PROVA SCRITTA DI STATISTICA
CLEA-CLEFIN-CLELI (COD. 5047 e 4038)
15 gennaio 2003

SOLUZIONI

Il nuovo direttore di una Banca di Credito Cooperativo si trova ad affrontare una vertenza di tipo sindacale che riguarda la presunta discriminazione reddituale dei lavoratori. La direzione della Banca, dotata di un'agenzia centrale e alcune periferiche, decide di discutere la questione con le organizzazioni sindacali sulla base di un campione casuale di dati amministrativi corrispondenti a 15 dipendenti. I dati sono riportati nella tabella seguente:

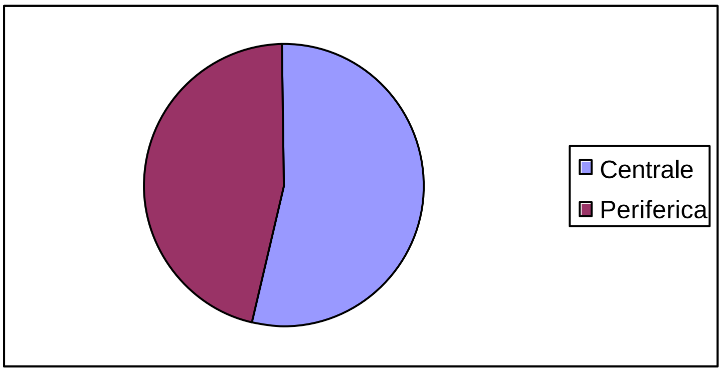
L Livello di studi del dipendente all'entrata in servizio
R Stipendio lordo annuo in migliaia di euro
S Sesso del dipendente
CA Collocazione agenzia
I Inquadramento del dipendente (livelli da I a V)

L	R	S	CA	I
laurea	20	M	Centrale	III
media superiore	20	M	Centrale	IV
media superiore	15	F	Periferica	I
laurea	25	F	Periferica	IV
media inferiore	20	M	Centrale	V
media inferiore	18	F	Periferica	IV
laurea	20	M	Periferica	IV
media inferiore	18	M	Centrale	III
media superiore media inferiore	18 15	M F	Periferica Periferica	II II
media superiore	20	F	Centrale	III
media inferiore	18	F	Centrale	IV
laurea	25	M	Centrale	V
laurea	25	M	Periferica	IV
media inferiore	15	F	Centrale	II

1. **(2 punti)**
Si dica, per ciascuna delle variabili considerate, quali strumenti è possibile utilizzare

	L	R	S	CA	I
Mediana	X	X			X
Grafico ad aste		X			
Coefficiente di variazione		X			
Moda	X	X	X	X	X

2. (1 punto)
Si fornisca una rappresentazione grafica della variabile CA.



3. (2 punti)
Si confronti la variabilità delle distribuzioni degli stipendi tra dipendenti maschi e femmine.

Per effettuare il confronto richiesto, determiniamo le distribuzioni degli stipendi dei dipendenti maschi e degli stipendi delle femmine e calcoliamo, per ciascuna di esse, il coefficiente di variazione.
La distribuzione degli stipendi tra i maschi è la seguente:

$$(R|S=M): \begin{cases} 18 & 2/8 \\ 20 & 4/8 \\ 25 & 2/8 \end{cases}$$

La distribuzione degli stipendi tra le femmine è invece

$$(R|S=F): \begin{cases} 15 & 3/7 \\ 18 & 2/7 \\ 20 & 1/7 \\ 25 & 1/7 \end{cases}$$

Si ha:
M(R|S=M)=20.75; M(R|S=F)=18; VAR(R|S=M)=6.6875; VAR(R|S=F)=11.4286.

Quindi $CV(R \mid S = M) = \dfrac{\sqrt{6.69}}{20.75} = 0.1246$ e $CV(R \mid S = F) = \dfrac{\sqrt{11.43}}{18} = 0.1878$, per cui è maggiore la variabilità degli stipendi tra le femmine.

4. (4 punti)
Si costruisca la tabella a doppia entrata di R e L. Si stabilisca se vi è indipendenza tra i due caratteri. Si costruisca quindi una tabella di massima connessione tenendo fissa la marginale del carattere R.

La tabella a doppia entrata di R ed L è la seguente:

R\L	M.I.	M.S.	L.
15	2/15	1/15	0
18	3/15	1/15	0
20	1/15	2/15	2/15
25	0	0	3/15

Non vi è indipendenza tra R ed L in quanto non tutte le frequenze relative congiunte sono uguali ai prodotti delle corrispondenti marginali (lo si verifica immediatamente notando che vi sono frequenze congiunte nulle).
Una tabella di massima connessione con la stessa distribuzione marginale di R di quella sopra è

R\L	M.I.	M.S.	L
15	3/15	0	0
18	0	4/15	0
20	0	0	5/15
25	3/15	0	0

5. (3 punti)

Sia X una variabile aleatoria con densità $f(x) = \begin{cases} \frac{3}{2}x^2 & -1 \le x \le 1 \\ 0 & altrove \end{cases}$ e sia *Y=3X-6*.

a. Si calcoli P(X<0).
b. Si calcoli il valore atteso di Y.
c. Si calcoli lo scarto quadratico medio di Y.

a. $P(X<0) = \int_{-\infty}^{0} f(x)dx = \int_{-1}^{0} \frac{3}{2}x^2 dx = \frac{1}{2}$.

b. $E(Y) = E(3X-6) = 3E(X) - 6 = 3\int_{-1}^{1} x \cdot \frac{3}{2}x^2 dx - 6 = 3 \cdot 0 - 6 = -6$.

c. $s(Y) = \sqrt{Var(Y)} = \sqrt{9Var(X)} = 3\sqrt{Var(X)} = 3\sqrt{E(X^2) - E(X)^2} = 3\sqrt{\int_{-1}^{1} \frac{3}{2}x^4 dx - 0} =$
$3\sqrt{\frac{3}{5}} = 2.3238$.

7. (3 punti)
Siano A e B due eventi stocasticamente indipendenti tali che $P(A \cup B) = 0.7$ e $P(\overline{A}) = 0.6$. Si determini $P(B)$.

Poiché A e B sono indipendenti, possiamo scrivere

$$P(A\cup B)=P(A)+P(B)-P(A\cap B)=P(A)+P(B)-P(A)P(B)=P(B)[1-P(A)]+P(A)$$

e, quindi, $P(B)=\dfrac{P(A\cup B)-P(A)}{1-P(A)}=\dfrac{0.7-0.4}{0.6}=0.5$.

8. (8 punti)

Un produttore di automobili dichiara che la velocità massima media m di un determinato modello di automobile è non inferiore a 180 Km/h. Volendo verificare questa affermazione, viene rilevata la velocità massima su un campione di 60 auto; si ottiene la realizzazione campionaria $(x_1, x_2, ..., x_{60})$ con $\sum x_i = 10920$ e $\sum x_i^2 = 2000000$.

a) Si fornisca una stima puntuale, ottenuta da uno stimatore non distorto, per la media incognita m della velocità massima.
b) Si fornisca una stima puntuale, ottenuta da uno stimatore non distorto, per la varianza incognita della velocità massima.
c) Si scriva la regione di rifiuto, di dimensione (approssimata) $a = 0.05$, per verificare l'ipotesi nulla $H_0 : m \le 180$ contro $H_1 : m > 180$.
d) Sulla base della realizzazione del campione ottenuta, si decida se rifiutare oppure no l'ipotesi nulla.
e) Si costruisca l'intervallo di confidenza per m di coefficiente di confidenza (approssimato) 0.99.

a) Uno stimatore non distorto di m è la media campionaria, per cui una stima puntuale di m è $\bar{x}=\dfrac{\sum x_i}{n}=\dfrac{10920}{60}=182$.
b) Uno stimatore non distorto della varianza s^2 è la varianza campionaria corretta, per cui una stima di s^2 è
$$s_c^2=\frac{n}{n-1}s^2=\frac{n}{n-1}\left(\frac{\sum x_i^2}{n}-\bar{x}^2\right)=\frac{60}{59}\left(\frac{2000000}{60}-182^2\right)=212.8813.$$
c) Poiché la varianza della popolazione è incognita e la numerosità del campione è sufficientemente elevata, una regione di rifiuto di dimensione approssimata 0.05 per verificare l'ipotesi considerata è
$$R=\left\{(x_1,...,x_{60}):\frac{\bar{x}-180}{s_c/\sqrt{60}}\ge 1.645\right\}.$$
d) Essendo $\dfrac{\bar{x}-180}{s_c/\sqrt{60}}=\dfrac{182-180}{14.5904/7.7459}=1.0618<1.645$, non si rifiuta l'ipotesi nulla.

e) L'intervallo richiesto è

$$\left(\overline{x} - z_{1-\frac{a}{2}} \cdot \frac{s_c}{\sqrt{n}}, \overline{x} + z_{1-\frac{a}{2}} \cdot \frac{s_c}{\sqrt{n}}\right) = \left(182 - 2.576 \cdot \frac{14.5904}{\sqrt{60}}, 182 + 2.576 \cdot \frac{14.5904}{\sqrt{60}}\right) =$$

$$(177.1479, 186.8521)$$

9. (4 punti)

a) Si fornisca la definizione di stimatore consistente (in senso forte o, equivalentemente, in media quadratica) per un parametro incognito q .

b) Sia T_n uno stimatore con $E(T_n) = q + \frac{1}{n}$ e $Var(T_n) = \frac{4q^2}{3n}$. Si dica se T_n è uno stimatore consistente per q .

a) Uno stimatore T_n si dice consistente in media quadratica per q se $\lim_{n \to +\infty} E\left[(T_n - q)^2\right] = 0$ per ogni q .

b) Poiché $\lim_{n \to +\infty} E\left[(T_n - q)^2\right] = 0$ vale se e solo se $\lim_{n \to +\infty} E(T_n) = q$ e $\lim_{n \to +\infty} Var(T_n) = 0$, lo stimatore considerato è consistente per q .