Cognome : **Nome :** **Matricola:**

Modulo: Laboratorio di Metodi Matematici e Statistici, M-Z

Appello del 21 Luglio 2017

Esercizio 1.
Un'azienda di trasporto extraurbano afferma che il 50% delle volte i suoi autobus arrivano con un ritardo inferiore ai 5 minuti, che il 40% delle volte il ritardo varia da 5 a 10 minuti e che solo il 10% delle volte il ritardo è superiore a 10 minuti. Un'associazione di consumatori, per verificare questa affermazione, ha effettuato 70 viaggi, ottenendo i seguenti risultati:

	Ritardo, in minuti			
	≤ 5	$5-10$	> 10	Totale
Numero di viaggi	28	35	7	70

a) Calcolare la media e la varianza campionarie del ritardo e rappresentare in un opportuno grafico la distribuzione di frequenza dei dati.
b) Quali valori ci si aspetterebbe di trovare nella tabella se l'affermazione dell'azienda fosse vera?
c) Eseguire un test per verificare se l'affermazione dell'azienda è vera, specificando le ipotesi nulla e alternativa e usando un livello di significatività 1%. Quale conclusione si raggiunge?
d) Senza fare ulteriori calcoli, si dica, se possibile, qual è l'esito del test di cui al punto c) con il livello 5%. Giustificare la risposta.

Soluzione.
a) La variabile *ritardo*, essendo definita come un tempo, viene considerata come una variabile (quantitativa) continua. Bisogna calcolare media e varianza di dati raggruppati in classi: si noti che la prima classe è chiusa (da 0 a 5 minuti), mentre l'ultima è aperta (ritardo > 10 minuti) e quindi per calcolare media e varianza bisogna "chiuderla". Per esempio, scegliamo di chiuderla a 60 minuti. Determiniamo, quindi, il valore centrale di ogni classe (2.5, 7.5 e 35 minuti) come valore di riferimento per i dati nella classe. La media campionaria vale quindi:

$$\bar{x_n} = \frac{28 \times 2.5 + 35 \times 7.5 + 7 \times 35}{70} = 8.25 \text{ minuti}$$

e quindi la varianza campionaria vale:

$$s_n^2 = \frac{28 \times (2.5 - 8.25)^2 + 35 \times (7.5 - 8.5)^2 + 7 \times (35 - 8.5)^2}{69} = 86.295 \Rightarrow s = 9.289 \text{ minuti.}$$

Per disegnare l'istrogramma, siccome le tre classi qui considerate hanno ampiezza diversa, determiniamo le densità di frequenza $h_i = \frac{n_i/n}{a_i}$ che valgono, rispettivamente $h_1 = \frac{25/70}{5} = 0.08,\ h_2 = \frac{35/70}{5} = 0.1,\ h_3 = \frac{7/70}{50} = 0.002$. Il grafico è quello di Figura 1. Se la terza classe fosse stata chiusa non a 60 minuti, ma a 15 minuti, allora non sarebbe stato necessario calcolare le densità, ma sarebbero bastate le frequenze (assolute o relative indifferentemente).
b) Sono presenti $n = 70$ osservazioni classificate in $k = 3$ classi. I valori riportati nel testo sono le frequenze osservate N_i. La ditta assegna alle classi di ritardo le probabilità p_i indicate nella tabella seguente, da cui conseguono i valori (le frequenze) attesi $E_i = n_i^* = N_i p_i$ pure indicati in tabella:

	Ritardo, in minuti			
	≤ 5	$5-10$	> 10	Totale
Probabilità p_i	0.5	0.4	0.1	1
Fr. attese $E_i = n_i^*$	35	28	7	70

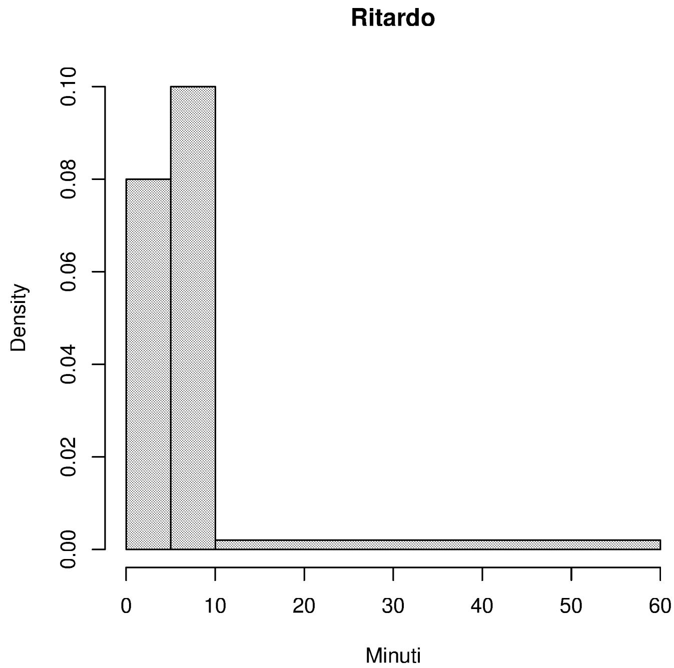

Figura 1: Esercizio 1 - Istogramma dei minuti di ritardo.

c) Un test di livello $\alpha = 0.05$ per l'ipotesi H_0 : "le probabilità delle varie classi sono p_i" contro l'alternativa H_1 che ciò non sia vero si esegue calcolando la statistica chi–quadrato

$$\chi^2 = \sum_{k=1}^{3} \frac{(N_i - E_i)^2}{E_i} = 3.15$$

e si rifiuta H_0 se risulta $\chi^2 > \chi^2_{\alpha,k-1}$. Nel nostro caso risulta $\chi^2_{0.01,2} = 9.210$ e non si può rifiutare l'ipotesi che la ditta dica la verità.

d) Siccome il valore della statistica test è inferiore al valore critico corrispondente al livello dell'1% e siccome il valore critico corrispondente al livello del 5% è inferiore al valore critico corrispondente al livello dell'1%, non è possibile dire senza fare calcoli cosa succede al livello del 5%, ma si deve confrontare ancora la statistica test con il valore critico al livello del 5%, che vale 5.9915. Si vede allora che nemmeno in questo caso si può rifiutare l'ipotesi nulla.

Esercizio 2.
I docenti di Statistica di un grande Ateneo vogliono informazioni sulla percentuale di studenti che superano l'esame di Statistica al primo tentativo. Per questo scelgono a caso 80 studenti che hanno già affrontato l'esame di Statistica e ottengono che in 7 hanno superato l'esame al primo tentativo.
a) Fornire una stima per la percentuale di studenti che hanno superato l'esame di Statistica al primo tentativo.
b) Calcolare un intervallo di confidenza di livello 95% per la percentuale di studenti che hanno superato l'esame di Statistica al primo tentativo. Quali ipotesi bisogna fare perchè il calcolo fatto sia corretto?
c) I docenti eseguono un test, con livello di significatività 5%, per dimostrare che la percentuale di studenti che hanno superato l'esame di Statistica al primo tentativo è inferiore al 10%. Qual è il risultato di tale test?
d) Senza fare calcoli, dire quale sarebbe il risultato del test se il livello fissato per il test fosse il 2%, giustificando la risposta.

Soluzione.
Sia p la percentuale in esame. Su $n = 80$ osservazioni se ne trovano $X = 7$ che corrispondono a studenti

che hanno superato l'esame alla prima prova.
a) La percentuale p viene stimata con $\widehat{p} = X/80 = 7/80 = 0.0875$.
b) I procedimenti approssimati che seguono sono giustificati poiché risulta $n > 30$, $n\widehat{p} = 7 > 5$, $n(1-\widehat{p}) = 73 > 5$. L'intervallo è

$$\widehat{p} \pm z_{\alpha/2}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}} = 0.0875 \pm 0.0619,$$

cioè l'intervallo $[2.56\%, 14.94\%]$, dato che $z_{\alpha/2} = z_{0.025} = 1.96$.
c) Un test di livello $\alpha = 0.05$ per l'ipotesi $H_0 : p = 0.10 = p_0$ contro l'alternativa $H_1 : p < 0.10$ si ottiene calcolando

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = -0.373$$

e rifiutando H_0 se risulta $Z < -z_\alpha = -z_{0.05} = -1.6449$. Poiché tale disuguaglianza non è verificata, non si rifiuta H_0 e si dichiara che la percentuale si studenti che passano l'esame di Statistica al primo tentativo è uguale o superiore al 10%.
d) Siccome abbassando il livello del test il valore critico diventa ancora più piccolo (nei negativi), non si può rifiutare l' ipotesi nulla nemmeno al livello del 2%.

Esercizio 3.
3. Un distributore automatico eroga bevande calde e fredde, a seconda delle scelte casuali e indipendenti degli utilizzatori. L'esperienza mostra che la probabilità che venga richiesta una bevanda calda è pari a 45%.
a) Calcolare la probabilità che ci siano esattamente 3 richieste di bevande calde su 6 bevande erogate.
b) Calcolare la probabilità che ci siano meno di 3 richieste di bevande calde su 10 bevande erogate.
c) Calcolare in modo approssimato la probabilità che, su 100 bevande erogate, il numero di bevande calde sia compreso tra 20 e 40.

Soluzione.
a)Sia X il numero di bevande calde su n richieste. Allora X ha distribuzione di Bernoulli $B(n, p)$ con parametro $p = 0.45$. a) Qui $n = 6$ e pertanto

$$P(X = 3) = \binom{6}{3} 0.45^3 0.55^3 = 0.3032.$$

b) Ora $n = 10$ e si ottiene

$$P(X < 3) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= 0.55^{10} + 10 \cdot 0.45 \cdot 0.55^9 + \binom{10}{2} 0.45^2 0.55^8 = 0.0996.$$

c) Ora $n = 100$ e poiché $np > 5$, $n(1-p) > 5$, la distribuzione di

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{X - 45}{\sqrt{24.75}}$$

si può approssimare con una normale standard. Perciò

$$P(20 \le X \le 40) = P(-5.03 \le Z \le -1.01) \simeq P(Z \le -1.01) = 0.156248.$$