1) Range (campo di variazione)

$$W = x_{max} - x_{min}$$

Misura puramente descrittiva e di solito poco informativa

Es. le altezze di 10 esemplari di una pianta sono:

10 22 33 44 46 51 67 74 79 85

W= 85-10 =75

le altezze di altri 10 esemplari sono invece

10 11 11 12 13 14 15 16 20 85

W = 85-10 = 75

2) Varianza

devianza

in una *popolazione*

$$\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N}$$

in un *campione* (varianza campionaria)

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

gradi di libertà

La varianza ha come unità di misura
il quadrato dell'unità utilizzata per la media

Dalla varianza può essere ricavata un'altra misura di variazione importante, la

DEVIAZIONE STANDARD

$$\sigma = \sqrt{\sigma^2}$$

Una formula alternativa per il calcolo della varianza

somma dei quadrati

quadrato della somma

$$s^2 = \frac{\sum x^2 - \frac{\sum (x)^2}{n}}{n-1}$$

x: 81 79 82 83 80 78 80 87 82

$$\sum x^2 = 81^2 + 79^2 + 82^2 + 83^2 + \ldots + 82^2 = 66316$$

$$(\sum x)^2 = (81 + 79 + 82 + 83 + \ldots + 82)^2 = (814)^2 = 662596$$

$$s^2 = \frac{\sum x^2 - \frac{\sum (x)^2}{n}}{n-1} \qquad s^2 = \frac{66316 - \frac{662596}{10}}{10-1} = 6.27$$

3) Coefficiente di variazione

$$CV = \frac{s}{\bar{x}} \cdot 100$$

Consente di esprimere in maniera ***relativa*** la variabilità di una distribuzione

FINO A QUI

Es. date le due serie di misure A e B:

	A	B
	12.40	125.80
	11.40	117.30
	15.10	121.90
	9.80	131.40
	10.30	111.90
	16.90	127.10
media =	12.65	122.57
s =	2.56	6.47
CV=	0.20	0.05

La probabilità
(cenni)

⇨ quantificazione di una 'possibilità'

⇨ frequenza attesa di un evento

Es. nel lancio di un dado, la frequenza attesa dell'uscita del numero "1" è 1/6.

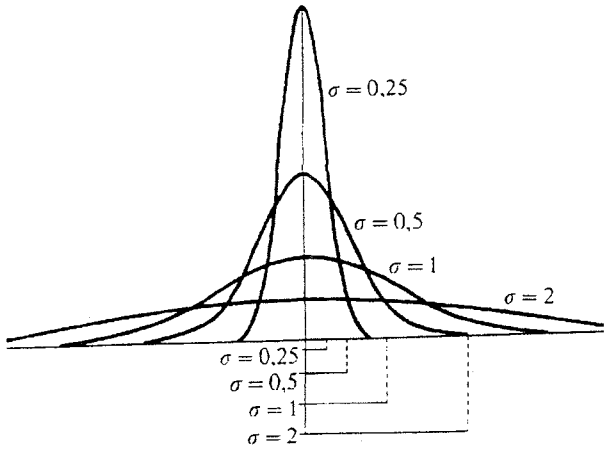
SPAZIO CAMPIONARIO (S) ⇨ insieme di tutti i possibili esiti di una prova (nell' esempio del dado =1,2,3,4,5,6).

EVENTO ELEMENTARE ⇨ ogni singolo elemento di S.

- ogni probabilità è un numero non negativo

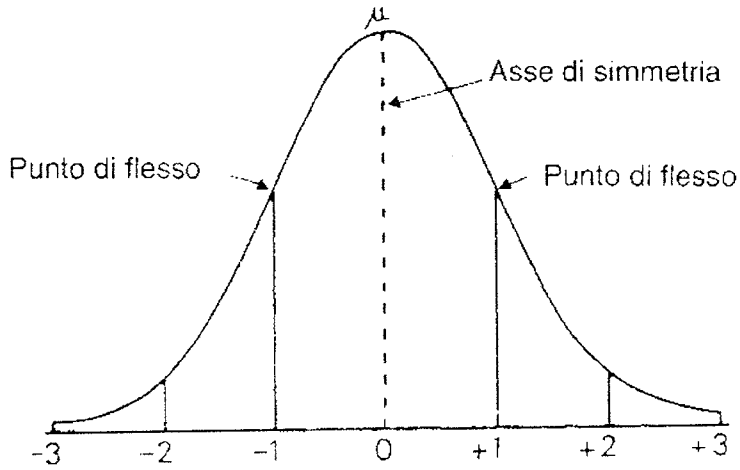
- la somma delle probabilità di tutti gli eventi elementari è uguale a 1.

Curva normale (Gaussiana)

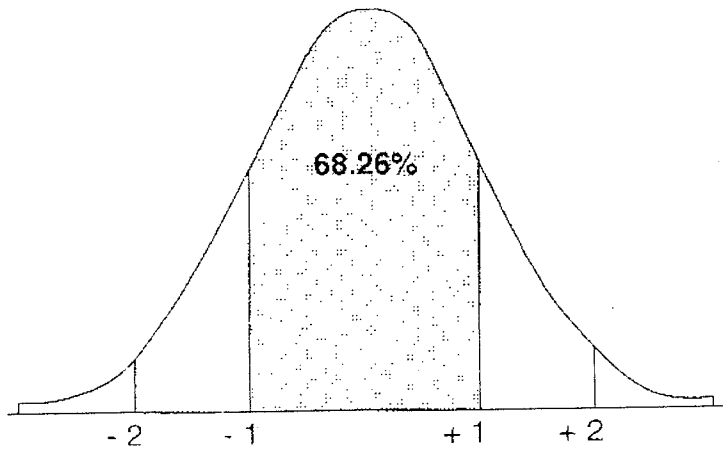


$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Proprietà della distribuzione normale



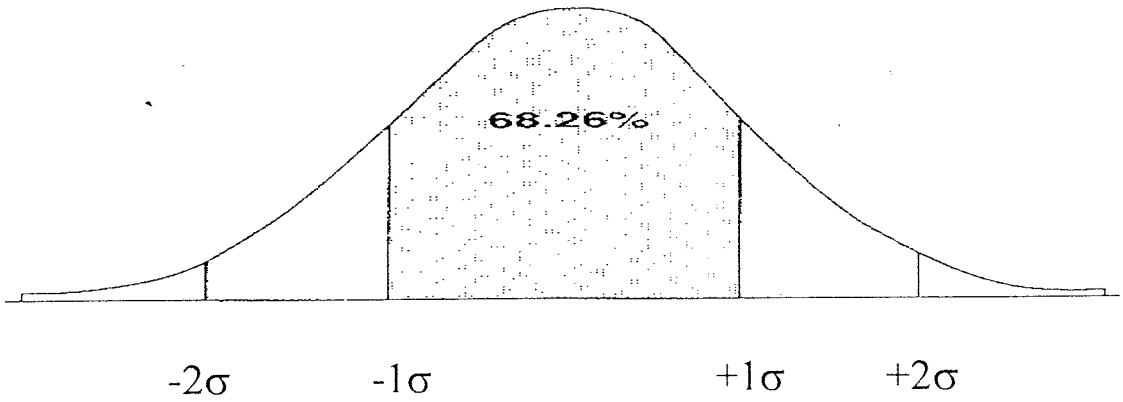
Distribuzione normale standardizzata: media=0 e dev standard=1



- il 68.26% di tutte le osservazioni è compreso in un intervallo delimitato da ± **1 σ** (z=±1);

- il 95.44% di tutte le osservazioni è compreso in un intervallo delimitato da ±**2 σ** (z=±2).

Questa proprietà e' generalizzabile



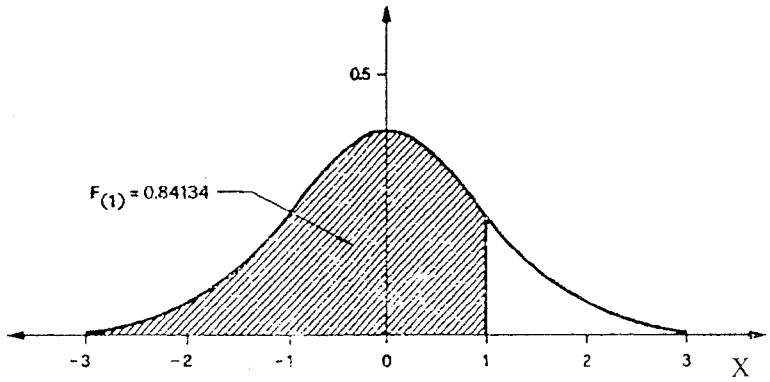
- il 95% di tutte le osservazioni è compreso in un intervallo delimitato da ±**1.96 σ** (z=±1.96);

- il 95% di tutte le osservazioni è compreso in un intervallo delimitato da ±**2.58 σ** (z=±2.58).

La probabilità che una qualsiasi osservazione estratta a caso da un campione cada al di fuori dell'intervallo determinato da ±1.96 s e ±2.58 s è pari rispettivamente a

***P*=0.05**
***P*=0.01**

E' possibile calcolare la probabilità che un certo evento x si verifichi (premesso che la distribuzione di x sia normale con media (μ) e deviazione standard (σ) note?



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

I valori di F(z) sono riportati in apposite tavole e si possono ottenere mediante la funzione DISTRIB.NORM.ST(z) di Excel.

Es. Se la statura media di una popolazione è 173.23 cm con deviazione standard 5.938, qual è la probabilità che un individuo abbia una statura > 179 cm?

$$z = \frac{179 - 172.23}{5.938} = 1.14$$

CORRELAZIONE

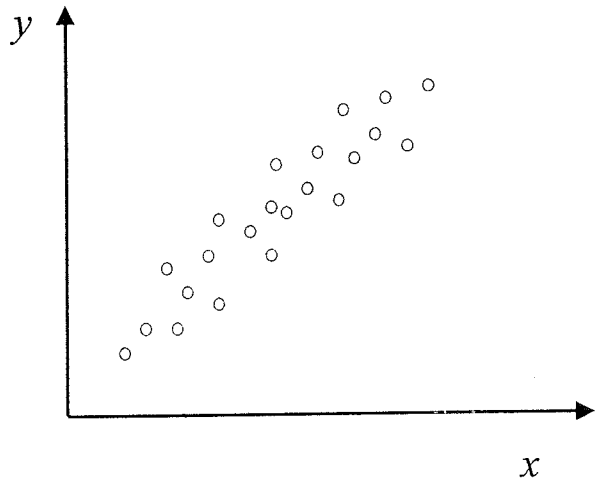
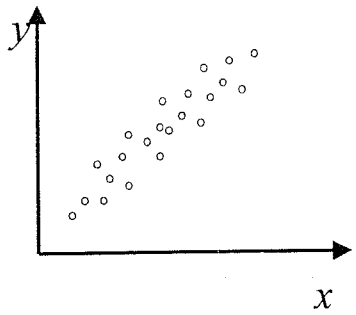
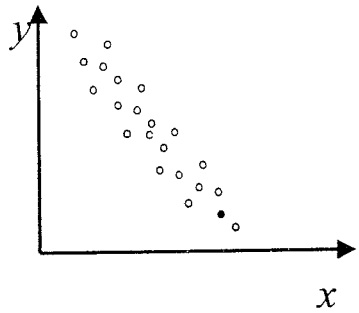


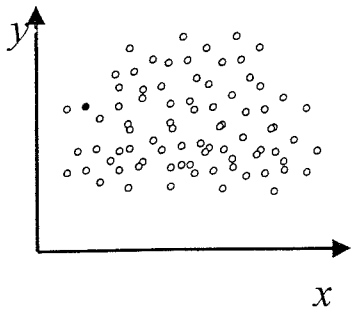
diagramma di dispersione

correlazione positiva

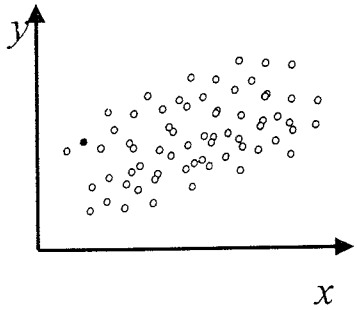


correlazione negativa





nessuna correlazione



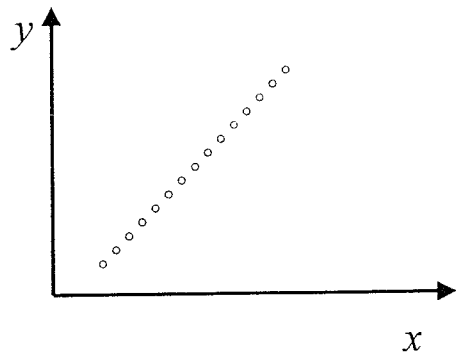
correlazione ?

Coefficiente di correlazione di *Pearson*
(parametrico)

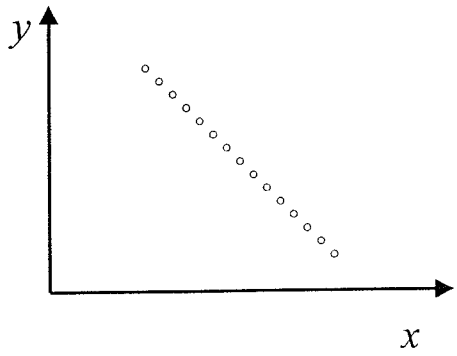
codevianza

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 (y-\bar{y})^2}}$$

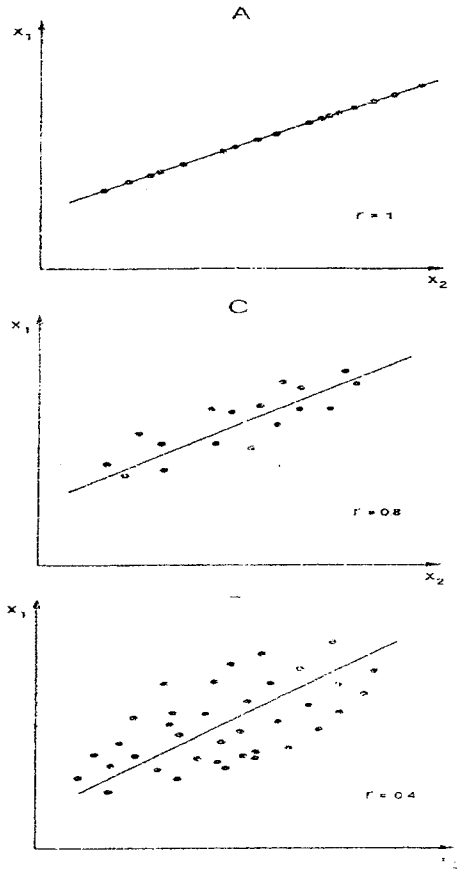
Può assumere valori compresi tra -1 e 1



correlazione positiva perfetta
r=1



correlazione negativa perfetta
r=-1



$r=1$

$r=0.8$

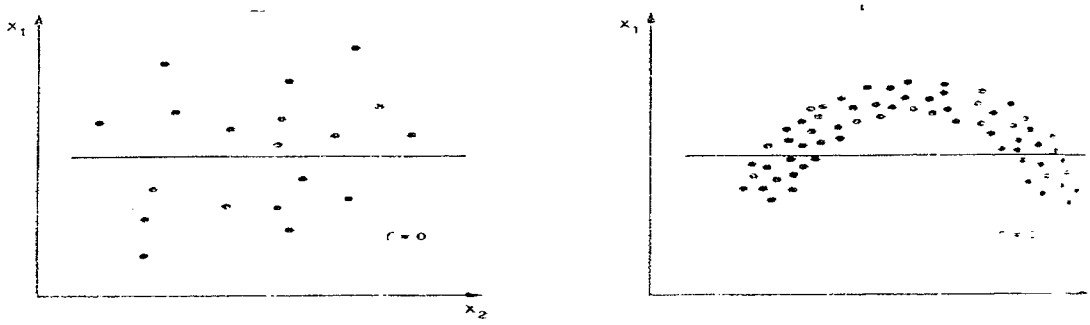
$r=0.4$

Alcune limitazioni e precauzioni

1) Le due variabili sono distribuite "normalmente"

2) Non viene identificata nessuna relazione di ***causa-effetto***

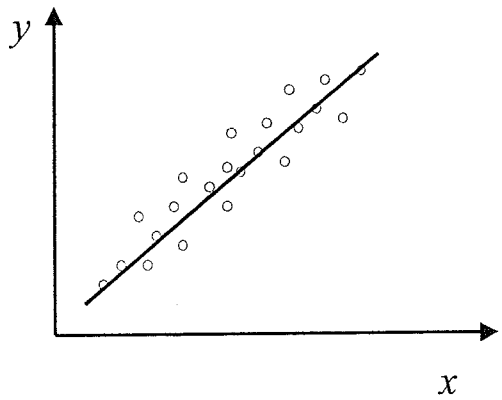
2) La relazione tra le due variabili deve essere "lineare"



REGRESSIONE

⇨ Identificare una funzione che descriva la relazione tra la variabile *y* ed *x*

variabile dipendente



variabile indipendente

Dal punto di vista matematico consiste nel determinare i valori dei coefficienti ***a*** e ***b*** della equazione della retta

$$y = a + b \cdot x$$

intercetta asse y

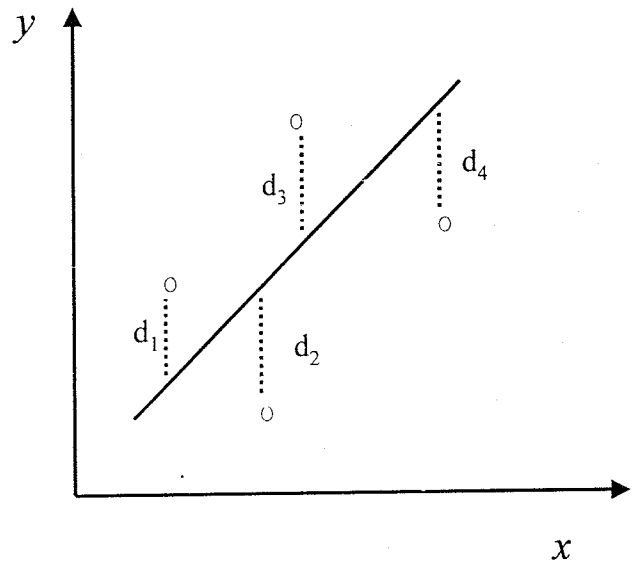
coefficiente angolare

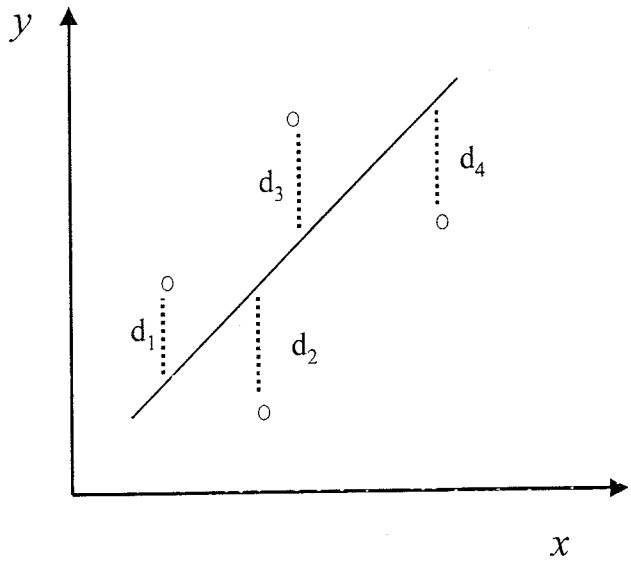
⇩

indica di quanto varia y al variare di una unità di x

Metodo dei minimi quadrati

⇨ La somma delle distanze verticali (deviazioni- scarti) di tutti i punti dalla linea deve essere la minima possibile





Poiche' alcune deviazioni sono negative ed altre positive si utilizzano gli *scarti* quadratici d^2