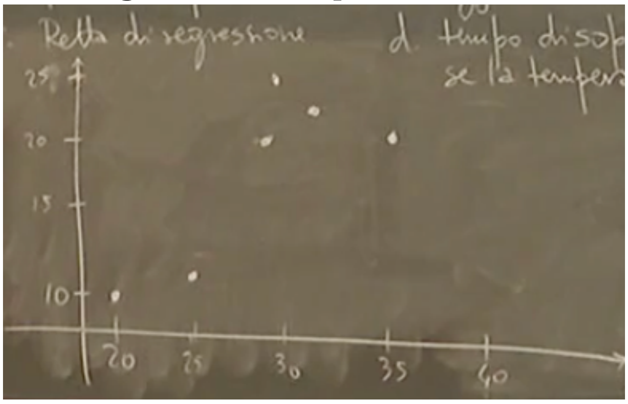
Regressione Lineare

Sunday, 11 June 2023 18:23

1) Relazione che intercorre tra la temperatura e il tempo di sopravvivenza di certi microrganismi.

Temperatura	Tempo
30	10
31	12
28	18
30	24
38	22
36	20

a. Fare il grafico di dispersione dei dati



b. Calcolare il coefficente di correlazione lineare

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{3136 - 5 * 28.333 * 17.6667}{\sqrt{\cdots}} \simeq 0.8228$$

$$S_{xy}^2 = \frac{1}{n-1}\sum(x_i - \bar{x})(y_i - \bar{y})$$

$$S_{xx} = \frac{1}{n-1}\sum(x_i - \bar{x})$$

$$S_{yy} = \frac{1}{n-1}\sum(y_i - \bar{y})$$

$$\bar{x} = \frac{1}{6}(20 + \cdots + 36) \simeq 28.333$$

$$\bar{y} = \frac{\breve{1}}{6}(10 + \cdots + 20) \simeq 17.667$$

$r_{xy} > 0.8$ quindi correlazione lineare forte significativa

c. Scrivere la retta di regressione e disegnarla
d. Calcolare il tempo di sopravvivenza previsto se temperatura=25

2)

$n \geq 2$ *con coppie* (x_i, y_i)

$$\sum x_i y_i = 0$$

$$s_x = \sum x_i - \bar{x} > 0$$

$$s_y = \sum y_i - \bar{y} > 0$$

$$r_{xy} = \sqrt{R_{xy}^2}$$

$$R_{xy}^2 = 1 - \frac{SS_r}{S_{yy}} = 1 - \frac{SS_r}{\sqrt{S_y}}$$

$$SS_r = \frac{S_{xx}S_{yy} - S_{xY}^2}{S_{xx}}$$

$$S_{xY}^2 = \sum x_i y_i - n\bar{x}\bar{y} = \sum -n\bar{x}\bar{y}$$

Siccome $\sum x_i y_i = 0$
Si suppone che $\bar{x} = 0$
Quindi $S_{xY}^2 = 0$

$$SS_r = \frac{S_{xx}S_{yy}}{S_{xx}} = S_{yy} > 0$$

$$R_{xy}^2 = 1 - \frac{S_{yy}}{S_{yy}} = 1 - 1 = 0$$

Quindi se $\bar{x} = 0 \rightarrow r_{xy} = 0$

3) Regressione lineare $y_i = \alpha + \beta x_i + e_1$
Si suppone che
H_0: $\beta \geq 1, H_1$: $\beta < 1$
Si rifiuta H_0 con livello di significità 3%
Quindi

- Si rifiuta H_0 a livello significavità 2% -> Non si dice per forza
- I dati non sono contraddizione h0 -> Falso siccome 3% è molto significativo

- $\bar{\alpha} = 1\%$ -> non per forza
- C'è evidenza empirica -> Esclusione

4)

x_i	y_i
6	4
7	3
8	5
4	2
5	3

$Y = \alpha + \beta x + \epsilon$

$\epsilon \sim N(0, \sigma^2)$

a. Trovare equazione retta regressione

$\alpha = \bar{y} - \hat{b}\bar{x}$

$$\beta = \frac{S_{xy}}{S_{xx}}$$

X	Y	X^2	Y^2	XY
6	4	36	16	24
7	3	49	9	21
8	5	64	25	40
4	2	16	4	8
5	3	25	9	15

$$\bar{x} = \frac{1}{5} * (6 + 7 + 8 + 4 + 5) = 6$$

$$\bar{y} = \frac{17}{5}$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2 = 10$$

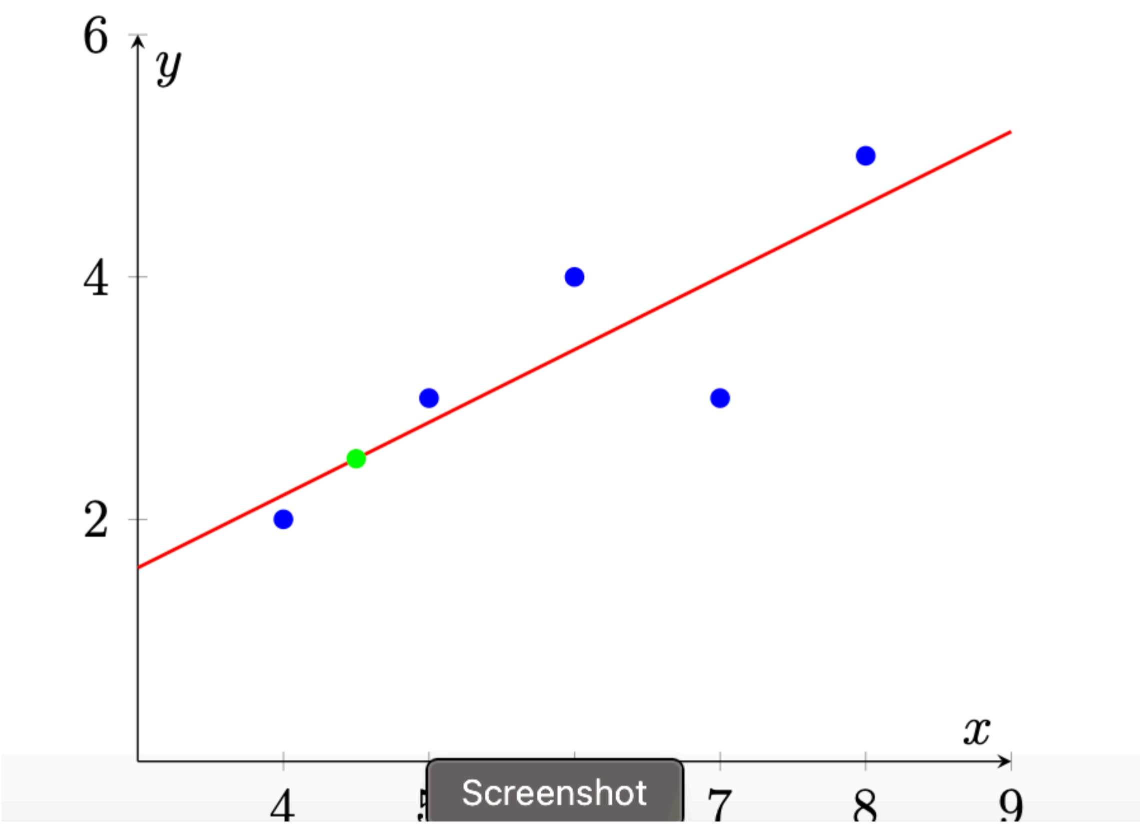
$$n\bar{x}^2 = 5 * 36 = 180$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 6$$

$$n\bar{x}\bar{y} = 5 * 6 * \frac{17}{5} = 102$$

$$S_{yy} = \sum y^2 - n\bar{y}^2 = \frac{26}{5}$$

$$n\bar{y}^2 = \frac{289}{5}$$
$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \frac{6}{10} = \frac{3}{5}$$
$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = \frac{17}{5} - \frac{18}{5} = -\frac{1}{5}$$
$$y = \hat{\alpha} + \hat{\beta}x = -\frac{1}{5} + \frac{3}{5}x$$



b. $x = 4.5$

$y = ?$

$$y = -\frac{1}{5} + \frac{3}{5} * \frac{9}{2} = -\frac{1}{5} + \frac{27}{10} = -\frac{2}{10} + \frac{27}{10} = \frac{25}{10}$$

Per 12000 non si può usare siccome è fuori dal range

c. $$SS_r = \frac{S_{xx}S_{YY} - S_{xY}^2}{S_{xx}} = \frac{47}{5}$$

$S_{xx} = 10$

$$S_{yy} = \frac{26}{5}$$

$S_{xy} = 6 = 36$

H_0: $\beta = 0$

H_1: $\beta \neq 0$

$$\left| \sqrt{\frac{S_{xx}(n-2)}{SS_r}}\,\hat{\beta} \right| > t_{n-2,\frac{\gamma}{2}}$$

$\gamma = 0.05$

$$\left|\sqrt{\frac{30}{\frac{47}{5}} * \frac{3}{5}}\right| > t_{3,0.025}$$

$t_{3,0.025} = 3.182$

$$30 * \frac{5}{47} * \frac{3}{5} = \frac{90}{47} \ngtr 3.18$$

Quindi rifiutiamo l'ipotesi

E quindi non è significiativo

5) N=100

$\bar{x}_n = 35$

$\sigma = 4$

a. Siccome livello confidenza 98%

$\alpha = 1 - 0.98 = 0.02$

Ci chiede intervallo per la media, quindi media incognita

$$\left(\bar{x}_n \pm z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right) = \left(35 \pm z_{0.01} * \frac{4}{10}\right)$$

Abbiamo notato un errore che abbiamo sempre fatto prima

Noi non abbiamo $z_{0.01}$

$\phi(z_{0.01}) = 1 - 0.01 = 0.99$

Quindi ora possiamo ricavare z

$$\phi(z_{0.01}) = \frac{2.32 + 2.33}{2} = 2.325$$

Siccome

$\phi(2.32) = 0.989, \phi(2.33) = 0.99$

$$\left(35 \pm 2.325 * \frac{4}{\sqrt{100}}\right)$$

b. Quanto deve essere numeroso il campione se si vuole ampliare l'intervallo di confidenza del 98% si dimezzi?

Aka se io voglio che quello sia il doppio, come deve cambiare n

Aka metto un *2

$$\left(35 \pm 2.325 * \frac{4}{\sqrt{100}} * \frac{1}{2}\right)$$

Ed ora lo porto dentro la radice

$$\left(35 \pm 2.325 * \frac{4}{\sqrt{100 * 2^2}}\right)$$

Quindi n deve quadruplicare

c. Quanti pasticcini occorre estrarre se si vuole che il peso medio discosti dal vero peso medio per meno di 1 grammo con probabilità 96%

$$\left(\bar{x}_n \pm z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}}\right)$$

Noi stiamo dicendo che

$$z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 1$$

Siccome facendo così x_n aumenterà di 1

Ed in più $\alpha = 0.04$

Quindi

$$z_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 1 \rightarrow 2.055 * \frac{4}{\sqrt{n}} = 1 \rightarrow 2.055 * 4 = \sqrt{n} \rightarrow \sqrt{n} = 8.22 \rightarrow n = 67.56$$

6) F

X	Y	X^2	Y^2	XY
20	195	400	38025	3900
25	190	625	38025	4750
29	188	841	35344	5452
31	185	961	34225	5735
45	163	2025	26569	7335

$$r = \sqrt{R^2}$$

$$y = \hat{\alpha} + \hat{\beta}x$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}}$$

$$S_{xx} = \sum x_i^2 - n\bar{x}^2$$

$$\bar{x} = 30$$

$$\bar{y} = \frac{921}{5}$$

$$n\bar{x}^2 = 5 * 900 = 4500$$

$$S_{xx} = 4852 - 4500 = 352$$

$$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 27172 - 27630 = -458$$

$$\hat{\beta} = \frac{-458}{352} = -1.30$$

921

$\alpha = \frac{}{5} + 1.30 * 30 = 223.2$
$y = 223.2 - 1.3x$

$h_0\text{: } \beta = 0$
$h_1\text{: } \beta \neq 0$
$\left| \sqrt{\frac{S_{xx}(n-2)}{SS_r}} \hat{\beta} \right| > t_{n-2,\frac{\gamma}{2}}$
$SS_r = * \, magic \;\; cami \, * = 18.9$

$\left| \sqrt{\frac{352 * 3}{18.9}} \widehat{1.30} \right| > t_{3,0.025}$
$8.5 > 3.3$
Quindi rifiuto l'ipotesi h_0

7)

X	Y	X^2	Y^2	XY
1	0.02	1	0.0004	0.02
2	0.03	4	0.0009	0.06
2.5	0.035	6.25	0.001225	0.0875
3	0.042	9	0.001764	0.126
3.5	0.05	12.25	0.0025	0.175
4	0.054	16	0.002916	0.216

$\bar{x} = 2.7$
$\bar{y} = 0.039$

$y = \alpha + \beta x$
$\alpha = \bar{y} - bx$
$\beta = S_{xy}/S_{xx}$
$S_{xx} = \sum x^2 - n\bar{x}^2 = 48.5 - 43.76 = 4.76$

$n\bar{x}^2 = 43.74$
(Errori di approssimazioni)

$S_{xy} = \sum x_i y_i - n\bar{x}\bar{y} = 0.68 - 0.63 = 0.05$

$S_{yy} = \sum y^2 - n\bar{y}^2 = 0.00081$

$$\beta = \frac{0.05}{4.76} = 0.0118$$
$\alpha = \cdots = 0.007$
$y = 0.007 + 0.0118 * x$

$S_{rr} = \cdots = 0.000007$
$h_0\colon\ \beta = 0$
$h_1\colon\ \beta \neq 0$
$$\left|\sqrt{\frac{S_{xx}(n-2)}{SS_r}}\beta\right| > t_{n-2,\frac{\alpha}{2}}$$
...
Si rifiuta, quindi è un buon modello
[Che esercizio di merda]