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

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

Appello del 24 Febbraio 2017

Esercizio 1.
Una ONG impegnata in un programma di aiuti allo sviluppo in un paese del Sudamerica ha selezionato un campione casuale di 1200 persone di cui ha rilevato l'età di uscita dal mondo dell'istruzione, ottenendo i seguenti dati, per classi:

Età	n_i
$(9, 12]$	60
$(12, 14]$	360
$(14, 18]$	630
$(18, 30]$	150

a) Calcolare la media e la varianza campionarie dell'età di uscita dal mondo dell'istruzione.
b) Rappresentare con un grafico opportuno la distribuzione di frequenza dell'età di uscita dal mondo dell'istruzione.
c) Indicare un valore rappresentativo per la moda, se unica, e la classe in cui cade la mediana.
d) Se l'ultima classe fosse estesa fino a 50 anni, lasciando invariate le frequenze, i valori di media, moda e mediana risulterebbero modificati? Giustificare la risposta.

Soluzione.
a) Per calcolare la media e la varianza ci servono i valori rappresentativi delle classi, dati dai punti medi x_i^*: 10.5, 13, 16 e 24. Si ha, pertanto,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{4} n_i \times x_i^* = \frac{60 \times 10.5 + 360 \times 13 + 630 \times 16 + 150 \times 24}{1200} = 15.825$$

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{k} n_i \times (x_i^* - \bar{x})^2 =$$

$$= \frac{60(10.5 - 15.825)^2 + 360(13 - 15.825)^2 + 630(16 - 15.825)^2 + 150(24 - 15.825)^2}{1199} = 12.192$$

b) Considerando l'età come una variabile quantitativa continua, il grafico corretto è dato dall'istogramma. Siccome le classi non sono tutte uguali in ampiezza, dobbiamo calcolare la densità di ciascuna classe (che verrà usata come altezza dei rettangoli dell'istogramma): $h_i = \frac{n_i}{l_i}$ o, alternativamente, $h_i' = \frac{f_i}{l_i}$ ove $f_i = \frac{n_i}{n}$ sono le frequenze relative. Con quest'ultima scelta si hanno i valori di h_i' pari a: 0.017, 0.15, 0.131, 0.01 rispettivamente.
c) La classe modale, quella a densità più elevata, è unica: $(12, 14]$ e, pertanto, la moda vale 13, il valore rappresentativo della classe. La mediana appartiene alla classe $(14, 18]$, essendo la classe che contiene il dato 600° ed anche il dato 601° ($(n + 1)/2$).
d) Siccome estendendo l'ultima classe la sua densità diminuirebbe, non ci sarebbero variazioni nella moda. Allo stesso modo, non varierebbe la mediana (che non è sensibile a variazioni dei valori che non mutino il riordinamento degli stessi), mentre aumenterebbe la media, dato che aumenterebbe il valore rappresentativo dell'ultima classe.

Esercizio 2

Nei test effettuati sui bambini dagli 8 ai 10 anni è risultato che la loro soglia di attenzione ha una distribuzione normale con media di 40 minuti e deviazione standard di 8 minuti.

a) Calcolare la probabilità che un bambino scelto a caso nella popolazione di riferimento abbia una soglia di attenzione superiore a 45 minuti.

b) Calcolare la probabilità che un bambino scelto a caso nella popolazione di riferimento abbia una soglia di attenzione tra 30 e 50 minuti.

c) Se viene dato un premio solo ai bambini la cui soglia di attenzione rientra nel 25% più alto, qual è la soglia minima che un bambino deve raggiungere (in minuti) per ricevere il premio?

d) Scelto a caso un campione di 10 bambini dalla popolazione di riferimento, calcolare la probabilità che al massimo 2 di loro abbiano una soglia di attenzione superiore a 45 minuti.

e) Scelto a caso un campione di 500 bambini dalla popolazione di riferimento, calcolare la probabilità che la media delle soglie di attenzione del campione sia inferiore a 40 minuti.

Soluzione.

Sia X la variabile casuale che descrive la soglia di attenzione in un bambino qualunque della popolazione di riferimento, si ha $X \sim \mathcal{N}(\mu = 40, \sigma^2 = 64)$. Pertanto, procedendo con le opportune standardizzazioni:

a) $P(X > 45) = P\left(\frac{X-40}{8} > \frac{45-40}{8}\right) = P(Z > 0.625) = 1 - P(Z \leq 0.625)$ e, approssimando 0.625 a 0.63), $= 1 - \Phi(0.63) = 1 - 0.7357 = 0.2643$

b) $P(30 < X \leq 50) = P\left(\frac{30-40}{8} < \frac{X-40}{8} \leq \frac{50-40}{8}\right) = P(-1.25 < Z \leq 1.25) = P(Z \leq 1.25) - P(Z < -1.25) = \Phi(1.25) - \Phi(-1.25) = 0.8943 - 0.1056 = 0.7887$

c) Serve il valore x tale che $P(X > x) = 0.25$. Quindi: il valore x tale che $P(X > x) = P\left(\frac{X-40}{8} > \frac{x-40}{8}\right) = P\left(Z > \frac{x-40}{8}\right) = 0.25$. Con riferimento alle tavole della distribuzione Normale Standard si deve cercare un valore z che limiti l'area sotto la gaussiana a 0.75 (=1-0.25), e quindi si ha

$$\frac{x-40}{8} = 0.675 = \text{ da cui } \quad x = 45.4$$

infatti, essendo 0.75 un valore intermedio tra 0.7486 e 0.7517, possiamo indicare come valore di z il valore 0.675, intermedio tra 0.67 e 0.68.

d) Con riferimento al punto a), la variabile Y che indica il numero di bambini con soglia di attenzione superiore a 45 minuti in un campione casuale di 10 bambini (quindi, 10 bambini "indipendenti") ha distribuzione $Y \sim \mathit{Binomiale}(10, 0.2643)$, pertanto basta calcolare $P(Y \leq 2) = P(Y = 0) + P(Y = 1) + P(Y = 2) = \binom{10}{0}0.2643^0(1 - 0.2643)^{10} + \binom{10}{1}0.2643^1(1 - 0.2643)^9 + \binom{10}{2}0.2643^2(1 - 0.2643)^8 = 0.0465 + 0.1669 + 0.2698 = 0.4832$

e) La variabile casuale $\bar{X}_{500}$ media campionaria di 500 soglie di attenzione "indipendenti" è una variabile ancora Normale con media $\mu_{500} = 40$ e varianza $\sigma^2_{500} = \frac{64}{500}$ e, pertanto, la probabilità cercata vale 0.5, essendo 40 minuti proprio la media della variabile.

Esercizio 3
Da uno studio sulle soglie di attenzione degli scolari italiani tra gli 8 ed i 10 anni condotto su un campione di 60 scolari di scuole private si è ottenuta una media campionaria di $\bar{x}_P = 39.3$ minuti con una deviazione standard di $s_P = 12.3$ minuti.
a) Calcolare un intervallo di confidenza del 95% per il valore medio della soglia di attenzione nella popolazione di riferimento. E' necessario formulare ipotesi sul modello probabilistico?
b) Il Responsabile Unico delle scuole private italiane sostiene che i dati raccolti dimostrano che gli scolari delle scuole private italiane sono in linea con la media europea, pari a 40 minuti. Il Responsabile Unico ha ragione? Giustificare la risposta.
c) Per un confronto sono state rilevate le soglie di attenzione su un campione di 320 scolari di pari età provenienti da scuole comunali, ottenendo una media campionaria di $\bar{x}_C = 40.7$ minuti ed una deviazione standard campionaria di $s_C = 8.5$ minuti. Possiamo affermare che nelle scuole private agli scolari raggiungono una soglia media di attenzione più bassa che non in quelle comunali? Giustificare la risposta.

Soluzione.
a) Non è necessario formulare un modello probabilistico per la soglia di attenzione di uno scolaro tra gli 8 ed i 10 anni essendo il campione di numerosità $n > 30$: vale, pertanto, il Teorema Centrale del Limite. L'intervallo di confidenza cercato è dato da:

$$\bar{x}_n \pm z_{0.025} \times \frac{s_n}{\sqrt{n}} \Longrightarrow \left(39.3 - 1.96 \times \frac{12.3}{\sqrt{60}}, 39.3 + 1.96 \times \frac{12.3}{\sqrt{60}}\right) = (36.188, 42.412)$$

ove il valore critico $z_{0.025}$ approssima il valore $t(59)_{0.025}$.
b) Per stabilire se il Responsabile abbia ragione potremmo eseguire un test al livello α per il valore della media μ della soglia di attenzione nella popolazione di riferimento, con ipotesi:

$$H_0 \ : \ \mu = \mu_0, \quad H_1 \ : \ \mu \neq \mu_0 \quad , \text{con} \quad \mu_0 = 40.$$

Osserviamo, tuttavia, che per $\alpha = 0.05$ questo è equivalente a considerare l'intervallo di confidenza trovato al punto a): l'intervallo contiene il valore $\mu_0 = 40$ e, pertanto, non possiamo rifiutare l'ipotesi nulla al livello del 5% : in altre parole, i dati non permettono di contraddire il Responsabile unico. Volendo procedere con il test, calcoliamo la statistica test:

$$\left|\frac{\bar{x}_n - 40}{\frac{s_n}{\sqrt{n}}}\right| = \left|\frac{39.3 - 40}{1.58792}\right| = 0.441.$$

Siccome questo valore non supera alcun valore critico, non possiamo rifiutare l'ipotesi nulla, a nessun livello di significatività.

c) Confrontiamo i due campioni indipendenti, quello delle scuole private con quello delle scuole comunali, nell'ipotesi che la varianza della soglia di attenzione sia uguale, ancorchè incognita, nei due campioni. Si deve sottoporre a verifica l'ipotesi nulla $H_0 \ : \ \mu_P = \mu_C$ contro l'alternativa $H_1 \ : \ \mu_P < \mu_C$, ove μ_P e μ_C indicano il valore della soglia di attenzione media nelle due popolazioni di riferimento. Osserviamo che le medie campionarie sono compatibili con l'alternativa e calcoliamo la statistica test, a partire dalla varianza *pooled*, s_p^2:

$$s_p^2 = \frac{59 \times s_P^2 + 319 \times s_C^2}{60 + 320 - 2} = \frac{59 \times 12.3^2 + 319 \times 8.5^2}{378} = 84.587 \Longrightarrow s_p = \sqrt{84.587} = 9.197$$

da cui

$$\frac{\bar{x}_P - \bar{x}_C}{s_p \times \sqrt{\frac{1}{n_P} + \frac{1}{n_C}}} = \frac{39.3 - 40.7}{9.197 \times \sqrt{\frac{1}{60} + \frac{1}{320}}} = -1.08$$

Siccome il valore della statistica test non risulta $< t(318)_{0.05} \approx z_{0.05} = -1.6449$ non è possibile rifiutare l'ipotesi nulla a favore dell'alternativa che nella scuola privata la soglia di attenzione sia inferiore rispetto alla scuola comunale.