Università di Cassino

Esercitazione di Statistica 1 del 28 novembre 2007

Dott.ssa Paola Costantini

Esercizio 1

Considerando il DATASET DIPENDENTI, si calcoli la correlazione tra i caratteri STIPENDIO PERCEPITO N. ANNI DI SERVIZIO. A tale riguardo si suddividano entrambi i caratteri in tre classi equiampie.

DATASET DIPENDENTI

ID	Stipendio percepito	Età	N. di anni di servizio	Qualifica funzionale	Regime di impiego	Genere	Stato Civile
1	2650	40	15	Operaio	Tempo pieno	M	Non coniugato
2	2600	43	5	Operaio	Part time	F	Vedovo
3	2050	35	6	Impiegato	Tempo pieno	F	Coniugato
4	3500	27	6	Dirigente	Part time	M	Non coniugato
5	1400	36	3	Dirigente	Collaboratori esterni	F	Vedovo
6	2400	30	12	Impiegato	Tempo pieno	M	Vedovo
7	1900	41	13	Operaio	Tempo pieno	F	Non coniugato
8	2100	35	4	Impiegato	Tempo pieno	M	Vedovo
9	2100	27	7	Operaio	Tempo pieno	F	Non coniugato
10	3050	38	18	Dirigente	Tempo pieno	F	Coniugato
11	2800	38	20	Operaio	Collaboratori esterni	M	Non coniugato
12	2950	41	11	Operaio	Collaboratori esterni	F	Non coniugato
13	1900	36	4	Dirigente	Collaboratori esterni	M	Vedovo
14	1650	29	11	Impiegato	Collaboratori esterni	F	Coniugato
15	2550	40	4	Impiegato	Collaboratori esterni	M	Non coniugato
16	2000	23	10	Impiegato	Tempo pieno	F	Coniugato
17	2150	26	8	Operaio	Collaboratori esterni	F	Coniugato
18	2900	41	9	Dirigente	Tempo pieno	M	Non coniugato
19	2450	35	12	Operaio	Collaboratori esterni	F	Coniugato
20	1950	31	8	Dirigente	Collaboratori esterni	M	Vedovo

Soluzione

1) L'ampiezza costante delle tre classi per il carattere "stipendio percepito" (espresso in migliaia di euro) si ottiene come:

range (STIPENDIO PERCEPITO) = 3500 – 1400 = 2100

L'ampiezza delle classi della prima distribuzione di frequenza è pari a:

Ampiezza **stipendio percepito** = range/3 = 2100/3 = 700

L'ampiezza costante delle tre classi per il carattere "N. ANNI DI SERVIZIO" si ottiene come:

range (N. ANNI DI SERVIZIO) = 20 – 3 = 17

L'ampiezza delle classi della prima distribuzione di frequenza è pari a:

Ampiezza **stipendio percepito** = range/3 = 17/3 = 5,67 (per cui avremo due classi di ampiezza 6 e una di ampiezza 5)

N. anni di servizio X \ Stipendio percepito Y	[1400, 2100]	]2100, 2800]	]2800, 3500]	**Totale**
[3,8]	6	3	1	**10**
]8,14]	3	2	2	**7**
]14,20]	1	1	1	**3**
Totale	**10**	**6**	**4**	**20**

Valori Centrali di y = 1750; 2450; 3150
Valori Centrali di x = 5,5; 11; 17

Media delle y = μ_y = 2450*6 +3150*4 = 44800/20 = **2240**
Media delle x = μ_x = 5,5*10 + 11*7 +17*3 = 183/20 = **9,15**
$\mu_{x,y}$ = 1750*5,5*6 + 1750*11*3 + 1750*17*1+…………………..+ 3150*17*1 = **21070**

Cov = $\sigma_{x,y}$ = $\mu_{x,y}$ - ($\mu_x * \mu_y$) = 21070 – (2240*9,15) = 574

Variaza delle x = σ_x^2 =
$$\frac{1}{N}\sum_{i=1}^{r}\left(\hat{x}_i-\mu_x\right)^2 n_i = (5{,}5 - 9{,}15)^2 \times 10 + (11 - 9{,}15)^2 \times 7 + (17 - 9{,}15)^2 \times 3 = 17{,}10$$

σ_x^2 = 17,10 σ_x = 4,13

Variaza delle y = σ_y^2 =
$$\frac{1}{N}\sum_{j=1}^{c}\left(\hat{y}_j-\mu_Y\right)^2 n_j = (1750 - 2240)^2 \times 10 + (2450 - 2240)^2 \times 6 + (3150 - 2240)^2 \times 4=$$
=298.900

σ_y^2 = 298900 σ_y = 546,7

Corr = $\rho_{x,y}$ = $\dfrac{Cov_{x,y}}{\sigma_x * \sigma_y}$ = $\dfrac{574}{546{,}7*4{,}13}$ = **0,25**

Esercizio n. 2

Un'azienda che vende fondi comuni di investiento ha raccolto informazioni tra i propri clienti circa la percentuale di capitale investito in fondi e il loro stato civile. I dati sono riportati nella seguente tabella, dove la percentuale in fondi (espressa in milioni di lire) è stata suddivisa in due classi.

Nella tabella sono riportate le frequenze relative:

fondi di investimento / Stato Civile	[0,50]	]50,100]	**Totale**
Sposato	0,4	0,24	**0,64**
Celibe	0,08	0,08	**0,16**
Nubile	0,04	0,16	**0,20**
Totale	**0,52**	**0,48**	**1**

Definiti i seguenti eventi:

A = Estrazione di un cliente che sia sposato;
B = Estrazione di un cliente che sia celibe;
C = Estrazione di un cliente che sia nubile;
D = Estrazione di un cliente che sia disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento;
E = Estrazione di un cliente che sia disposto ad investire almeno 50 milioni del proprio capitale in fondi di investimento;

determinare le probabilità:

1) degli eventi elementari, ossia: P(A), P(B), P(C), P(D), P(E);
2) che un cliente sia sposato o celibe;
3) che un cliente sia sposato e disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento;
4) che un cliente sia celibe e disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento;
5) che un cliente sia celibe o disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento;
6) che un cliente sia celibe e disposto ad investire almeno 50 milioni del proprio capitale in fondi di investimento;
7) che un cliente sia sposato, sapendo che e disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento;
8) che un cliente sia disposto ad investire al massimo 50 milioni del proprio capitale in fondi di investimento, sapendo che è sposato.

Soluzione

1) $P(A) = 0{,}64$; $P(B) = 0{,}16$; $P(C) = 0{,}2$; $P(D)= 0{,}52$;
$P(E) = 0{,}48$

2) $P(A\cup B) = P(A) + P(B) - P(A\cap B)= 0{,}64+0{,}16 - 0 = 0{,}80$
3) $P(A\cap D) = 0{,}4$
4) $P(B\cap D) = 0{,}08$
5) $P(B\cup D) = P(B) + P(D) - P(B\cap D)= 0{,}16 + 0{,}52 - 0{,}08 = 0{,}6$
6) $P(B\cap E) = 0{,}08$
7) $P(A|D) = \dfrac{P(A \cap D)}{P(D)} = \dfrac{0{,}4}{0{,}52} = 0{,}76$

8) $P(D|A) = \dfrac{P(A \cap D)}{P(A)} = \dfrac{0{,}4}{0{,}64} = 0{,}625$

Esercizio n. 3

Nella tabella seguente sono riportate le numerosità delle birre vendute distinte per tipologia.

(frequenze assolute)

Dimensione / **Tipologia di birra**	Piccola	Media	Grande	**Totale**
Bionda	7	9	25	**41**
Rossa	12	10	9	**31**
Scura	17	8	3	**28**
Totale	**36**	**27**	**37**	**100**

Legenda

BIRRA BIONDA	B
BIRRA ROSSA	R
BIRRA SCURA	S
BIRRA PICCOLA	P
BIRRA MEDIA	M
BIRRA GRANDE	G

A) Calcolare le seguenti probabilità:

1) che un bevitore abbia scelto una birra rossa;
2) che un bevitore abbia scelto una birra scura;
3) che un bevitore abbia scelto una birra media;
4) che un bevitore abbia scelto una birra rossa piccola;
5) che un bevitore abbia scelto una birra media, sapendo che è bionda;
6) che un bevitore abbia scelto una birra scura, sapendo che è grande;

B) Verificare se la tipologia di birra è <u>stocasticamente indipendente</u> dalla dimensione o meno.

Soluzione

A)

1) P(R) = 31/100= 0,31
2) P(S) = 28/100= 0,28
3) P(M) = 27/100= 0,27
4) $P(R \cap P)$ = 10/100 = 0,1
5) $P(M|B) = \frac{P(M \cap B)}{P(B)} = \frac{9/100}{41/100} = \frac{9}{41} = 0,22$
6) $P(S|G) = \frac{P(S \cap G)}{P(G)} = \frac{3/100}{37/100} = \frac{3}{37} = 0,08$

B)

Se i due caratteri sono indipendenti devono essere verificate le seguenti uguaglianze:

- se la tipologia non dipende dalla dimensione:

P(B|P) = P(B|M) = P(B|G) = P(B)	P(B|P) = 7/36 = 0,19 P(B|M) = 9/27 = 0,33 P(B|G) = 25/37 = 0,67 P(B) = 41/100 = 0,41
P(R|P) = P(R|M) = P(R|G) = P(R)	P(R|P) = 12/36 = 0,33 P(R|M) = 10/27 = 0,37 P(R|G) = 9/37 = 0,24 P(R) = 31/100 = 0,31
P(S|P) = P(S|M) = P(S|G) = P(S)	P(S|P) = 17/36 = 0,47 P(S|M) = 8/27 = 0,29 P(S|G) = 3/37 = 0,08 P(S) = 28/100 = 0,28

- se la dimensione non dipende dalla tipologia:

P(P|B) = P(P|R) = P(P|S) = P(P)	P(P|B) = 7/41 = 0,17 P(P|R) = 12/31 = 0,39 P(P|S) = 17/28 = 0,61 P(P) = 36/100 = 0,36
P(M|B) = P(M|R) = P(M|S) = P(M)	P(M|B) = 9/41 = 0,22 P(M|R) = 10/31 = 0,32 P(M|S) = 8/28 = 0,28 P(M) = 27/100 = 0,27
P(G|B) = P(G|R) = P(G|S) = P(G)	P(G|B) = 25/41 = 0,61 P(G|R) = 9/31 = 0,29 P(G|S) = 3/28 = 0,11 P(G) = 37/100 = 0,37

Le uguaglianze non sono soddisfatte, quindi i due caratteri non sono stocasticamente indipendenti.

Esercizio 4

Misurare il grado di dipendenza tra la tipologia di birra e la dimensione del boccale in cui viene bevuta.

Soluzione

Trattandosi di due caratteri qualitativi, procediamo al calcolo del χ^2.

Tabella delle frequenze osservate:

Dimensione / **Tipologia di birra**	Piccola	Media	Grande
Bionda	7	9	25
Rossa	12	10	9
Scura	17	8	3

Tabella delle frequenze teoriche: $\hat{n}_{ij} = \dfrac{n_{i.} \times n_{.j}}{N}$:

Dimensione / **Tipologia di birra**	Piccola	Media	Grande
Bionda	14,76	11,07	15,17
Rossa	11,16	8,37	11,47
Scura	10,08	7,56	10,36

Sostituendo nella formula si ha:

$$\chi^2 = \frac{(7\text{-}14{,}76)^2}{14{,}76} + \frac{(9\text{-}11{,}07)^2}{11{,}07} + \frac{(25\text{-}15{,}17)^2}{b15{,}17} + \frac{(12\text{-}11{,}16)^2}{11{,}16} + \ldots\ldots + \frac{(3\text{-}10{,}36)^2}{10{,}36} = \mathbf{21{,}685}$$

Quindi si ha:

$$\phi^2 = \frac{\chi^2}{N} = \frac{21{,}685}{100} = \mathbf{0{,}217}$$

Tale valore va confrontato con l'intervallo [0, 2], in quanto

$$0 \le \varphi^2 \le \min(r-1;\ c-1)$$

Pertanto possiamo affermare che *esiste un certo grado di connessione* (e peraltro avremmo già potuto sostenerlo guardando che le frequenze teoriche sono diverse da quelle osservate), ma si tratta di un legame piuttosto debole.

In altre parole, a seconda del tipo di birra, viene preferito un certo tipo di boccale.

Infine calcoliamo l' **Indice T di Tchuprov**

$$\mathbf{T} = \frac{\phi^2}{\min\{r-1, c-1\}} = \frac{\chi^2}{n \times \min\{r-1, c-1\}} = \mathbf{0{,}1085}$$

per $0 \leq \mathbf{T} \leq 1$ si conferma che si tratta di un legame debole.