}_7 = 6665.143$ e $s_7 = 708.5907$ al posto delle variabili casuali campionarie $\bar{X}_7$ e S_7 otteniamo l'intervallo di previsione

$$\left(\bar{x}_7 - t_{6,0.025}s_7\sqrt{1 + \frac{1}{7}}, \bar{x}_7 + t_{6,0.025}s_7\sqrt{1 + \frac{1}{7}}\right) = (4811.503, 8518.783)$$

B3. Sono stati rilevati alcuni dati su 102 case vendute di recente in una contea della Florida. Seguono i valori di sintesi dei dati osservati (indichiamo con v_i i valori relativi al prezzo di vendita (in migliaia di dollari) e con w_i quelli relativi alla superficie calpestabile (in piedi quadrati):

$$\sum_{i=1}^{102} v_i = 158430 \qquad \sum_{i=1}^{102} v_i^2 = 280526700 \qquad \sum_{i=1}^{102} w_i = 13134.8 \qquad \sum_{i=1}^{102} w_i^2 = 2028396 \qquad \sum_{i=1}^{102} v_i w_i = 23056172$$

Costruire un modello regressione per stabilire una relazione tra prezzo di vendita e superficie.

(a) Stabilire in quale delle due variabili è opportuno sia la variabile indipendente x e quale la variabile dipendente y.

(b) Stimare i parametri della regressione.

(c) Costruire un intervallo di confidenza al 95% per il coefficiente angolare β.

(d) Un agente immobiliare sostiene che il prezzo di vendita aumenta di 100 dollari per ogni piede quadrato in più. Sulla base dei risultati ottenuti ai punti precedenti, ritieni che l'affermazione sia fondata? Giustifica la tua conclusione in modo esauriente.

---

argomentare tutti i passaggi in modo adeguato

(a) Chiaramente il prezzo di vendita dipende dalla superficie e non viceversa, quindi x deve essere la superficie e Y il prezzo di vendita.
(b) Dai dati otteniamo

$$\bar{x} = 128.7725 \quad \bar{y} = 1553.235 \quad S_{xx} = 336994.2 \quad S_{YY} = 34447632 \quad S_{xY} = 2654737$$

quindi $B = 7.877691$, $A = 538.805$ e la retta di regressione stimata è $y = 538.805 + 7.877691x$.
(c) L'intervallo di confidenza per β è

$$B \mp t_{\alpha/2,n-2} \sqrt{\frac{SS_R}{(n-2)S_{xx}}}.$$

Dal momento che $t_{0.025,100} = 1.984$ e che $SS_R = S_{YY} - S_{xY}^2/S_{xx} = 13534435$, l'intervallo è

$$7.877691 \mp 1.984\sqrt{\frac{13534435}{100 \cdot 336994.2}} = (6.620357, 9.135024).$$

(d) I dati dicono che l'incremento è circa 8000 dollari per ogni piede quadro aggiuntivo.