

Esempio

In un paese è noto che la statura media degli abitanti è 175 cm. Sulla base di cambiamenti nello stile di vita un ricercatore ipotizza una sua variazione in 178 cm. Assumendo che la v.c. statura sia distribuita come una normale, $X \sim N(\mu, \sigma^2)$ con varianza nota ed utilizzando la statistica media campionaria consideriamo sotto l'ipotesi nulla la statistica test:

$$Z = \frac{\bar{X} - 175}{\sigma / \sqrt{n}} \quad N(0,1)$$

a seconda della ipotesi alternativa si hanno le seguenti regioni di rifiuto:

$$H_1 : \mu > 175 \quad Z \geq z_\alpha$$

$$H_1 : \mu < 175 \quad Z \leq -z_\alpha$$

$$H_1 : \mu \neq 175 \quad |Z| \geq z_{\alpha/2}$$

es. per $\sigma^2 = 300$ e per $n=20$ avessimo $\bar{x}=177,5$

$$z = \frac{177,5 - 175}{\sqrt{300 / 20}} = 0.645$$

con un livello $\alpha=0.05$ i valori critici sarebbero $z_\alpha=1.645$ e $z_{\alpha/2}=1.96$.

In nessuno dei casi si potrebbe rifiutare l'ipotesi nulla.

Test per la media - esempio

Test sulla statura media in un collettivo:

$$H_0 : \mu = 175 \text{ contro } H_1 : \mu > 175$$

Si assume che la statura sia una variabile casuale $X \sim N(\mu, \sigma^2)$ **con varianza ignota.**

Si estrae un campione di 10 giovani e si trova:

$$\bar{x} = 181,5$$

$$S^2 = 95,5067$$

$$t = \frac{181,5 - 175}{\sqrt{95,5067/10}} = 2,103$$

Ponendo $\alpha = 0,05$ si ottiene dalla t-Student con 9 gradi di libertà:

$$t_{0,05} = 1,8331$$

e quindi si rifiuterà l'ipotesi nulla poiché $t \geq 1,8331$

Test per una proporzione - esempio

Si vuole verificare che nel 2002 la percentuale degli occupati in Italia nel settore agricolo è la stessa del 1991 pari a 8,4%:

$$H_0 : \pi = 0,084 \text{ contro } H_1 : \pi \leq 0,084$$

Si estrae un campione di 1000 occupati.

Dei mille estratti, 53 sono occupati nel settore agricolo, pertanto:

Il valore della statistica test è: $\bar{x} = 0,053$

Ponendo $\alpha = 0,01$ si ottiene dalla normale standardizzata:

$$z = \frac{0,053 - 0,084}{\sqrt{(0,084)(0,916)/1000}} = -3,534 \qquad -z_{0,01} = -2,326$$

e quindi si rifiuterà l'ipotesi nulla poiché $z < -2,326$

Il p-value corrispondente al valore osservato della statistica test è 0,0002. Tale valore mostra una forte evidenza contro l'ipotesi nulla.

Test per la varianza - esempio

Si ipotizzi che la spesa delle famiglie sia una variabile casuale Normale. Si vuole verificare:

$$H_0 : \sigma^2 = 22500 \text{ contro } H_1 : \sigma^2 \neq 22500$$

Si estrae un campione di 61 famiglie.

Dal campione, si osservano: $\bar{x} = 2010$ $S^2 = 20000$

Il valore della statistica test è: $\frac{60 \cdot 20000}{22500} = 53,33$

Ponendo $\alpha = 0,05$ si ottiene dal Chi-quadrato con $n-1=60$ gradi di libertà:

$$\chi^2_{0,975} = 40,482 \qquad \chi^2_{0,025} = 83,298$$

quindi il valore osservato non cade nella regione di rifiuto.

Pertanto non si rifiuterà l'ipotesi nulla, ossia i dati osservati non sono difforni all'ipotesi che la varianza è pari a 22500.

Esercizio 5

Un docente vuole valutare gli effetti sul grado di preparazione dei propri studenti derivanti dall'adozione di un nuovo approccio didattico. A tal fine verifica se il voto medio conseguito all'esame finale dagli studenti frequentanti il corso è maggiore rispetto a quello degli studenti degli anni precedenti. Da un'analisi preliminare è noto che il voto all'esame conseguito dagli studenti degli anni precedenti è pari a 23,8. Relativamente ad un campione di 5 studenti che hanno seguito il corso secondo il nuovo approccio didattico, i voti conseguiti all'esame sono stati i seguenti:

23 25 28 26 27

Stabilire, assumendo un livello di significatività $\alpha=0.05$, se il nuovo approccio conduce ad un miglioramento del grado di preparazione degli studenti.

$$\mu_0 = 23,8$$

$$n = 5$$

$$\alpha = 0,05$$

$$\mu_1 = ?$$

23 25 28 26 27 $\rightarrow$ dobbiamo calcolare $\bar{x}$ e S^2

$$\bar{x} = \frac{23+25+28+26+27}{5} = 25,8$$

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2 = \frac{1}{4} (23-25,8)^2 + (25-25,8)^2 \dots = 3,7$$

TEST MONODIREZIONALE

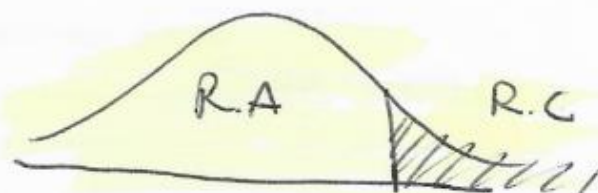
H_0 ipotesi nulla
 $\mu_1 = \mu_0$

accetteremo H_0 e accetteremo H_1
se $t > t_{\alpha}^{n-1}$

H_1 $\mu_1 > \mu_0$

Usiamo t perché σ ignota e $n < 30$

$$t = \frac{\bar{x} - \mu_0}{\sqrt{S/n}} = \frac{25,8 - 23,8}{\sqrt{\frac{3,7}{5}}} = \frac{2}{0,86} \approx 2,326$$



$$t_{\alpha}^{n-1} = t_{0,05}^4 = 2,132$$

$2,326 > 2,132$? Sì! quindi rifiuto H_0

Test per due campioni indipendenti - esempio

Si vuole comparare il tasso di occupazione π_1 dei laureati in Legge con il tasso di occupazione π_2 dei laureati in Economia:

$$H_0 : \pi_1 = \pi_2 \text{ contro } H_1 : \pi_1 < \pi_2$$

Sono estratti due campioni di persone laureate nell'anno precedente:

$$n_1 = 80 \text{ laureati in Legge} \quad n_2 = 120 \text{ laureati in Economia}$$

risultano rispettivamente 59 occupati nel primo campione e 93 nel secondo. Pertanto: $\bar{x}_1 = 0,738$ $\bar{x}_2 = 0,775$ $\bar{x}_p = 0,76$

$\bar{X}_p$ lo stimatore congiunto della proporzione:

$$\bar{X}_p = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$z = \frac{0,738 - 0,775}{\sqrt{0,76(1 - 0,76)(1/80 + 1/120)}} = -0,6$$

Ponendo $\alpha = 0,05$ si ottiene il valore della normale standardizzata:

$$-z_{0,05} = -1,645$$

e l'ipotesi nulla non può essere rifiutata.

mpione casuale di 83 diplomati di un noto istituto superiore della provincia di CS è stato se ritenessero adeguata la preparazione conseguita nel proprio istituto superiore di nza rispetto agli studi universitari intrapresi.

Indirizzo studi universitari	Preparazione adeguata	Preparazione Non adeguata	Totale
Scientifico	35	15	50
Non scientifico	28	5	33
Totale	63	20	83

- a) Applicare un metodo inferenziale per stabilire se l' istituto scolastico in questione è in grado o meno di garantire una preparazione adeguata rispetto ai due indirizzi di studi universitari;
- b) Commentare i risultati ottenuti.

TEST SUI PROPORZIONI TRA 2 CAMPIONI IID

$$H_0: \pi_1 \geq \pi_2$$

$$H_1: \pi_1 \neq \pi_2$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{X_p(1-X_p) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

X_p = STIMATORE CONGIUNTO DELLE PROPORZIONI

$$X_p = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2}{n_1 + n_2}$$

$$\bar{X}_1 = \frac{35}{50} = 0,7$$

$$n_1 = 50$$

$$n_2 = 33$$

$$\bar{X}_2 = \frac{28}{33} = 0,85$$

$$X_p = 0,76$$

$$\bar{X}_p = \frac{50(0,7) + 33(0,85)}{50 + 33}$$

$$Z = \frac{0,7 - 0,85}{\sqrt{0,76(0,24) \left(\frac{1}{50} + \frac{1}{33} \right)}} = \frac{-0,15}{0,096} = -1,56$$

$$|Z| > Z_{\frac{\alpha}{2}}$$

α LO FISSIAMO AL 0,05

$|1,56| = 1,56$? NO $\rightarrow$ NON RIFIAMO H_0

Test di indipendenza - esempio

In un'indagine di marketing (139 interviste) si è chiesto di indicare la preferenza tra 3 alimenti liquidi e 3 alimenti solidi da consumare al mattino a colazione:

	Biscotti	Merendina	Fette biscottate	Totale
Caffelatte	45	8	5	58
Tè	7	5	31	43
Succo di frutta	5	27	6	38
Totale	57	40	42	139

La statistica test vale: $\chi^2 = 101,57$ (calcolando l'indice chi-quadrato)

Ponendo $\alpha = 0,01$ il valore critico per la distribuzione Chi-quadrato con $(3-1)(3-1)=4$ gradi di libertà è

$$\chi^2_{0,01} = 13,227$$

Sirifiuta quindi l'ipotesi nulla di indipendenza tra i due caratteri.

Una nota società di mercato ha effettuato una indagine campionaria per verificare la propensione degli italiani al consumo di bevande alcoliche. Sono stati rilevati indipendentemente un campione casuale di 100 uomini e un campione casuale di 100 donne per i quali si è ottenuto:

<i>Attitudine al consumo di bevande alcoliche</i>	Maschi	Femmine		totale
SI	85	70		155
NO	15	30		45
totale	100	100		200

Verificare se la differenza nell' Attitudine al consumo di bevande alcoliche tra i due sessi è statisticamente significativa al livello $\alpha=0.02$.

<i>Attitudine al consumo di bevande alcoliche</i>	Maschi	Femmine	Totale
SI	85	70	155
NO	15	30	45
Totale	100	100	200

Calcolo delle n_{ij}^*

<i>Attitudine al consumo di bevande alcoliche</i>	Maschi	Femmine
SI	77,5	77,5
NO	22,5	22,5

$$\chi^2 = 6,45$$



$$\chi^2_{0,02} = 5,412$$

Valore critico del Chi-Quadro con 1 grado di libertà

Poiché $\chi^2 > \chi^2_{0,02}$ si può concludere che si rifiuta l'ipotesi di indipendenza tra l'attitudine al consumo di bevande alcoliche e il sesso del campione analizzato

Un'associazione per i diritti dei consumatori ha rilevato il prezzo di vendita per Kg in euro di un certo prodotto ortofrutticolo su diversi punti vendita presenti in una grande città. In particolare, su un campione casuale di 11 punti vendita situati in centro, sono stati ottenuti una media campionaria e varianza corretta rispettivamente di 1,2 e 0,06. Le analoghe statistiche campionarie calcolate su un campione di 21 punti vendita situati in periferia, sono risultate pari a 1,05 e 0,1. Supponendo che la variabile rilevata si distribuisca normalmente e svolgendo preventivamente se necessario il test di omoschedasticità verificare se la differenza tra prezzi medi risulti statisticamente significativa al livello $\alpha=0,01$.

$n_1 = 11$ $\bar{x}_1 = 1,2$ $S_1^2 = 0,06$ $\alpha = 0,01$
 $n_2 = 21$ $\bar{x}_2 = 1,05$ $S_2^2 = 0,1$

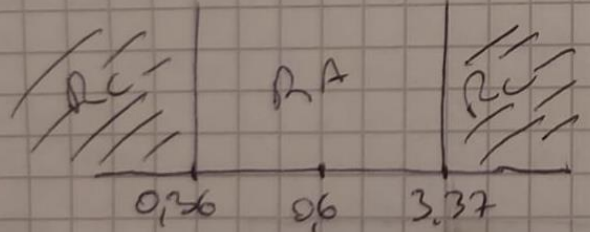
TEST OMOSEDASTICITÀ:

$$\frac{S_1^2}{S_2^2} = \frac{0,06}{0,1} = 0,6$$

$H_0: \sigma_1^2 = \sigma_2^2$ $H_2: \sigma_1^2 \neq \sigma_2^2$

R.e. $F > F_{\frac{\alpha}{2}} = F_{0,005}^{10,20} = 3,37$

$F \leq F_{1-\frac{\alpha}{2}} = F_{0,995}^{10,20} = \frac{1}{F_{0,005}^{20,10}} = \frac{1}{2,77} = 0,36$



NON RIFIUTO H_0

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

UTILIZZIAMO CASO IN
CUI LE VARIANZE
SONO CONSIDERATE UGUALI
NON AVENDO RIFERIMENTO
LE POSE DI DI OROSCOPF BASINCA

$$S_p = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

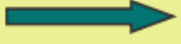
$$S_p = \frac{(10)(0,06) + 20(0,1)}{30} = 0,087$$

$$T = \frac{1,2 - 1,05}{\sqrt{0,087 \left(\frac{1}{11} + \frac{1}{21} \right)}} = \frac{0,15}{0,11} = 1,36$$

$$R.C. = |t|, t \in \frac{1}{2} n_1 + n_2 - 2$$

$$|1,36|, t \in 0,0130 = 2,25 \quad \text{No} \Rightarrow \text{NON RIFIUTO } H_0$$

L'approccio del p-value alla verifica di ipotesi

Per l'ampia diffusione dei pacchetti statistici è venuto affermandosi un altro modo per evidenziare il risultato del test: quello di calcolare il **p-value**. Poiché il valore di α è arbitrario una ipotesi rifiutata al livello $\alpha=0.1$ potrebbe invece essere accettata con $\alpha=0.01$ 

p-value : rappresenta la probabilità di osservare un valore della statistica test uguale o più estremo del valore ottenuto a partire dal campione, quando l'ipotesi nulla H_0 è vera.

Viene anche chiamato **livello di significatività osservato**, poiché coincide con **il più piccolo livello di significatività in corrispondenza del quale H_0 è rifiutata**. Non è una quantità fissata come α , ma che misura l'evidenza fornita dai dati contro l'ipotesi nulla: minore è il valore del p-value, più è forte l'evidenza contro l'ipotesi nulla.

In pratica la regola decisionale:

- se $\text{p-value} \geq \alpha$ non rifiuto H_0
- se $\text{p-value} < \alpha$ rifiuto H_0

Test per una proporzione - esempio

Si vuole verificare che nel 2002 la percentuale degli occupati in Italia nel settore agricolo è la stessa del 1991 pari a 8,4%:

$$H_0 : \pi = 0,084 \text{ contro } H_1 : \pi \neq 0,084$$

Si estrae un campione di 1000 occupati.

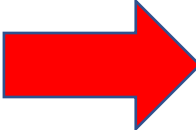
Dei mille estratti, 53 sono occupati nel settore agricolo, pertanto:

Il valore della statistica test è: $\bar{x} = 0,053$

Ponendo $\alpha = 0,01$ si ottiene dalla normale standardizzata:

$$z = \frac{0,053 - 0,084}{\sqrt{(0,084)(0,916)/1000}} = -3,534 \qquad -z_{0,01} = -2,326$$

e quindi si rifiuterà l'ipotesi nulla poiché $z < -2,326$

 **Il p-value corrispondente al valore osservato della statistica test è 0,0002. Tale valore mostra una forte evidenza contro l'ipotesi nulla.**

Esempio

In un paese è noto che la statura media degli abitanti è 175 cm. Sulla base di cambiamenti nello stile di vita un ricercatore ipotizza una sua variazione in 178 cm. Assumendo che la v.c. statura sia distribuita come una normale, $X \sim N(\mu, \sigma^2)$ con varianza nota ed utilizzando la statistica media campionaria consideriamo sotto l'ipotesi nulla la statistica test:

$$Z = \frac{\bar{X} - 175}{\sigma / \sqrt{n}} \sim N(0,1)$$

a seconda della ipotesi alternativa si hanno le seguenti regioni di rifiuto:

$$H_1: \mu > 175 \quad Z \geq z_\alpha$$

$$H_1: \mu < 175 \quad Z \leq -z_\alpha$$

$$H_1: \mu \neq 175 \quad |Z| \geq z_{\alpha/2}$$

es. per $\sigma^2 = 300$ e per $n=20$ avessimo $\bar{x}=177,5$

$$z = \frac{177,5 - 175}{\sqrt{300/20}} = 0.645$$

con un livello $\alpha=0.05$ i valori critici sarebbero $z_\alpha=1.645$ e $z_{\alpha/2}=1.96$.

In nessuno dei casi si potrebbe rifiutare l'ipotesi nulla.

Nell'esempio precedente piuttosto che rifiutare o accettare l'ipotesi che $\mu=175$ ad un prefissato valore di α è preferibile riportare un *p-value* che misuri l'evidenza contro l'ipotesi H_0 .

Poiché $z=0.645$ se l'ipotesi alternativa fosse $H_1: \mu > 175$ risulterebbe:

$$P\text{-value} = p(Z > 0.645) = 1 - \Phi(0.645) = 0.26$$

Non c'è una forte evidenza nei dati contro l'ipotesi nulla: poiché il test è unilaterale l'ipotesi nulla sarà rifiutata per valori di α maggiori di 0.26!

Vantaggio: il p-value può essere confrontato con diversi possibili valori di α

Test sulla statura media in un collettivo:

$$H_0 : \mu = 175 \text{ contro } H_1 : \mu > 175$$

Si assume che la statura sia una variabile casuale $X \sim N(\mu, \sigma^2)$ con varianza ignota.

Si estrae un campione di 10 giovani e si trova:

$$\bar{x} = 181,5$$

$$S^2 = 95,5067$$

$$t = \frac{181,5 - 175}{\sqrt{95,5067/10}} = 2,103$$

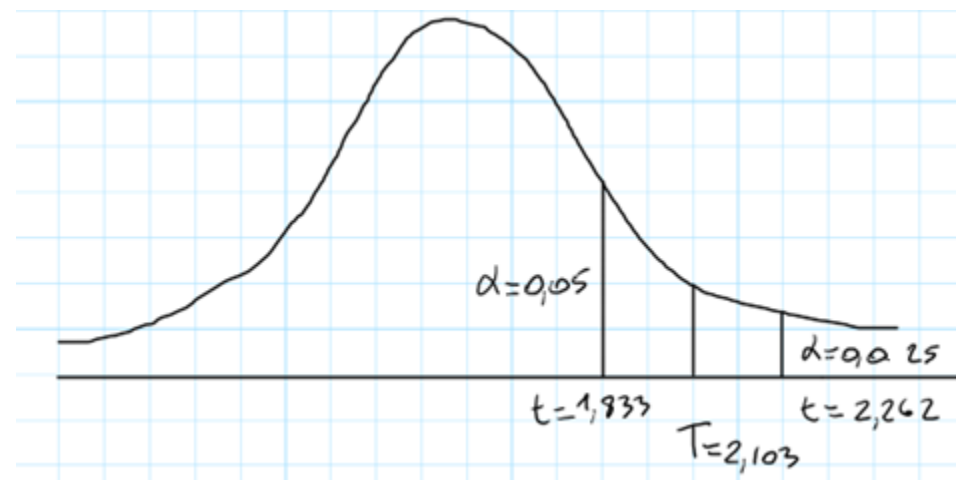
Ponendo $\alpha = 0,05$ si ottiene dalla t-Student con 9 gradi di libertà:

$$t_{0,05} = 1,8331$$

e quindi si rifiuterà l'ipotesi nulla poiché $t \geq 1,8331$.

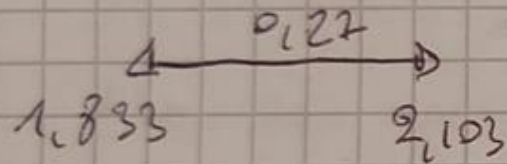
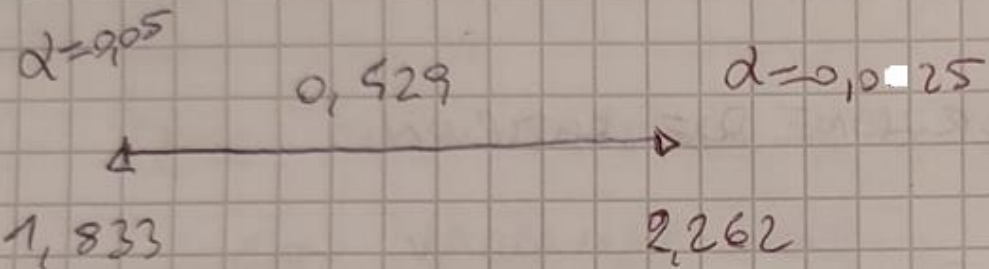
Il p-value corrispondente al valore osservato della statistica test è $P(T \geq 2,103 / H_0 \text{ è vera}) = 0,0324$. $\implies$ forte evidenza empirica contro l'ipotesi nulla la quale potrebbe essere accettata soltanto ad un livello di significatività α minore di 0.0324

$$\text{p-value} = P(T > t_{oss}) = P(T > 2,103)$$



Per un livello di significatività $\alpha = 0,05$, dalle tavole statistiche, troviamo il valore critico

$$t_{\frac{\alpha}{2}, n-1} = t_{\frac{0,05}{2}, 10-1} = t_{0,025, 9} = 2,262$$



P-VAWE

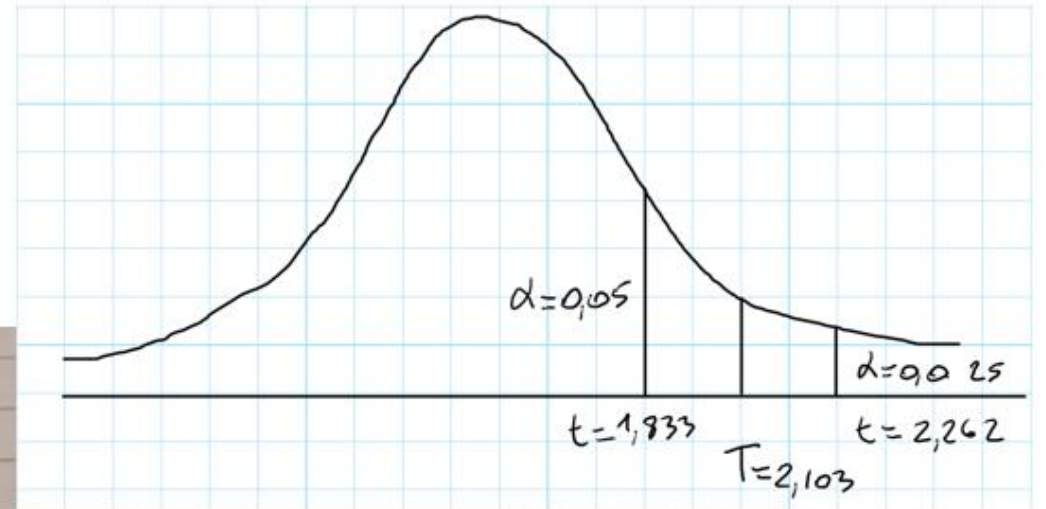
0,27 = 63% delle distanze tra
1,833 e 2,262

0,05 e 0,025

$$\frac{0,27}{0,429} = 0,63 \Rightarrow 63\%$$

ovvero $63\% \cdot 0,025 = 0,01575$

$$PVAWE = 0,05 - 0,01575 = 0,03425$$



REGRESSIONE E INFERENZA SULLA REGRESSIONE

La tabella seguente riporta 14 osservazioni sulla resistenza a trazione della carta (R), espressa in newton per centimetro quadrato e la percentuale di fibra di legno (F) presente nella pasta con la quale la carta è prodotta:

R	4.3	7.6	13.8	16.6	18.0	20.7	23.3	23.4	26.3	27.5	29.0	31.8	36.6	35.9
F	1.0	1.5	2.0	3.0	4.0	4.5	5.0	5.5	6.0	6.5	7.0	8.0	9.0	10.0

- Determinare la retta di regressione stimata che spiega la resistenza a trazione della carta in funzione della percentuale di fibra di legno;
- Calcolare l'indice di determinazione;
- Verificare a livello di significatività del 5% se il coefficiente di pendenza sia nullo contro l'ipotesi alternativa che sia crescente, successivamente, verificare al livello di significatività del 5% se la retta di regressione passa per l'origine.

ESERCIZIO SULLA REGRESSIONE

1) DETERMINARE LA RETTA DI REGRESSIONE

$$Y = \beta_0 + \beta_1 X$$

$$\hat{\beta}_1 = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

Y : VARIABILE DIPENDENTE (Y) = RESISTENZA A TAZIONE DEL CARRA

X : VARIABILE INDIPENDENTE (X) = PERCENTUALE DI FIBRA DI LENO

$$\bar{X} = 5,21 \quad \bar{Y} = 22,49$$

$$\begin{aligned} \text{Cov}(X, Y) &= \frac{1}{n} \sum (x_i - \mu_X)(y_i - \mu_Y) \Rightarrow \frac{1}{n} \sum (x_i y_i) - \mu_X \mu_Y \\ &= \frac{1985,8}{14} - (5,21)(22,49) = 24,67 \end{aligned}$$

$$\text{Var}(X) = \frac{1}{n} \sum (x_i - \mu_X)^2 = \frac{1}{n} \sum x_i^2 - \mu_X^2 = \frac{480}{14} - (5,21)^2 = 7,09$$

$$\hat{\beta}_1 = \frac{24,67}{7,09} = 3,47$$

$$\hat{\beta}_0 = 22,49 - 3,47(5,21) = 4,41$$

$$\hat{Y} = 4,41 + 3,47 X$$

5) INDICE DI DETERMINAZIONE

$$R^2 = [\rho(x, y)]^2$$

$$[\rho(x, y)]^2 = \left[\frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \right]^2 = \frac{24,67}{(2,67)(9,39)} = (0,98)^2 = 0,97$$

$$\sigma_x^2 = 7,09 \Rightarrow \sigma_x = 2,67$$

$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \mu_y)^2 = \frac{1}{14} \sum y_i^2 - \mu_y^2 = \frac{1}{14} 8313,74 - (22,49)^2 = 88,23$$

$$\Rightarrow \sigma_y = 9,39$$

$$0 \leq R^2 \leq 1$$

5) INDICE DI DETERMINAZIONE

$$R^2 = [\rho(x, y)]^2$$

$$[\rho(x, y)]^2 = \left[\frac{\text{cov}(x, y)}{\sigma_x \sigma_y} \right]^2 = \frac{24,67}{(2,67)(9,39)} = (0,98)^2 = 0,97$$

$$\sigma_x^2 = 7,09 \Rightarrow \sigma_x = 2,67$$

$$\sigma_y^2 = \frac{1}{n} \sum (y_i - \mu_y)^2 = \frac{1}{14} \sum y_i^2 - \mu_y^2 = \frac{1}{14} 8313,74 - (22,49)^2 = 88,23$$

$$\Rightarrow \sigma_y = 9,39$$

$$0 \leq R^2 \leq 1$$

c) TEST SU $\hat{\beta}_1$ PENDENZA DELLA RETTA O INCLINAZIONE

$$H_0: \hat{\beta}_1 = 0$$

$$H_1: \hat{\beta}_1 > 0$$

$$t = \frac{\hat{\beta}_1}{S(\hat{\beta}_1)} \sim t_{n-2}$$

BISOGNA CALCOLARE $S(\hat{\beta}_1)$

$$S^2 = \frac{\sum e_i^2}{n-2} = \frac{43,02}{12} = 3,6$$

$$\begin{aligned}\sum e_i^2 &= \sum y_i^2 - n\bar{y}^2 - \hat{\beta}_1^2 (\sum x_i^2 - n\bar{x}^2) \\ &= 8313,74 - 14(505,6) - 12(99,36) = \\ &= 8313,74 - 7078,4 - 1192,32 = 43,02\end{aligned}$$

$$S(\hat{\beta}_1) = \sqrt{\frac{S^2}{n s_x^2}} = \sqrt{\frac{3,6}{99,36}} = 0,19$$

$$S(\beta_0) = \sqrt{s^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{n s_x^2} \right]} = \sqrt{3,6 \left[\frac{1}{14} + \frac{5,21^2}{99,36} \right]} = 1,11$$

$$t = \frac{3,57}{0,19} = 18,26$$

$$R_{\alpha} = \left\{ t > t_{\frac{\alpha}{2}, n-2} \right\} \Rightarrow \left\{ 18,26 > 1,782 \right\}$$

$$\downarrow$$

$$t_{0,05}^{12} = 1,782$$

REFUSO H_0

Se $H_0: \hat{\beta}_1 = 0$ $H_1: \hat{\beta}_1 \neq 0$ $R_{\alpha} = \left\{ |t| > t_{\frac{\alpha}{2}, n-2} \right\}$

TEST su $\hat{\beta}_0 \Rightarrow t = \frac{\hat{\beta}_0}{S(\hat{\beta}_0)} = \frac{4,41}{1,11} \approx 4$

$$R_{\alpha} = \left\{ |t| > t_{\frac{\alpha}{2}, n-2} \right\} \Rightarrow \left\{ |4| > t_{0,025}^{12} \right\}$$

$4 > 2,179 \Rightarrow$ RIFIUTO H_0 , LA RETTA

NON PASSA PER L'ORIGINE

Per un certo appartamento scelto casualmente sono stati rilevati per 26 settimane consecutive, i dati relativamente alla temperatura esterna in gradi centigradi ed il consumo di gas per il riscaldamento in decimetri cubi. Indicando con x_i le osservazioni sulla temperatura esterna e con y_i quelle sul consumo di gas per il riscaldamento si ha:

$$\sum x_i = 139,100; \sum y_i = 3496,285; \sum x_i^2 = 950,510; \sum y_i^2 = 497247,408; \sum x_i y_i = 16408,193$$

Si vuole spiegare mediante un modello di regressione lineare, il consumo del gas in funzione della temperatura esterna.

- a) Stimare la retta di regressione e commentare il risultato;
- b) Calcolare il coefficiente di determinazione e commentare il risultato;
- c) Sottoporre a test d'ipotesi che il consumo del gas non dipenda dalla temperatura esterna e determinare il p-value del test;
- d) Si costruisca la tavola ANOVA, si commentino i risultati e si confrontino con quelli ottenuti nel punto precedente

$$n=26 \quad \sum x_i = 139,100 \quad \sum y_i = 3496,205 \quad \sum x_i^2 = 950,510 \quad \sum y_i^2 = 447247,406$$

$$\sum x_i y_i = 16408,193$$

c) Stimare retta di regressione

$$\beta_1 = \frac{\text{cov } xy}{\text{var } x} = \frac{\frac{\sum xy}{n} - \bar{x} \bar{y}}{\frac{\sum x_i^2}{n} - \bar{x}^2} = \frac{\frac{16408,193}{26} - (5,35 \cdot 134,47)}{\frac{950,510}{26} - 5,35^2} = \frac{-88,33}{7,94} = -11,12$$

$$\bar{x} = \frac{\sum x_i}{n} = 5,35$$

$$\bar{y} = \frac{\sum y_i}{n} = 134,47$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x} = 134,47 - (-11,12) (5,35) = 143,96$$

b) Calcular R^2

$$R^2 = \left(\frac{\text{cov}_{xy}}{\sigma_x \sigma_y} \right)^2 = \begin{cases} \text{cov}_{xy} = -88,33 \\ \sigma_x = 2,82 \\ \sigma_y^2 = \frac{\sum y_i^2}{n} - \bar{y}^2 = 1042,72 \rightarrow \sigma_y = 32,30 \end{cases}$$

$$R^2 = \left(\frac{-88,33}{2,82 \cdot 32,30} \right)^2 = 0,94$$

c) ~~Calcular~~ $s^2 = \frac{\sum y_i^2 - n\bar{y}^2 - \hat{\beta}_1^2 (\sum x_i^2 - n\bar{x}^2)}{n-2} = \frac{n\sigma_y^2 - \hat{\beta}_1^2 n\sigma_x^2}{n-2}$

$$= \frac{(26 \cdot 1042,72) - (-11,12)^2 (26 \cdot 7,94)}{24} = 65,98$$

$$S(\hat{\beta}_1) = \sqrt{\frac{s^2}{n\sigma_x^2}} = \sqrt{\frac{65,98}{26 \cdot 7,94}} = 0,56$$

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 < 0$$

$$R.c \quad t \leq -t_{\alpha}^{n-2}$$

$$t_1 = \frac{-11,12}{0,56} = -19,86$$

$$t_{0,05}^{24} = 1,719$$

$t_1 \leq t_{\alpha}^{n-2}$ quindi rifiutiamo H_0
il cui valore dipende da T

$$p\text{-value } P(t \leq -19,86) \text{ dove } P(t \geq 19,86) = p\text{-value} < 0,0005$$

tante o 0

d) costruire la ANOVA

Somma dei quadrati	q.d.f.	Media dei quadrati	F
SQR	1	$MQR = SQR/1$	$F = \frac{MQR}{MRE}$
SQE	n-2	$MRE = SQE/n-2$	
SQT	n-1		

$$SQT = \sum (y_i - \bar{y})^2 = n s^2 y = 27110,72$$

$$R^2 = \frac{SQR}{SQT} \rightarrow SQR = R^2 SQT = 25484,08 \rightarrow MQR$$

$$SQE = SQT - SQR = 1626,64 \rightarrow MRE = \frac{1626,64}{24} = 67,78$$

$$F = \frac{25484,08}{67,78} = 374,80 \sim F_{(1;24)}$$

R.C. $F > F_{0,05}^{1,24}$

Test $H_0: \beta_1 = 0$
 $H_1: \beta_1 < 0$

$$F > F_{\alpha}^{1;n-2} \rightarrow 4,26 \text{ rifiuto } H_0$$



Un ricercatore che lavora alla dipendenza di una industria produttrice di lampadine elettriche afferma di aver trovato un nuovo tipo di filamento che prolunga la durata delle lampadine. Poiché il nuovo filamento è considerevolmente molto più costoso di quello attualmente in uso, l'industria intende, prima di adottarlo, aver il confronto di una verifica sperimentale. Viene allora formulata l'ipotesi nulla di uguale durata media tra lampadine contro l'ipotesi alternativa che la durata media delle lampadine del vecchio filamento siano inferiori rispetto al nuovo. Per verificare l'ipotesi nulla, vengono osservati due campioni dei due tipi di lampadine entrambi di ampiezza 31. Le medie e le varianze corrette dei due campioni risultano essere:

$$\bar{x}_1 = 1.180,05$$

$$\bar{x}_2 = 1.195,16$$

$$s_1^2 = 124,34$$

$$s_2^2 = 118,13$$

Ammettendo che le varianze delle popolazioni di origine dei due campioni siano uguali, verificare al livello di significatività $\alpha=0,01$ se la qualità del vecchio filamento sia inferiore rispetto al nuovo.



Si vuole verificare l'ipotesi che il daltonismo sia indipendente dal sesso al livello di significatività del 1%, i dati ricavati da un'indagine hanno prodotto i seguenti risultati.

	Maschi	Femmine	Totale
Normali	654	771	1425
Daltonici	66	9	75
Totale	720	780	1500



Un produttore di fili di rame per impieghi elettronici assicura che, in media, la percentuale di impurità presente nel suo prodotto è pari al 5%. La percentuale di impurità ha una distribuzione normale con scarto quadratico medio di 0,033 in un campione di 138 fili di rame, la media campionaria è risultata $\bar{x} = 0,054$.

- a) Verificare l'affermazione del produttore al livello di significatività del 10%;
- b) Verificare l'esito del test al livello di significatività del 5%;
- c) Determinare il p-value



Per un campione casuale di 120 studenti dell'Università di Firenze si rileva un'età media di 20,2 anni mentre per un campione casuale di 100 studenti dell'Università di Roma il valore riscontrato è 21 anni. Nell'ipotesi che le due popolazioni di riferimento di distribuiscano Normalmente con varianze, rispettivamente, pari a 9,44 e 10,25 si verifichi ad un livello di significatività $\alpha=0,05$ l'uguaglianza nell'età tra gli studenti dei due Atenei.