SOLUZIONI

della Prova Scritta di Statistica del 9 febbraio 2005

Modalità A1 e A2

Esercizio 1

punto a

Ricordiamo che le frequenze congiunte relative di ciascuna coppia di modalità (x_i^*, y_j^*) si ottengono dalla relazione:

$$p_{XY}(x_i^*, y_j^*) = \frac{n_{XY}(x_i^*, y_j^*)}{N}.$$

Modalità A1

Pertanto, dividendo tutte le frequenze della tabella per $N = 200$, si ottiene:

$X \backslash Y$	0	1	2	
R. libri	0.5	0.1	0	0.6
R. per.	0	0.2	0.1	0.3
D. gen.	0	0.05	0.05	0.1
	0.5	0.35	0.15	1

Modalità A2

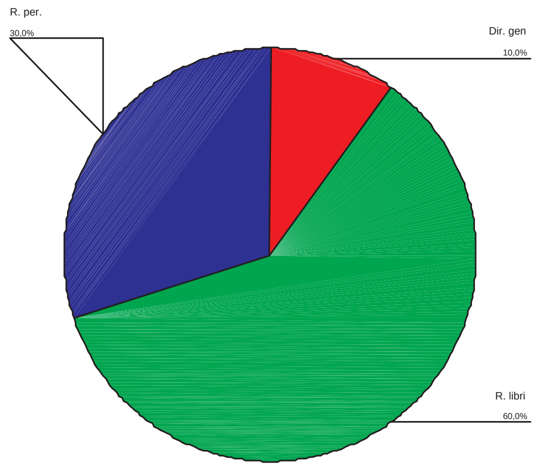
Pertanto, dividendo tutte le frequenze della tabella per $N = 250$, si ottiene:

$X \backslash Y$	Milano	Bergamo	Brescia	
0	0.2	0	0.1	0.3
1	0.16	0	0.3	0.46
2	0.04	0.2	0	0.24
	0.4	0.2	0.4	1

punto b

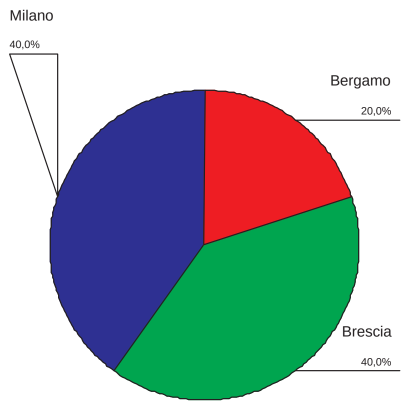
Modalità A1

Il carattere X è di tipo qualitativo nominale; la rappresentazione grafica piu' appropriata è dunque il diagramma a torta:



Modalità A2

Il carattere Y è di tipo qualitativo nominale; la rappresentazione grafica piu' appropriata è dunque il diagramma a torta:



punto c

Modalità A1

Utilizzando la distribuzione di frequenze marginale di Y evidenziata nella tabella a doppia entrata di cui sopra, si ottiene:

$$\begin{aligned}
\mu(Y) &= (0)(0.5) + (1)(0.35) + (2)(0.15) = 0.65 \\
\sigma^2(Y) &= (0)^2(0.5) + (1)^2(0.35) + (2)^2(0.15) - (0.65)^2 = 0.5272 \\
\sigma(Y) &= 0.7263
\end{aligned}$$

Modalità A2

Utilizzando la distribuzione di frequenze marginale di X evidenziata nella tabella a doppia entrata di cui sopra, si ottiene:

$$\begin{aligned}\mu(X) &= (0)(0.3) + (1)(0.46) + (2)(0.24) = 0.94 \\ \sigma^2(X) &= (0)^2(0.3) + (1)^2(0.46) + (2)^2(0.24) - (0.94)^2 = 0.5364 \\ \sigma(X) &= \sqrt{0.5364} = 0.7324\end{aligned}$$

punto d

Modalità A1

Perché si abbia indipendenza regressiva di Y da X, la funzione di regressione di Y su X deve essere costante e coincidere con $\mu(Y)$. Poiché

$$\mu(Y|X = \text{R. libri}) = (0)\frac{0.5}{0.6} + (1)\frac{0.1}{0.6} = 0.1667$$

è diverso da $\mu(Y) =$ 0.65 , Y non è regressivamente indipendente da X.

Modalità A2

Perché si abbia indipendenza regressiva di X da Y, la funzione di regressione di X su Y deve essere costante e coincidere con $\mu(X)$. Poiché

$$\mu(X|Y = \text{Milano}) = (0)\frac{0.2}{0.4} + (1)\frac{0.16}{0.4} + (2)\frac{0.04}{0.4} = 0.6$$

è diverso da $\mu(X) = 0.94$, X non è regressivamente indipendente da Y.

punto e

Ricordiamo che un indice relativo di connessione è dato da

$$\tilde{\varphi}^2(X,Y) = \frac{\varphi^2(X,Y)}{\min\{(h-1),(k-1)\}} \qquad \text{dove } \varphi^2(X,Y) = \sum_{i=1}^{h}\sum_{j=1}^{k} \frac{\left[p_{XY}(x_i^*, y_j^*) - p_X(x_i^*)\, p_Y(y_j^*)\right]^2}{p_X(x_i^*)\, p_Y(y_j^*)}$$

Modalità A1

Nella prossima tabella riportiamo i valori $p_X(x_i^*)\, p_Y(y_j^*)$ (frequenze congiunte in ipotesi di indipendenza statistica):

$X\backslash Y$	0	1	2	
R. libri	0.3	0.21	0.09	0.6
R. per.	0.15	0.105	0.045	0.3
D. gen.	0.05	0.035	0.015	0.1
	0.5	0.35	0.15	1

Riportiamo ora i valori $\frac{\left[p_{XY}(x_i^*, y_j^*) - p_X(x_i^*)\, p_Y(y_j^*)\right]^2}{p_X(x_i^*)\, p_Y(y_j^*)}$

$X\backslash Y$	0	1	2
R. libri	0.1333	0.0576	0.09
R. per.	0.15	0.0859	0.0672
D. gen.	0.05	0.0064	0.0817

Sommando ora i valori nella tabella, si ottiene $\tilde{\varphi}^2(X,Y) = 0.7221$; osservando inoltre che, nel nostro caso, $h = k = 3$, si ottiene

$$\tilde{\varphi}^2(X,Y) = \frac{0.7221}{2} = 0.36105.$$

Modalità A2

Nella prossima tabella riportiamo i valori $p_X(x_i^*)\, p_Y(y_j^*)$ (frequenze congiunte in ipotesi di indipendenza statistica):

$X\backslash Y$	Milano	Bergamo	Brescia	
0	0.12	0.06	0.12	0.3
1	0.184	0.092	0.184	0.46
2	0.096	0.048	0.096	0.24
	0.4	0.2	0.4	1

Riportiamo ora i valori $\frac{\left[p_{XY}(x_i^*,y_j^*) - p_X(x_i^*)\, p_Y(y_j^*)\right]^2}{p_X(x_i^*)\, p_Y(y_j^*)}$

$X\backslash Y$	Milano	Bergamo	Brescia
0	0.0533	0.06	0.0033
1	0.0031	0.092	0.0731
2	0.0327	0.4813	0.096

Sommando ora i valori nella tabella, si ottiene $\tilde{\varphi}^2(X,Y) = 0.8948$; osservando inoltre che, nel nostro caso, $h = k = 3$, si ottiene

$$\tilde{\varphi}^2(X,Y) = \frac{0.8943}{2} = 0.4474.$$

Esercizio 2

punto a

Ricordiamo che la probabilità dell'unione di due eventi può essere ottenuta come $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Modalità A1

In base alla definizione classica di probabilità, otteniamo

$$\begin{aligned} P(A) &= \frac{20}{200} = 0.1 \\ P(B) &= \frac{100}{200} = 0.5 \\ P(A \cap B) &= \frac{20}{200} = 0.1 \end{aligned}$$

da cui si ottiene

$$P(A \cup B) = 0.1 + 0.5 - 0.1 = 0.5.$$

Modalità A2

In base alla definizione classica di probabilità, otteniamo

$$\begin{aligned} P(A) &= \frac{175}{250} = 0.7 \\ P(B) &= \frac{100}{250} = 0.4 \\ P(A \cap B) &= \frac{50}{250} = 0.2 \end{aligned}$$

da cui si ottiene

$$P(A \cup B) = 0.7 + 0.4 - 0.2 = 0.9.$$

punto b

Modalità A1/A2

Due eventi A e B si definiscono *stocasticamente indipendenti* se

$$P(A \cap B) = P(A)\, P(B).$$

punto c

Modalità A1

Si osservi che, dei 200 dipedenti dell'azienda, 20 lavorano nella direzione generale. Sia inoltre D_i l'evento "l'$i-$esimo dipendente estratto lavora nella direzione generale" ($i = 1, 2, 3, 4$). La probabilità richiesta può essere dunque calcolata come segue:

$$P(\bar{D}_1 \cap \bar{D}_2 \cap D_3 \cap D_4) = \frac{180}{200}\frac{179}{199}\frac{20}{198}\frac{19}{197} = 0.0079.$$

Modalità A2

Si osservi che, dei 250 operai dell'azienda, 100 appartengono allo stabilimento di Milano. Sia inoltre M_i l'evento "l'$i-$esimo dipendente estratto appartiene allo stabilimento di Milano'" ($i = 1, 2, 3$). La probabilità richiesta può essere dunque calcolata come segue:

$$P(M_1 \cap \bar{M}_2 \cap M_3) = \frac{100}{250}\frac{150}{249}\frac{99}{248} = 0.0962.$$

Esercizio 3

punto a

Modalità A1/A2

Sia X una popolazione con $\mathrm{E}(X) = \mu$ e sia $(X_1, \ldots X_n)$ un campione bernoulliano da essa estratto. Ricordiamo che, per le proprietà del campione bernoulliano, gli elementi del campione sono identicamente distribuiti come X e dunque $\mathrm{E}(X_i) = \mu$ ($i = 1, \ldots, n$). Considerata la media campionaria $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, si può dunque calcolare

$$
\begin{aligned}
\mathrm{E}(\bar{X}) &= \mathrm{E}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \\
&= \frac{1}{n}\mathrm{E}\left\{\sum_{i=1}^{n} X_i\right\} = \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \\
&= \frac{1}{n}\sum_{i=1}^{n}\mu = \\
&= \frac{1}{n}n\mu = \mu
\end{aligned}
$$

il che dimostra che la media campionaria è uno stimatore non distorto di μ.

punto b

Modalità A1

Ricordiamo che lo scarto quadratico medio di $\bar{X}$ è dato da $\frac{\sigma}{\sqrt{n}}$. Nel nostro caso si avrà dunque

$$
\frac{\sqrt{1.8}}{\sqrt{n}} \leq 0.2 \qquad \Rightarrow \qquad \sqrt{n} \geq \frac{\sqrt{1.8}}{0.2} \qquad \Rightarrow \qquad n \geq \frac{1.8}{(0.2)^2} = 45
$$

Modalità A2

Ricordiamo che lo scarto quadratico medio di $\bar{X}$ è dato da $\frac{\sigma}{\sqrt{n}}$. Nel nostro caso si avrà dunque

$$
\frac{\sqrt{0.6}}{\sqrt{n}} \leq 0.1 \qquad \Rightarrow \qquad \sqrt{n} \geq \frac{\sqrt{0.6}}{0.1} \qquad \Rightarrow \qquad n \geq \frac{0.6}{(0.1)^2} = 60
$$

punto c

Ricordiamo che l'errore di seconda specie consiste nell'accettare H_0 quando questa è falsa. Ricordiamo inoltre che, essendo la popolazione Normale, $\bar{X} \sim \mathcal{N}\left(\mu; \frac{\sigma^2}{n}\right)$.

Modalità A1

Possiamo pertanto calcolare (indicando con Z una variabile casuale distribuita secondo la legge legge Normale standardizzata e con $\Phi(\cdot)$ la sua funzione di ripartizione):

$$
\begin{aligned}
\beta &= \Pr\{\bar{X} > 84.63|\mu = 84.5\} = \\
&= 1 - \Pr\{\bar{X} \leq 84.63|\mu = 84.5\} = \\
&= 1 - \Pr\left\{Z \leq \frac{84.63 - 84.5}{\sqrt{\frac{1.8}{50}}}\right\} = \\
&= 1 - \Phi\,(0.69) = \\
&= 1 - 0.7549 = 0.2451
\end{aligned}
$$

Modalità A2

Possiamo pertanto calcolare (indicando con Z una variabile casuale distribuita secondo la legge legge Normale standardizzata e con $\Phi(\cdot)$ la sua funzione di ripartizione):

$$\begin{aligned}\beta &= \Pr\{\bar{X} \leq 25.15|\mu = 25.25\} = \\ &= \Pr\left\{Z \leq \frac{25.15 - 25.25}{\sqrt{\frac{0.6}{60}}}\right\} = \\ &= \Phi(-1) = \\ &= 1 - \Phi(1) = \\ &= 1 - 0.8413 = 0.1587\end{aligned}$$

punto d

Ricordiamo che l'espressione dell'intervallo di confidenza per μ (essendo σ^2 non noto) è

$$\left(\bar{x} - t_{1-\alpha/2}^{(n-1)}\frac{s_c}{\sqrt{n}} \quad ; \quad \bar{x} + t_{1-\alpha/2}^{(n-1)}\frac{s_c}{\sqrt{n}}\right)$$

Modalità A1

Nel nostro caso si ha:

$$\begin{aligned}n &= 5 \\ \bar{x} &= \frac{85.1 + 85.0 + 84.6 + 84.9 + 85.3}{5} = 84.98 \\ s_c^2 &= \frac{(85.1 - 84.98)^2 + (85.0 - 84.98)^2 + (84.6 - 84.98)^2 + (84.9 - 84.98)^2 + (85.3 - 84.98)^2}{4} = \\ &= 0.067 \\ t_{0.95}^{(4)} &= 2.132\end{aligned}$$

e dunque

$$\left(84.98 - 2.132\frac{\sqrt{0.067}}{\sqrt{5}} \quad ; \quad 84.98 + 2.132\frac{\sqrt{0.067}}{\sqrt{5}}\right)$$

ossia $(84.7232; 85.2268)$.

Modalità A2

Nel nostro caso si ha:

$$\begin{aligned}n &= 6 \\ \bar{x} &= \frac{25.1 + 24.9 + 25.5 + 24.8 + 25.3 + 25.6}{6} = 25.2 \\ s_c^2 &= \frac{(25.1 - 25.2)^2 + (24.9 - 25.2)^2 + (25.5 - 25.2)^2 + (25.3 - 25.2)^2 + (25.6 - 25.2)^2}{5} = \\ &= 0.104 \\ t_{0.975}^{(5)} &= 2.571\end{aligned}$$

e dunque

$$\left(25.2 - 2.571\frac{\sqrt{0.104}}{\sqrt{6}} \quad ; \quad 25.2 + 2.571\frac{\sqrt{0.104}}{\sqrt{6}}\right)$$

ossia $(24.8615; 25.5385)$.

punto e

Modalità A1

Per il collegamento tra gli intervalli di confidenza e i test bilaterali, poichè il valore $\mu_0 = 85$ cade nell'intervallo determinato al punto precedente, l'ipotesi nulla viene accettata.

Modalità A2

Per il collegamento tra gli intervalli di confidenza e i test bilaterali, poichè il valore $\mu_0 = 25$ cade nell'intervallo determinato al punto precedente, l'ipotesi nulla viene accettata.

punto f

Modalità A1/A2

Rispetto al punto precedente, il valore di α è diminuito. Ciò significa che si vuole essere piu' "conservativi" nei confronti dell'ipotesi nulla, che pertanto deve continuare ad essere accettata. Da un diverso punto di vista, un minore valore di α rende piu' ampio l'intervallo di confidenza di cui al punto e); dunque il valore μ_0 cade nuovamente nell'intervallo e l'ipotesi nulla viene necessariamente accettata.