Corso di Statistica - Prof. Fabio Zucca	Appello 0 - 1 luglio 2017		
Nome e cognome:		**Matricola:**	
			12957

Esercizio 1 In uno studio per la rilevazione dell'inquinamento atmosferico si sono registrati le seguenti concentrazioni in $\mu g \cdot m^{-3}$

$$2.2 \quad 1.8 \quad 3.1 \quad 2.0 \quad 2.4 \quad 2.0 \quad 2.1 \quad 1.2$$

Sapendo che $\sum_i x_i = 16.8$ e $\sum_i x_i^2 = 37.3$ si proceda allo svolgimento dei seguenti punti. Si supponga che i campioni provengano da famiglie normali.

1. Stimare la media e la varianza con degli stimatori non distorti.
2. Sotto l'ipotesi di normalità si determini un intervallo di confidenza per la media al 95%.
3. È plausibile che $\mu \neq 2$ al 5%? Stimare il P-value fornendone un intervallo.
4. In una seconda zona si misurano 10 concentrazioni ottenendo $\sum_i y_i = 20$ e $\sum_i y_i^2 = 42$. Sotto l'ipotesi di omoschedasticità, é plausibile l'ipotesi $\mu_X > \mu_Y$ al 5%? Stimare il P-value fornendone un intervallo.

Soluzione.

1. La stima della media si ottiene utilizzando la media campionaria: $\bar{x}_8 = n^{-1}\sum_{i=0}^{n} x_i = 8^{-1}\cdot 16.8 = 2.1$. Mentre la varianza campionaria $s_8^2 = (n-1)^{-1}\sum_{i=0}^{n} x_i^2 - n(n-1)^{-1}\bar{x}_8^2 = 7^{-1}\cdot 37.3 - 8\cdot 7^{-1}\cdot 2.1^2 = 0.2885714286$.

2. L'intervallo di confidenza per la media a varianza incognita ha come estremi $\bar{x}_n \pm t_{(1+\alpha)/2}(n-1)s_n/\sqrt{n}$. Essendo $t_{0.975}(7) = 2.364623$ e $\sqrt{s_8^2/8} = 0.1899247972$ si ricavano gli estremi $2.1 \pm 2.364623\cdot\sqrt{0.2885714286/8} = 2.1 \pm 0.4491005437$. L'intervallo è quindi $[1.650899456, 2.549100544]$.

3. Sia $H_0\colon \mu = 2 =: \mu_0$ e $H_1\colon \mu \neq 2$. Dal punto recedente si ottiene subito che $2 \in [1.650899456, 2.549100544]$ quindi non si può rifiutare H_0 al 5%. Di fatto la regione di accettazione del test a livello β è

$$\bar{x}_n - t_{1-\beta/2}(n-1)s_n/\sqrt{n} < \mu_0 < \bar{x}_n + t_{1-\beta/2}(n-1)s_n/\sqrt{n}$$

che, essendo $\beta = 0.05 = 1 - 0.95 = \alpha$ risulta equivalente all'intervallo di confidenza calcolato al punto precedente. Il P-value soddisfa l'uguaglianza $t_{1-\bar{\beta}/2}(7) = |\sqrt{8}(\bar{x}_8 - \mu_0)/s_8| = 0.1\cdot\sqrt{8}/\sqrt{0.2885714286} = 0.5265241901$. Essendo

$$t_{0.5}(7) = 0 < t_{1-\bar{\beta}/2}(7) = 0.5265241901 < t_{0.8}(7) = 0.89603$$

si ha

$$0.5 < 1 - \bar{\beta}/2 < 0.8 \Longrightarrow 0.4 < \bar{\beta}.$$

Il valore approssimato restituito dal calcolatore è 0.6926050448.

4. Si tratta di svolgere un test del tipo $H_0 : \mu_X \leq \mu_Y$ contro $H_1 : \mu_X > \mu_Y$. Dai dati a disposizione si ha $\bar{y}_{10} = 20/10 = 2$ e $s_{10}^2(Y) = (n-1)^{-1}\sum_{i=0}^{n} y_i^2 - n(n-1)^{-1}\bar{y}_{10}^2 = 9^{-1}\cdot 42 - 10\cdot 9^{-1}\cdot 2^2 \approx 0.2222222$. Da cui si calcola la varianza combinata $s^2 = (7\cdot s_8^2(X) + 9\cdot s_9^2(Y))/16 = 0.25125$ e la statistica $t = (\bar{x}_8 - \bar{y}_{10})/\sqrt{s^2(1/7+1/9)} = 0.4205868653$. La regione di rifiuto è $t > t_{1-\alpha}(16) = 1.7458836763$; l'ipotesi nulla non può pertanto essere rifiutata. Dalle tabelle si vede che $t < t_{0.8}(16) = 0.864667$; inoltre $t > 0 = t_{0.5}(16)$. Quindi $\alpha \in (0.2, 0.5)$ (il valore approssimato restituito dal calcolatore è 0.3398256969).

 8994

Esercizio 2 Un treno ha n carrozze; i numeri dei passeggeri delle carrozze sono variabili binomiali $B(20, 1/4)$.

1. Qual è il numero medio di passeggeri per carrozza? E il numero medio di passeggeri del treno?
2. Qual è la probabilità che una carrozza fissata sia vuota? Qual è la probabilità che una carrozza fissata sia piena?
3. Qual è il numero medio di carrozze vuote? Ed il numero medio di carrozze piene?
4. Quante carrozze deve avere come minimo il treno affinchè la probabilità di avere almeno una carrozza vuota sia superiore a 0.5?
5. Si supponga che il numero di passeggeri della carrozza 1 in un certo istante sia distribuito come una binomiale $B(20, 1/4)$. Se dopo quell'istante vedo entrare 18 persone nella carrozza 1 e non li vedo uscire (si suppone che abbiano trovato posto), qual è la probabilità che la carrozza fosse vuota prima del loro ingresso?
6. Se un treno ha 50 carrozze, qual è la probabilità che vi siano almeno 240 passeggeri?

Soluzione. Siano $\{X_i\}_{i=1}^n$ le variabili che contano i passeggeri delle carrozze. Sono i.i.d. con legge $B(20, 1/4)$. Si potrebbe mostrare facilmente che il numero totale di passeggeri $T_n := \sum_{i=1}^n X_i \sim B(20n, 1/4)$, ma non utilizzeremo questo risultato. Sia $Y_i := 1$ se $X_i > 0$ ed $Y_i : 0$ se $X_i = 0$ (formalmente $Y_i := \mathbb{1}_{X_i>0}$). Sono variabili $B(\mathbb{P}(X_i > 0))$ i.i.d.

1. $\mathbb{E}[X_i] = 20 \cdot 1/4 = 5$. $\mathbb{E}[T_n] = \sum_{i=1}^n \mathbb{E}[X_i] = 5n$.

2. Sia $Y_i := 1$ se $X_i = 0$ ed $Y_i := 0$ se $X_i > 0$ (formalmente $Y_i := \mathbb{1}_{X_i=0}$). Sono variabili $B(\mathbb{P}(X_i = 0))$ i.i.d.; l'evento $\{Y_i = 1\} \equiv \{X_i = 0\}$ è il "successo" e ha probabilità $p_v := \mathbb{P}(X_i = 0) = (3/4)^{20} \approx 0.003171212$.

   Analogamente, sia $Z_i := 1$ se $X_i = 20$ ed $Z_i := 0$ se $X_i < 20$ (formalmente $Y_i := \mathbb{1}_{X_i=20}$). Sono variabili $B(\mathbb{P}(X_i = 20))$ i.i.d.; l'evento $\{Z_i = 1\} \equiv \{X_i = 20\}$ è il "successo" e ha probabilità $p_p := \mathbb{P}(X_i = 20) = 1/4^{20} \approx 0.909494702 \cdot 10^{-12}$.

3. Il numero di carrozze vuote è $\sum_{i=1}^n Y_i$ da cui si ha il numero medio di carrozze vuote $np_p = n \cdot 0.003171212$. Il numero di carrozze piene è $\sum_{i=1}^n Z_i$ da cui si ha il numero medio di carrozze piene $np_v = n \cdot 0.909494702 \cdot 10^{-12}$.

4. La probabilità di avere almeno una carrozza vuota è

$$\mathbb{P}\Big(\bigcup_{i=1}^n \{Y_i = 1\}\Big) = 1 - \mathbb{P}\Big(\bigcap_{i=1}^n \{Y_i = 0\}\Big) = 1 - (1 - (3/4)^{20})^n.$$

   Pertanto $1 - (1 - (1/4)^{20})^n \geq 0.5$ se e solo se $(1 - (3/4)^{20})^n \leq 0.5$ che equivale a $n \geq \log(0.5)/\log(1 - (3/4)^{20}) \approx 218.228095052$. Quindi $n \geq 219$.

5. Sia X_1 il numero di passeggeri. L'evento A "18 persone hanno trovato posto" equivale a $\{X_1 \leq 2\}$, mentre l'evento B "la carrozza è vuota" equivale a $\{X_1 = 0\}$. Pertanto la probabilità richiesta è

$$\begin{aligned}\mathbb{P}(B|A) = \mathbb{P}(X_1 = 0|X_1 \leq 2) &= \frac{\mathbb{P}(X_1 = 0, X_1 \leq 2)}{\mathbb{P}(X_1 \leq 2)} = \frac{\mathbb{P}(X_1 = 0)}{\mathbb{P}(X_1 \leq 2)} \\ &= \frac{(3/4)^{20}}{(3/4)^{20} + 20 \cdot (3/4)^{19}(1/4) + 190 \cdot (3/4)^{18}(1/4)^2} = \frac{0.0031712119}{0.0912604325} = 0.0347490347.\end{aligned}$$

6. Si potrebbe utilizzare il fatto che il numero di passeggeri del treno ha distribuzione $B(1000, 1/4)$ e quindi l'approssimazione normale visto che $1000/4 \geq 5$ e $1000 \cdot 3/4 \geq 5$.

   Alternativamente consideriamo il numero totale dei passeggeri del treno $T := T_{50} = \sum_{i=1}^{50} X_i$ e applichiamo l'approssimazione gaussiana essendo il numero di addendi superiore a 30. Osserviamo che $\mathbb{E}[T] = 50 \cdot 20 \cdot (1/4) = 250$ e $\mathrm{var}(T) = 50 \cdot 20 \cdot (1/4)(3/4) = 375/2 = 187.5$. Pertanto, se $Y \sim \mathcal{N}(250, 187.5)$, si ha

$$\begin{aligned}\mathbb{P}(T \geq 240) = \mathbb{P}(T \geq 239.5) \approx \mathbb{P}(Y \geq 239.5) &= \mathbb{P}\Big(\frac{Y - 250}{\sqrt{187.5}} \geq \frac{239.5 - 250}{\sqrt{187.5}}\Big) \\ &= 1 - \Phi\Big(-0.766811581\Big) = \Phi\Big(0.766811581\Big) \approx 0.7784032251.\end{aligned}$$