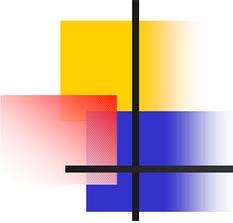


Foglio di calcolo (2)

- Applicazioni statistiche
- Grafici di correlazione
- I stogrammi



Semplici calcoli statistici

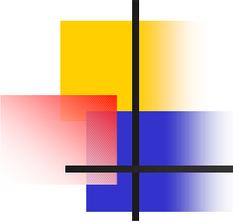
Una serie di dati (per esempio le misure di un ossido nell'esercitazione 3) possono essere analizzate dal punto di vista statistico.

In particolare si puo' calcolare la

Media (average)

=average(A2:A259)

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$$



Indici di dispersione

Varianza (Variance)

=var(A2:A259)

Deviazione standard

standard deviation

=std(A2:A259)

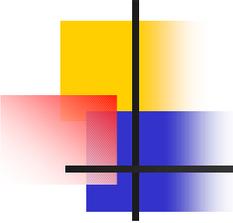
Range

=max(A2:A259)- min(A2:A259)

$$s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$$

$$s = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2}$$

$$R = x_{\max} - x_{\min}$$



Indici di forma

asimmetria (skewness)

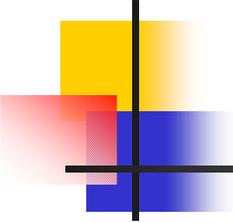
=`skew(A2:A259)`

$$a = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - \bar{x}}{s} \right)^3$$

$a=0$ simmetria della distribuzione

$a>0$ coda verso destra

$a<0$ coda verso sinistra



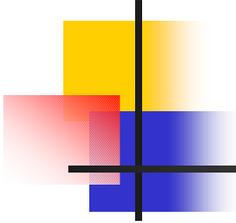
La correlazione

La correlazione

Per sapere se esiste un 'legame' tra due caratteri quantitativi, e cioè se uno di essi esercita un'influenza sull'altro, ad esempio il peso delle persone e la loro altezza, si utilizzano gli indici di **correlazione**, i quali danno anche una misura di questo 'legame'.

Quando la dipendenza tra due variabili è lineare si parla di correlazione lineare. L'indice usato è detto indice di correlazione di Bravais-Person

$$r = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{N s_x s_y} \quad -1 < r < 1$$



La correlazione

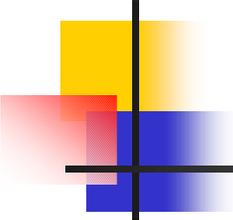
per $r=1$ si ha il massimo di correlazione diretta

per $r=-1$ si ha il massimo di correlazione inversa

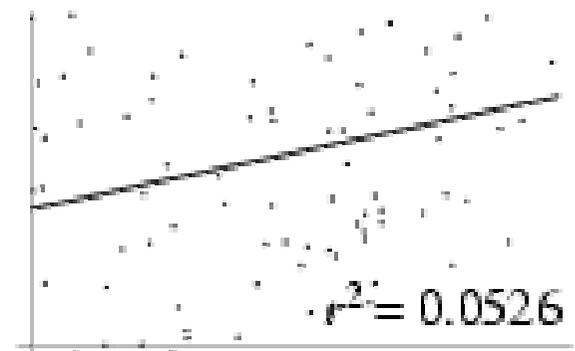
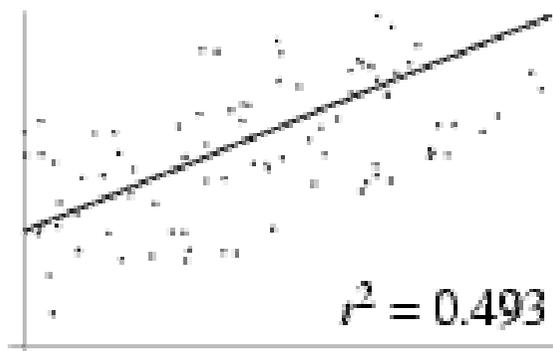
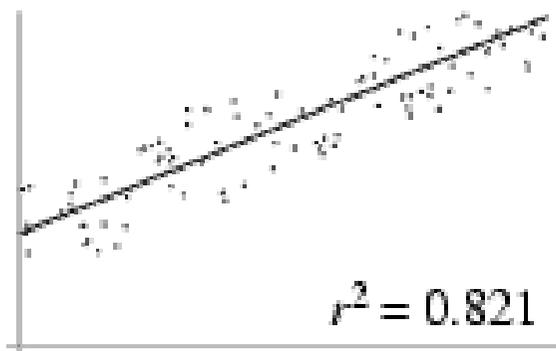
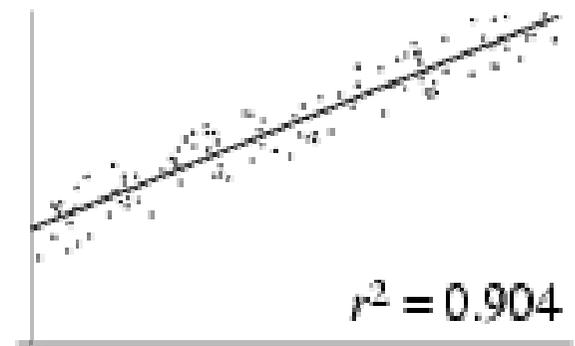
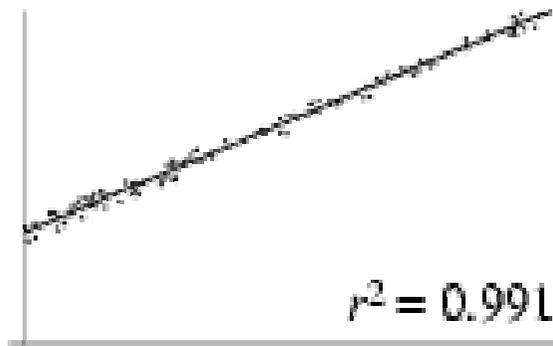
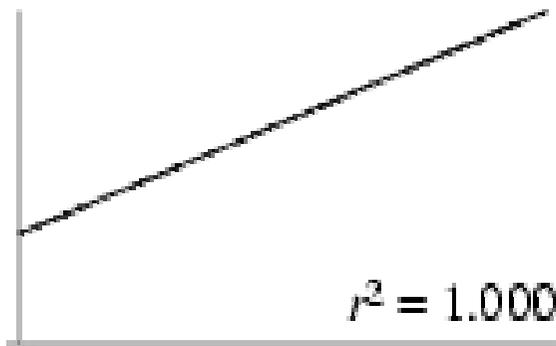
per $r=0$ non si ha correlazione

La correlazione si dice diretta se ai valori crescenti di una variabile corrispondono valori pure crescenti dell'altra variabile, ad esempio reddito e consumi, altezza e peso, SiO_2 e Al_2O_3 nell'esercitazione

La correlazione si dice inversa se ai valori crescenti di una variabile corrispondono valori decrescenti dell'altra variabile, ad esempio altitudine e pressione atmosferica, SiO_2 e MgO nell'esercitazione.



La correlazione



Frequenze ed Istogrammi

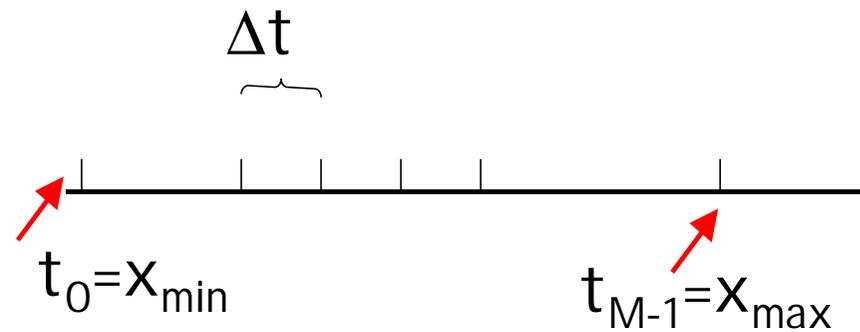
Un istogramma è il grafico di una serie di dati in funzione di una classe di intervalli possibili.

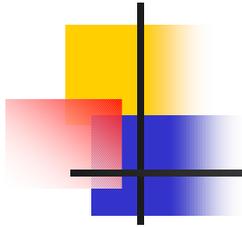
Si considerano N dati da analizzare. Questi dati saranno compresi in un intervallo $[x_{\min}, x_{\max}]$

Dividiamo l'intervallo in M intervallini tali che

$$x_{\min} = t_0, t_1, \dots, t_{M-1} = x_{\max}$$

$$t_k = K\Delta t = K \frac{x_{\max} - x_{\min}}{(M - 1)}$$



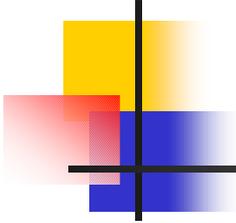


I valori t_k sono gli estremi delle classi.

Ogni dato x_i cade in un intervallo $S_k = [t_{k-1}, t_k)$
(classe)

Diciamo V_j la frequenza assoluta dell'intervallo S_j .

**Un istogramma è il grafico delle frequenze assolute
in funzione delle classi.**



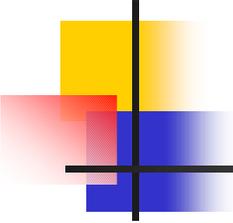
Frequenze

La funzione FREQUENZA calcola la frequenza di occorrenza dei valori di un intervallo e restituisce una matrice verticale di numeri.

Sintassi:

`FREQUENZA(matrice_ dati; matrice classi)`

tale istruzione viene inserita come formula matrice dopo aver selezionato un intervallo di celle adiacenti nel quale dovrà apparire il risultato. Il numero di elementi nella matrice restituita **è maggiore di un'unità** rispetto al numero di elementi contenuti in matrice_ classi.



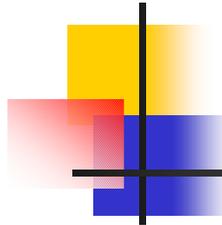
Formula matrice

Una formula in forma di matrice può eseguire più calcoli e restituire uno o più risultati.

Procedura:

selezionare la cella o le celle in cui si desidera
immettere la formula, creare la formula e premere

CTRL+ MAI USC+I NVIO



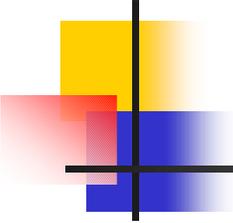
I stogramma

Per disegnare un istogramma

Si fissano sull'asse delle ascisse i valori delle classi e, in corrispondenza, si

disegna una barra la cui altezza è pari alla frequenza.

L'altezza ha la stessa unità di misura della probabilità teorica



Scelta del numero di classi

Per costruire un diagramma delle frequenze bisogna discretizzare in modo opportuno il range dei valori assunti dalla variabile.

Qual è la scelta ottimale?

Regola empirica:

Numero di intervalli = \sqrt{N}