STATISTICA (I MODULO - STATISTICA DESCRITTIVA)
Soluzione esercitazione 6

Esercizio A. Alcune delle quantità utilizzate nei calcoli sono riportate nella tabella seguente:

i	x_i	y_i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	$\hat{y}_i$	$(\hat{y}_i - \bar{y})^2$
1	17.270	2128.020	0.867	24822.515	146.740	2156.285	16715.125
2	18.006	2430.049	0.038	20873.712	-28.227	2258.451	735.527
3	13.428	1623.967	22.785	437720.680	3158.087	1622.965	439047.578
4	14.123	1799.352	16.633	236409.524	1982.986	1719.440	320504.985
5	28.942	3844.136	115.361	2429122.911	16739.955	3776.510	2222898.253
6	17.135	1696.326	1.137	347210.407	628.357	2137.545	21911.918
7	12.537	1649.746	32.085	404274.225	3601.555	1499.283	618250.493
8	24.170	3112.977	35.624	684599.655	4938.472	3114.095	686450.245
Totale	145.611	18284.573	224.532	4585033.628	31167.925	18284.573	4326514.123

1. I parametri della retta di regressione $y = \beta_0 + \beta_1 x$ determinati mediante il metodo dei minimi quadrati sono:

$$\beta_1 = \frac{31167.925}{224.532} = 138.81$$

$$\beta_0 = 2285.6 - 138.81 \times 18.201 = -240.88$$

dove $N = 8$, $\bar{x} = 145.611/8 = 18.201$, $\bar{y} = 18284.573/8 = 2285.6$, $C_{XY} = 31167.925$, $D_X = 224.532$.

2. Il diagramma di dispersione contenente la retta di regressione del fatturato (y) rispetto agli investimenti (x) è riportato nella Figura 1.

Le due linee tratteggiate hanno coordinate pari alle medie, quindi la loro intersezione identifica il baricentro della distribuzione. Si può verificare graficamente che la retta di regressione passa per il baricentro della nuvola dei punti.

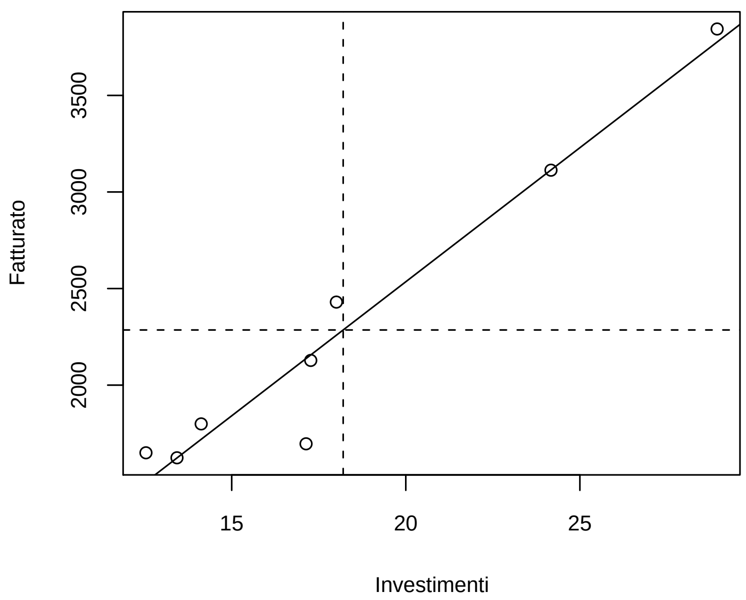

Figura 1.

3. L'indice r^2 può essere calcolato come

$$r^2 = \frac{D_S}{D_Y} = \frac{4326514.123}{4585033.628} = 0.9436$$

oppure come

$$r^2 = \frac{C_{XY}^2}{D_X D_Y} = \frac{31167.925^2}{224.532 \times 4585033.628} = 0.9436$$

In quest'ultimo caso il calcolo risulta più agevole in quanto non richiede il calcolo dei valori empirici $\hat{y}_i$ e, quindi, della devianza spiegata D_S.

Dal valore di r^2 si deduce che la bontà di adattamento della retta di regressione ai dati è molto buona. La percentuale della variabilità del fatturato spiegata dalla retta di regressione è pari al 94.36%.

4. L'indice di correlazione è pari a

$$r = \frac{31167.925}{\sqrt{224.532 \times 4585033.628}} = 0.9714$$

da cui si deduce che le due variabili sono fortemente correlate in maniera positiva. Tale risultato è anche supportato dall'inclinazione (positiva) della retta di regressione.

5. L'errore medio di predizione si può calcolare come segue:

$$\sigma_{RL} = \sigma_y \sqrt{1 - r^2} = 757.05\sqrt{1 - 0.9436} = 179.79$$

dove $\sigma_y = \sqrt{4585033.628/8} = 757.05$. Tale valore fornisce una misura dell'errore medio che si commette attribuendo al fatturato i valori forniti dalla retta di regressione.

Esercizio B.

– Per la serie delle esportazioni Alcune delle quantità utilizzate sono calcolate nella tabella seguente:

Anno	x_i (Anno−1997)	y_i (Esportazioni)	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1998	1	220105	20.25	3389102656	261972.0
1999	2	221040	12.25	3281112961	200483.5
2000	3	260413	6.25	320696464	44770.0
2001	4	272990	2.25	28419561	7996.5
2002	5	269064	0.25	85692049	4628.5
2003	6	264616	0.25	187827025	-6852.5
2004	7	284413	2.25	37112464	9138.0
2005	8	299923	6.25	466646404	54005.0
2006	9	332013	12.25	2882830864	187922.0
2007	10	358633	20.25	6450017344	361404.0
Totale	55	2783210	82.50	17129457792	1125467.0

1. I parametri della retta di regressione per le esportazioni in funzione dell'anno sono i seguenti:

$$\beta_1 = \frac{1125467.0}{82.50} = 13642$$

$$\beta_0 = 278321 - 13642 \times 5.5 = 203290$$

dove $N = 10$, $\bar{x} = 55/10 = 5.5$, $\bar{y} = 2783210/10 = 278321$, $C_{XY} = 1125467.0$, $D_X = 82.50$.

2. La serie storica e la retta di regressione calcolata al punto precedente sono rappresentate nel grafico a sinistra della Figura 2.

3. Il valore previsto per le esportazioni nel 2008 è il seguente:

$$y_{2008} = 203290 + 13642 \times (2008 - 1997) = 353352$$

con un errore medio di predizione risulta pari a:

$$\sigma_{RL} = \sigma_y \sqrt{1 - r^2} = 41388\sqrt{1 - 0.8963} = 13328$$

dove $\sigma_y = \sqrt{17129457792/10} = 41388$ e $r^2 = 1125467.0^2/(82.50 \times 17129457792) = 0.8963$.

– Per la serie delle importazioni, alcune delle quantità utilizzate sono calcolate nella tabella seguente:

Anno	x_i (Anno−1997)	y_i (Importazioni)	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1998	1	195625	20.25	6534281058	363757
1999	2	207015	12.25	4822594136	243057
2000	3	258507	6.25	322306618	44882
2001	4	263757	2.25	161363668	19054
2002	5	261226	0.25	232071709	7617
2003	6	262998	0.25	181222752	-6731
2004	7	285634	2.25	84164111	13761
2005	8	309292	6.25	1077946790	82080
2006	9	352465	12.25	5776775226	266018
2007	10	368080	20.25	8394242724	412290
Totale	55	2764599	82.50	27586968793	1445786

4. I parametri della retta di regressione per le importazioni in funzione dell'anno sono i seguenti:

$$\begin{aligned}\beta_1 &= \frac{1445786}{82.50} = 17525 \\ \beta_0 &= 276459.9 - 17525 \times 5.5 = 180072\end{aligned}$$

dove $N = 10$, $\bar{x} = 55/10 = 5.5$, $\bar{y} = 2764599/10 = 276459.9$, $C_{XY} = 1445786$, $D_X = 82.50$.

La serie storica delle importazioni e la retta di regressione calcolata sono rappresentate nel grafico a destra della Figura 2.

Il valore previsto per le importazioni nel 2008 è il seguente:

$$y_{2008} = 180072 + 17525 \times (2008 - 1997) = 372847$$

con un errore medio di predizione risulta pari a:

$$\sigma_{RL} = \sigma_y \sqrt{1 - r^2} = 52523\sqrt{1 - 0.9184} = 15004$$

dove $\sigma_y = \sqrt{27586968793/10} = 52523$ e $r^2 = 1445786^2/(82.50 \times 27586968793) = 0.9184$.

5. Sia le esportazioni che le importazioni presentano trend crescenti, quindi i coefficienti angolari delle corrispondenti rette di regressione sono positivi. Inoltre, in entrambi i casi la bontà di adattamento della retta ai dati risulta elevata.

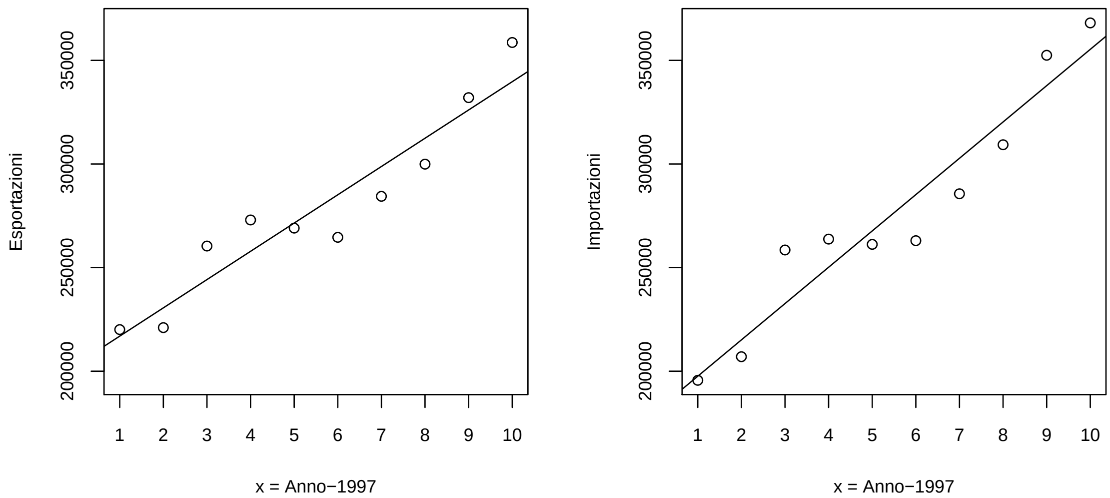

Figura 2.

Esercizio C.

1. Il grafico di dispersione del costo degli interventi di manutenzione rispetto all'età dell'auto è riportato in Figura 4. Le linee tratteggiate indicano le medie delle due variabili.

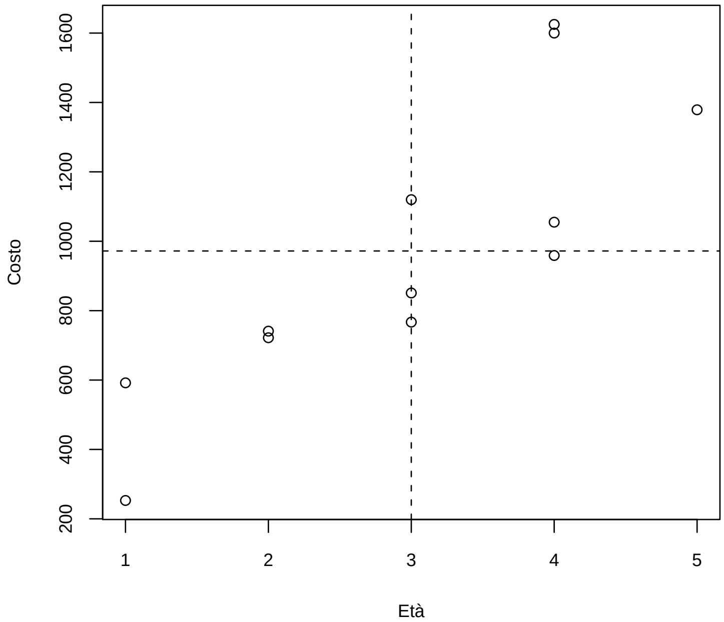

Figura 3.

La tabella seguente riporta le quantità necessarie per il calcolo del coefficiente di correlazione di Bravais:

	x_i	$(x_i - \bar{x})^2$	y_i	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})^2$
1	1	4	253	516961	1438
2	1	4	592	144400	760
3	2	1	741	53361	231
4	2	1	722	62500	250
5	3	0	851	14641	0
6	3	0	767	42025	0
7	3	0	1120	21904	0
8	4	1	1055	6889	83
9	4	1	959	169	-13
10	4	1	1625	426409	653
11	4	1	1600	394384	628
12	5	4	1379	165649	814
Totale	36	18	11664	1849292	4844

Quindi, $N = 12$, $\bar{x} = 36/12 = 3$, $\bar{y} = 11664/12 = 972$, $C_{XY} = 4844$, $D_X = 18$, $D_Y = 1849292$, da cui

$$r = \frac{C_{XY}}{\sqrt{D_X D_Y}} = \frac{4844}{\sqrt{18 \times 1849292}} = 0.8396$$

2. Il coefficiente di correlazione indica una forte relazione lineare positiva tra costi di manutenzione e età dell'auto. Quindi, è verificata l'aspettativa per la quale ad auto più vecchie sono associati maggiori costi di manutenzione.