Compito scritto dell'esame di Statistica e analisi dei dati 14 giugno 2024	Prof. Giuseppe Boccignone	Corso di Laurea
Cognome:	Nome:	Matricola:

Istruzioni

- Il tempo riservato alla prova scritta e' di 2 ore. Durante la prova e' possibile consultare il formulario ed utilizzare la calcolatrice. Non e' possibile consultare libri, appunti, cellulari.
- Ogni foglio deve riportare il numero di matricola
- In ogni esercizio occorre indicare chiaramente, per ogni risposta, il numero della domanda corrispondente
- Riportare lo svolgimento degli esercizi per esteso (quando l'esercizio richiede piu' passaggi di calcolo, non sara' preso in considerazione se riporta solo le soluzioni). Se una serie di calcoli coinvolge una o piu' frazioni semplici (numeratore e denominatore interi), per chiarezza, si svolgano i calcoli mantenendo tali numeri in forma frazionaria fin dove possibile (non li si converta nelle loro approssimazioni con virgola e decimali: solo il risultato finale sara' eventualmente rappresentato in quest'ultima forma).

Problemi

ESERCIZIO 1.
Si stima che il 30% degli adulti negli Stati Uniti siano obesi, che il 3% siano diabetici e che il 2% siano sia obesi che diabetici. Determina la probabilitá che un individuo scelto casualmente:

(a) sia diabetico se é obeso;

Soluzione: Definiamo i seguenti eventi

- $O =$ `"un individuo scelto casualmente sia obeso"`
- $D =$ `"un individuo scelto casualmente sia di diabetico"`

Dai dati del problema:

$$P(O) = 0.30, P(D) = 0.03, P(O \cap D) = 0.02$$

Pertanto:

$$P(D \mid O) = \frac{P(O \cap D)}{P(O)} = \frac{0.02}{0.3} = 0.067$$

(b) sia obeso se é diabetico.

Soluzione:

$$P(O \mid D) = \frac{P(O \cap D)}{P(D)} = \frac{0.02}{0.03} = 0.667$$

ESERCIZIO 2. Un sensore di temperatura ha un tasso medio di malfunzionamento di un guasto su 8 anni.

(a) Qual é la probabilitá che funzioni per almeno 10 anni?

Soluzione:
Si assuma una distribuzione esponenziale negativa del tempo di attesa T del malfunzionamento, $T \sim Exp(t; \lambda)$.
Il parametro della distribuzione esponenziale é il rate

$$\lambda = \frac{1}{8}$$

La probabilitá di non avere malfunzionamenti per almeno dieci anni é la probabilitá di sopravvivenza $P(T \geq t) = S(t) = e^{-\lambda t}$, ovvero

$$P(T \geq 10) = S(10) = e^{-\frac{1}{8} \cdot 10} = e^{-\frac{5}{4}} \approx 0.29$$

ESERCIZIO 3.
L'altezza dei figli di un uomo alto 180 centimetri é normalmente distribuita con media 182 e varianza 10. Qual é la probabilitá che il figlio dell'uomo:

(a) abbia un'altezza compresa tra 180 e 184 cm?

Soluzione:

Abbiamo per la VA X che denota l'altezza dei figli :

$$X \sim \mathcal{N}(x; \mu = 182, \sigma^2 = 10)$$

Poiché $\sigma = \sqrt{10}$:

$$P(180 \leq X \leq 184) = P(\frac{180-\mu}{\sigma} \leq Z \leq \frac{184-\mu}{\sigma}) = P(-\frac{2}{\sqrt{10}} \leq Z \leq \frac{2}{\sqrt{10}}) = 2\Phi(0.63) - 1 = 2 \cdot 0.7357 - 1 \approx 47\%$$

dove $\Phi(0.63) = 0.7357$ é stato letto nella tabella della CDF della Normale standard

(b) superi i 185 cm di altezza?

Soluzione:

$$P(X > 185) = 1 - P(X \leq 185) = 1 - P(Z \leq \frac{185-\mu}{\sigma}) = 1 - P(Z \leq \frac{3}{\sqrt{10}}) = 1 - \Phi(0.95) = 1 - 0.8289 \approx 17\%$$

Esercizio 4. Si vuole stimare la proporzione di elettori che approva l'operato del capo del governo. Su un campione di 150 persone intervistate, 90 si sono dichiarate favorevoli.

(a) Determinare un intervallo di confidenza con grado di fiducia del 95% per la proporzione degli elettori favorevoli al capo del governo

Soluzione:

L' intervallo di confidenza bilaterale per la proporzione p di una popolazione bernoulliana, con grado di fiducia $(1-\alpha)100\% = 95\%$, dunque $\alpha = 0.05$, e campione grande, si scrive in approssimazione normale

$$\hat{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

La taglia del campione é $n = 150$

La proporzione campionaria dei favorevoli é

$$\hat{p} = \frac{x}{n} = \frac{90}{150} = 0.6$$

Quindi, con $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

$$0.6 - 1.96\sqrt{\frac{0.6(1-0.6)}{150}} < p < 0.6 + 1.96\sqrt{\frac{0.6(1-0.6)}{150}}$$

$$\implies 0.52 < p < 0.68$$

Pertanto, la percentuale dei favorevoli, con un grado di fiducia del 95%, é compresa in una forchetta tra il 52% e il 68%

Esercizio 5. La precisione di una macchina che produce componenti di dimensioni specificate viene controllata con periodiche verifiche a campione: la dimensione media richiesta é 3.5 mm, con una varianza di 0.22 mm^2.

(a) Valutare mediante un test di ipotesi (con livello di significativitá del 5%) se il processo é da ritenersi sotto controllo oppure no, quando la media riscontrata su un campione di 60 pezzi é 3.42 mm

Soluzione:

Abbiamo che

$$n = 60, \bar{x} = 3.42\ mm.$$

Il campione é ampio e $\sigma^2 = 0.22\ mm^2$ é nota.

Possiamo assumere che $\bar{X}$ sia distribuita con legge normale $\bar{X} \sim \mathcal{N}(\mu_0 = 3.5, \frac{\sigma^2}{n})$.

Si utilizza un test di ipotesi a due code e si sceglie come ipotesi nulla di ritenere che il processo sia sotto controllo e non sia quindi necessario alcun intervento; l' ipotesi alternativa é che il processo sia fuori controllo:

$$\begin{cases} H_0 & : \mu = 3.5 = \mu_0 \\ H_1 & : \mu \neq 3.5. \end{cases} \tag{1}$$

Il valore della statistica test é

$$Z_{\bar{X}} = \frac{\bar{3}.42 - 3.5}{\sqrt{0.22}/\sqrt{60}} = -1.32$$

La regione di rifiuto é

$$|Z_{\bar{X}}| > z_{\frac{\alpha}{2}}$$

con $z_{\frac{\alpha}{2}} = z_{0.025}$. Mediante la tavola della Normale standard, $z_{0.025} = 1.96$.
Dunque

$$|Z_{\bar{X}} = -1.32| \leq 1.96$$

ovvero Il valore $Z_{\bar{X}} = -1.32$ non appartiene alla regione di rifiuto, quindi l' ipotesi nulla non viene rifiutata al livello di significativitá del 5%.
Il processo é da ritenersi sotto controllo.

Esercizio 6.
Vengono sottoposti a confronto i consumi delle autovetture A e B alla velocitá costante di $120Km/h$. Si ritiene che i consumi dei due tipi di autovetture possa essere descritto da variabili aleatorie con distribuzione normale con la stessa varianza. L'auto di tipo A in 20 prove consuma mediamente $6.5litri/100Km$, quella di tipo B in 22 prove consuma mediamente $6.6litri/100Km$. Le relative varianze campionarie sono rispettivamente di 0.30 e 0.28.

(a) Ragionando con gli intervalli di confidenza, possiamo ritenere che le due autovetture abbiano lo stesso consumo medio al livello di significativitá del 5%?

Soluzione:
Abbiamo che:

- $\bar{x}_A = 6.5, s_A^2 = 0.30, n_A = 20$
- $\bar{x}_B = 6.6, s_B^2 = 0.28, n_B = 22$

Siamo nel caso di varianze ignote ma eguali, e la differenza fra medie é valutabile mediante una statistica t-Student, dove i gradi di libertá ν sono dati da $\nu = n_A + n_B - 2 = 40$, $\alpha = 0.05$, per cui, $t_{\nu,\frac{\alpha}{2}} = t_{40,0.025} \approx 2.021$
La deviazione standard pooled vale:

$$S_{pool} = \sqrt{\frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2}} = 0.5381$$

L'intervallo di confidenza per la differenza fra le medie $\mu_A - \mu_B$ a livello $(100)(1-\alpha) = (100)(1-0.05) = 95\%$ vale pertanto

$$\bar{x}_A - \bar{x}_B \pm t_{40,0.025} S_{pool} \sqrt{\frac{1}{n_A} + \frac{1}{n_B}} = -0.1 \pm 0.33586$$

ovvero l'IC ha estremi $[-0.43596; 0.23596]$ per cui si puó ragionevolmente ritenere che le due autovetture abbiano differenza in consumo medio trascurabile.
Equivalentemente, si osserva che la statistica della differenza fra medie normalizzata

$$\left| \frac{\bar{x}_A - \bar{x}_B}{S_{pool}\sqrt{\frac{1}{n_A} + \frac{1}{n_B}}} \right| = 0.612 < t_{40,0.025} = 2.021$$