

# STATISTICA

Prova scritta di esonero del 10 maggio 2008

*Soluzione - versione A*

**Quesito 1** Nella seguente tabella è contenuta la distribuzione del reddito annuo in migliaia di euro delle famiglie residenti in una certa zona

classi di reddito	frequenze relative
0 - 15	0.24
15 - 25	0.27
25 - 40	0.29
40 - 100	0.20

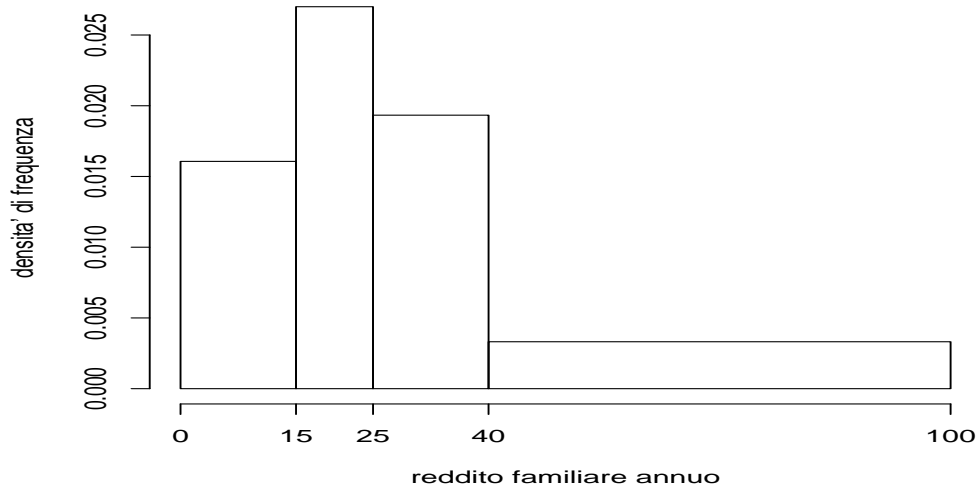
- i) Rappresentare graficamente la distribuzione.
- ii) Individuare la classe modale.
- iii) Calcolare il valore di un opportuno indice di posizione e di variabilità.
- iv) Sulla base delle informazioni disponibili e senza fare i conti, ritenete che il valore della mediana della distribuzione debba risultare inferiore, uguale, o superiore a quello della media aritmetica? Perché?
- v) Il valore della differenza interquartile risulta 22. Qual è il significato di tale indicatore?

*Soluzione:*

- i) La rappresentazione grafica idonea per la distribuzione in classi di un carattere quantitativo continuo è l'istogramma. Dal momento che le classi non hanno tutte la stessa ampiezza è necessario preliminarmente calcolare le densità di frequenza delle classi.

classi di reddito	frequenze relative	ampiezze classi	densità di frequenza
0 - 15	0.24	15	$\frac{0.24}{15} = 0.016$
15 - 25	0.27	10	$\frac{0.27}{10} = 0.027$
25 - 40	0.29	15	$\frac{0.29}{15} = 0.019$
40 - 100	0.20	60	$\frac{0.20}{60} = 0.003$

L'istogramma risulta



- ii) La classe modale è quella a cui corrisponde la densità di frequenza più elevata, quindi la classe 15 - 25.
- iii) Indicando con  $x$  il carattere considerato, con  $\bar{x}_i$  il valore centrale della classe  $x_{i-1} - x_i$  e con  $f_i$  la frequenza relativa corrispondente, sulla base delle informazioni disponibili, possiamo calcolare valori approssimati della media aritmetica e della varianza della distribuzione attraverso le formule

$$\bar{x} \simeq \sum_{i=1}^4 \bar{x}_i f_i \quad s^2 \simeq \sum_{i=1}^4 \bar{x}_i^2 f_i - \bar{x}^2.$$

Dopo aver effettuato i calcoli intermedi

classi	$f_i$	$\bar{x}_i$	$\bar{x}_i f_i$	$\bar{x}_i^2$	$\bar{x}_i^2 f_i$
0 - 15	0.24	7.5	1.8	56.25	13.50
15 - 25	0.27	20	5.4	400.00	108.00
25 - 40	0.29	32.5	9.4	1056.25	306.31
40 - 100	0.20	70	14	4900.00	980.00
totale	1		30.6		1407.81

otteniamo  $\bar{x} = 30.6$  e  $s^2 = 1407.81 - 30.6^2 = 471.45$ . Il valore dello scarto quadratico medio risulta  $s = \sqrt{471.45} = 21.71$ .

- iv) La mediana si trova nella classe 15 - 25 poiché più del 50% dei dati (in realtà il 51%) è minore o al più uguale all'estremo superiore di tale classe, pari a 25. Quindi la mediana è sicuramente minore della media.

Possiamo anche osservare che la distribuzione presenta una marcata asimmetria positiva. La presenza di valori del reddito annuo più elevati rispetto alla maggioranza degli altri (il 20% delle famiglie ha un reddito tra 40mila e 100 mila euro, mentre l'80% ha un reddito inferiore a 40mila) ci porta a ritenere che il valore della media aritmetica sia più elevato del valore della mediana. In effetti quest'ultimo risulta pari a 24.6 (mila euro).

- v) La differenza interquartile dà un'indicazione sulla variabilità della metà centrale dei dati. Nel nostro caso abbiamo che, se escludiamo il 25% dei valori più piccoli ed il 25% dei valori più elevati, il 50% dei valori restanti del reddito familiare annuo, quelli *centrali* variano in un intervallo di ampiezza 22.

## Quesito 2

- i) Siano  $A$  e  $B$  due eventi di  $S$ , come si calcola la probabilità di  $A \cup B$ ?
- ii) Siano  $E_1, E_2, \dots, E_n$  un insieme finito di  $n$  eventi elementari ed equiprobabili. Come si calcola la probabilità di  $C = E_1 \cup E_2 \cup \dots \cup E_k$ ?

*Soluzione:* Si vedano i paragrafi 6.4 e 6.5 del libro di testo.

**Quesito 3** Da un'indagine svolta dall'ATAC, la società di trasporti locale, risulta che il 18% dei residenti nel comune di Roma usa per gli spostamenti i mezzi pubblici. Consideriamo un campione di numerosità  $n = 10$ , ad elementi indipendenti, estratto in modo casuale tra i residenti a Roma ed indichiamo con  $X$  il numero di persone che nel campione usa il mezzo pubblico.

- i) Qual è la distribuzione di  $X$ ?
- ii) Qual è la probabilità che almeno una tra le persone nel campione usi abitualmente il mezzo pubblico?
- iii) Qual è il numero atteso di persone nel campione che usano per gli spostamenti i mezzi pubblici?
- iv) Quanto vale la varianza di  $X$ ?
- v) Se invece considerassimo un campione di numerosità  $n = 1000$ , quale sarebbe, con buona approssimazione, la probabilità che  $X$  assuma un valore al più pari al valore atteso? Perché?

*Soluzione:*

- i) La variabile  $X$  rappresenta il numero di “successi” (in cui per successo si intende il fatto che la persona considerata usa per gli spostamenti i mezzi pubblici) in  $n = 10$  prove indipendenti, ciascuna con probabilità di successo  $\pi$  pari a 0.18, quindi

$$X \sim \text{Binomiale}(10, 0.18).$$

- ii) La probabilità richiesta è

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0} 0.18^0 (1 - 0.18)^{10} = 0.86.$$

- iii) La media di  $X$  è

$$E(X) = n\pi = 10 \cdot 0.18 = 1.8$$

- iv) La varianza di  $X$  è

$$\text{var}(X) = n\pi(1 - \pi) = 10 \cdot 0.18(1 - 0.18) = 1.48.$$

- v) Per il teorema del limite centrale, sappiamo che, se  $X \sim \text{Binomiale}(n, \pi)$  e  $n$  è sufficientemente grande (diciamo tale che  $n\pi(1 - \pi) > 10$ ), la distribuzione della variabile casuale  $\frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}}$  è ben approssimata dalla distribuzione della variabile casuale  $N(0, 1)$ , che indichiamo con  $Z$ . Per la simmetria rispetto al valore atteso della distribuzione Normale, nel nostro caso avremo che

$$P(X < n\pi) \simeq P(Z < 0) = 0.5$$

**Quesito 4** Un chiosco nel parco vende panini al prosciutto. Il peso di ciascuna rosetta (vuota) ha distribuzione normale di media 100 grammi e scarto quadratico medio 3 grammi, mentre il prosciutto che la imbottisce ha distribuzione normale di media 90 grammi e scarto quadratico medio 4 grammi. Il peso delle rosette è indipendente da quello del prosciutto.

- i) Mario ha appena acquistato un panino al chiosco
- qual è la probabilità che il peso della rosetta sia compreso fra 98 e 102 grammi?
  - Sappiamo che è pari a 0.0062 la probabilità che il peso del prosciutto sia maggiore di  $y$  grammi. Qual è il valore di  $y$ ?
  - Qual è la probabilità che un panino imbottito pesi al più 200 grammi?
  - Qual è la probabilità che il peso del prosciutto sia maggiore di quello del pane?
- ii) Quali sono le analogie e le differenze fra la variabile casuale normale standard e la variabile casuale  $t$  di Student con  $n$  gradi di libertà?

*Soluzione:*

- i) Indichiamo con  $X$  la variabile casuale che si riferisce al peso di ciascuna rosetta vuota e con  $Y$  la variabile casuale che si riferisce al peso del prosciutto che la imbottisce. Sappiamo che  $X \sim N(100, 3^2)$  e  $Y \sim N(90, 4^2)$ .

- a) La probabilità richiesta vale

$$\begin{aligned} P(98 < X < 102) &= P\left(\frac{98 - 100}{3} < Z < \frac{102 - 100}{3}\right) = \\ &= P(-0.67 < Z < 0.67) = \Phi(0.67) - \Phi(-0.67) = \\ &= \Phi(0.67) - [1 - \Phi(0.67)] = 2\Phi(0.67) - 1 \end{aligned}$$

ma  $\Phi(0.67) = 0.7486$ , allora

$$P(98 < X < 102) = 2 \cdot 0.7486 - 1 = 0.4972$$

- b) Vogliamo ricavare il valore di  $y$  tale che  $P(Y > y) = 0.0062$ . Dal momento che

$$P(Y > y) = P\left(Z > \frac{y - 90}{4}\right) = 1 - \Phi\left(\frac{y - 90}{4}\right) = 0.0062$$

allora

$$\Phi\left(\frac{y - 90}{4}\right) = 1 - 0.0062 = 0.9938$$

da cui

$$\frac{y - 90}{4} = 2.5$$

e  $y = 2.5 \cdot 4 + 90 = 100$ .

- c) Il peso del panino imbottito è dato da  $X + Y$ , con  $X$  ed  $Y$  variabili casuali indipendenti. Allora  $E(X + Y) = E(X) + E(Y) = 100 + 90 = 190$  e  $var(X + Y) = var(X) + var(Y) = 9 + 16 = 25$ . Per la proprietà riproduttiva della Normale, si ha che  $X + Y \sim N(190, 25)$ . Risulta

$$P(X + Y < 200) = P\left(Z < \frac{200 - 190}{5}\right) = P(Z < 2) = \Phi(2) = 0.9772$$

- d) E' richiesto di calcolare  $P(Y > X) = P(Y - X > 0)$ . Per quanto detto in precedenza  $E(Y - X) = 90 - 100 = -10$ ,  $var(Y - X) = 16 + 9 = 25$  e  $Y - X \sim N(-10, 25)$ . Allora

$$P(Y > X) = P\left(Z > \frac{0 - (-10)}{5}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

ii) Si veda il paragrafo 10.5 del libro di testo.

**Quesito 5** A un esame di ammissione a un corso di inglese, gli studenti vengono classificati in “beginner”, “intermediate” e “advanced”. A questo corso partecipano studenti provenienti sia dal liceo scientifico sia dal liceo classico. La probabilità che uno studente provenga dal liceo classico è 0.35. La probabilità che uno studente sia classificato come “beginner” è 0.40, mentre la probabilità che sia classificato come “advanced” è 0.15. La probabilità che uno studente provenga dal liceo classico e sia classificato come beginner è 0.20, mentre la probabilità che uno studente provenga dal liceo scientifico e sia classificato come “advanced” è 0.10.

Sia  $Y$  una variabile casuale che assume valore 1 se lo studente classificato come “beginner”, 2 se lo studente classificato come “intermediate” e 3 se lo studente classificato come “advanced”. Sia  $X$  una variabile casuale che assume valore 1 se lo studente proviene dal liceo classico e 0 altrimenti.

- i) Costruire la distribuzione di probabilità congiunta di  $X$  e  $Y$ .
- ii) Determinare la distribuzione condizionata di  $Y$  quando  $X = 0$  e quando  $X = 1$ .
- iii) E' possibile affermare che le due variabili casuali siano indipendenti? Motivare la risposta.

*Soluzione:*

i) Sulla base delle informazioni disponibili siamo in grado di ricostruire le seguente tabella

X \ Y	1	2	3	
0			0.10	
1	0.20			0.35
	0.40		0.15	

Tenendo conto dei vincoli che devono essere soddisfatti dagli elementi di una distribuzione di probabilità doppia, si ricava

X \ Y	1	2	3	
0	0.20	0.35	0.10	0.65
1	0.20	0.10	0.05	0.35
	0.40	0.45	0.15	1

ii) La distribuzione di probabilità di  $Y$  dato  $X = 0$  risulta

y	1	2	3	
$P(Y = y X = 0)$	$\frac{0.2}{0.65}$	$\frac{0.35}{0.65}$	$\frac{0.1}{0.65}$	1

cioè

y	1	2	3	
$P(Y = y X = 0)$	0.31	0.54	0.15	1

mentre la distribuzione di probabilità di  $Y$  dato  $X = 1$  risulta

y	1	2	3	
$P(Y = y X = 1)$	$\frac{0.2}{0.35}$	$\frac{0.1}{0.35}$	$\frac{0.05}{0.35}$	1

cioè

y	1	2	3	
$P(Y = y X = 1)$	0.57	0.29	0.14	1

iii) Dal momento che la distribuzione di  $Y$  condizionata a  $X = 0$  risulta diversa dalla distribuzione di  $Y$  condizionata a  $X = 1$  concludiamo che le due variabili casuali  $X$  e  $Y$  non risultano indipendenti.

# STATISTICA

Prova scritta di esonero del 10 maggio 2008

*Soluzione - versione B*

**Quesito 1** Nella seguente tabella è contenuta la distribuzione del reddito annuo in migliaia di euro delle famiglie residenti in una certa zona

classi di reddito	frequenze relative
0 - 10	0.24
10 - 20	0.27
20 - 40	0.29
40 - 100	0.20

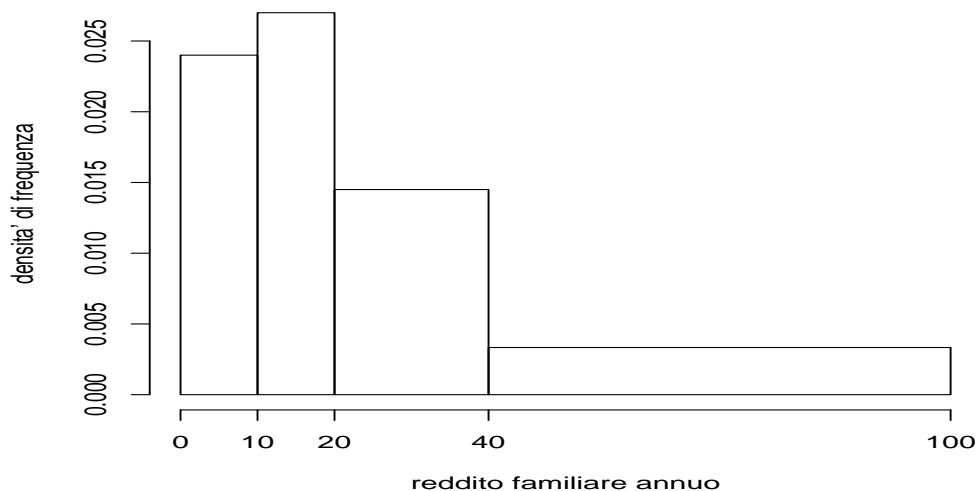
- i) Rappresentare graficamente la distribuzione.
- ii) Individuare la classe modale.
- iii) Calcolare il valore di un opportuno indice di posizione e di variabilità.
- iv) Sulla base delle informazioni disponibili e senza fare i conti, ritenete che il valore della mediana della distribuzione debba risultare inferiore, uguale, o superiore a quello della media aritmetica? Perché?
- v) Il valore della differenza interquartile risulta 26. Qual è il significato di tale indicatore?

*Soluzione:*

- i) La rappresentazione grafica idonea per la distribuzione in classi di un carattere quantitativo continuo è l'istogramma. Dal momento che le classi non hanno tutte la stessa ampiezza è necessario preliminarmente calcolare le densità di frequenza delle classi.

classi di reddito	frequenze relative	ampiezze classi	densità di frequenza
0 - 10	0.24	10	$\frac{0.24}{10} = 0.024$
10 - 20	0.27	10	$\frac{0.27}{10} = 0.027$
20 - 40	0.29	20	$\frac{0.29}{20} = 0.015$
40 - 100	0.20	60	$\frac{0.20}{60} = 0.003$

L'istogramma risulta



- ii) La classe modale è quella a cui corrisponde la densità di frequenza più elevata, quindi la classe 10 - 20.
- iii) Indicando con  $x$  il carattere considerato, con  $\bar{x}_i$  il valore centrale della classe  $x_{i-1} - x_i$  e con  $f_i$  la frequenza relativa corrispondente, sulla base delle informazioni disponibili, possiamo calcolare valori approssimati della media aritmetica e della varianza della distribuzione attraverso le formule

$$\bar{x} \simeq \sum_{i=1}^4 \bar{x}_i f_i \quad s^2 \simeq \sum_{i=1}^4 \bar{x}_i^2 f_i - \bar{x}^2.$$

Dopo aver effettuato i calcoli intermedi

classi	$f_i$	$\bar{x}_i$	$\bar{x}_i f_i$	$\bar{x}_i^2$	$\bar{x}_i^2 f_i$
0 - 10	0.24	5	1.2	25	6.00
10 - 20	0.27	15	4.1	225	60.75
20 - 40	0.29	30	8.7	900	261.00
40 - 100	0.20	70	14.0	4900	980.00
totale	1		28.0		1307.75

otteniamo  $\bar{x} = 28$  e  $s^2 = 1307.75 - 28^2 = 523.75$ . Il valore dello scarto quadratico medio risulta  $s = \sqrt{523.75} = 22.89$ .

- iv) La mediana si trova nella classe 10 - 20 poiché più del 50% dei dati (in realtà il 51%) è minore o al più uguale all'estremo superiore di tale classe, pari a 20. Quindi la mediana è sicuramente minore della media.

Possiamo anche osservare che la distribuzione presenta una marcata asimmetria positiva. La presenza di valori del reddito annuo più elevati rispetto alla maggioranza degli altri (il 20% delle famiglie ha un reddito tra 40mila e 100 mila euro, mentre l'80% ha un reddito inferiore a 40mila) ci porta a ritenere che il valore della media aritmetica sia più elevato del valore della mediana. In effetti quest'ultimo risulta pari a 19.6 (mila euro).

- v) La differenza interquartile dà un'indicazione sulla variabilità della metà centrale dei dati. Nel nostro caso abbiamo che, se escludiamo il 25% dei valori più piccoli ed il 25% dei valori più elevati, il 50% dei valori restanti del reddito familiare annuo, quelli *centrali* variano in un intervallo di ampiezza 26.

## Quesito 2

- i) Siano  $A$  e  $B$  due eventi di  $S$ , come si calcola la probabilità di  $A \cup B$ ?
- ii) Siano  $E_1, E_2, \dots, E_n$  un insieme finito di  $n$  eventi elementari ed equiprobabili. Come si calcola la probabilità di  $C = E_1 \cup E_2 \cup \dots \cup E_k$ ?

*Soluzione:* Si vedano i paragrafi 6.4 e 6.5 del libro di testo.

**Quesito 3** Da un'indagine svolta dall'ATAC, la società di trasporti locale, risulta che il 56% dei residenti nel comune di Roma usa per gli spostamenti un mezzo privato. Consideriamo un campione di numerosità  $n = 5$ , ad elementi indipendenti, estratto in modo casuale tra i residenti a Roma ed indichiamo con  $X$  il numero di persone che nel campione usa un mezzo privato.

- i) Qual è la distribuzione di  $X$ ?
- ii) Qual è la probabilità che almeno una tra le persone nel campione usi abitualmente un mezzo privato?
- iii) Qual è il numero atteso di persone nel campione che usano per gli spostamenti un mezzo privato?
- iv) Quanto vale la varianza di  $X$ ?
- v) Se invece considerassimo un campione di numerosità  $n = 1000$ , quale sarebbe, con buona approssimazione, la probabilità che  $X$  assuma un valore almeno pari al valore atteso? Perché?

*Soluzione:*

- i) La variabile  $X$  rappresenta il numero di “successi” (in cui per successo si intende il fatto che la persona considerata usa per gli spostamenti un mezzo privato) in  $n = 5$  prove indipendenti, ciascuna con probabilità di successo  $\pi$  pari a 0.56, quindi

$$X \sim \text{Binomiale}(5, 0.56).$$

- ii) La probabilità richiesta è

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0} 0.56^0 (1 - 0.56)^5 = 0.98.$$

- iii) La media di  $X$  è

$$E(X) = n\pi = 5 \cdot 0.56 = 2.8$$

- iv) La varianza di  $X$  è

$$\text{var}(X) = n\pi(1 - \pi) = 5 \cdot 0.56(1 - 0.56) = 1.23.$$

- v) Per il teorema del limite centrale, sappiamo che, se  $X \sim \text{Binomiale}(n, \pi)$  e  $n$  è sufficientemente grande (diciamo tale che  $n\pi(1 - \pi) > 10$ ), la distribuzione della variabile casuale  $\frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}}$  è ben approssimata dalla distribuzione della variabile casuale  $N(0, 1)$ , che indichiamo con  $Z$ . Per la simmetria rispetto al valore atteso della distribuzione Normale, nel nostro caso avremo che

$$P(X > n\pi) \simeq P(Z > 0) = 0.5$$

**Quesito 4** Un chiosco nel parco vende panini al prosciutto. Il peso di ciascuna rosetta (vuota) ha distribuzione normale di media 90 grammi e scarto quadratico medio 3 grammi, mentre il prosciutto che la imbottisce ha distribuzione normale di media 80 grammi e scarto quadratico medio 4 grammi. Il peso delle rosette è indipendente da quello del prosciutto.

- i) Mario ha appena acquistato un panino al chiosco
- qual è la probabilità che il peso della rosetta sia compreso fra 87 e 93 grammi?
  - Sappiamo che è pari a 0.1056 la probabilità che il peso del prosciutto sia maggiore di  $y$  grammi. Qual è il valore di  $y$ ?
  - Qual è la probabilità che un panino imbottito pesi almeno 180 grammi?
  - Qual è la probabilità che il peso del pane sia maggiore di quello del prosciutto?
- ii) Quali sono le analogie e le differenze fra la variabile casuale normale standard e la variabile casuale  $t$  di Student con  $n$  gradi di libertà?

*Soluzione:*

- i) Indichiamo con  $X$  la variabile casuale che si riferisce al peso di ciascuna rosetta vuota e con  $Y$  la variabile casuale che si riferisce al peso del prosciutto che la imbottisce. Sappiamo che  $X \sim N(90, 3^2)$  e  $Y \sim N(80, 4^2)$ .

- a) La probabilità richiesta vale

$$\begin{aligned} P(87 < X < 93) &= P\left(\frac{97-90}{3} < Z < \frac{93-90}{3}\right) == \\ &= P(-1 < Z < 1) = \Phi(1) - \Phi(-1) = \\ &= \Phi(1) - [1 - \Phi(1)] = 2\Phi(1) - 1 \end{aligned}$$

ma  $\Phi(1) = 0.8413$ , allora

$$P(87 < X < 93) = 2 \cdot 0.8413 - 1 = 0.6826$$

- b) Vogliamo ricavare il valore di  $y$  tale che  $P(Y > y) = 0.1056$ . Dal momento che

$$P(Y > y) = P\left(Z > \frac{y-80}{4}\right) = 1 - \Phi\left(\frac{y-80}{4}\right) = 0.1056$$

allora

$$\Phi\left(\frac{y-80}{4}\right) = 1 - 0.1056 = 0.8944$$

da cui

$$\frac{y-80}{4} = 1.25$$

e  $y = 1.25 \cdot 4 + 80 = 85$ .

- c) Il peso del panino imbottito è dato da  $X + Y$ , con  $X$  ed  $Y$  variabili casuali indipendenti. Allora  $E(X + Y) = E(X) + E(Y) = 90 + 80 = 170$  e  $var(X + Y) = var(X) + var(Y) = 9 + 16 = 25$ . Per la proprietà riproduttiva della Normale, si ha che  $X + Y \sim N(170, 25)$ . Risulta

$$P(X + Y > 180) = P\left(Z > \frac{180-170}{5}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

- d) E' richiesto di calcolare  $P(X > Y) = P(X - Y > 0)$ . Per quanto detto in precedenza  $E(X - Y) = 90 - 80 = 10$ ,  $var(X - Y) = 16 + 9 = 25$  e  $X - Y \sim N(10, 25)$ . Allora

$$P(X > Y) = P\left(Z > \frac{0-10}{5}\right) = P(Z > -2) = \Phi(2) = 0.9772$$

- ii) Si veda il paragrafo 10.5 del libro di testo.

**Quesito 5** A un esame di ammissione a un corso di inglese, gli studenti vengono classificati in “beginner”, “intermediate” e “advanced”. A questo corso partecipano studenti provenienti sia dal liceo scientifico sia dal liceo classico. La probabilità che uno studente provenga dal liceo scientifico è 0.60. La probabilità che uno studente sia classificato come “beginner” è 0.45, mentre la probabilità che sia classificato come “advanced” è 0.15. La probabilità che uno studente provenga dal liceo scientifico e sia classificato come “advanced” è 0.10, mentre la probabilità che uno studente provenga dal liceo classico e sia classificato come “beginner” è 0.25.

Sia  $Y$  una variabile casuale che assume valore 1 se lo studente è classificato come “beginner”, 2 se lo studente è classificato come “intermediate” e 3 se lo studente è classificato come “advanced”. Sia  $X$  una variabile casuale che assume valore 1 se lo studente proviene dal liceo classico e 0 altrimenti.

- i) Costruire la distribuzione di probabilità congiunta di  $X$  e  $Y$ .
- ii) Determinare la distribuzione condizionata di  $Y$  quando  $X = 0$  e quando  $X = 1$ .
- iii) E' possibile affermare che le due variabili casuali siano indipendenti? Motivare la risposta.

*Soluzione:*

- i) Sulla base delle informazioni disponibili siamo in grado di ricostruire la seguente tabella

$X \backslash Y$	1	2	3	
0			0.10	0.60
1	0.25			
	0.45		0.15	

Tenendo conto dei vincoli che devono essere soddisfatti dagli elementi di una distribuzione di probabilità doppia, si ricava

$X \backslash Y$	1	2	3	
0	0.20	0.30	0.10	0.60
1	0.25	0.10	0.05	0.40
	0.45	0.40	0.15	1

- ii) La distribuzione di probabilità di  $Y$  dato  $X = 0$  risulta

$y$	1	2	3	
$P(Y = y X = 0)$	$\frac{0.2}{0.6}$	$\frac{0.3}{0.6}$	$\frac{0.1}{0.6}$	1

cioè

$y$	1	2	3	
$P(Y = y X = 0)$	0.33	0.50	0.17	1

mentre la distribuzione di probabilità di  $Y$  dato  $X = 1$  risulta

$y$	1	2	3	
$P(Y = y X = 1)$	$\frac{0.25}{0.40}$	$\frac{0.1}{0.4}$	$\frac{0.05}{0.40}$	1

cioè

$y$	1	2	3	
$P(Y = y X = 1)$	0.625	0.250	0.125	1

- iii) Dal momento che la distribuzione di  $Y$  condizionata a  $X = 0$  risulta diversa dalla distribuzione di  $Y$  condizionata a  $X = 1$  concludiamo che le due variabili casuali  $X$  e  $Y$  non risultano indipendenti.

# STATISTICA

Prova scritta di esonero del 10 maggio 2008

*Soluzione - versione C*

**Quesito 1** Nella seguente tabella è contenuta la distribuzione del reddito annuo in migliaia di euro delle famiglie residenti in una certa zona

classi di reddito	frequenze relative
0 - 10	0.20
10 - 20	0.29
20 - 40	0.31
40 - 100	0.20

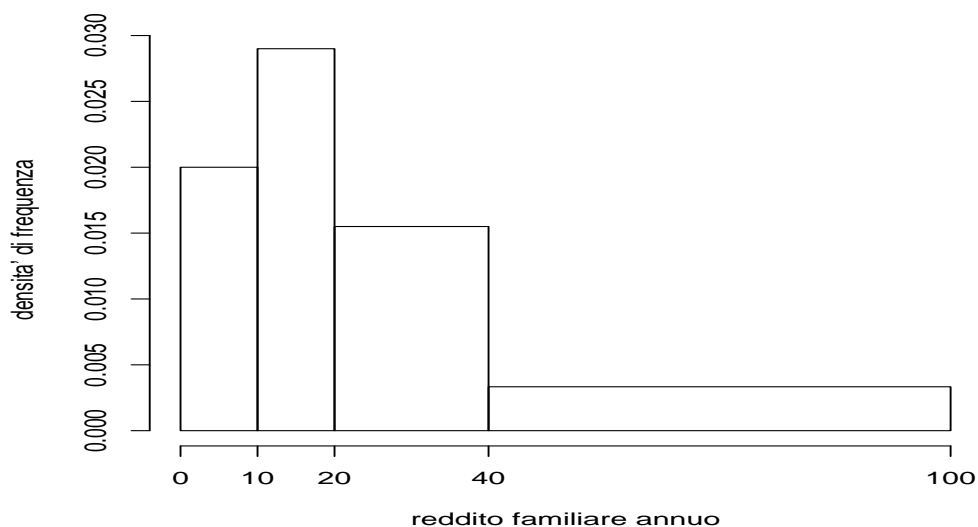
- i) Rappresentare graficamente la distribuzione.
- ii) Individuare la classe modale.
- iii) Calcolare il valore di un opportuno indice di posizione e di variabilità.
- iv) Sulla base delle informazioni disponibili e senza fare i conti, ritenete che il valore della mediana della distribuzione debba risultare inferiore, uguale, o superiore a quello della media aritmetica? Perché?
- v) Il valore della differenza interquartile risulta 25. Qual è il significato di tale indicatore?

*Soluzione:*

- i) La rappresentazione grafica idonea per la distribuzione in classi di un carattere quantitativo continuo è l'istogramma. Dal momento che le classi non hanno tutte la stessa ampiezza è necessario preliminarmente calcolare le densità di frequenza delle classi.

classi di reddito	frequenze relative	ampiezze classi	densità di frequenza
0 - 10	0.20	10	$\frac{0.20}{10} = 0.020$
10 - 20	0.29	10	$\frac{0.29}{10} = 0.029$
20 - 40	0.31	20	$\frac{0.31}{20} = 0.016$
40 - 100	0.20	60	$\frac{0.20}{60} = 0.003$

L'istogramma risulta



- ii) La classe modale è quella a cui corrisponde la densità di frequenza più elevata, quindi la classe 10 - 20.
- iii) Indicando con  $x$  il carattere considerato, con  $\bar{x}_i$  il valore centrale della classe  $x_{i-1} - x_i$  e con  $f_i$  la frequenza relativa corrispondente, sulla base delle informazioni disponibili, possiamo calcolare valori approssimati della media aritmetica e della varianza della distribuzione attraverso le formule

$$\bar{x} \simeq \sum_{i=1}^4 \bar{x}_i f_i \quad s^2 \simeq \sum_{i=1}^4 \bar{x}_i^2 f_i - \bar{x}^2.$$

Dopo aver effettuato i calcoli intermedi

classi	$f_i$	$\bar{x}_i$	$\bar{x}_i f_i$	$\bar{x}_i^2$	$\bar{x}_i^2 f_i$
0 - 10	0.20	5	1.0	25	5.00
10 - 20	0.29	15	4.4	225	65.25
20 - 40	0.31	30	9.3	900	279.00
40 - 100	0.20	70	14.0	4900	980.00
totale	1		28.7		1329.25

otteniamo  $\bar{x} = 28.7$  e  $s^2 = 1329.25 - 28.7^2 = 505.56$ . Il valore dello scarto quadratico medio risulta  $s = \sqrt{505.56} = 22.48$ .

- iv) Il 49% dei dati è minore o al più uguale all'estremo superiore della seconda classe, cioè a 20, ed il 31% dei dati appartenenti alla classe 20 - 40 presenta valori sparsi all'interno della classe stessa. Quindi il 64.5% (con  $64.5 = 49 + 31/2$ ) dei dati ha un valore minore di 30, il punto medio della terza classe. Queste considerazioni ci portano a concludere che la mediana si posiziona all'inizio della classe 20 - 40, prima della media, che si trova vicina al centro della classe stessa.

Possiamo anche osservare che la distribuzione presenta una marcata asimmetria positiva. La presenza di valori del reddito annuo più elevati rispetto alla maggioranza degli altri (il 20% delle famiglie ha un reddito tra 40mila e 100 mila euro, mentre l'80% ha un reddito inferiore a 40mila) ci porta a ritenere che il valore della media aritmetica sia più elevato del valore della mediana. In effetti quest'ultimo risulta pari a 20.34 (mila euro).

- v) La differenza interquartile dà un'indicazione sulla variabilità della metà centrale dei dati. Nel nostro caso abbiamo che, se escludiamo il 25% dei valori più piccoli ed il 25% dei valori più elevati,

il 50% dei valori restanti del reddito familiare annuo, quelli *centrali* variano in un intervallo di ampiezza 25.

### Quesito 2

- i) Siano  $A$  e  $B$  due eventi di  $S$ , come si calcola la probabilità di  $A \cup B$ ?
- ii) Siano  $E_1, E_2, \dots, E_n$  un insieme finito di  $n$  eventi elementari ed equiprobabili. Come si calcola la probabilità di  $C = E_1 \cup E_2 \cup \dots \cup E_k$ ?

*Soluzione:* Si vedano i paragrafi 6.4 e 6.5 del libro di testo.

**Quesito 3** Da un'indagine svolta dall'ATAC, la società di trasporti locale, risulta che il 16% dei residenti nella provincia di Roma usa per gli spostamenti i mezzi pubblici. Consideriamo un campione di numerosità  $n = 10$ , ad elementi indipendenti, estratto in modo casuale tra i residenti nella provincia di Roma ed indichiamo con  $X$  il numero di persone che nel campione usa il mezzo pubblico.

- i) Qual è la distribuzione di  $X$ ?
- ii) Qual è la probabilità che almeno una tra le persone nel campione usi abitualmente il mezzo pubblico?
- iii) Qual è il numero atteso di persone nel campione che usano per gli spostamenti i mezzi pubblici?
- iv) Quanto vale la varianza di  $X$ ?
- v) Se invece considerassimo un campione di numerosità  $n = 1000$ , quale sarebbe, con buona approssimazione, la probabilità che  $X$  assuma un valore almeno pari al valore atteso? Perché?

*Soluzione:*

- i) La variabile  $X$  rappresenta il numero di "successi" (in cui per successo si intende il fatto che la persona considerata usa per gli spostamenti i mezzi pubblici) in  $n = 10$  prove indipendenti, ciascuna con probabilità di successo  $\pi$  pari a 0.16, quindi

$$X \sim \text{Binomiale}(10, 0.16).$$

- ii) La probabilità richiesta è

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{10}{0} 0.16^0 (1 - 0.16)^{10} = 0.83.$$

- iii) La media di  $X$  è

$$E(X) = n\pi = 10 \cdot 0.16 = 1.6$$

- iv) La varianza di  $X$  è

$$\text{var}(X) = n\pi(1 - \pi) = 10 \cdot 0.16(1 - 0.16) = 1.34.$$

- v) Per il teorema del limite centrale, sappiamo che, se  $X \sim \text{Binomiale}(n, \pi)$  e  $n$  è sufficientemente grande (diciamo tale che  $n\pi(1 - \pi) > 10$ ), la distribuzione della variabile casuale  $\frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}}$  è ben approssimata dalla distribuzione della variabile casuale  $N(0, 1)$ , che indichiamo con  $Z$ . Per la simmetria rispetto al valore atteso della distribuzione Normale, nel nostro caso avremo che

$$P(X > n\pi) \simeq P(Z > 0) = 0.5$$

**Quesito 4** Un chiosco nel parco vende panini al prosciutto. Il peso di ciascuna rosetta (vuota) ha distribuzione normale di media 100 grammi e scarto quadratico medio 4 grammi, mentre il prosciutto che la imbottisce ha distribuzione normale di media 90 grammi e scarto quadratico medio 3 grammi. Il peso delle rosette è indipendente da quello del prosciutto.

- i) Mario ha appena acquistato un panino al chiosco
- qual è la probabilità che il peso della rosetta sia compreso fra 98 e 102 grammi?
  - Sappiamo che pari a 0.0013 la probabilità che il peso del prosciutto sia maggiore di  $y$  grammi. Qual è il valore di  $y$ ?
  - Qual è la probabilità che un panino imbottito pesi almeno 200 grammi?
  - Qual è la probabilità che il peso del pane sia maggiore di quello del prosciutto?
- ii) Quali sono le analogie e le differenze fra la variabile casuale normale standard e la variabile casuale  $t$  di Student con  $n$  gradi di libertà?

*Soluzione:*

- i) Indichiamo con  $X$  la variabile casuale che si riferisce al peso di ciascuna rosetta vuota e con  $Y$  la variabile casuale che si riferisce al peso del prosciutto che la imbottisce. Sappiamo che  $X \sim N(100, 4^2)$  e  $Y \sim N(90, 3^2)$ .

- a) La probabilità richiesta vale

$$\begin{aligned} P(98 < X < 102) &= P\left(\frac{98 - 100}{4} < Z < \frac{102 - 100}{4}\right) == \\ &= P(-0.5 < Z < 0.5) = \Phi(0.5) - \Phi(-0.5) = \\ &= \Phi(0.5) - [1 - \Phi(0.5)] = 2\Phi(0.5) - 1 \end{aligned}$$

ma  $\Phi(0.5) = 0.6915$ , allora

$$P(98 < X < 102) = 2 \cdot 0.6915 - 1 = 0.3830$$

- b) Vogliamo ricavare il valore di  $y$  tale che  $P(Y > y) = 0.0013$ . Dal momento che

$$P(Y > y) = P\left(Z > \frac{y - 90}{3}\right) = 1 - \Phi\left(\frac{y - 90}{3}\right) = 0.0013$$

allora

$$\Phi\left(\frac{y - 90}{3}\right) = 1 - 0.0013 = 0.9987$$

da cui

$$\frac{y - 90}{3} = 3$$

e  $y = 3 \cdot 3 + 90 = 99$ .

- c) Il peso del panino imbottito è dato da  $X + Y$ , con  $X$  ed  $Y$  variabili casuali indipendenti. Allora  $E(X + Y) = E(X) + E(Y) = 100 + 90 = 190$  e  $var(X + Y) = var(X) + var(Y) = 16 + 9 = 25$ . Per la proprietà riproduttiva della Normale, si ha che  $X + Y \sim N(190, 25)$ . Risulta

$$P(X + Y > 200) = P\left(Z > \frac{200 - 190}{5}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$$

d) E' richiesto di calcolare  $P(X > Y) = P(X - Y > 0)$ . Per quanto detto in precedenza  $E(X - Y) = 80 - 70 = 10$ ,  $var(X - Y) = 16 + 9 = 25$  e  $X - Y \sim N(10, 25)$ . Allora

$$P(X > Y) = P\left(Z > \frac{0 - 10}{5}\right) = P(Z > -2) = \Phi(2) = 0.9772$$

ii) Si veda il paragrafo 10.5 del libro di testo.

**Quesito 5** A un esame di ammissione a un corso di inglese, gli studenti vengono classificati in "beginner", "intermediate" e "advanced". A questo corso partecipano studenti provenienti sia dal liceo scientifico sia dal liceo classico. La probabilità che uno studente provenga dal liceo classico è 0.40. La probabilità che uno studente sia classificato come "advanced" è 0.15, mentre la probabilità che sia classificato come "intermediate" è 0.40. La probabilità che uno studente provenga dal liceo scientifico e sia classificato come "intermediate" è 0.30, mentre la probabilità che uno studente provenga dal liceo classico e sia classificato come "advanced" è 0.05.

Sia  $Y$  una variabile casuale che assume valore 1 se lo studente è classificato come "beginner", 2 se lo studente è classificato come "intermediate" e 3 se lo studente è classificato come "advanced". Sia  $X$  una variabile casuale che assume valore 0 se lo studente proviene dal liceo classico e 1 altrimenti.

- i) Costruire la distribuzione di probabilità congiunta di  $X$  e  $Y$ .
- ii) Determinare la distribuzione condizionata di  $Y$  quando  $X = 0$  e quando  $X = 1$ .
- iii) E' possibile affermare che le due variabili casuali siano indipendenti? Motivare la risposta.

*Soluzione:*

i) Sulla base delle informazioni disponibili siamo in grado di ricostruire la seguente tabella

X \ Y	Y			
	1	2	3	
0			0.05	0.40
1		0.30		
		0.40	0.15	

Tenendo conto dei vincoli che devono essere soddisfatti dagli elementi di una distribuzione di probabilità doppia, si ricava

X \ Y	Y			
	1	2	3	
0	0.25	0.10	0.05	0.40
1	0.20	0.30	0.10	0.60
	0.45	0.40	0.15	1

ii) La distribuzione di probabilità di  $Y$  dato  $X = 0$  risulta

y	1	2	3	
$P(Y = y   X = 0)$	$\frac{0.25}{0.40}$	$\frac{0.10}{0.40}$	$\frac{0.05}{0.40}$	1

cioè

y	1	2	3	
$P(Y = y   X = 0)$	0.625	0.250	0.125	1

mentre la distribuzione di probabilità di  $Y$  dato  $X = 1$  risulta

$y$	1	2	3	
$P(Y = y X = 1)$	$\frac{0.20}{0.60}$	$\frac{0.30}{0.60}$	$\frac{0.10}{0.60}$	1

cioè

$y$	1	2	3	
$P(Y = y X = 1)$	0.33	0.50	0.17	1

- iii) Dal momento che la distribuzione di  $Y$  condizionata a  $X = 0$  risulta diversa dalla distribuzione di  $Y$  condizionata a  $X = 1$  concludiamo che le due variabili casuali  $X$  e  $Y$  non risultano indipendenti.

# STATISTICA

Prova scritta di esonero del 10 maggio 2008

*Soluzione - versione D*

**Quesito 1** Nella seguente tabella è contenuta la distribuzione del reddito annuo in migliaia di euro delle famiglie residenti in una certa zona

classi di reddito	frequenze relative
0 - 15	0.20
15 - 25	0.29
25 - 40	0.31
40 - 100	0.20

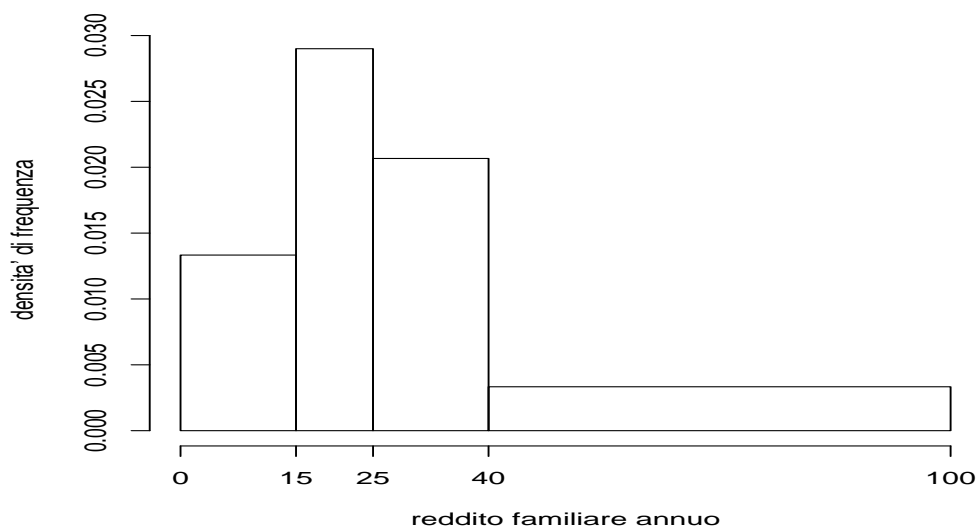
- i) Rappresentare graficamente la distribuzione.
- ii) Individuare la classe modale.
- iii) Calcolare il valore di un opportuno indice di posizione e di variabilità.
- iv) Sulla base delle informazioni disponibili e senza fare i conti, ritenete che il valore della mediana della distribuzione debba risultare inferiore, uguale, o superiore a quello della media aritmetica? Perché?
- v) Il valore della differenza interquartile risulta 21. Qual è il significato di tale indicatore?

*Soluzione:*

- i) La rappresentazione grafica idonea per la distribuzione in classi di un carattere quantitativo continuo è l'istogramma. Dal momento che le classi non hanno tutte la stessa ampiezza è necessario preliminarmente calcolare le densità di frequenza delle classi.

classi di reddito	frequenze relative	ampiezze classi	densità di frequenza
0 - 15	0.20	15	$\frac{0.20}{15} = 0.013$
15 - 25	0.29	10	$\frac{0.29}{10} = 0.029$
25 - 40	0.31	15	$\frac{0.31}{15} = 0.021$
40 - 100	0.20	60	$\frac{0.20}{60} = 0.003$

L'istogramma risulta



- ii) La classe modale è quella a cui corrisponde la densità di frequenza più elevata, quindi la classe 15 + 25.
- iii) Indicando con  $x$  il carattere considerato, con  $\bar{x}_i$  il valore centrale della classe  $x_{i-1} + x_i$  e con  $f_i$  la frequenza relativa corrispondente, sulla base delle informazioni disponibili, possiamo calcolare valori approssimati della media aritmetica e della varianza della distribuzione attraverso le formule

$$\bar{x} \simeq \sum_{i=1}^4 \bar{x}_i f_i \quad s^2 \simeq \sum_{i=1}^4 \bar{x}_i^2 f_i - \bar{x}^2.$$

Dopo aver effettuato i calcoli intermedi

classi	$f_i$	$\bar{x}_i$	$\bar{x}_i f_i$	$\bar{x}_i^2$	$\bar{x}_i^2 f_i$
0 + 15	0.20	7.5	1.5	56.25	11.25
15 + 25	0.29	20	5.8	400.00	116.00
25 + 40	0.31	32.5	10.1	1056.25	327.44
40 + 100	0.20	70	14.0	4900.00	980.00
totale	1		31.4		1434.69

otteniamo  $\bar{x} = 31.4$  e  $s^2 = 1434.69 - 31.4^2 = 448.73$ . Il valore dello scarto quadratico medio risulta  $s = \sqrt{448.73} = 21.18$ .

- iv) Il 49% dei dati è minore o al più uguale all'estremo superiore della seconda classe, cioè a 25, ed il 31% dei dati appartenenti alla classe 25 + 40 presenta valori sparsi all'interno della classe stessa. Quindi il 64.5% (con  $64.5 = 49 + 31/2$ ) dei dati ha un valore minore di 32.5, il punto medio della terza classe. Queste considerazioni ci portano a concludere che la mediana si posiziona all'inizio della classe 25 + 40, prima della media, che si trova vicina al centro della classe stessa.

Possiamo anche osservare che la distribuzione presenta una marcata asimmetria positiva. La presenza di valori del reddito annuo più elevati rispetto alla maggioranza degli altri (il 20% delle famiglie ha un reddito tra 40mila e 100 mila euro, mentre l'80% ha un reddito inferiore a 40mila) ci porta a ritenere che il valore della media aritmetica sia più elevato del valore della mediana. In effetti quest'ultimo risulta pari a 25.34 (mila euro).

- v) La differenza interquartile dà un'indicazione sulla variabilità della metà centrale dei dati. Nel nostro caso abbiamo che, se escludiamo il 25% dei valori più piccoli ed il 25% dei valori più elevati,

il 50% dei valori restanti del reddito familiare annuo, quelli *centrali* variano in un intervallo di ampiezza 21.

### Quesito 2

- i) Siano  $A$  e  $B$  due eventi di  $S$ , come si calcola la probabilità di  $A \cup B$ ?
- ii) Siano  $E_1, E_2, \dots, E_n$  un insieme finito di  $n$  eventi elementari ed equiprobabili. Come si calcola la probabilità di  $C = E_1 \cup E_2 \cup \dots \cup E_k$ ?

*Soluzione:* Si vedano i paragrafi 6.4 e 6.5 del libro di testo.

**Quesito 3** Da un'indagine svolta dall'ATAC, la società di trasporti locale, risulta che il 59% dei residenti nella provincia di Roma usa per gli spostamenti un mezzo privato. Consideriamo un campione di numerosità  $n = 5$ , ad elementi indipendenti, estratto in modo casuale tra i residenti nella provincia di Roma ed indichiamo con  $X$  il numero di persone che nel campione usa un mezzo privato.

- i) Qual è la distribuzione di  $X$ ?
- ii) Qual è la probabilità che almeno una tra le persone nel campione usi abitualmente un mezzo privato?
- iii) Qual è il numero atteso di persone nel campione che usano per gli spostamenti un mezzo privato?
- iv) Quanto vale la varianza di  $X$ ?
- v) Se invece considerassimo un campione di numerosità  $n = 1000$ , quale sarebbe, con buona approssimazione, la probabilità che  $X$  assuma un valore al più pari al valore atteso? Perché?

*Soluzione:*

- i) La variabile  $X$  rappresenta il numero di “successi” (in cui per successo si intende il fatto che la persona considerata usa per gli spostamenti un mezzo privato) in  $n = 5$  prove indipendenti, ciascuna con probabilità di successo  $\pi$  pari a 0.59, quindi

$$X \sim \text{Binomiale}(5, 0.59).$$

- ii) La probabilità richiesta è

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \binom{5}{0} 0.59^0 (1 - 0.59)^5 = 0.99.$$

- iii) La media di  $X$  è

$$E(X) = n\pi = 5 \cdot 0.59 = 2.95$$

- iv) La varianza di  $X$  è

$$\text{var}(X) = n\pi(1 - \pi) = 5 \cdot 0.59(1 - 0.59) = 1.21.$$

- v) Per il teorema del limite centrale, sappiamo che, se  $X \sim \text{Binomiale}(n, \pi)$  e  $n$  è sufficientemente grande (diciamo tale che  $n\pi(1 - \pi) > 10$ ), la distribuzione della variabile casuale  $\frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}}$  è ben approssimata dalla distribuzione della variabile casuale  $N(0, 1)$ , che indichiamo con  $Z$ . Per la simmetria rispetto al valore atteso della distribuzione Normale, nel nostro caso avremo che

$$P(X < n\pi) \simeq P(Z < 0) = 0.5$$

**Quesito 4** Un chiosco nel parco vende panini al prosciutto. Il peso di ciascuna rosetta (vuota) ha distribuzione normale di media 90 grammi e scarto quadratico medio 4 grammi, mentre il prosciutto che la imbottisce ha distribuzione normale di media 80 grammi e scarto quadratico medio 3 grammi. Il peso delle rosette è indipendente da quello del prosciutto.

- i) Mario ha appena acquistato un panino al chiosco
- qual è la probabilità che il peso della rosetta sia compreso fra 87 e 93 grammi?
  - Sappiamo che è pari a 0.0013 la probabilità che il peso del prosciutto sia maggiore di  $y$  grammi. Qual è il valore di  $y$ ?
  - Qual è la probabilità che un panino imbottito pesi al più 180 grammi?
  - Qual è la probabilità che il peso del prosciutto sia maggiore di quello del pane?
- ii) Quali sono le analogie e le differenze fra la variabile casuale normale standard e la variabile casuale  $t$  di Student con  $n$  gradi di libertà?

*Soluzione:*

- i) Indichiamo con  $X$  la variabile casuale che si riferisce al peso di ciascuna rosetta vuota e con  $Y$  la variabile casuale che si riferisce al peso del prosciutto che la imbottisce. Sappiamo che  $X \sim N(90, 4^2)$  e  $Y \sim N(80, 3^2)$ .

- a) La probabilità richiesta vale

$$\begin{aligned} P(87 < X < 93) &= P\left(\frac{87 - 90}{4} < Z < \frac{93 - 90}{4}\right) == \\ &= P(-0.75 < Z < 0.75) = \Phi(0.75) - \Phi(-0.75) = \\ &= \Phi(0.75) - [1 - \Phi(0.75)] = 2\Phi(0.75) - 1 \end{aligned}$$

ma  $\Phi(0.75) = 0.7734$ , allora

$$P(87 < X < 93) = 2 \cdot 0.7734 = 0.5468$$

- b) Vogliamo ricavare il valore di  $y$  tale che  $P(Y > y) = 0.0013$ . Dal momento che

$$P(Y > y) = P\left(Z > \frac{y - 80}{3}\right) = 1 - \Phi\left(\frac{y - 80}{3}\right) = 0.0013$$

allora

$$\Phi\left(\frac{y - 80}{3}\right) = 1 - 0.0013 = 0.9987$$

da cui

$$\frac{y - 80}{3} = 3$$

e  $y = 3 \cdot 3 + 80 = 89$ .

- c) Il peso del panino imbottito è dato da  $X + Y$ , con  $X$  ed  $Y$  variabili casuali indipendenti. Allora  $E(X + Y) = E(X) + E(Y) = 90 + 80 = 170$  e  $var(X + Y) = var(X) + var(Y) = 16 + 9 = 25$ . Per la proprietà riproduttiva della Normale, si ha che  $X + Y \sim N(170, 25)$ . Risulta

$$P(X + Y < 180) = P\left(Z < \frac{180 - 170}{5}\right) = P(Z < 2) = \Phi(2) = 0.97725$$

- d) E' richiesto di calcolare  $P(Y > X) = P(Y - X > 0)$ . Per quanto detto in precedenza  $E(Y - X) = 70 - 80 = -10$ ,  $var(Y - X) = 9 + 16 = 25$  e  $Y - X \sim N(-10, 25)$ . Allora

$$P(Y > X) = P\left(Z > \frac{0 - (-10)}{5}\right) = P(Z > 2) = 1 - \Phi(2) = 1 - 0.97725 = 0.02275$$

- ii) Si veda il paragrafo 10.5 del libro di testo.

**Quesito 5** A un esame di ammissione a un corso di inglese, gli studenti vengono classificati in "beginner", "intermediate" e "advanced". A questo corso partecipano studenti provenienti sia dal liceo scientifico sia dal liceo classico. La probabilità che uno studente provenga dal liceo scientifico è 0.65. La probabilità che uno studente sia classificato come "beginner" è 0.40, mentre la probabilità che sia classificato come "intermediate" è 0.45. La probabilità che uno studente provenga dal liceo classico e sia classificato come beginner è 0.20, mentre la probabilità che uno studente provenga dal liceo scientifico e sia classificato come "intermediate" è 0.35.

Sia  $Y$  una variabile casuale che assume valore 1 se lo studente classificato come "beginner", 2 se lo studente classificato come "intermediate" e 3 se lo studente classificato come "advanced". Sia  $X$  una variabile casuale che assume valore 0 se lo studente proviene dal liceo classico e 1 altrimenti.

- i) Costruire la distribuzione di probabilità congiunta di  $X$  e  $Y$ .  
 ii) Determinare la distribuzione condizionata di  $Y$  quando  $X = 0$  e quando  $X = 1$ .  
 iii) E' possibile affermare che le due variabili casuali siano indipendenti? Motivare la risposta.

*Soluzione:*

- i) Sulla base delle informazioni disponibili siamo in grado di ricostruire le seguente tabella

		Y			
		1	2	3	
X	0	0.20			
	1		0.35		0.65
		0.40	0.45		

Tenendo conto dei vincoli che devono essere soddisfatti dagli elementi di una distribuzione di probabilità doppia, si ricava

		Y			
		1	2	3	
X	0	0.20	0.10	0.05	0.35
	1	0.20	0.35	0.10	0.65
		0.40	0.45	0.15	1

- ii) La distribuzione di probabilità di  $Y$  dato  $X = 0$  risulta

y	1	2	3	
$P(Y = y X = 0)$	$\frac{0.2}{0.35}$	$\frac{0.1}{0.35}$	$\frac{0.05}{0.35}$	1

cioè

y	1	2	3	
$P(Y = y X = 1)$	0.57	0.29	0.14	1

mentre la distribuzione di probabilità di  $Y$  dato  $X = 1$  risulta

y	1	2	3	
$P(Y = y X = 0)$	$\frac{0.2}{0.65}$	$\frac{0.35}{0.65}$	$\frac{0.1}{0.65}$	1

cioè

y	1	2	3	
$P(Y = y X = 0)$	0.31	0.54	0.15	1

- iii) Dal momento che la distribuzione di  $Y$  condizionata a  $X = 0$  risulta diversa dalla distribuzione di  $Y$  condizionata a  $X = 1$  concludiamo che le due variabili casuali  $X$  e  $Y$  non risultano indipendenti.