

- L'incertezza va arrotondata alla 1° cifra significativa.
- **N.B.** Se la 1° cifra significativa è 1, allora si può arrotondare alla 2° cifra significativa.
- La misura deve avere l'ultima cifra significativa dello stesso ordine di grandezza dell'incertezza.

• MISURE $\left\{ \begin{array}{l} \text{DIPENDENTI: } \frac{\delta A}{A} = \frac{\delta B}{B} + \frac{\delta C}{C} \\ \text{INDIPENDENTI: } \frac{\delta A}{|A|} = \sqrt{\left(\frac{\delta B}{B}\right)^2 + \left(\frac{\delta C}{C}\right)^2} \end{array} \right.$

- **MIGLIOR STIMA:** valore centrale delle varie misure

- Parametri statistici caratteristici del campione $\left\{ \begin{array}{l} \text{indici di posizione centrale (media, moda, mediana)} \\ \text{indici di dispersione (varianza, dispersione standard)} \end{array} \right.$

- **MEDIA:** $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ per i singoli dati

$$\bar{x} = \sum_{i=1}^n \frac{x_i^* \cdot f_i}{n} = \sum_{i=1}^n x_i^* p_i \quad \text{dopo la suddivisione in classi}$$

percentuale di occorrenza/frequenza relativa

$$x_i^* = \frac{x_A + x_B}{2}$$

- **MODA:** valore con la massima frequenza di occorrenza
- **MEDIANA:** valore che divide in due parti uguali l'istogramma o i dati.

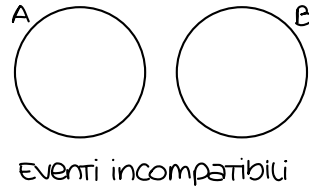
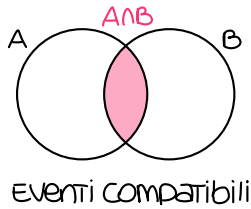
- **VARIANZA:** $s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$ per i singoli dati ($x_i - \bar{x}$)²
- $s^2 = \frac{\sum_{i=1}^n f_i (x_i^* - \bar{x}^*)^2}{n-1}$ dopo la suddivisione in classi

- **DISPERSIONE STANDARD:** $s = \sqrt{s^2}$

- **SKEWNESS:** $\text{skew} = \frac{3(\bar{x} - x_{\text{mediana}})}{\sigma}$
- $\left\{ \begin{array}{l} > 0 \quad \text{coda a destra} \quad \curvearrowright \\ = 0 \quad \text{simmetria} \\ < 0 \quad \text{coda a sinistra} \quad \curvearrowleft \end{array} \right.$

- **PROBABILITÀ:** $p(E) = \frac{n \text{ casi favorevoli}}{n \text{ casi possibili}}$

- **TEOREMA DELLA PROBABILITÀ TOTALE:** $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



- **TEOREMA DELLA PROBABILITÀ COMPOSTA:** $P(A \cap B) = P(A) \cdot P(B|A)$

Eventi indipendenti: $P(B|A) = P(B)$

Eventi dipendenti: $P(B|A) \neq P(B)$

- **VARIABILE CASUALE:** definisce un risultato ALEATORIO (esito sconosciuto a priori)

- **VARIABILE**
 - DISCRETA:** può assumere solo un set di valore assegnato [lancio di un dado, estrazione carte, ...]
 - CONTINUA:** può assumere un set infinito di valori [lunghezza di un tavolo, ...]

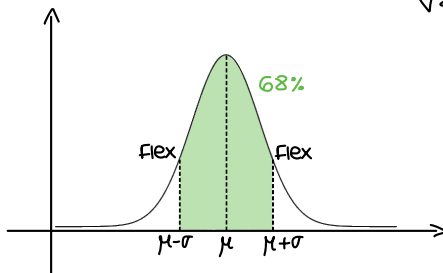
| | DISCRETA | CONTINUA |
|------------------------------|---------------------------------------|--|
| PROBABILITÀ DI IMPEDIMENTO | $Pr(x = x_i) = p(x)$ | $Pr(x \leq x_i) = P(x)$ |
| MISURA DI POSIZIONE CENTRALE | $E(x) = \sum x_i p(x_i)$ | $E(x) = \int_{-\infty}^{+\infty} x p(x) dx$ |
| VARIANZA | $V(x) = \sum [(x_i - E(x))^2 p(x_i)]$ | $V(x) = \int_{-\infty}^{+\infty} (x - E(x))^2 p(x) dx$ |

$$V(x) = E(x^2) - [E(x)]^2$$

- **PROPRIETÀ DEL VALORE ATTESO**

- $E(A \cdot x) = A \cdot E(x)$
- $E(A + x) = A + E(x)$
- $E(A + Bx) = A + BE(x)$

- **DISTRIBUZIONE NORMALE:** $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$



$$x \sim N(\mu; \sigma^2)$$

- **DISTRIBUZIONE NORMALE STANDARD:** $p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$

$$z = \frac{x - \mu}{\sigma}$$

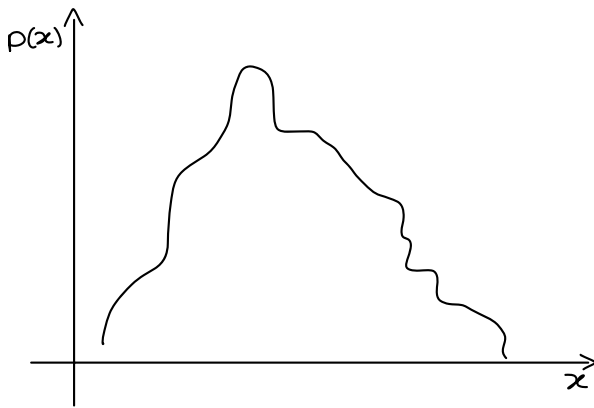
$$E(x) = \mu$$

$$V(x) = \sigma^2 = 1$$

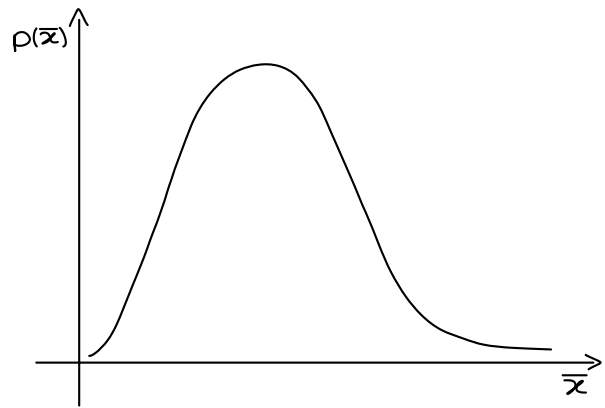
$$z \sim N(0; 1)$$

- **TEOREMA DEL LIMITE CENTRALE:** Data una qualsiasi popolazione con valore atteso μ e varianza σ^2 , estraendo da essa n dati indipendenti tra loro x_1, \dots, x_n
 $\Rightarrow \bar{x} = \frac{\sum x_i}{n}$ tende ad essere distribuita al crescere di n

come una distribuzione normale con valore atteso μ e varianza $\frac{\sigma^2}{n}$.



\Rightarrow

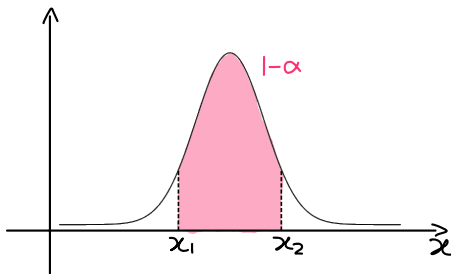


$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$\bar{x} \sim N\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

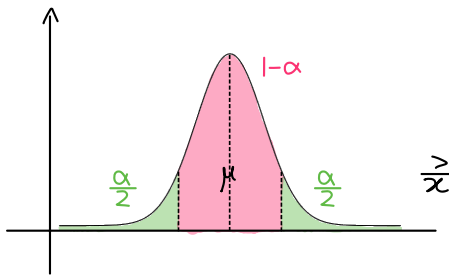
- **INTERVALLI FIDUCIARI:** intervalli di valori al cui interno abbiamo una determinata confidenza/fiducia (in termini di probabilità) che sia contenuto il valore vero di una determinata variabile considerata.

- **VARIABILE CONTINUA:** $P(x = x_i) = 0$
 $P(x_1 < x < x_2) \neq 0$



α : probabilità che la variabile NON sia inclusa nell'intervallo

$$\bar{x} \sim N\left(\mu; \frac{\sigma^2}{n}\right)$$



$$p(\bar{x}-d < \mu < \bar{x}+d) = 1-\alpha$$

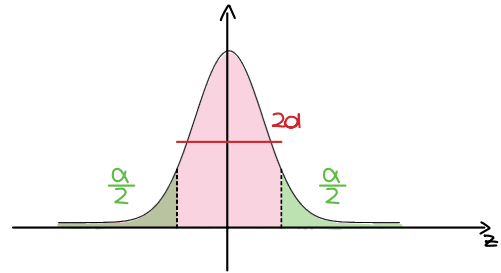
Possiamo ricondurre la distribuzione normale a una distribuzione normale standard mediante un cambio di variabile.

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$p\left(z_{1-\frac{\alpha}{2}} < \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$p\left(\bar{x} - z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\mu = \bar{x} \pm z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad \text{con affidabilità } (1-\alpha)$$



PARAMETRI DELLA POPOLAZIONE: definiti partendo da misure campionarie

- valore atteso
- varianza

PROPRIETÀ FONDAMENTALI DI UNO STIMATORE

- **correttezza**: valore atteso dello stimatore \rightarrow parametro della popolazione
- **efficienza**: varianza dello stimatore \rightarrow minima possibile
- **consistenza**: precisione e affidabilità della stima devono crescere al crescere della numerosità del campione

TEST STATISTICO

LE PASSAGGI DEL TEST

1. - formulazione delle ipotesi

H_0 : ipotesi nulla
 H_1 : ipotesi alternativa

esaurive e
mutualmente esclusive

2. - definizione della statistica coinvolta

3. - definizione del livello di significatività di rigetto α

I: rigetto H_0 se vera

II: non rigetto H_0 se falsa

4. - raccolta dati

5. - decisione

$$1.- H_0: \mu \leq 150 \text{ ppm}$$

$$H_1: \mu > 150 \text{ ppm}$$

$$2.- \bar{x} \sim N\left(\mu; \frac{\sigma^2}{n}\right) \quad \text{TLC}$$

3.- TEST ad una coda (destra)

$$z_{CR} = z_{1-\alpha} = z_{0,95} = 1,644$$

$$4.- \bar{x} = \frac{3648}{23} = 158,61$$

$$\sigma^2 = 455,16$$

$$\sigma = 21,33$$

$$\bar{x}_{CR} = \mu + z_{CR} \frac{\sigma}{\sqrt{n}} = 150 + 1,644 \frac{21,33}{\sqrt{23}} = 157,31 \text{ ppm}$$

$$5.- \bar{x} > \bar{x}_{CR}$$

\Rightarrow RIGETTO H_0

L

• **VARIABILE CHI-QUADRO (χ^2)**: rappresenta una distribuzione



Analizzare le deviazioni $O_i - E_i$



confrontare

frequenze di appartenenza osservate in un intervallo O_i , $i = 1, \dots, n$ intervalli
frequenze di occorrenza previste E_i

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

L'unico parametro che caratterizza questa distribuzione è il NUMERO DI GRADI DI LIBERTÀ (gdl)

$$E(\chi^2) = \text{gdl} = k - c$$

numero di vincoli, parametri calcolati
numero di intervalli

$$c = 1 \quad \Rightarrow \quad \text{numero totale di osservazioni in ogni intervallo}$$

$$c = 2 \quad \Rightarrow \quad \bar{x}, \sigma^2$$

N.B. Per un numero basso di gdl avremo una distribuzione con un massimo verso zero, mentre all'aumentare dei gdl la distribuzione diventa più "spanciata"

• **TEST STATISTICO**

FACILMENTE RICORDABILE

1.- formulazione delle ipotesi

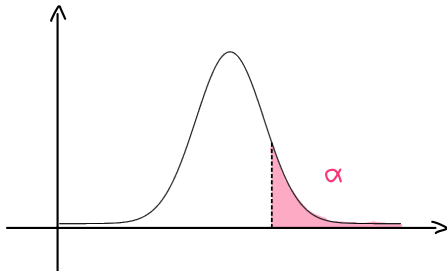
H_0 : distribuzione normale

H_1 : NO distribuzione normale

} esaustive e mutualmente esclusive

2.- definizione della statistica coinvolta: distribuzione chi-quadro

3.- definizione del livello di significatività di rigetto α



A sinistra del valore critico NON rigetto H_0 , mentre a destra del valore critico rigetto H_0 .

4.- raccolta dati

- definisco degli intervalli in maniera arbitraria (numero di intervalli maggiore di almeno un'unità rispetto al numero dei vincoli)
- conto O_i ed eseguo il confronto con E_i
- calcolo il valore di χ^2

NON FACILMENTE RICORDABILE

5.- decisione: confronto χ^2 e χ^2_{CR}

$$\chi^2 > \chi^2_{CR} \quad \Rightarrow \quad \text{rigetto}$$

$$\chi^2 < \chi^2_{CR} \quad \Rightarrow \quad \text{NON rigetto}$$

Γ

1.- H_0 : monete NON TRUCCATE H_1 : monete TRUCCATE

2.- $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

3.- $\alpha = 0,05$

4.-

| n° C | p | O _i | E _i | (O _i - E _i) ² /E _i |
|------|-------|----------------|----------------|---|
| 0 | 1/32 | 21 | 31,25 | 3,362 |
| 1 | 5/32 | 179 | 156,25 | 3,3124 |
| 2 | 10/32 | 223 | 312,5 | 25,63 |
| 3 | 10/32 | 293 | 312,5 | 1,21 |
| 4 | 5/32 | 199 | 156,25 | 11,69 |
| 5 | 1/32 | 85 | 31,25 | 92,45 |
| | 1 | 1000 | 1000 | 137,67 |

5.- $gdl = 6 - 1 = 5$

$\chi^2_{0,95;5} = 11,07$

$\chi^2 > \chi^2_{cr} \Rightarrow \text{Rigetto } H_0$

$$\alpha \quad E_i \quad \frac{(O_i - E_i)^2}{E_i}$$

calcolare p: $p = P(\chi^2_5 > 137,67)$

⇐ Nelle tabelle NON si trova questo valore, ma si trova $\chi^2_{0,995;5} = 20,52$
 \Rightarrow il valore di p è sicuramente $< 0,1\%$

L

• DIFFERENZA: $D = \bar{x}_1 - \bar{x}_2$ Se vale il TLC per \bar{x}_1 e \bar{x}_2 , dunque $\bar{x}_1 \sim N(\mu_1; \frac{\sigma_1^2}{n_1})$ e $\bar{x}_2 \sim N(\mu_2; \frac{\sigma_2^2}{n_2})$

$\Rightarrow D \sim N(\mu_1 - \mu_2; \sigma_D^2)$

$$\begin{cases} n_1 > 30 \\ n_2 > 30 \end{cases}$$

$$\sigma_D^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

Numero arbitrario

$$\begin{cases} n_1 < 30 \\ n_2 < 30 \end{cases}$$

$$\sigma_D^2 = \frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

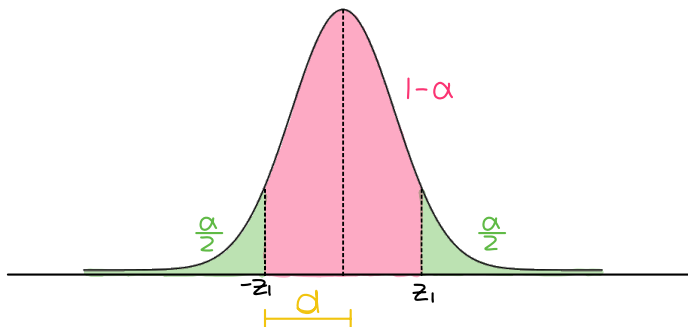
$$P[(\bar{x}_1 - \bar{x}_2) - a \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + a] = 1 - \alpha$$

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_D} \quad \text{D.N.S.}$$

$$P[-z_1 \leq \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_D} \leq z_1] = 1 - \alpha$$

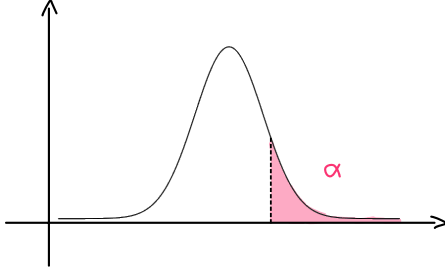
$$P[(\bar{x}_1 - \bar{x}_2) - z_1 \sigma_D \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + z_1 \sigma_D] = 1 - \alpha$$

$$\mu_1 - \mu_2 = (\bar{x}_1 - \bar{x}_2) \pm \sigma_D z_{1-\frac{\alpha}{2}}$$



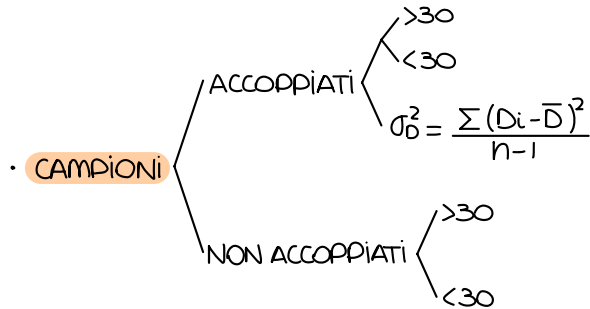
TEST STATISTICO

1. - formulazione delle ipotesi
- $$\begin{aligned} H_0: \mu_1 \leq \mu_2 &\Rightarrow \mu_1 - \mu_2 \leq 0 \\ H_1: \mu_1 > \mu_2 &\Rightarrow \mu_1 - \mu_2 > 0 \end{aligned} \quad \left. \vphantom{\begin{aligned} H_0: \mu_1 \leq \mu_2 \\ H_1: \mu_1 > \mu_2 \end{aligned}} \right\} \begin{array}{l} \text{esaustive e} \\ \text{mutualmente esclusive} \end{array}$$
2. - definizione della statistica coinvolta: $D \sim N(\mu_1 - \mu_2; \sigma_D^2)$
3. - definizione del livello di significatività di rigetto α



A sinistra del valore critico NON rigetto H_0 , mentre a destra del valore critico rigetto H_0 .

4. - raccolta dati: calcolare $\bar{x}_1, \bar{x}_2, \sigma_D^2$
5. - decisione: confronto $z_1, z_{1-\alpha}$



┌

$$\begin{aligned} 1. - H_0: \mu_1 &\leq \mu_2 \\ H_1: \mu_1 &> \mu_2 \end{aligned}$$

$$\begin{aligned} 4. - \bar{x}_1 - \bar{x}_2 &= 21,57 \\ z &= \frac{\bar{D}}{\sigma_D} = 1,27 \end{aligned}$$

$$2. - \bar{D} = \bar{x}_1 - \bar{x}_2 \sim N(\mu_D; \sigma_D^2)$$

$$5. - z < z_{0,95} \Rightarrow \text{NON rigetto } H_0$$

$$3. - \alpha = 0,05$$

$$z_{0,95} = 1,645$$

└

VARIABILI CASUALI

$$x_1, x_2, \dots, x_n$$

$$\begin{cases} x_1 \leq x_1 \\ x_2 \leq x_2 \\ \vdots \\ x_n \leq x_n \end{cases}$$

$$P(E) = P(x_1 \leq x_1 \text{ AND } x_2 \leq x_2 \dots \text{ AND } x_n \leq x_n) = P(x_1 \leq x_1) \cdot P(x_2 \leq x_2) \cdot \dots \cdot P(x_n \leq x_n)$$

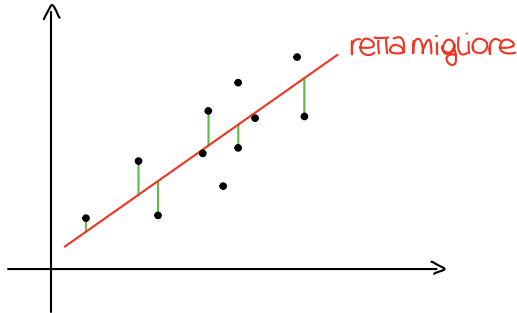


EVENTI INDIPENDENTI

- **COVARIANZA**: $\sigma_{xy} = E[(x - \mu_x)(y - \mu_y)] = E(xy) - E(x)E(y) = 0 \Leftrightarrow$ VARIABILI INDIPENDENTI

$$\sigma_{xy} = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{n-1}$$

- **COEFFICIENTE DI CORRELAZIONE LINEARE**: $R = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \begin{cases} 0 & \Rightarrow \text{variabili indipendenti} \\ \pm 1 & \Rightarrow \text{variabili correlate linearmente} \end{cases}$
 $y = mx + q$



SOMMA DEI QUADRATI DEGLI SCARTI
 $SSE = \sum (y_i - \hat{y}_i)^2 = \sum [y_i - (b_0 + b_1 x)]^2$

$$\text{min. SSE : } \begin{cases} \frac{\partial SSE}{\partial b_0} = 0 \\ \frac{\partial SSE}{\partial b_1} = 0 \end{cases}$$

$$\begin{cases} b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x^2 - (\sum x)^2} \\ b_0 = \frac{\sum x^2 \sum y_i - \sum x_i \sum x_i y_i}{n \sum x^2 - (\sum x)^2} = \bar{y} - b_1 \bar{x} \end{cases}$$

$$\sigma_{b_0} = \sigma_y \cdot \sqrt{\frac{\sum x^2}{n \sum x^2 - (\sum x)^2}}$$

$$\sigma_{y_p} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$$

$$\sigma_{b_1} = \sigma_y \cdot \sqrt{\frac{n}{n \sum x^2 - (\sum x)^2}}$$

$$\beta_0 = b_0 \pm z_{1-\frac{\alpha}{2}} \sigma_{b_0}$$

$$\beta_1 = b_1 \pm z_{1-\frac{\alpha}{2}} \sigma_{b_1}$$

- **RESIDUI**: $e_i = 0 \pm \sigma_y \cdot z_{1-\frac{\alpha}{2}}$

$$e_i = y_i - \hat{y}$$

- **REGRESSIONE LINEARE MULTIPLA**: $y = b_0 + b_1 x_1 + b_2 x_2$

$$\begin{cases} \frac{\partial SSE}{\partial b_0} = 0 \\ \frac{\partial SSE}{\partial b_1} = 0 \\ \frac{\partial SSE}{\partial b_2} = 0 \end{cases}$$

$$\begin{cases} n b_0 + b_1 \sum x_{1,i} + b_2 \sum x_{2,i} = \sum y_i \\ b_0 \sum x_{1,i} + b_1 \sum x_{1,i}^2 + b_2 \sum x_{1,i} x_{2,i} = \sum y_i x_{1,i} \\ b_0 \sum x_{2,i} + b_1 \sum x_{1,i} x_{2,i} + b_2 \sum x_{2,i}^2 = \sum y_i x_{2,i} \end{cases}$$

$$\sigma_{y_p} = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-3}}$$

- **COEFFICIENTE DI CORRELAZIONE**: $SSTO = SSE + SSR$

$$\sum (y_i - \bar{y}_i)^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y}_i)^2$$

$$R^2 = \frac{\text{Deviazione spiegata}}{\text{Deviazione totale}} = \frac{\sum (\hat{y}_i - \bar{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} = \frac{\sum (y_i - \bar{y}_i)^2 - \sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2}$$

- **F-TEST**: confronto tra variabilità

A n_a
B n_b

$$F_{n_a-1; n_b-1} = \frac{\sigma_A^2}{\sigma_B^2}$$

$$F_{v_1; v_2; 1-\alpha} = \frac{1}{F_{v_1; v_2; \alpha}}$$

- **TEST STATISTICO**

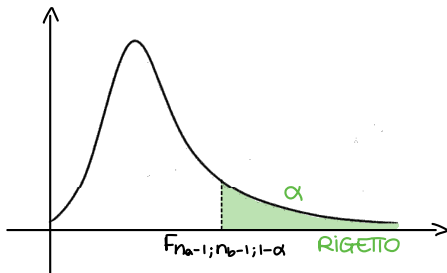
FORMULAZIONE DELL'IPOTESI

1. - formulazione delle ipotesi

$H_0: \sigma_A^2 \leq \sigma_B^2$
 $H_1: \sigma_A^2 > \sigma_B^2$ } esaustive e mutualmente esclusive

2. - definizione della statistica coinvolta: $F_{n_a-1; n_b-1} = \frac{\sigma_A^2}{\sigma_B^2}$

3. - definizione del livello di significatività di rigetto α



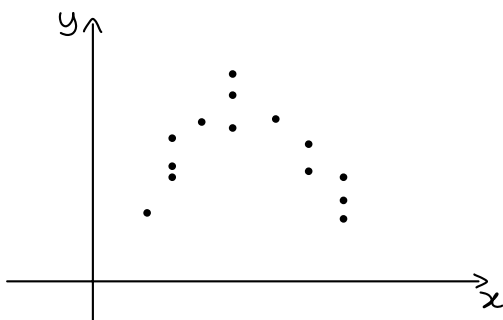
ACQUISIZIONE DATI

4. - raccolta dati: calcolare $\frac{\sigma_A^2}{\sigma_B^2}$

$$\sigma_y^2 = \frac{SSE}{N-m}$$

5. - decisione: confronto $F_{n_a-1; n_b-1}$, $F_{n_a-1; n_b-1; 1-\alpha}$

- **OSSERVAZIONE**: Posso verificare con F-test che l'incertezza di un modello, usato per fare un fitting o una regressione su dei dati sperimentali, NON è superiore rispetto all'incertezza insita nei dati sperimentali.



N : numero totale delle osservazioni
= 14

n : numero di osservazioni distinte
= 7

p_i : numero di prove replicate per ogni punto

$$y_{ij} - \bar{y} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \hat{y}_i) + (\hat{y}_i - \bar{y})$$

prova replicata

$$\sum \sum (y_{ij} - \bar{y})^2 = \sum \sum (y_{ij} - \bar{y}_i)^2 + \sum p_i (\bar{y}_i - \hat{y}_i)^2 + \sum p_i (\hat{y}_i - \bar{y})^2$$

| | | | | |
|------------------|-----|------|------|-----|
| | SST | SSEE | SSLF | SSR |
| Gradi di libertà | N-1 | N-n | n-m | m-1 |

$$SSE = SSEE + SSLF$$

$$\frac{\frac{SSLF}{n-m}}{\frac{SSEE}{N-n}} \sim F_{n-m; N-n}$$

┌

$$1.- H_0: \frac{\frac{SSLF}{n-m}}{\frac{SSEE}{N-n}} \leq 1$$

$$H_1: \frac{\frac{SSLF}{n-m}}{\frac{SSEE}{N-n}} > 1$$

$$2.- \frac{\frac{SSLF}{n-m}}{\frac{SSEE}{N-n}} \sim F_{n-m; N-n}$$

$$3.- 1-\alpha = 0,95$$

$$m=3$$

$$N=10$$

$$n=6$$

$$n-m=3$$

$$N-n=4$$

$$F_{3;4;0,95} = 6,6$$

4.-

$$\frac{SSLF}{n-m} = 152,34$$

$$\frac{SSEE}{N-n} = 32,75$$

$$\frac{\frac{SSLF}{n-m}}{\frac{SSEE}{N-n}} = 4,652$$

$$5.- 4,662 < 6,600 \Rightarrow \text{Non rigetto } H_0$$

$$x_i \quad y_i \quad p_i$$

$$x_i \quad y_{im} \quad p_i \quad \hat{y} \quad (y_{im} - \hat{y})^2$$

$$p_i (y_{im} - \hat{y})^2 \quad p_i (y_i - y_{im})^2$$

└