Università di Cassino

Corso di Statistica

Esercitazione del 18 marzo 2010

Simona Balzano

ESERCIZIO 1

Sia X la variabile casuale “numero di auto per famiglia”, così distribuita:

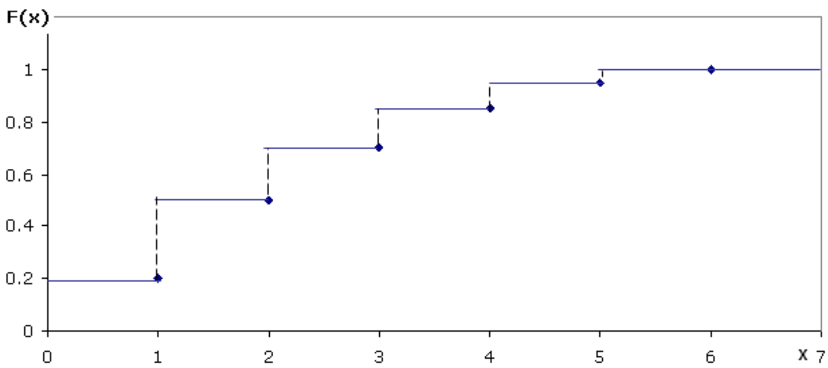
x_i	p_i
1	0,2
2	0,3
3	0,2
4	0,15
5	0,1
6	0,05
	1

a) rappresentare graficamente la funzione di ripartizione di X;
b) determinare la probabilità che una famiglia abbia fino a 4 auto;
c) determinare la probabilità che una famiglia abbia da 2 a 4 auto;
d) standardizzare X e verificare che la v.c. standardizzata abbia media nulla e varianza unitaria.

Soluzione

a)

x_i	p_i	$F(x_i)$
1	0,2	0,2
2	0,3	0,5
3	0,2	0,7
4	0,15	0,85
5	0,1	0,95
6	0,05	1
	1	



b)

$P(X \leq 4) = F(4) = 0,85$

c)

$P(2 \leq X \leq 4) = F(4) - F(1) = 0,85 - 0,2 = 0,65$

Nota: "Da 2 a 4 auto" vuol dire che anche il 2 deve essere incluso nella probabilità, quindi dalla $F(4) = P(X \leq 4)$ va sottratta la $P(X \leq 1)$. $F(4) - F(2)$ avrebbe significato $P(2 < X \leq 4)$, cioè la probabilità che una famiglia abbia da "3" a 4 auto (escludendo il valore 2).

c)

Per standardizzare X bisogna calcolare il suo valore atteso e la sua varianza.

Valore atteso:

$$E(X) = \sum_{i=1}^{k} x_i p_i = (1 \times 0,2) + (2 \times 0,3) + \cdots + (6 \times 0,05) = 2,8$$

Varianza:

$$E[X - E(X)]^2 = \sum_{i=1}^{k} (x_i - E(X))^2 p_i =$$

$$= (1 - 2,8)^2 \times 0,2 + (2 - 2,8)^2 \times 0,3 + \cdots + (6 - 2,8)^2 \times 0,05 = 2,06$$

La variabile standardizzata Z presenta i seguenti valori, ottenuti dalla relazione:

$$z_i = \frac{x_i - \mu}{\sigma}$$

z_i	p_i
-1.25	0,2
-0.56	0,3
0.14	0,2
0.84	0,15
1.53	0,1
2.23	0,05
	1

Valore atteso:

$$E(Z) = \sum_{i=1}^{k} z_i p_i = (-1,25 \times 0,2) + (-0,56 \times 0,3) + \cdots + (2,23 \times 0,05) = 0$$

Varianza:

$$E[Z - E(Z)]^2 = \sum_{i=1}^{k} (z_i - E(Z))^2 p_i = \sum_{i=1}^{k} z_i^2 p_i$$

$$= (-1,25)^2 \times 0,2 + (-0,56)^2 \times 0,3 + \cdots + (2,23)^2 \times 0,05 = 1$$

ESERCIZIO 2

Al termine di un corso universitario il docente pianifica un esame consistente in un test scritto più una prova orale. Per valutare la coerenza del sistema di valutazione ipotizzato, egli utilizza due variabili casuali X = "voto finale" ed Y = "numero di errori commessi al test", per le quali definisce la seguente distribuzione di probabilità congiunta:

Voto (X) \ Errori (Y)	1	2	3	**totale**
4	0.485	0.27	0.165	**0.92**
5	0.015	0.03	0.035	**0.08**
totale	**0.5**	**0.3**	**0.2**	**1**

Determinare:

a) i valori attesi condizionati di X e di Y
b) la covarianza ed il coefficiente di correlazione tra X ed Y
c) sulla base delle quantità calcolate si può concludere che il sistema di valutazione pensato sia coerente?

Soluzione

r = numero di valori che può assumere X
c = numero di valori che può assumere Y

a)
Valori attesi di X condizionati ad Y (voto finale atteso per chi commetterà 1 errore, 2 errori e 3 errori al test):

$$E\left[X \mid Y = y_j\right] = \sum_{i=1}^{r} x_i \times P\left(X = x_i \mid Y = y_j\right)$$

in cui:

$$P\left(X = x_i \mid Y = y_j\right) = \frac{P\left(x_i \cap y_j\right)}{P\left(y_j\right)} = \frac{p_{ij}}{p_{\bullet j}}$$

Quindi:

$$E\left[X \mid Y = 1\right] = 4 \times P\left(X = 4 \mid Y = 1\right) + 5 \times P\left(X = 5 \mid Y = 1\right) =$$

$$= 4 \times \frac{p_{11}}{p_{\bullet 1}} + 5 \times \frac{p_{21}}{p_{\bullet 1}} =$$

$$= 4 \times \frac{0{,}485}{0{,}5} + 5 \times \frac{0{,}015}{0{,}5} = 4{,}03$$

$$E\left[X \mid Y=2\right] = 4 \times P\left(X=4 \mid Y=2\right) + 5 \times P\left(X=5 \mid Y=2\right) =$$
$$= 4 \times \frac{p_{12}}{p_{\bullet 2}} + 5 \times \frac{p_{22}}{p_{\bullet 2}} =$$
$$= 4 \times \frac{0,27}{0,3} + 5 \times \frac{0,003}{0,3} = 4,1$$

$$E\left[X \mid Y=3\right] = 4 \times P\left(X=4 \mid Y=3\right) + 5 \times P\left(X=5 \mid Y=3\right) =$$
$$= 4 \times \frac{p_{13}}{p_{\bullet 3}} + 5 \times \frac{p_{23}}{p_{\bullet 3}} =$$
$$= 4 \times \frac{0,165}{0,2} + 5 \times \frac{0,035}{0,2} = 4,175$$

Valori attesi di Y condizionati ad X (numero atteso di errori per chi avrà un voto pari a 4 e a 5):

$$E\left[Y \mid X=x_0\right] = \sum_{j=1}^{c} y_j \times P\left(Y=y_j \mid X=x_0\right)$$

in cui:

$$P\left(Y=y_j \mid X=x_i\right) = \frac{P\left(y_j \cap x_i\right)}{P\left(x_i\right)} = \frac{p_{ij}}{p_{i\bullet}}$$

Quindi:

$$E\left[Y \mid X=4\right] = 1 \times P\left(Y=1 \mid X=4\right) + 2 \times P\left(Y=2 \mid X=4\right) + 3 \times P\left(Y=3 \mid X=4\right) =$$
$$= 1 \times \frac{p_{11}}{p_{1\bullet}} + 2 \times \frac{p_{12}}{p_{1\bullet}} + 3 \times \frac{p_{13}}{p_{1\bullet}} =$$
$$= 1 \times \frac{0,485}{0,92} + 2 \times \frac{0,27}{0,92} + 3 \times \frac{0,165}{0,92} = 1,65$$

$$E\left[Y \mid X=5\right] = 1 \times P\left(Y=1 \mid X=5\right) + 2 \times P\left(Y=2 \mid X=5\right) + 3 \times P\left(Y=3 \mid X=5\right) =$$
$$= 1 \times \frac{p_{21}}{p_{2\bullet}} + 2 \times \frac{p_{22}}{p_{2\bullet}} + 3 \times \frac{p_{23}}{p_{2\bullet}} =$$
$$= 1 \times \frac{0,015}{0,08} + 2 \times \frac{0,03}{0,08} + 3 \times \frac{0,035}{0,08} = 2,25$$

Dall'esame dei valori attesi condizionati si può concludere che:

- ci si aspetta un numero di errori maggiore da chi ha un voto finale più alto
- ci si aspetta un voto il voto finale più alto da ci commette più errori

quindi che il sistema di valutazione non è coerente.

b)

La covarianza tra le v.c. X ed Y si calcola come:

$$\sigma_{XY} = \sum_{i=1}^{k}\sum_{j=1}^{h}\left(x_i - E(X)\right)\left(y_j - E(Y)\right)p_{ij} = E(X,Y) - E(X)E(Y)$$

in cui:

$$\begin{aligned} E(X,Y) &= \sum_{i=1}^{r}\sum_{j=1}^{c} x_i y_j p_{ij} = \\ &= 4\times 1\times 0,485 + 4\times 2\times 0,27 + 4\times 3\times 0,165 + \\ &\quad +5\times 1\times 0,015 + 5\times 2\times 0,03 + 5\times 3\times 0,035 = 6,98 \end{aligned}$$

mentre:

$$E(X) = \sum_{i=1}^{r} x_i p_{i\bullet} = 4\times 0,92 + 5\times 0,08 = 4,08$$

e:

$$E(Y) = \sum_{i=1}^{c} y_j p_{\bullet j} = 1\times 0,5 + 2\times 0,3 + 3\times 0,2 = 1,7$$

> Nota:
>
> esiste una relazione tra i valori attesi E(X) ed E(Y) ed i valori attesi condizionati, infatti:
>
> $$E(X) = \sum_{j=1}^{c} E\left(X \mid Y = y_j\right)p_{\bullet j} = 4,03\times 0,5 + 4,1\times 0,3 + 4,175\times 0,2 = 4,08$$
>
> e:
>
> $$E(Y) = \sum_{i=1}^{r} E\left(Y \mid X = x_i\right)p_{i\bullet} = 1,65\times 0,92 + 2,25\times 0,08 = 1,7$$

La covarianza è uguale a:

$$\sigma_{XY} = E(X,Y) - E(X)E(Y) = 6,98 - 4,08\times 1,7 = 0,044$$

Il coefficiente di correlazione è dato da:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Quindi:

$$\sigma_X = \sqrt{\sum_{i=1}^{r} \left(x_i - E(X)\right)^2 p_{i\bullet}} = 0,27$$

e:

$$\sigma_Y = \sqrt{\sum_{j=1}^{c} \left(y_j - E(Y)\right)^2 p_{\bullet j}} = 1,15$$

da cui:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0,044}{0,27 \times 1,15} = 0,14$$

ESERCIZIO 3.

Un investitore possiede un portafoglio composto per metà da fondi azionari e per metà da fondi obbligazionari. Per ogni anno, il rendimento atteso E(X) dei fondi azionari è del 6% con scarto quadratico medio 0,03, mentre il rendimento atteso E(Y) dei fondi obbligazionari è del 4% con scarto quadratico medio 0,02. La covarianza fra i rendimenti dei due fondi è 0,0006. Si vuole calcolare il rendimento atteso e lo scarto quadratico medio dell'intero portafoglio.

Soluzione

Il rendimento del portafoglio P può essere considerato una combinazione lineare delle variabili casuali X (rendimento dei fondi azionari) ed Y (rendimento dei fondi obbligazionari):

$$P = 0,5\ X + 0,5\ Y$$

Ricordando che per una combinazione lineare Z di p variabili X_i:

$$Z = \sum_{i=1}^{p} a_i X_i$$

(i = 1, …, p) valgono le seguenti regole:

$$E(Z) = \sum_{i=1}^{p} a_i E(X_i)$$

e:

$$Var(Z) = \sum_{i=1}^{p} a_i^2 Var(X_i) + 2\sum_{i=1}^{n} \sum_{j \neq i} a_i a_j Cov(X_i, X_j),$$

si ha:

$$E(P) = 0,5 \times E(X) + 0,5 \times E(Y) = 0,5 \times 0,06 + 0,5 \times 0,04 = 0,05$$

e:

$$\begin{aligned} Var(P) &= 0,5^2 \times Var(X) + 0,5^2 \times Var(Y) + 2 \times 0,5 \times 0,5 \times Cov(X, Y) = \\ &= 0,25 \times 0,009 + 0,25 \times 0,004 + 2 \times 0,25 \times 0,0006 = 0,0035 \end{aligned}$$

da cui:

$$\sigma_P = \sqrt{0,0035} = 0,06$$