

Statistica Inferenziale

Soluzioni 1. Distribuzioni campionarie

Esercizio A.

a) Sia C la variabile aleatoria che descrive il costo nel prossimo volo. Allora $P(C > 576,15) = P\left(\frac{C}{6,00 \cdot 25} > \frac{576,15}{6,00 \cdot 25}\right) = P(\chi_1^2 > 3,841)$, dove con χ_1^2 si è indicata una variabile aleatoria con distribuzione chi-quadrato con un grado di libertà. Dalle tavole si ricava che $P(\chi_1^2 > 3,841) = 0,05$.

b) Sia $C_T = \sum_{i=1}^{12} C_i$ il costo aggiuntivo complessivo di 12 voli. Essendo il costo di ogni volo indipendente da quello degli altri voli, $\frac{C_T}{6,00 \cdot 25}$ si distribuisce come una chi-quadrato con 12 g.d.l. Da cui si ricava che $P(C_T < 3.500,55) = P\left(\frac{C_T}{6,00 \cdot 25} < \frac{3.500,55}{6,00 \cdot 25}\right) = P(\chi_{12}^2 < 23,337) = 1 - 0,025 = 0,975$.

c) Sia G_i il guadagno dell' i -esimo volo, allora $E(G_i) = 1.500$ e $\text{Var}(G_i) = 50^2$. Sia $D = G_i - G_{i+1}$ la differenza del guadagno in due voli successivi. Essa è distribuita come una v. a. normale con valore atteso pari a zero. Inoltre, essendo G_i e G_{i+1} indipendenti, $\text{Var}(D) = \text{Var}(G_i) + \text{Var}(G_{i+1}) = 70,711^2$. Dalla simmetria della normale, $P(|D| > 120) = 2 \cdot P\left(\frac{D}{70,711} > \frac{120}{70,711}\right) = 2 \cdot (1 - \Phi(1,697)) = 2 \cdot (1 - 0,95513) = 0,08974$. Il valore 0,95513 è stato ottenuto per interpolazione lineare ponendo

$$\frac{\Phi(1,697) - \Phi(1,690)}{1,697 - 1,690} = \frac{\Phi(1,70) - \Phi(1,69)}{0,01}.$$

Sostituendo si ottiene $(\Phi(1,697) - 0,9633)/0,007 = (0,9554 - 0,9545)/0,01$. Da cui si ricava $\Phi(1,697) = 0,9545 + 0,01 \cdot 0,0009/0,007 = 0,95513$.

d) $G_T = \sum_{i=1}^6 G_i$ è la somma dei guadagni di 6 voli. $E(G_T) = 9.000$ e $\text{Var}(G_T) = 6 \cdot \text{Var}(G_i) = 122,474^2$. Da cui si ricava che $P(G_T > 9.500) = P(Z > 4,082) = 1 - \Phi(4,082) = 1 - 1 = 0$ (dove si è indicato con Z una variabile aleatoria con distribuzione normale standardizzata).

e) Ponendo $U_T = \sum_{i=1}^6 U_i$ si ha che U_T è una v. a. normale con $E(U_T) = 720$ e $\text{Var}(U_T) = 216^2$. Inoltre, $C_T/(6 \cdot \text{Var}(T))$ è una v. a. chi-quadrato con 6 g.d.l. Da cui si ricava che $R \cdot \sqrt{6 \cdot 6 \cdot \text{Var}(T)}/\sqrt{\text{Var}(U_T)}$ si distribuisce come una t di Student con 6 g.d.l., per cui $P(R > 13,99) = P(R \sqrt{6 \cdot 6 \cdot 25}/216 > 1,943)$. Dalle tavole si ricava che la probabilità richiesta è data da $P(t > 1,943) = 0,05$, dove con t si è indicata una v. a. con distribuzione t di Student.

f) Il 90-esimo percentile di una t di Student con 6 g.d.l. è pari a 1,44. Ovvero possiamo scrivere

$$P\left(R \cdot \frac{\sqrt{6 \cdot 6 \cdot \text{Var}(T)}}{\sqrt{\text{Var}(U_T)}} < 1,44\right) = 0,90,$$

da cui $r = 1,44 \sqrt{\text{Var}(U_T)}/\sqrt{6 \cdot 6 \cdot \text{Var}(T)} = 10,368$.

Esercizio B.

a) Considerando $n = 30$, la probabilità richiesta è data da

$$P(|\bar{X}_A - \mu_A| \geq 0,5 \cdot \sqrt{\sigma_A^2}) = P\left(\frac{|\bar{X}_A - \mu_A|}{\sqrt{\sigma_A^2/n}} \geq 0,5 \cdot \sqrt{n}\right) = 2 \cdot (1 - \Phi(2,7386)) = 0,0062.$$

b) Considerando che $n = 30$, possiamo scrivere

$$0,95 = P\left(\left|\frac{\bar{X}_A - \mu_A}{\sqrt{S_A^2}}\right| \leq y\right) = P\left(\frac{\frac{|\bar{X}_A - \mu_A|}{\frac{\sigma_A}{\sqrt{n}}}}{\sqrt{\frac{S_A^2}{\sigma_A^2}}} \leq y \cdot \sqrt{n}\right) = P(|t_{29}| \leq y \cdot \sqrt{n}).$$

Dalle tavole si trova che $y \cdot \sqrt{n} = 2,045$ da cui $y = 2,045/\sqrt{30} = 0,3734$.

c) Considerando che $n = 30$, possiamo scrivere

$$0,95 = P\left(\frac{|\bar{X}_A - \mu_A|}{\sqrt{D_A^2}} \leq z\right) = P\left(\frac{\frac{\sqrt{n}|\bar{X}_A - \mu_A|}{\sigma_A}}{\sqrt{\frac{D_A^2}{\sigma_A^2}}} \leq z \cdot n\right) = P(|t_{30}| \leq z \cdot n).$$

Dalle tavole si ricava che $z \cdot n = 2,042$ da cui $z = 2,042/30 = 0,0681$.

d) Possiamo scrivere

$$P[(\bar{X}_B - \bar{X}_A) - (\mu_B - \mu_A) \leq 2] = P\left[\frac{(\bar{X}_B - \bar{X}_A) - (\mu_B - \mu_A)}{\sqrt{\frac{\sigma_A^2}{30} + \frac{\sigma_B^2}{20}}} \leq \frac{2}{\sqrt{\frac{\sigma_A^2}{30} + \frac{\sigma_B^2}{20}}}\right],$$

da cui, indicando con Z una variabile aleatoria normale standardizzata, si ha che la probabilità richiesta è data da $\Phi(Z \leq 1,92154) = 0,97267$.

e) Possiamo scrivere

$$0,05 = P(S_A^2 \geq d) = P\left(\frac{S_A^2}{\sigma_A^2} \cdot 29 \geq \frac{d}{\sigma_A^2} \cdot 29\right) = \left(\chi_{29}^2 \geq \frac{d}{10} \cdot 29\right).$$

Dalle tavole si trova che $29 \cdot d/30 = 42,557$ da cui $d = 14,6748$.

Esercizio C.

a) Sia X_A la v. a. che conta il numero di clienti attivi nel primo campione di numerosità $n_A = 50$ e sia $\hat{P}_A = \frac{X_A}{n_A}$. Per il teorema del limite centrale, \hat{P}_A ha approssimativamente distribuzione normale con media p_A e varianza $\frac{p_A(1-p_A)}{n_A}$. Analogamente sia X_B la v. a. che conta il numero di clienti attivi nel secondo campione di numerosità $n_B = 50$ e sia $\hat{P}_B = \frac{X_B}{n_B}$. Per il teorema del limite centrale \hat{P}_B ha approssimativamente distribuzione normale con media p_B e varianza $\frac{p_B(1-p_B)}{n_B}$. Quindi $\hat{P}_B - \hat{P}_A$ ha approssimativamente distribuzione normale con media $p_B - p_A$ e varianza pari alla somma delle due varianze asintotiche. Si vuole la probabilità di osservare uno scostamento fra le due frequenze campionarie superiore a quello osservato, cioè si vuole

$$P\left(\hat{P}_B - \hat{P}_A > \frac{27}{50} - \frac{23}{50}\right),$$

nel caso in cui $p_A = p_B = p$. In tale caso, approssimativamente, $\frac{\hat{P}_B - \hat{P}_A}{\sqrt{\frac{2p(1-p)}{50}}} \sim N(0;1)$ dove il

valore di p non si conosce. Per un'estensione del teorema del limite centrale, l'approssimazione alla distribuzione normale continua a valere anche sostituendo a p nella espressione di sopra la stima \hat{p} . Avendo noi $\hat{p} = (\hat{p}_A 50 + \hat{p}_B 50)/100 = 50/100 = 0,5$ si ha $P(\hat{P}_B - \hat{P}_A > 0,54 - 0,46) = P\left(\frac{\hat{P}_B - \hat{P}_A}{\sqrt{2 \cdot \frac{0,5^2}{50}}} > \frac{0,54 - 0,46}{\sqrt{2 \cdot \frac{0,5^2}{50}}}\right) \approx P(Z > 0,8) = 1 - \Phi(0,8) = 0,2119$, dove con Z si è indicata una v. a. con distribuzione normale standardizzata.