ESERCIZIO 8.1

Dall'osservazione di un campione di 8 abitazioni risulta la seguente relazione tra volume (in metri cubi) e spesa per riscaldamento (in Euro):

volume	240	300	180	320	270	150	200	420
spesa	80	90	70	100	85	65	80	120

a) Determinare la retta di regressione tra le due variabili;
b) misurare la bontà dell'adattamento;
c) sapendo che $s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 12{,}15$, verificare la significatività dei risultati al livello del 95%.

Soluzione

a)
I coefficienti di regressione vengono determinati attraverso le quantità riportate nella tabella:

	volume	spesa					
	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
	240	80	-20	-6,25	400	39,06	125
	300	90	40	3,75	1.600	14,06	150
	180	70	-80	-16,25	6.400	264,06	1.300
	320	100	60	13,75	3.600	189,06	825
	270	85	10	-1,25	100	1,56	-12,5
	150	65	-110	-21,25	12.100	451,56	2.337,5
	200	80	-60	-6,25	3.600	39,06	375
	420	120	160	33,75	25.600	1.139,06	5.400
totali	2.080	690	0	0	53.400	2.137,5	10.500
	$\bar{x} = 260$	$\bar{y} = 86{,}25$			$\sigma_x^2 = 6.675$	$\sigma_y^2 = 267{,}2$	$\sigma_{xy}^2 = 1312{,}5$

$$b = \frac{\sigma_{xy}^2}{\sigma_x^2} = \frac{1312{,}5}{6675} = 0{,}197$$

$$a = \bar{y} - b\bar{x} = 86{,}25 - 0{,}197 \times 260 = 35{,}12$$

L'equazione della retta di regressione è:

$$y = 35{,}12 + 0{,}197\ x$$

b)
La bontà dell'adattamento è misurata dall'indice di determinazione lineare, che può essere calcolato come:

$$R^2 = \rho^2 = \left(\frac{\sigma_{xy}^2}{\sigma_x \sigma_y}\right)^2 = \left(\frac{1.312,5}{\sqrt{6.675 \times 267,2}}\right)^2 = 0,97$$

oppure:

$$R^2 = \frac{dev_{reg}}{dev_Y} = 1 - \frac{dev_{RES}}{dev_Y} = 1 - \frac{(n-2)s^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2} = 1 - \frac{6 \times 12,15}{2137,5} = 0,97$$

c)
La significatività delle stime può essere valutata relativamente sia ai singoli coefficienti α e β sia all'indice di determinazione lineare R^2.

Significatività di α

Ipotesi:

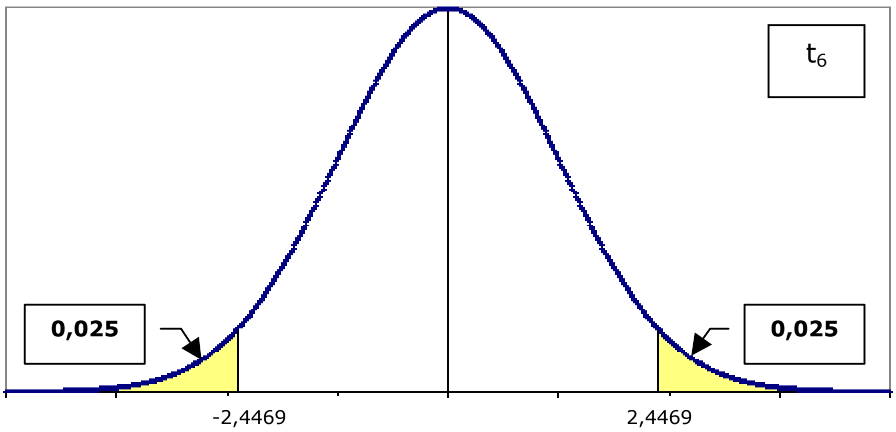
$H_0 : \alpha = 0$
$H_1 : \alpha \neq 0$

Statistica test:

$$X_{test} = \frac{a - \alpha}{s_a} \sim t_{n-2}$$

in cui $s_a^2 = \dfrac{s^2}{n}\left[1 + \dfrac{\left(\overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right]$

Valori critici:
$t_{0,025;6} = \pm\, 2,4469$



Regola di decisione:

$$-2{,}4469 \le x_{test} \le 2{,}4469 \quad \Rightarrow \quad \text{si accetta } H_0$$
$$x_{test} < -2{,}4469;\, x_{test} > 2{,}4469 \quad \Rightarrow \quad \text{si rifiuta } H_0$$

Valore test:

- $$s_a^2 = \frac{s^2}{n}\left[1 + \frac{\left(\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right] = \frac{12{,}15}{8} \times \left[1 + \frac{260^2}{53400}\right] = 1{,}52 \times 2{,}27 = 3{,}44$$

- $$X_{test} = \frac{a - \alpha}{s_a} = \frac{35{,}12}{\sqrt{3{,}44}} = 18{,}93$$

Decisione:

$x_{test} > 2{,}4469 \quad \Rightarrow$ **Rifiutiamo H_0**

Significatività di β

Ipotesi:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

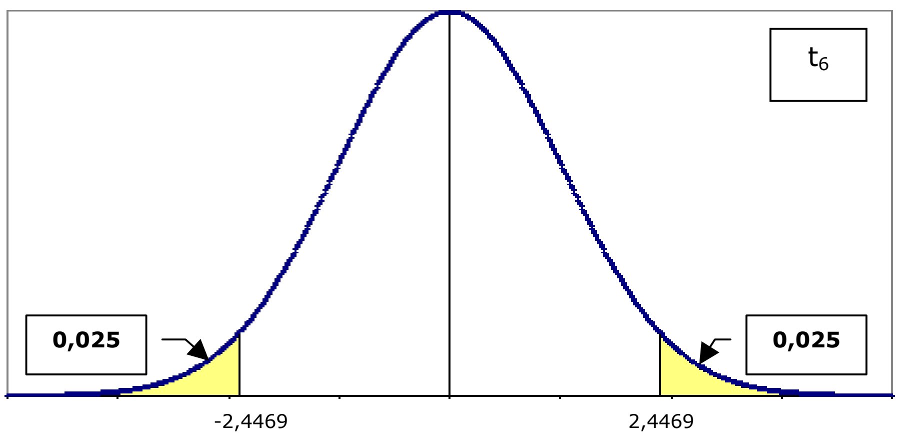
Statistica test:

$$X_{test} = \frac{b - \beta}{s_b} \sim t_{n-2}$$

in cui $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

Valori critici:

$t_{0{,}025;6} = \pm\, 2{,}4469$



Regola di decisione:

$$-2{,}4469 \le x_{test} \le 2{,}4469 \quad \Rightarrow \quad \text{si accetta } H_0$$
$$x_{test} < -2{,}4469;\, x_{test} > 2{,}4469 \quad \Rightarrow \quad \text{si rifiuta } H_0$$

Valore test:

- $$s_b^2 = s^2 \frac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \frac{12{,}15}{53.400} = 0{,}000228$$
- $$X_{test} = \frac{b-\beta}{s_b} = \frac{0{,}197-0}{\sqrt{0{,}000228}} = 13{,}04$$

Decisione:

$$x_{test} > 2{,}4469 \quad \Rightarrow$$ **Rifiutiamo H_0**

Significatività dell'indice R^2

Ipotesi:

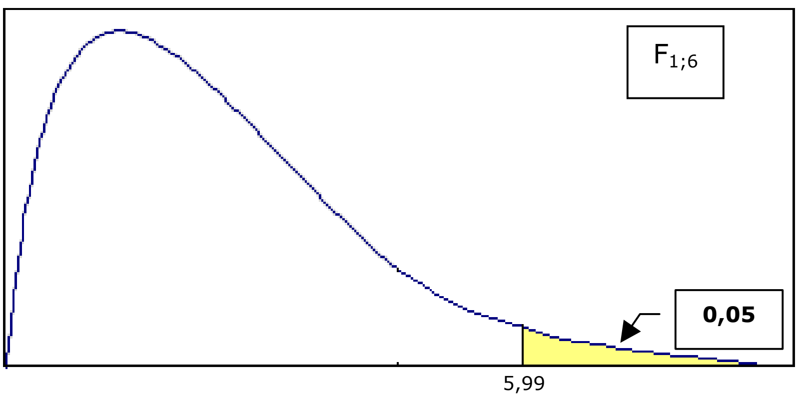
$H_0 : R^2 = 0$
$H_1 : R^2 > 0$

Statistica test:

$$X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} \sim F_{1;n-2}$$

Valore critico:

$F_{0,05;1;6} = 5{,}99$



Regola di decisione:

$x_{test} \leq 5,99 \quad \Rightarrow$ si accetta H_0

$x_{test} > 5,99 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $dev_{REG} = dev_Y - dev_{RES\_} = dev_Y - (n-2)\, s^2 = 2137,5 - (6 \times 12,5) = 2062,5$

- $X_{test} = \dfrac{dev_{REG}}{dev_{RES}/(n-2)} = \dfrac{2062,5}{12,15} = 169,75$

Decisione:

$x_{test} > 5,99 \quad \Rightarrow \quad \boxed{\textbf{Rifiutiamo } H_0}$

ESERCIZIO 8.2

A 10 operai di un'industria di automobili sono stati chiesti gli anni di lavoro (X) e lo stipendio percepito (Y) in un anno (in migliaia di Euro). Si sono ottenuti i seguenti risultati:

$\bar{x} = 11,3$ $\quad \bar{y} = 20,2$ $\quad codev_{XY} = 441,4$ $\quad dev_X = 346,1$ $\quad dev_Y = 593,6$

$dev_{REG} = 510,2$ $\quad s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 14,24$

a) Si stimi la retta di regressione tra X e Y;
b) si verifichi la significatività del modello al livello del 99%.

Soluzione

a)
I coefficienti di regressione vengono stimati come:

$$b = \frac{cod_{XY}}{dev_X} = \frac{441,4}{346,1} = 1,27$$

$$a = \bar{y} - b\bar{x} = 20,2 - 1,27 \times 11,3 = 5,85$$

La retta di regressione è definita dall'equazione:

$$y = 5,85 + 1,27\,x$$

b)

Significatività di α

Ipotesi:

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

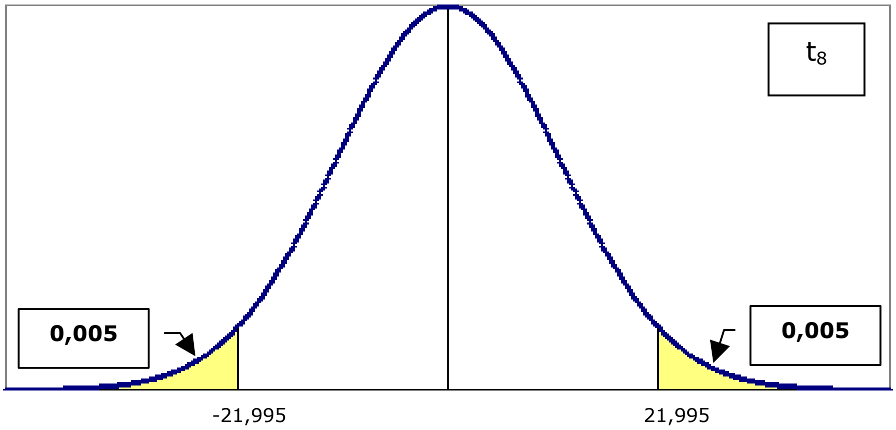
Statistica test:

$$X_{test} = \frac{a-\alpha}{s_a} \sim t_{n-2}$$

in cui $s_a^2 = \frac{s^2}{n}\left[1+\frac{\left(\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right]$

Valori critici:

$t_{0,005;8} = \pm\, 21,995$



Regola di decisione:

$-21,995 \le x_{test} \le 21,995 \quad \Rightarrow$ si accetta H_0

$x_{test} < -21,995; x_{test} > 21,995 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $s_a^2 = \frac{s^2}{n}\left[1+\frac{\left(\bar{x}\right)^2}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\right] = \frac{14,24}{10}\times\left[1+\frac{11,3^2}{346,1}\right] = 0,52$
- $X_{test} = \frac{a-\alpha}{s_a} = \frac{5,58}{\sqrt{0,52}} = 7,738$

Decisione:

$-21,995 \le x_{test} \le 21,995 \quad \Rightarrow$ **Accettiamo H_0**

Significatività di β

Ipotesi:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

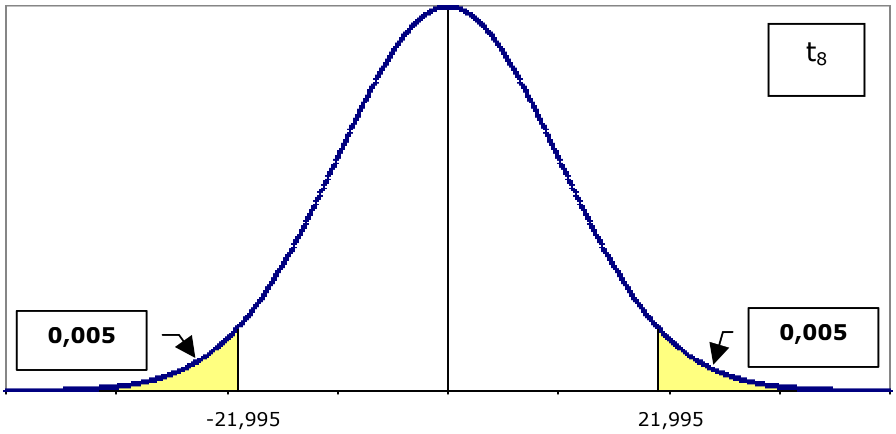
Statistica test:

$$X_{test} = \frac{b - \beta}{s_b} \sim t_{n-2}$$

in cui $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$

Valori critici:

$t_{0,005;8} = \pm\ 21,995$



Regola di decisione:

$$-21,995 \leq x_{test} \leq 21,995 \quad \Rightarrow \quad \text{si accetta } H_0$$
$$x_{test} < -21,995; x_{test} > 21,995 \quad \Rightarrow \quad \text{si rifiuta } H_0$$

Valore test:

- $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2} = \dfrac{14,24}{346,1} = 0,041$

- $X_{test} = \dfrac{b - \beta}{s_b} = \dfrac{1,27 - 0}{\sqrt{0,041}} = 6,27$

Decisione:

$-21,995 \leq x_{test} \leq 21,995 \quad \Rightarrow$ **Accettiamo H_0**

Significatività del modello

La bontà dell'adattamento del modello è misurata dall'indice di determinazione lineare:

$$R^2 = \frac{dev_{REG}}{dev_Y} = \frac{510,2}{593,6} = 0,86$$

Ipotesi:

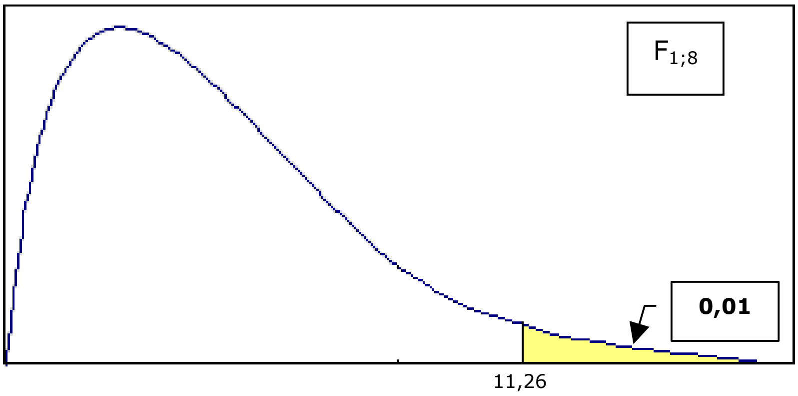
$H_0 : R^2 = 0$
$H_1 : R^2 > 0$

Statistica test:

$$X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} \sim F_{1;n-2}$$

Valore critico:

$F_{0,01;1;8} = 11,26$



Regola di decisione:

$x_{test} \le 11,26 \quad \Rightarrow$ si accetta H_0
$x_{test} > 11,26 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $dev_{RES} = dev_Y - dev_{REG} = 593,6 - 510,2 = 83,4$
- $X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} = \frac{510,2}{83,4/8} = 48,94$

Decisione:

$x_{test} > 11,26 \quad \Rightarrow$ **Rifiutiamo H_0**

ESERCIZIO 8.3

In una gara di nuoto sono stati rilevati l'altezza (in centimetri) di 7 nuotatori ed il tempo (in secondi) impiegato da ciascuno di essi per nuotare 100 metri a stile libero:

Tempo	57	79	62	77	67	73
Altezza	190	168	185	170	182	177

a) Stimare i coefficienti di regressione tra X e Y;
b) misurare la bontà dell'adattamento;
c) verificare la significatività dei risultati al livello del 99%, sapendo che $s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 = 3,08$.

Soluzione

a)
I coefficienti di regressione vengono stimati attraverso le quantità:

	Altezza	**Tempo**					
	x	y	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
	190	57	11,86	-12,29	140,59	150,94	-145,67
	168	79	-10,14	9,71	102,88	94,37	-98,53
	185	62	6,86	-7,29	47,02	53,08	-49,96
	170	77	-8,14	7,71	66,31	59,51	-62,82
	182	67	3,86	-2,29	14,88	5,22	-8,82
	177	73	-1,14	3,71	1,31	13,80	-4,24
	175	70	-3,14	0,71	9,88	0,51	-2,24
totali	**1.247**	**485**	**0,00**	**0,00**	**382,86**	**377,43**	**-372,29**
	$\bar{x}$	$\bar{y}$			σ_x^2	σ_y^2	σ_{xy}^2
	178,14	69,29			54,69	53,92	-53,18

$$b = \frac{\sigma_{xy}^2}{\sigma_x^2} = \frac{-46,54}{47,86} = -0,97$$

$$a = \bar{y} - b\bar{x} = 69,29 + 0,97 \times 178,14 = 242,51$$

L'equazione della retta di regressione è:

$$y = 242,51 - 0,97\ x$$

b)
La bontà dell'adattamento è misurata dall'indice di determinazione lineare, che può essere calcolato come:

$$R^2 = \rho^2 = \left(\frac{\sigma_{xy}^2}{\sigma_x \sigma_y}\right)^2 = \left(\frac{-53,18}{\sqrt{54,69 \times 53,92}}\right)^2 = 0,96$$

oppure:

$$R^2 = \frac{dev_{reg}}{dev_Y} = 1 - \frac{dev_{RES}}{dev_Y} = 1 - \frac{(n-2)s^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{5 \times 3,08}{377,43} = 0,96$$

c)
La significatività delle stime può essere valutata relativamente sia ai singoli coefficienti α e β sia all'indice di determinazione lineare R^2.

Significatività di α

Ipotesi:

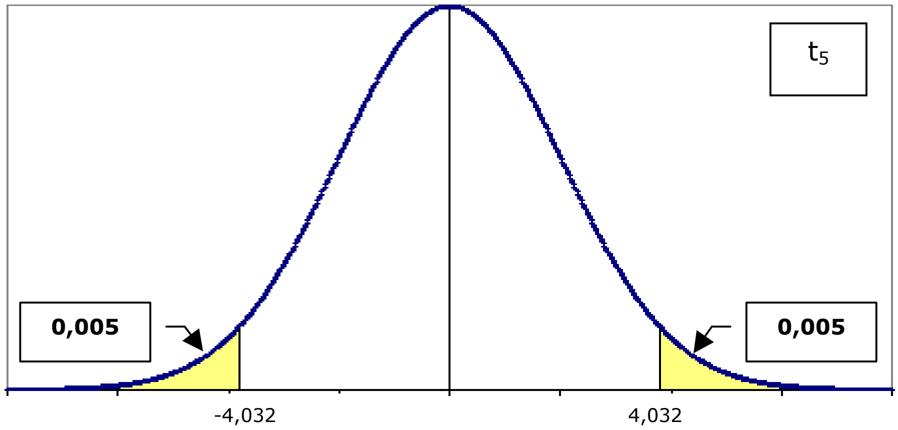
$H_0 : \alpha = 0$
$H_1 : \alpha \neq 0$

Statistica test:

$$X_{test} = \frac{a - \alpha}{s_a} \sim t_{n-2}$$

in cui $s_a^2 = \frac{s^2}{n}\left[1 + \frac{(\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$

Valori critici:
$t_{0,005;5} = \pm 4,032$



Regola di decisione:

$$-4{,}032 \le x_{test} \le 4{,}032 \quad \Rightarrow \text{ si accetta } H_0$$
$$x_{test} < -4{,}032;\, x_{test} > 4{,}032 \quad \Rightarrow \text{ si rifiuta } H_0$$

Valore test:

- $$s_a^2 = \frac{s^2}{n}\left[1 + \frac{(\bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = \frac{3{,}08}{7} \times \left[1 + \frac{178{,}14^2}{382{,}86}\right] = 36{,}97$$

- $$X_{test} = \frac{a - \alpha}{s_a} = \frac{242{,}5}{\sqrt{36{,}97}} = 39{,}88$$

Decisione:

$x_{test} > 4{,}032 \quad \Rightarrow$ **Rifiutiamo H_0**

Significatività di β

Ipotesi:

$H_0 : \beta = 0$
$H_1 : \beta \neq 0$

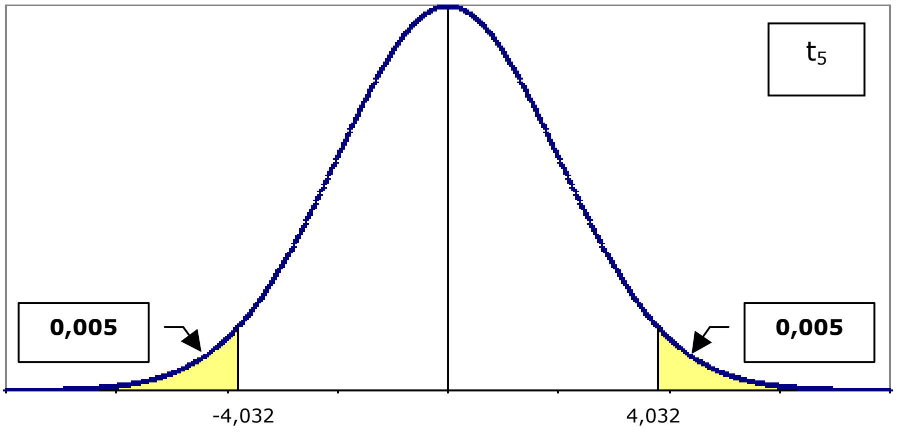
Statistica test:

$$X_{test} = \frac{b - \beta}{s_b} \sim t_{n-2}$$

in cui $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

Valori critici:

$t_{0,005;5} = \pm\, 4{,}032$



Regola di decisione:

$$-4{,}032 \leq x_{test} \leq 4{,}032 \quad \Rightarrow \quad \text{si accetta } H_0$$
$$x_{test} < -4{,}032;\, x_{test} > 4{,}032 \quad \Rightarrow \quad \text{si rifiuta } H_0$$

Valore test:

- $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \dfrac{3{,}08}{382{,}86} = 0{,}008$
- $X_{test} = \dfrac{b-\beta}{s_b} = \dfrac{-0{,}97-0}{\sqrt{0{,}008}} = -10{,}845$

Decisione:

$x_{test} < -4{,}032 \quad \Rightarrow$ **Rifiutiamo H_0**

Significatività dell'indice R^2

Ipotesi:

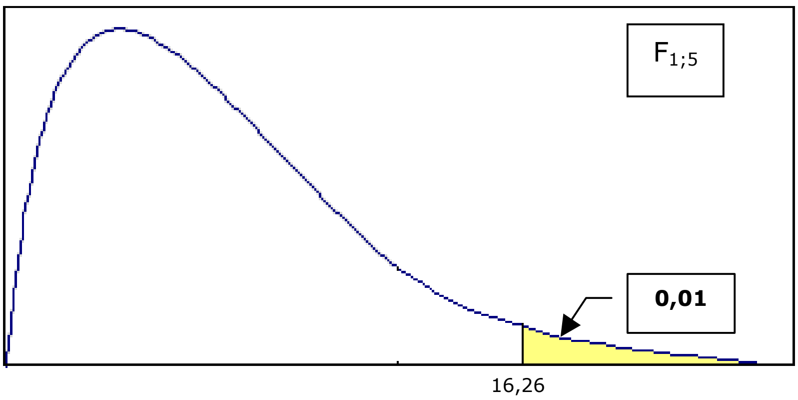
$H_0 : R^2 = 0$
$H_1 : R^2 > 0$

Statistica test:

$$X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} \sim F_{1;n-2}$$

Valore critico:

$F_{0,01;1;5} = 16{,}26$



Regola di decisione:

$x_{test} \leq 16,26 \quad \Rightarrow$ si accetta H_0
$x_{test} > 16,26 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $dev_{REG} = dev_Y - dev_{RES\_} = dev_Y - (n-2)\, s^2 = 377,43 - (5 \times 3,08) = 362,01$

- $X_{test} = \dfrac{dev_{REG}}{dev_{RES} / (n-2)} = \dfrac{362,01}{15,43} = 117,36$

Decisione:

$x_{test} > 16,26 \quad \Rightarrow$ $\boxed{\textbf{Rifiutiamo } H_0}$

ESERCIZIO 8.4

Si supponga di voler spiegare i secondi di ritardo di un aereo Y per mezzo del numero di passeggeri X, avendo calcolato le seguenti statistiche su 12 voli:

$$\sum_{i=1}^{n} x_i = 4.395; \quad \sum_{i=1}^{n} y_i = 5.940; \quad \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = 410.575;$$

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 149.156{,}25; \quad \sum_{i=1}^{n}(y_i - \bar{y})^2 = 1.251.500; \quad \sum_{i=1}^{n} e_i^2 = 121.330{,}59.$$

a) Si stimi la retta di regressione tra X e Y;
b) si verifichi la significatività del modello al livello del 95%.

<u>Soluzione</u>

a)

Innanzitutto calcoliamo le medie di x ed y:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{4.395}{12} = 366{,}25$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{5.940}{12} = 495.$$

A questo punto è possibile stimare i coefficienti di regressione come:

$$b = \frac{codev_{XY}}{dev_X} = \frac{410.575}{149.156{,}25} = 2{,}75$$

$$a = \bar{y} - b\bar{x} = 495 - 2{,}75 \times 366{,}25 = -513{,}16$$

La retta di regressione è definita dall'equazione:

$$y = -513{,}16 + 2{,}75\ x$$

b)

Significatività di α

Ipotesi:

$H_0 : \alpha = 0$

$H_1 : \alpha \neq 0$

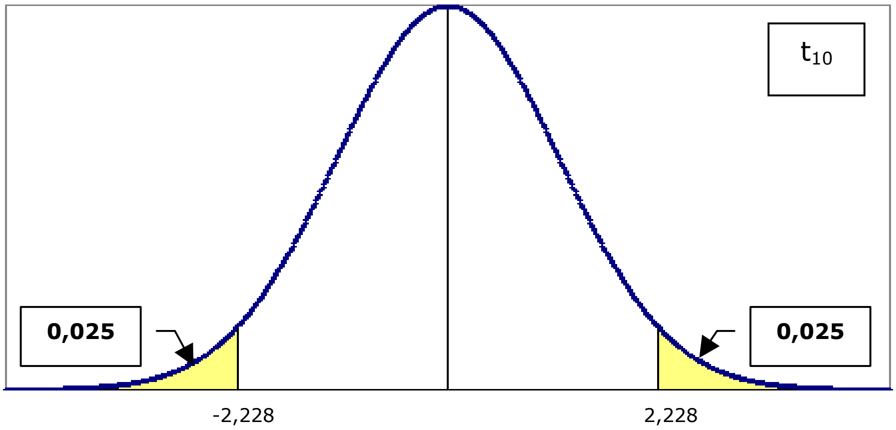
Statistica test:

$$X_{test} = \frac{a - \alpha}{s_a} \sim t_{n-2}$$

in cui $$s_a^2 = \frac{s^2}{n}\left[1 + \frac{\left(\overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right]$$

Valori critici:

$t_{0,025;10} = \pm\, 2{,}228$



Regola di decisione:

$-2{,}228 \le x_{test} \le 2{,}228 \quad \Rightarrow$ si accetta H_0

$x_{test} < -2{,}228;\ x_{test} > 2{,}228 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $$s_a^2 = \frac{s^2}{n}\left[1 + \frac{\left(\overline{x}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}\right] = \frac{12.133{,}06}{12} \times \left[1 + \frac{366{,}25^2}{149.156{,}25}\right] = 1.920{,}38$$
- $$X_{test} = \frac{a - \alpha}{s_a} = \frac{-513{,}16}{\sqrt{1.920{,}38}} = -11{,}71$$

Decisione:

$x_{test} < -2{,}228 \quad \Rightarrow$ **Rifiutiamo H_0**

Significatività di β

Ipotesi:

$H_0 : \beta = 0$

$H_1 : \beta \neq 0$

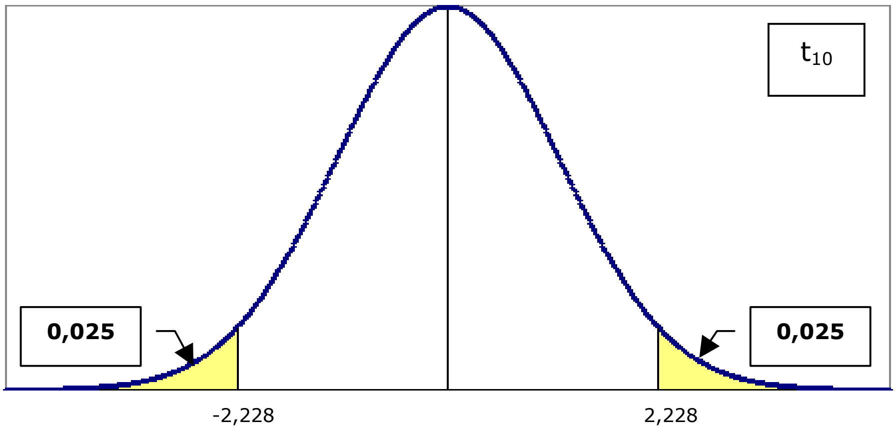
Statistica test:

$$X_{test} = \frac{b-\beta}{s_b} \sim t_{n-2}$$

in cui $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

Valori critici:

$t_{0,025;10} = \pm 2{,}228$



Regola di decisione:

$-2{,}228 \leq x_{test} \leq 2{,}228 \quad \Rightarrow$ si accetta H_0

$x_{test} < -2{,}228; x_{test} > 2{,}228 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $s_b^2 = s^2 \dfrac{1}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2} = \dfrac{12.133{,}06}{149.156{,}25} = 0{,}0813$

- $X_{test} = \dfrac{b-\beta}{s_b} = \dfrac{2{,}75-0}{\sqrt{0{,}0813}} = 9{,}65$

Decisione:

$x_{test} > 2{,}228 \quad \Rightarrow$ **Rifiutiamo H_0**

Significatività del modello

La bontà dell'adattamento del modello è misurata dall'indice di determinazione lineare:

$$R^2 = \frac{dev_{REG}}{dev_Y} = 1 - \frac{dev_{RES}}{dev_Y} = 1 - \frac{121.330,59}{12 * 104.291,67} = 0,9$$

Ipotesi:

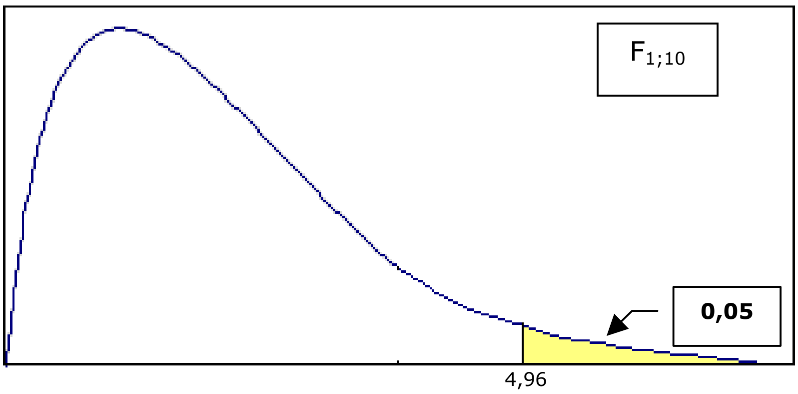
$H_0 : R^2 = 0$
$H_1 : R^2 > 0$

Statistica test:

$$X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} \sim F_{1;n-2}$$

Valore critico:

$F_{0,05;1;10} = 4,96$



Regola di decisione:

$x_{test} \leq 4,96 \quad \Rightarrow$ si accetta H_0
$x_{test} > 4,96 \quad \Rightarrow$ si rifiuta H_0

Valore test:

- $dev_{REG} = dev_Y - dev_{RES} = n * var_Y - n * s^2 = 1.251.500 - 121.330,59 = 1.130.169,41$

- $$X_{test} = \frac{dev_{REG}}{dev_{RES}/(n-2)} = \frac{1.130.169,41}{121.330,59/10} = 93,1479$$

Decisione:

$x_{test} > 4,96 \quad \Rightarrow$ **Rifiutiamo H_0**