Corso di Statistica - Prof. Fabio Zucca	Appello 0 - 3 luglio 2015		
Nome e cognome:		**Matricola:**	
			12957

Esercizio 1 La I prova in itinere di Statistica si è svolta nelle aule A, B e C con 78, 74 e 97 studenti rispettivamente. La temperatura delle prime due aule era paragonabile mentre quella dell'aula C era sensibilmente più alta. I dati riguardanti i voti del primo esercizio sono sintetizzati nella tabella seguente

	A	B	C
Ampiezza campione	78	74	97
Somma dei voti	620.6375	588.8375	744.485
Somma quadrati dei voti	5223.5042	4981.6383	6167.5135

Si supponga di essere in regime di omoschedasticità.

1. Calcolare medie e varianze campionarie dei campioni.
2. Trovare l'intervallo di confidenza bilatero al 95% per la media dei voti dell'aula A.
3. Vi sono evidenze statistiche per affermare al 5% che il voto atteso di uno studente dell'aula A è differente da quello di uno studente dell'aula B? Descrivere il test utilizzato e stimarne il P-value.
4. Vi sono evidenze statistiche per affermare all'1% che la temperatura abbia penalizzato gli studenti dell'aula C rispetto a quelli della A? Descrivere il test utilizzato e stimarne il P-value.
5. **(Domanda bonus)** Il docente non vuole correre il rischio di sottovalutare gli effetti negativi della temperatura in aula C rispetto alla A. Proporre un test adeguato, svolgerlo al 10% e stimarne il P-value.

Soluzione.

1. La media campionaria è $\bar{x}_n = n^{-1}\sum_{i=1}^n x_i$. La varianza è $s_n^2 = (n-1)^{-1}\sum_{i=1}^n x_i^2 - n/(n-1)\bar{x}_n^2$. I dati richiesti sono riassunti nella seguente tabella

AULA	A	B	C
Media campionaria	7.9569	7.9573	7.6751
Varianza campionaria	3.70337	4.0562	4.7241

2. Si sa che gli estremi dell'intervallo sono

$$\begin{aligned} x^{\pm} &= \bar{x}_{78} \pm t_{(1+0.95)/2}(77)\cdot s_{78}/\sqrt{78} = 7.9569 \pm 1.99125\cdot\sqrt{3.70337/78} \\ &= 7.9569 \pm 0.43389 \\ &= \begin{cases} 7.523 \\ 8.39078 \end{cases} \end{aligned}$$

3. Chiamiamo X e Y le variabili relative alle aule A e B rispettivamente. Osserviamo che i dati sono sufficientemente numerosi per applicare i test per le variabili normali. Per avere evidenze statistiche che $\mu_X = \mu_Y$ si utilizza un test di confronto tra due medie a varianze incognite che si suppongono uguali. Si sceglie dunque $H_0 : \mu_X \neq \mu_Y$ contro $H_1 : \mu_X = \mu_Y$. Si osservi che i gradi di libertà $n+m-2 = n_X + n_Y - 2 = 150 > 120$ pertanto si possono approssimare i quantili della t-student con quelli di una variabile normale standard.

   La statistica per il test è

$$t = \frac{\bar{x}_n - \bar{y}_m}{s\sqrt{1/n + 1/m}} = -0.001166$$

   essendo la varianza combinata $s^2 = ((n-1)s_X^2 + (m-1)s_Y^2)/(n+m-2) = 3.87508$ ($n = 78$ ed $m = 74$). Il P-value si calcola, in maniera interpolata, come $\bar{\alpha} = 2\Phi(|t|) - 1 = 0.00093$ essendo $\Phi(|t|) = 0.500465$. La zona di rifiuto a livello $\alpha = 0.05$ è $|t| < t_{(1+\alpha)/2}(n+m-2) = t_{0.975}(150) \approx q_{0.975} = 1.95996$. Sia utilizzando la regione di rifiuto, che semplicemente osservando che $\bar{\alpha} < 0.05$ si conclude che vi sono evidenze statistiche per affermare che i due valori attesi sono uguali.

4. Chiamiamo ora X e Y le variabili relative alle aule A e C rispettivamente. Osserviamo ancora che i dati sono sufficientemente numerosi per applicare i test per le variabili normali. Per avere evidenze statistiche che $\mu_X > \mu_Y$ si utilizza un test di confronto tra due medie a varianze incognite che si suppongono uguali. Si sceglie dunque $H_0 : \mu_X \leq \mu_Y$ contro $H_1 : \mu_X > \mu_Y$. Si osservi che, come in precedenza, i gradi di libertà $n+m-2 = n_X + n_Y - 2 = 173 > 120$ pertanto si possono approssimare i quantili della t-student con quelli di una variabile normale standard.

   La statistica per il test è sempre

$$t = \frac{\bar{x}_n - \bar{y}_m}{s\sqrt{1/n + 1/m}} = 0.89667$$

   essendo la varianza combinata $s^2 = ((n-1)s_X^2 + (m-1)s_Y^2)/(n+m-2) = 4.26979$ ($n = 78$ ed $m = 97$). Il P-value si calcola, in maniera interpolata, come $\bar{\alpha} = 1 - \Phi(t) = 0.184947$ essendo $\Phi(t) = 0.81505284$. La zona di rifiuto a livello $\alpha = 0.01$ è $t > t_{1-\alpha}(n+m-2) = t_{0.99}(150) \approx q_{0.99} = 2.32635$. Sia utilizzando la regione di rifiuto, che semplicemente osservando che $\bar{\alpha} > 0.05$ si conclude che non vi sono evidenze statistiche per affermare che l'aula C sia stata penalizzata.

5. Se il docente non vuole correre il rischio di penalizzare l'aula C nel caso la temperatura abbia avuto effetti negativi significativi, vuol dire che vorrebbe evitare di credere che la temperatura non abbia avuto effetti negativi nel caso gli effetti ci siano stati. In altre parole si vuole vedere se ci sono evidenze statistiche per concludere che la temperatura non abbia penalizzato gli studenti dell'aula C. Il test é analogo al caso precedente, ma con ipotesi invertite: $H_0 : \mu_X \geq \mu_Y$ contro $H_1 : \mu_X < \mu_Y$. La statistica è identica al caso precedente. Il P-value si calcola, in maniera interpolata, come $\bar{\alpha} = \Phi(t) = 0.81505284$. La zona di rifiuto a livello $\alpha = 0.1$ è $t < t_\alpha(n+m-2) = t_{0.1}(150) \approx -q_{0.9} = -1.2815516$. Sia utilizzando la regione di rifiuto, che semplicemente osservando che $\bar{\alpha} > 0.1$ si conclude che non vi sono evidenze statistiche per affermare che l'aula C NON sia stata penalizzata.

Corso di Statistica - Prof. Fabio Zucca	Appello 0 - 3 luglio 2015		
Nome e cognome:		**Matricola:**	
			12957

Esercizio 2 Un treno ha n carrozze; i numeri dei passeggeri delle carrozze sono variabili binomiali $B(20, 3/4)$.

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?
3. Qual è la probabilità che una carrozza fissata sia piena?
4. Qual è il numero medio di carrozze vuote? Ed il numero medio di carrozze piene?
5. Quante carrozze deve avere come minimo il treno affinchè la probabilità di avere almeno una carrozza piena sia superiore a 0.5?
6. **Domanda Bonus**. Si supponga che il numero di passeggeri della carrozza 1 in un certo istante sia $B(20, 3/4)$. Se in quell'istante vedo entrare due persone prima di me nella carrozza 1 e non li vedo uscire (si suppone che abbiano trovato posto), qual è la probabilità che entrando subito dopo trovi posto anche io nella stessa carrozza?

Soluzione. Siano $\{X_i\}_{i=1}^n$ le variabili che contano i passeggeri delle carrozze. Sono i.i.d. con legge $B(20, 3/4)$. Si potrebbe mostrare facilmente che il numero totale di passeggeri $T_n := \sum_{i=1}^n X_i \sim B(20n, 3/4)$, ma non utilizzeremo mai 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. $\mathbb{E}[X_i] = 20 \cdot 3/4 = 15$. $\mathbb{E}[T_n] = \sum_{i=1}^n \mathbb{E}[X_i] = 15n$.

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 $p_v := \mathbb{P}(X_i = 0) = 1/4^{20} \approx 0.909494702 \cdot 10^{-12}$.

3. 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 $p_p := \mathbb{P}(X_i = 20) = (3/4)^{20} \approx 0.003171212$.

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

5. La probabilità di avere almeno una carrozza piena è

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

   Pertanto $1 - (1 - (3/4)^{20})^n \geq 0.5$ 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 $n \geq 219$.

6. Sia X_1 il numero di passeggeri L'evento A "le due persone hanno trovato posto" equivale a $\{X_1 \leq 18\}$, mentre l'evento B "le due persone hanno trovato posto e trovo posto anche io" equivale a $\{X_1 \leq 17\}$. Pertanto la probabilità richiesta è

$$\begin{aligned}\mathbb{P}(B|A) = \mathbb{P}(X_1 \leq 17 | X_1 \leq 18) &= \frac{\mathbb{P}(X_1 \leq 17, X_1 \leq 18)}{\mathbb{P}(X_1 \leq 18)} = \frac{\mathbb{P}(X_1 \leq 17)}{\mathbb{P}(X_1 \leq 18)} \\ &= \frac{1 - (3/4)^{20} - 20 \cdot (3/4)^{19}(1/4) - 190 \cdot (3/4)^{18}(1/4)^2}{1 - (3/4)^{20} - 20 \cdot (3/4)^{19}(1/4)} = \frac{0.9087395675}{0.9756873751} = 0.9313839563.\end{aligned}$$