PROVA SCRITTA DI STATISTICA
CLEA-CLEFIN-CLELI (COD. 4038)
6 novembre 2002
SOLUZIONI

L'Associazione Industriali di una provincia lombarda ha commissionato un'indagine per studiare le esigenze del mercato del lavoro locale. Allo scopo sono stati somministrati 2000 questionari alle industrie della provincia. I dati per 15 di queste aziende sono raccolti nella seguente tabella:

L Livello di studi prevalente del personale impiegato a tempo indeterminato
A Ampiezza misurata attraverso il numero di dipendenti
S Settore prevalente di attività
CA Numero di contratti di lavoro diversi da quelli a tempo indeterminato
V Volume di affari per lavoratore a tempo indeterminato (in migliaia di euro)

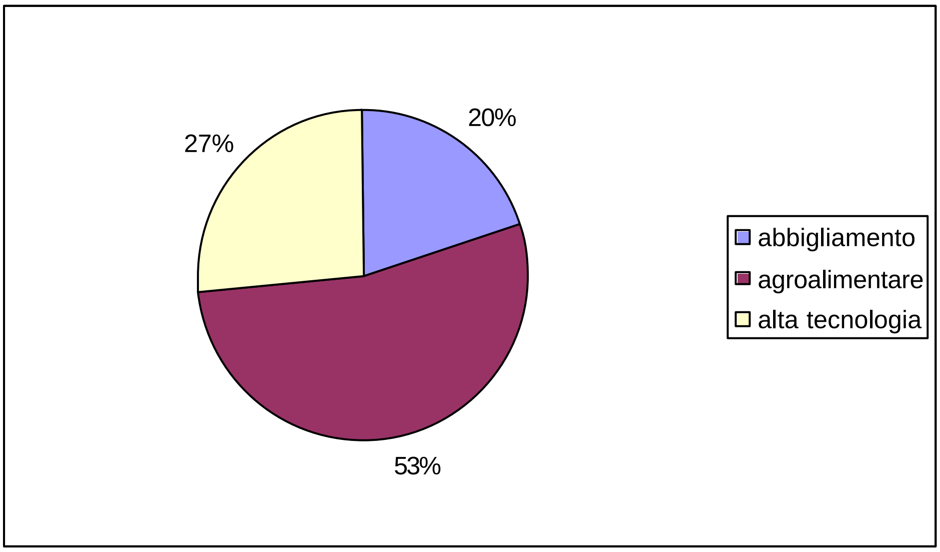
L	A	S	CA	V
laurea	25	alta tecnologia	6	228
media superiore	40	abbigliamento	4	62
media superiore	15	agroalimentare	10	80
laurea	20	alta tecnologia	3	600
media inferiore	45	agroalimentare	5	78
media inferiore	34	abbigliamento	0	44
laurea	150	agroalimentare	130	62
media inferiore	76	agroalimentare	9	224
media superiore	60	abbigliamento	4	36
media inferiore	35	agroalimentare	5	85
media superiore	21	alta tecnologia	5	548
media inferiore	62	agroalimentare	50	112
laurea	45	agroalimentare	40	83
laurea	10	alta tecnologia	0	600
media inferiore	30	agroalimentare	15	95

ESERCIZIO 1. (2 punti) Si dica, per ciascuna delle variabili considerate, quali tra gli strumenti indicati è possible utilizzare:

	L	A	S	CA	V
Media		X		X	X
Mediana	X	X		X	X
Q1 (I quartile)	X	X		X	X
Box-plot (grafico a scatola e baffi)		X		X	X

ESERCIZIO 2. (2 punti) Si fornisca una opportuna rappresentazione grafica per la variabile S.

Una rappresentazione grafica opportuna per rappresentare S è il diagramma a torta:



ESERCIZIO 3. (3 punti) Si costruisca la funzione di regressione di V su S e si dica se V è regressivamente indipendente da S.

La funzione di regressione di V su S è:

$$m_{V|S}(s) = \begin{cases} 494 & s = a.t. \\ 47.3333 & s = ab. \\ 102.375 & s = ag. \end{cases}$$

V non è regressivamente indipendente da S in quanto la funzione di regressione di V su S non è costante.

ESERCIZIO 4. (2 punti) Si costruisca la tabella a doppia entrata relativa alle variabili L ed S. Si dica se esse sono oppure no statisticamente indipendenti.

L S	Alta tecn.	Abbigl.	Agroalim.	
Med. Inf.	0	1/15	5/15	6/15
Med. Sup.	1/15	2/15	1/15	4/15
Laurea	3/15 4/15	0 3/15	2/15 8/15	5/15

L ed S non sono statisticamente indipendenti in quanto non tutte le frequenze relative congiunte coincidono con i prodotti delle corrispondenti frequenze relative marginali; ad esempio Fr(L=Med. Inf, S=Alta tecn.)=0 differisce dal prodotto di Fr(L=Med. Inf) e Fr(S=Alta tecn.).

ESERCIZIO 5. (4 punti) Siete spirati. Vi trovate di fronte a due porte identiche, una delle quali per l'Inferno (*Inf*) e l'altra per il Paradiso (*Par*). Siete destinati ad andare all'Inferno -- chi non lo è -- ma S. Pietro, per il vostro impegno profuso nel corso di Statistica, decide di proporvi comunque una via di salvezza per il Paradiso e vi lascia guardare dietro una delle due porte. S.Pietro vi informa che dietro la porta per l'Inferno vi sono un angelo e un diavolo. Dietro la porta per il Paradiso ci sono due angeli. Scegliendo a caso una delle due porte (*P(Inf)=P(Par)=0.5*), la aprite e riuscite a scorgere solo una sagoma.

a) Con quale probabilità scorgete un angelo?

b) Se avete scorto un angelo, con quale probabilità avete aperto la porta per l'Inferno?

Indichiamo con A l'evento "la porta che viene aperta è per l'inferno" e con B l'evento "la sagoma vista è di un angelo". In base ai dati forniti, si ha *P(A)*=0.5, $P(B \mid A) = 0.5$ e $P(B \mid \overline{A}) = 1$.

a) La probabilità di B si può calcolare utilizzando il teorema delle probabilità totali; precisamente $P(B) = P(B \mid A)P(A) + P(B \mid \overline{A})P(\overline{A}) = 0.5 * 0.5 + 1 * 0.5 = 0.75$.

b) La probabilità richiesta si può calcolare utilizzando la formula di Bayes; precisamente $P(A \mid B) = \dfrac{P(B \mid A)P(A)}{P(B)} = \dfrac{0.5 * 0.5}{0.75} = 0.3333$.

ESERCIZIO 6. (4 punti) Si consideri una variabile aleatoria X la cui funzione di ripartizione è:

$$F(x) = \begin{cases} 0 & x < -2 \\ 1/4 & -2 \le x < 0 \\ 2/4 & 0 \le x < 2 \\ 1 & x \ge 2 \end{cases}$$

a) Si determini la funzione di probabilità di X.

b) Si calcoli il valore atteso della variabile aleatoria *Y=3X-2*.

c) Si calcoli $P(X \ge 1)$.

a) La funzione di probabilità di X è $p_X(x) = \begin{cases} 1/4 & x = -2 \; ; \; x = 0 \\ 2/4 & x = 2 \\ 0 & altrove \end{cases}$

b) Poiché E(X)=-2*1/4+0*1/4+2*2/4=1/2, si ha E(Y)=3*E(X)-2=-1/2.

c) $P(X \ge 1) = 1/2$.

ESERCIZIO 7. (7 punti) E' noto che il prezzo medio di vendita di un determinato articolo nei negozi della periferia di una città è pari a 8. Indicato con m il prezzo medio (incognito) dello stesso articolo nei negozi del centro della città, si vuole verificare se m è differente da 8. Allo scopo si rileva il prezzo dell'articolo in un campione di 7 negozi del centro. Indicati con $x_1, x_2, ..., x_7$ i prezzi di vendita rilevati nei negozi del centro, si ha:

$$\sum_{i=1}^{7} x_i = 63, \; \sum_{i=1}^{7} x_i^2 = 581,$$

a) Si fornisca una stima del prezzo medio di vendita dell'articolo nei negozi del centro.
b) Si fornisca una stima della varianza del prezzo nei negozi del centro.
c) Supponendo normale la distribuzione del prezzo di vendita dell'articolo nei negozi del centro si stabilisca se, al livello $a = 0.05$, è da rifiutare oppure no l'ipotesi $m = 8$ contro l'alternativa $m \neq 8$.

a) Una stima (ottenuta da uno stimatore non distorto) del prezzo medio di vendita dell'articolo nei negozi del centro è $\bar{x} = \frac{63}{7} = 9.$

b) Una stima (ottenuta da uno stimatore non distorto) della varianza è

$$s_c^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{n}{n-1}\left(\frac{\sum x_i^2}{n} - \bar{x}^2\right) = \frac{7}{6}\left(\frac{581}{7} - 81\right) = 2.33.$$

c) Per verificare $H_0 : m = 8$ contro $H_1 : m \neq 8$, avendo la popolazione distribuzione normale ed essendo la varianza incognita, usiamo la seguente regione di rifiuto (di livello 0.05):

$$R = \left\{ (x_1,\ldots,x_7) : \frac{|\bar{x} - 8|}{\frac{s_c}{\sqrt{n}}} > t_{n-1,1-a/2} \right\} = \left\{ (x_1,\ldots,x_7) : \frac{|\bar{x} - 8|}{\frac{s_c}{\sqrt{7}}} > 2.447 \right\}.$$

Poiché $\frac{|\bar{x} - 8|}{\frac{s_c}{\sqrt{7}}} = 1.7322$ è minore di 2.447, non si rifiuta l'ipotesi nulla.

ESERCIZIO 8. (4 punti) Si consideri una popolazione statistica X con distribuzione gaussiana, media incognita m e varianza nota s^2.

a) Si scriva l'espressione dell'intervallo di confidenza per m, di coefficiente di confidenza $1-a$, basato su un campione di numerosità n la cui realizzazione ha fornito il valore $\bar{x}$ per la media campionaria.

b) Si dica se, con lo stesso valore $\bar{x}$ e con $a' > a$, l'ampiezza dell'intervallo risulta maggiore o minore della precedente; si giustifichi la risposta.

a) L'intervallo è: $\left(\bar{x} - z_{1-a/2}\frac{s}{\sqrt{n}}, \bar{x} + z_{1-a/2}\frac{s}{\sqrt{n}}\right)$, essendo $z_{1-a/2}$ il quantile di ordine $1-a/2$ di una variabile aleatoria con distribuzione normale standard.

b) Essendo l'ampiezza dell'intervallo con coefficiente di confidenza $1-a$ uguale a $A(a) = 2z_{1-a/2}\frac{s}{\sqrt{n}}$ e poiché, per $a' > a$, si ha $z_{1-a'/2} < z_{1-a/2}$, risulta $A(a') < A(a)$; ovvero, l'ampiezza dell'intervallo con coefficiente di confidenza $1-a'$ è inferiore a quella dell'intervallo con coefficiente di confidenza $1-a$ quando $a' > a$.

ESERCIZIO 9. **(2 punti)** Si dica quali operazioni è necessario effettuare per ottenere, con EXCEL, gli indicatori mediana e varianza di una *n*-upla di dati numerici.

Per determinare mediana e varianza di n numeri, si può selezionare “funzione”dal menù “inserisci” e, quindi, selezionare “statistiche” come categoria e “var.pop.” (e “mediana”) come nome funzione. Si seleziona quindi la serie di numeri e si clicca OK.