Probabilità e Statistica Matematica – Soluzioni   21. 1. 2008

1. Ricordiamo che il quantile $q_p(x)$ al $(100p)\%$ di n dati numerici ordinati $x_{(1)}, \dots, x_{(n)}$ è definito come

$$q_p(x) := x_{(\lfloor\alpha\rfloor)} + (\alpha - \lfloor\alpha\rfloor)(x_{(\lfloor\alpha\rfloor+1)} - x_{(\lfloor\alpha\rfloor)}),$$

dove $\alpha = 1 + (n-1)p$. Si ottengono:

(a) media $\overline{x} = 5$; mediana $q_{0,5} = 5$; primo quartile $q_{0,25} = 2,5$; terzo quartile $q_{0,75} = 7,5$; $s^2 = (0-\overline{x})^2 + (10-\overline{x})^2 = 50$; $s = 5\sqrt{2} \approx 7,071$.

Utilizzando il foglio elettronico Excel, gli indici q_p si possono calcolare con le funzioni PERCENTILE($\{0;10\}$;p), ($p = 0,5;\ 0,25;\ 0,75$) o QUARTILE($\{0;10\}$;k), ($k = 2,1,3$) e VAR($\{0;10\}$), DEV.ST($\{0;10\}$).

(b) media $\overline{x} = 90$; mediana $q_{0,5} = 89$; primo quartile $q_{0,25} = 77$; terzo quartile $q_{0,75} = 100$; $s^2 = 292,4375$; $s \approx 17,10080407$.

		9	7			
	8	8	5			
	7	7	5			
8	5	5	4	8		
5	4	4	2	7	5	
5	4	4	2	2	5	9
0	2	2	0	0	0	2
6	**7**	**8**	**9**	**10**	**11**	**12**

Il coefficiente di asimmetria è positivo (la "coda" è a destra) e uguale a 0,360246 (calcolato con la funzione ASIMMETRIA() di Excel).

2. Lo scarto di ogni dato dalla media è $\frac{1}{2}$. Di conseguenza la varianza è

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{1}{2}\right)^2 = \frac{n}{4(n-1)}.$$

3.
$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{i=1}^{n}\frac{x_i - \overline{x}}{s_x} = \frac{1}{ns_x}\sum_{i=1}^{n}(x_i - \overline{x}) = \frac{1}{ns_x}(n\overline{x} - n\overline{x}) = 0,$$

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n} y_i^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{x_i - \overline{x}}{s_x}\right)^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{(n-1)s_x^2} = \frac{(n-1)s_x^2}{(n-1)s_x^2} = 1.$$

4.
$$s_x^2 = \frac{1}{2}\left[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + (x_3 - \overline{x})^2\right], \quad s^2_{[x_1,x_2],[x_3]} = \frac{1}{2}\left[2(\frac{x_1 + x_2}{2} - \overline{x})^2 + (x_3 - \overline{x})^2\right]$$

$$s_x^2 - s^2_{[x_1,x_2],[x_3]} = \frac{1}{2}\left[(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 - 2(\frac{x_1 + x_2}{2} - \overline{x})^2\right] = \frac{1}{4}(x_1 - x_2)^2.$$

Quindi, sostituendo x_1 e x_2 con la loro media, si commette un errore che è zero se e solo se $x_1 = x_2$.

5. Con $\hat{y}_i - \overline{y} = a(x_i - \overline{x})$ e
$y_i - \hat{y}_i = (y_i - \overline{y}) - (\hat{y}_i - \overline{y}) = (y_i - \overline{y}) - (ax_i + b - a\overline{x} - b) = (y_i - \overline{y}) - a(x_i - \overline{x})$
si ottiene

(a)

$$\begin{aligned}
& \sum_{i=1}^{n}(\hat{y}_i - \overline{y})(y_i - \hat{y}_i) = \sum_{i=1}^{n} a(x_i - \overline{x})[(y_i - \overline{y}) - a(x_i - \overline{x})] = \\
= \ & a\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) - a^2\sum_{i=1}^{n}(x_i - \overline{x})^2 = \\
= \ & a\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) - a\frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \\
= \ & 0.
\end{aligned}$$

(b)

$$\begin{aligned}
DT &= \sum_{i=1}^{n}(y_i - \overline{y})^2 = \sum_{i=1}^{n}[(\hat{y}_i - \overline{y}) - (y_i - \hat{y}_i)]^2 = \\
&= \sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2 + \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + 2\sum_{i=1}^{n}(\hat{y}_i - \overline{y})(y_i - \hat{y}_i) = \\
&= DS + DR + 0,
\end{aligned}$$

si veda (a).

(c)

$$\begin{aligned}
R^2 &= \frac{DS}{DT} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{\sum_{i=1}^{n}(ax_i + b - a\overline{x} - b)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \frac{a^2\sum_{i=1}^{n}(x_i - \overline{x})^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = \\
&= \frac{a^2 s_x^2}{s_y^2} = \frac{s_{xy}^2 s_x^2}{s_x^4 s_y^2} = \left(\frac{s_{xy}}{s_x s_y}\right)^2 = r_{xy}^2.
\end{aligned}$$