Media geometrica

La media geometrica di un insieme di n valori positivi x_1, x_2, …, x_n di un carattere quantitativo X è pari alla radice n-esima del prodotto dei singoli valori:

$$\bar{x}_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

Esempio sulla media geometrica

Capitale iniziale: 10.000€

Investimento in obbligazioni a tasso variabile

Anno	Tasso di interesse	Tasso di interesse %
I	0,015	1,5
II	0,020	2,0
III	0,072	7,2
IV	0,090	9,0
V	0,074	7,4
VI	0,045	4,5

Qual è il tasso di interesse medio annuo?

$$\overline{x}_g = \sqrt[6]{(1{,}015)(1{,}02)(1{,}072)(1{,}09)(1{,}074)(1{,}045)} = 1{,}05229$$

Proprietà della media geometrica

1- Il prodotto dei valori $x_1, x_2, \ldots, x_n$ assunti da un insieme di unità statistiche è pari alla potenza n-esima della media geometrica:

$$\boxed{x_1 \cdot x_2 \cdot \ldots \cdot x_n = \left(\overline{x}_g\right)^n}$$

2- Il logaritmo della media geometrica è uguale alla media aritmetica dei logaritmi:

$$\boxed{\log \overline{x}_g = \frac{1}{n}\sum_{i=1}^{n}\log(x_i)}$$

Media armonica

La media armonica di un insieme di n valori $x_1, x_2, \ldots, x_n$ di un carattere quantitativo X è definita da:

$$\boxed{\bar{x}_a = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}}$$

Media potenziata di ordine r

Si definisce **media potenziata di ordine r** la radice r-esima della media aritmetica delle potenze r-esime delle osservazioni:

$$\bar{x}_r = \left(\frac{1}{n} \sum_{i=1}^{n} (x_i)^r \right)^{\frac{1}{r}}$$

$r \to -\infty$	$\Rightarrow$	$\bar{x}_r \to x_1$	= valore minimo della variabile X
$r = -1$	$\Rightarrow$	$\bar{x}_r \to \bar{x}_a$	= media armonica della variabile X
$r \to 0$	$\Rightarrow$	$\bar{x}_r \to \bar{x}_G$	= media geometrica della variabile X
$r = 1$	$\Rightarrow$	$\bar{x}_r \to \bar{x}$	= media aritmetica della variabile X
$r \to +\infty$	$\Rightarrow$	$\bar{x}_r \to x_n$	= valore massimo della variabile X

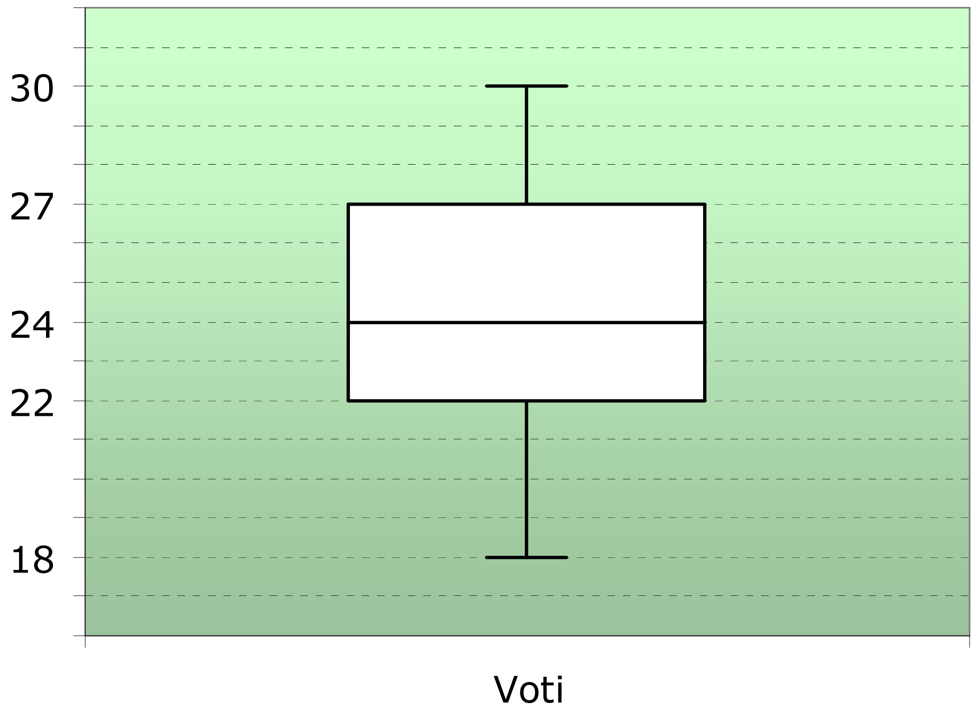
Box-plot

X	n_i
18	3
19	5
20	7
21	5
22	10
23	6
24	10
25	6
26	11
27	6
28	5
29	10
30	4
Totale	88

Q_0 =
Q_1 =
Q_2 =
Q_3 =
Q_4 =

Differenza interquartile Iqr =

Q_2-(Iqr·1,5) =
Q_2+(Iqr·1,5) =



Confronti fra dati statistici

- Differenza assoluta

$$\boxed{x_2 - x_1}$$

- Divario relativo

$$\boxed{x_2 / x_1}$$

- Variazione relativa

$$\boxed{\frac{x_2 - x_1}{x_1}}$$

I rapporti statistici

- Composizione: rapporti di parte al tutto
- Coesistenza: rapporti tra le frequenze di modalità alternative di uno stesso carattere
- Derivazione: rapporti tra il numero di eventi in un intervallo di tempo prefissato e la numerosità della popolazione che li ha generati

Rapporto di composizione

Sport praticato	Numero di praticanti
Calcio	2.518
Atletica leggera	455
Footing, jogging, podismo	290
Ginnastica, attrezzistica, danza	2.490
Pallacanestro, pallavolo	1.307
Nuoto, pallanuoto, tuffi	1.595
Tennis	1.034
Sport invernali	1.082
Caccia	328
Pesca	261
Altro	2.134
Totale	13.494

$$n_{Calcio}/n_{Totale}=2.518/13.494=0,1866$$

Su 100 atleti 18,66 praticano calcio

Esempio sui rapporti di composizione

	Fumatori	Non fumatori
M	23	20
F	18	10

Quanti sono gli individui maschi? 23+20=43

E le femmine? 18+10=28

E nel complesso? 43+28=41+30=71

Esempio sui rapporti di composizione

	Fumatori	Non fumatori	Totale
M	23	20	43
F	18	10	28
Totale	41	30	71

Qual è la composizione per sesso dell'intero gruppo?

$$\frac{\text{n° Maschi}}{\text{n° Maschi} + \text{n° Femmine}} = \frac{43}{43+28} = 0{,}6056$$

$$\frac{\text{n° Femmine}}{\text{n° Maschi} + \text{n° Femmine}} = \frac{28}{43+28} = 0{,}3944$$

Esempio sui rapporti di composizione

	Fumatori	Non fumatori	Totale
M F	23 18	20 10	43 28
Totale	41	30	71

Qual è la composizione del gruppo rispetto al fumo?

$$\frac{\text{n}^\circ\ \text{Fumatori}}{\text{n}^\circ\ \text{Fumatori} + \text{n}^\circ\ \text{Non fumatori}} = \frac{41}{41+30} = 0{,}5775$$

$$\frac{\text{n}^\circ\ \text{Non fumatori}}{\text{n}^\circ\ \text{Fumatori} + \text{n}^\circ\ \text{Non fumatori}} = \frac{30}{41+30} = 0{,}4225$$

Esempio sui rapporti di composizione

	Fumatori	Non fumatori	Totale
M	23	20	43
F	18	10	28
Totale	41	30	71

Fumano di più i maschi o le femmine?

$$\frac{\text{n° Maschi fumatori}}{\text{n° Maschi (Fumatori e Non fumatori)}} = \frac{23}{23+20} = 0{,}5349$$

$$\frac{\text{n° Femmine fumatrici}}{\text{n° Femmine (Fumatrici e Non fumatrici)}} = \frac{18}{18+10} = 0{,}6429$$

Esempio sui rapporti di composizione

	Fumatori	Non fumatori	Totale
M	23	20	43
F	18	10	28
Totale	41	30	71

Tra i fumatori, è più facile trovare un maschio o una femmina?

$$\frac{\text{n° Maschi fumatori}}{\text{n° Fumatori (Maschi e Femmine)}} = \frac{23}{23+18} = 0{,}5610$$

$$\frac{\text{n° Femmine fumatrici}}{\text{n° Fumatori (Maschi e Femmine)}} = \frac{18}{23+18} = 0{,}4390$$

Rapporto di coesistenza

Sport praticato	Numero di praticanti
Calcio	2.518
Atletica leggera	455
Footing, jogging, podismo	290
Ginnastica, attrezzistica, danza	2.490
Pallacanestro, pallavolo	1.307
Nuoto, pallanuoto, tuffi	1.595
Tennis	1.034
Sport invernali	1.082
Caccia	328
Pesca	261
Altro	2.134
Totale	13.494

$$n_{Calcio}/n_{Tennis}=2.518/1.034=2,4352$$

Su 100 tennisti ci sono 243,52 calciatori

Esempio sui rapporti di coesistenza

Sesso	Nati vivi
M	322.536
F	302.269
Totale	624.805

Rapporto di mascolinità

$$\frac{\text{n° Nati Vivi Maschi}}{\text{n° Nati Vivi Femmine}} = \frac{322.536}{302.269} = 1{,}07$$

Rapporto di femminilità

$$\frac{\text{n° Nati Vivi Femmine}}{\text{n° Nati Vivi Maschi}} = \frac{302.269}{322.536} = 0{,}94$$

Esempio sui rapporti di coesistenza

Sesso	Pop. residente
M	27.889.911
F	29.509.197
Totale	57.399.108

Rapporto di mascolinità

$$\frac{\text{Pop. Res. Maschile}}{\text{Pop. Res. Femminile}} = \frac{27.889.911}{29.509.197} = 0,945$$

Esempio sui rapporti di coesistenza

Classi di età	Pop. residente
0|--|14	460.102
15|--|64	2.702.125
65 e oltre	736.943
Totale	3.899.170

Indice di vecchiaia

$$I_v = \frac{Pop._{65\ e\ +}}{Pop._{0|--|14}} \cdot 100 = \frac{736.943}{460.102} \cdot 100 = 160{,}17$$

Indice di dipendenza degli anziani

$$I_{d.a.} = \frac{Pop._{65\ e\ +}}{Pop._{15|--|64}} \cdot 100 = \frac{736.943}{2.702.125} \cdot 100 = 27{,}27$$

Rapporto di derivazione

Regione	Nati vivi nel 1991	Popolazione residente al 20-10-91
Emilia-Romagna	27.746	3.899.170
Abruzzo	12.320	1.243.690

$$\frac{27.746}{3.899.170} \cdot 1000 = 7{,}12$$

$$\frac{12.320}{1.243.690} \cdot 1000 = 9{,}91$$

In Emilia-Romagna ci sono 7,12 nati vivi ogni 1000 abitanti

In Abruzzo ci sono 9,91 nati vivi ogni 1000 abitanti

Rapporto di derivazione

Per confrontare la propensione a trascorrere le vacanze all’estero dei turisti di due nazioni A e B, si deve tener conto del fatto che le due nazioni hanno popolazioni numericamente diverse. Si dovrà quindi individuare una grandezza che esprima, per una qualsiasi nazione, quanti turisti si sono recati all’estero ogni 1000 abitanti.

$$\frac{\text{Turisti andati all'estero}}{\text{Popolazione}} \cdot 1000$$

Calcolo di un numero indice

	04-feb	05-feb	06-feb	07-feb	08-feb	09-feb	10-feb	11-feb	12-feb	13-feb
PIACENZA	64	44	51	75	75	94	112	106	96	26
PARMA	63	51	47	54	66	103	92	73	68	40
REGGIO NELL'EMILIA	60	46	43	53	53	78	76	71	65	50
MODENA	48	38	54	52	51	97	99	82	96	58
BOLOGNA	70	44	37	59	60	82	69	48	86	60
FERRARA	51	33	43	50	55	106	128	123	150	89
RAVENNA	38	35	46	59	68	93	78	46	59	49
FORLI'	54	n.d.	28	n.d.	n.d.	56	76	52	48	44
RIMINI	70	42	41	66	76	86	90	62	41	40

Livelli PM10	
Superiore al limite di legge (al 2005)	>50
Entro il limite di legge	0-50
Dato non disponibile	n.d.

$$ {}_bI_t = \frac{q_t}{q_b} $$

$$ {}_{4feb}I_{5feb} = \frac{PM10_{5feb}}{PM10_{4feb}} = \frac{44}{70} = 0{,}629 $$

Fonte: www.liberiamolaria.it

Numeri indici semplici

- Numeri indici a base fissa
- Numeri indici a base mobile

	04-feb	05-feb	06-feb	07-feb	08-feb	09-feb	10-feb	11-feb	12-feb	13-feb
BOLOGNA	**70**	**44**	**37**	**59**	**60**	**82**	**69**	**48**	**86**	**60**

$$ {}_{4feb}I_{5feb} = \frac{PM10_{5feb}}{PM10_{4feb}} = \frac{44}{70} = 0{,}629 $$

$$ {}_{4feb}I_{6feb} = \frac{PM10_{6feb}}{PM10_{4feb}} = \frac{37}{70} = 0{,}529 $$

...

$$ {}_{4feb}I_{13feb} = \frac{PM10_{13feb}}{PM10_{4feb}} = \frac{60}{70} = 0{,}857 $$

Serie dei numeri indice con base *4 febbraio (%)*

t	04-feb	05-feb	06-feb	07-feb	08-feb	09-feb	10-feb	11-feb	12-feb	13-feb
${}_{4feb}I_t$ (%)	100	62.9	52.9	84.3	85.7	117.1	98.6	68.6	122.9	85.7

Serie dei numeri indice a base mobile

Se interessa studiare le variazioni relative di *Q* da un tempo *t* -1 a quello successivo *t*, si divide ogni valore q_t per il precedente q_{t-1}, e si ottiene la serie dei numeri indice a base mobile

$$ {}_{t-1}I_t = \frac{q_t}{q_{t-1}} $$

Numero indice a base mobile riferito al tempo *t*

Numero indice a base mobile delle concentrazioni di PM10

$$ {}_{t-1}I_t = \frac{PM10_t}{PM10_{t-1}} $$

	04-feb	05-feb	06-feb	07-feb	08-feb	09-feb	10-feb	11-feb	12-feb	13-feb
BOLOGNA	**70**	**44**	**37**	**59**	**60**	**82**	**69**	**48**	**86**	**60**

$$ {}_{4feb}I_{5feb} = \frac{PM10_{5feb}}{PM10_{4feb}} = \frac{44}{70} = 0{,}629 $$

$$ {}_{5feb}I_{6feb} = \frac{PM10_{6feb}}{PM10_{5feb}} = \frac{37}{44} = 0{,}841 $$

...

$$ {}_{12feb}I_{13feb} = \frac{PM10_{13feb}}{PM10_{12feb}} = \frac{60}{86} = 0{,}698 $$

Serie dei numeri indice a base mobile (%)

t	04-feb	05-feb	06-feb	07-feb	08-feb	09-feb	10-feb	11-feb	12-feb	13-feb
${}_{t-1}I_{t}$ (%)	-	62.9	84.1	159.5	101.7	136.7	84.2	69.6	179.2	69.8

Proprietà dei numeri indici semplici

- Identità $\quad {}_{t}I_{t} = 1$
- Reversibilità delle basi $\quad {}_{b}I_{t} = \dfrac{1}{{}_{t}I_{b}}$
- Circolarità $\quad {}_{s}I_{t} \cdot {}_{r}I_{s} = {}_{r}I_{t}$

La variabilità

Il valor medio fornisce una sintesi della distribuzione di un carattere.

Accanto agli indici di posizione considerati fino a ora, introduciamo altri indicatori il cui proposito è misurare la “attitudine a variare” del fenomeno oggetto di studio.

L’attitudine di un carattere quantitativo X ad assumere valori differenti tra le unità componenti un insieme statistico è chiamata variabilità.

La variabilità

S1: cm 169	T1: cm 150
S2: cm 170	T2: cm 170
S3: cm 171	T3: cm 190

La terna {S1,S2,S3} ha minore variabilità della terna {T1,T2,T3}.

La variabilità

La variabilità costituisce una caratteristica degli insiemi statistici e può essere descritta mediante indicatori che godano di particolari proprietà:

- una misura di variabilità deve annullarsi quando, e solo quando, tutte le unità del collettivo presentano il medesimo stato di grandezza del carattere
- una misura di variabilità deve assumere valori crescenti all'aumentare della variabilità

La variabilità

Gli indicatori comunemente adoperati possono essere distinti in tre categorie fondamentali:

- indicatori che misurano la diversità fra due particolari termini della distribuzione o fra due quartili (intervallo di variabilità, differenza interquartile)
- indicatori che misurano la dispersione dei valori osservati xi attorno a un valor medio (scostamenti medi)
- indicatori che misurano le disuguaglianze a due a due fra tutti i valori individuali (differenze medie)

Alcune misure di variabilità

Sia $x_1 \leq x_2 \leq \ldots \leq x_n$ l'insieme delle osservazioni del carattere X

- Intervallo di variabilità

$$\boxed{I_v = x_n - x_1}$$

- Differenza interquartile

$$\boxed{W = x_{3/4} - x_{1/4}}$$

Intervallo di variabilità e Differenza interquartile

Voti	Studenti
18	3
19	5
20	7
21	5
22	10
23	6
24	10
25	6
26	11
27	6
28	5
29	10
30	4
Totale	88

Intervallo di variabilità

x_1 =

x_n =

$\Rightarrow I_v = 30\text{-}18 = 12$

Differenza interquartile

$x_{1/4}$ =

$x_{3/4}$ =

$\Rightarrow W = 27\text{-}22 = 5$

Indicatori di variabilità

S1: cm 169	T1: cm 150
S2: cm 170	T2: cm 170
S3: cm 171	T3: cm 190

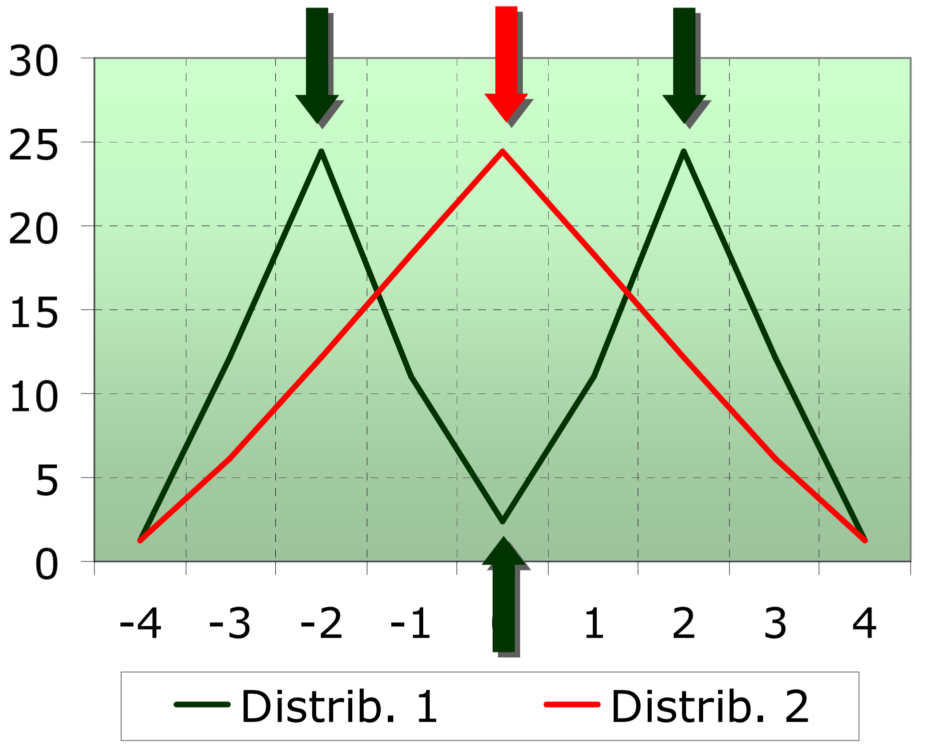
(S1,S2,S3)	(T1,T2,T3)
$\|169-170\|=1$	$\|150-170\|=20$
$\|169-171\|=2$	$\|150-190\|=40$
$\|170-169\|=1$	$\|170-150\|=20$
$\|170-171\|=2$	$\|170-190\|=20$
$\|171-169\|=1$	$\|190-150\|=20$
$\|171-170\|=2$	$\|190-170\|=20$

Consideriamo la serie delle differenze in valore assoluto tra ciascuna unità statistica e le altre. Le due somme, pari a 9 e 140, indicano che il carattere “statura” presenta nei riguardi del gruppo (S1,S2,S3) variabilità minore che non nei riguardi del gruppo (T1,T2,T3).

Indicatori di variabilità

X	Distrib. 1	Distrib. 2
-4	1,2	1,2
-3	12,2	6,1
-2	24,4	12,2
-1	11,0	18,3
0	2,4	24,4
1	11,0	18,3
2	24,4	12,2
3	12,2	6,1
4	1,2	1,2

$\overline{x}_1 = 0 \quad \overline{x}_2 = 0$



La varianza

La varianza di un insieme di n valori osservati $x_1, x_2, \ldots, x_n$ di una variabile X con media aritmetica $\overline{x}$ è data da:

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2$$

Esempio

X = Ore di allenamento settimanale
n = 6 atleti che si preparano a una gara

Dati relativi a 4 differenti situazioni

{	8	10,5	9	8	12	9,5	}
{	9,5	9,5	9,5	9,5	9,5	9,5	}
{	8,5	10,5	9,5	9,5	9,5	9,5	}
{	0	0	0	0	57	0	}

Determinare la media aritmetica.

La media aritmetica è sempre 9,5

Esempio

Per i medesimi protocolli elementari, calcolare intervallo di variazione, devianza, varianza e deviazione standard.

	{	8	10,5	9	8	12	9,5	}	°
Iv	{	9,5	9,5	9,5	9,5	9,5	9,5	}	=57
Dev()	{	8,5	10,5	9,5	9,5	9,5	9,5	}	7,5
V(X)	{	0	0	0	0	57	0	}	25
s_x		√2=1,4142		0	√0,33=0,57				21,24

$\bar{x} = 9,5$

Esempio

In un campione di 128 uomini adulti sono stati rilevati:

X = Circonferenza del torace (in cm)

Y = Peso corporeo (in kg)

$x_{i-1} \dashv x_i$	n_i
$80 \dashv 90$	34
$90 \dashv 110$	38
$110 \dashv 130$	56
	128

$y_{i-1} \dashv y_i$	n_i
$55 \dashv 65$	23
$65 \dashv 75$	38
$75 \dashv 90$	34
$90 \dashv 110$	33
	128

a) Misurare la variabilità di X e Y mediante la deviazione standard

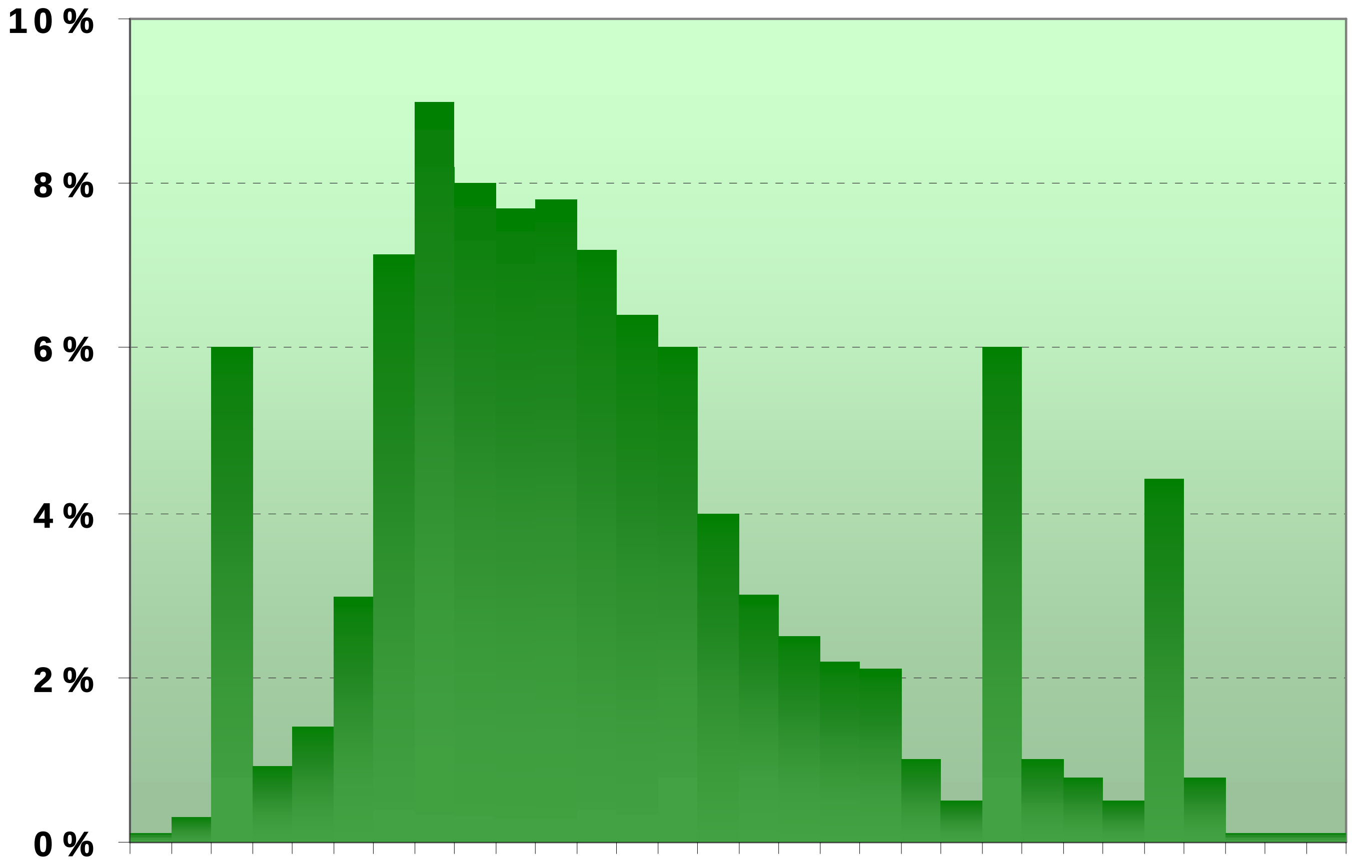
b) E' maggiore la variabilità di X o quella di Y?

Coefficiente di variazione

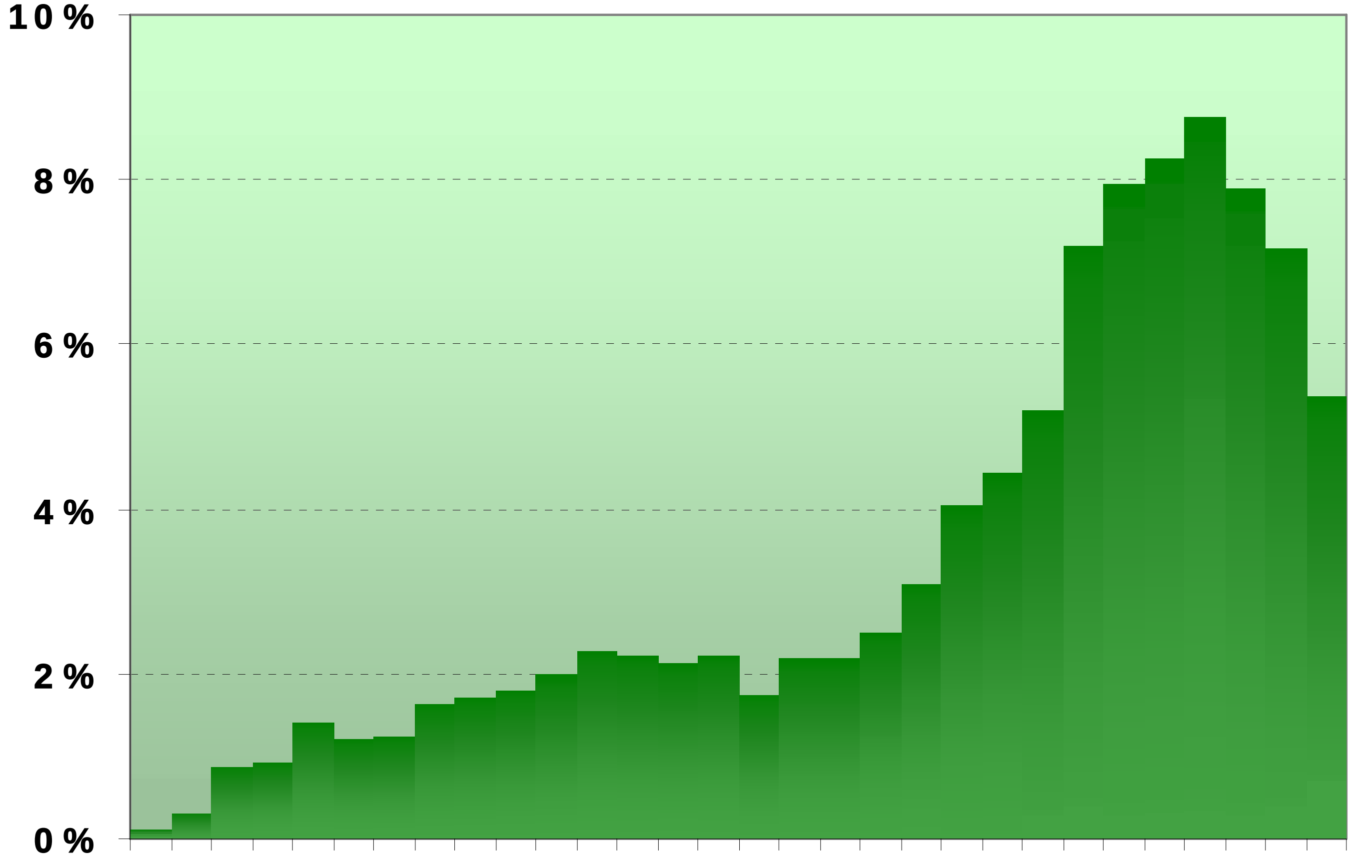
$$Cv = \frac{\sigma}{\bar{x}} = \frac{{}^{2}S(X)}{\bar{x}}$$

$$Cvn = \frac{Cv}{\sqrt{n-1}}$$

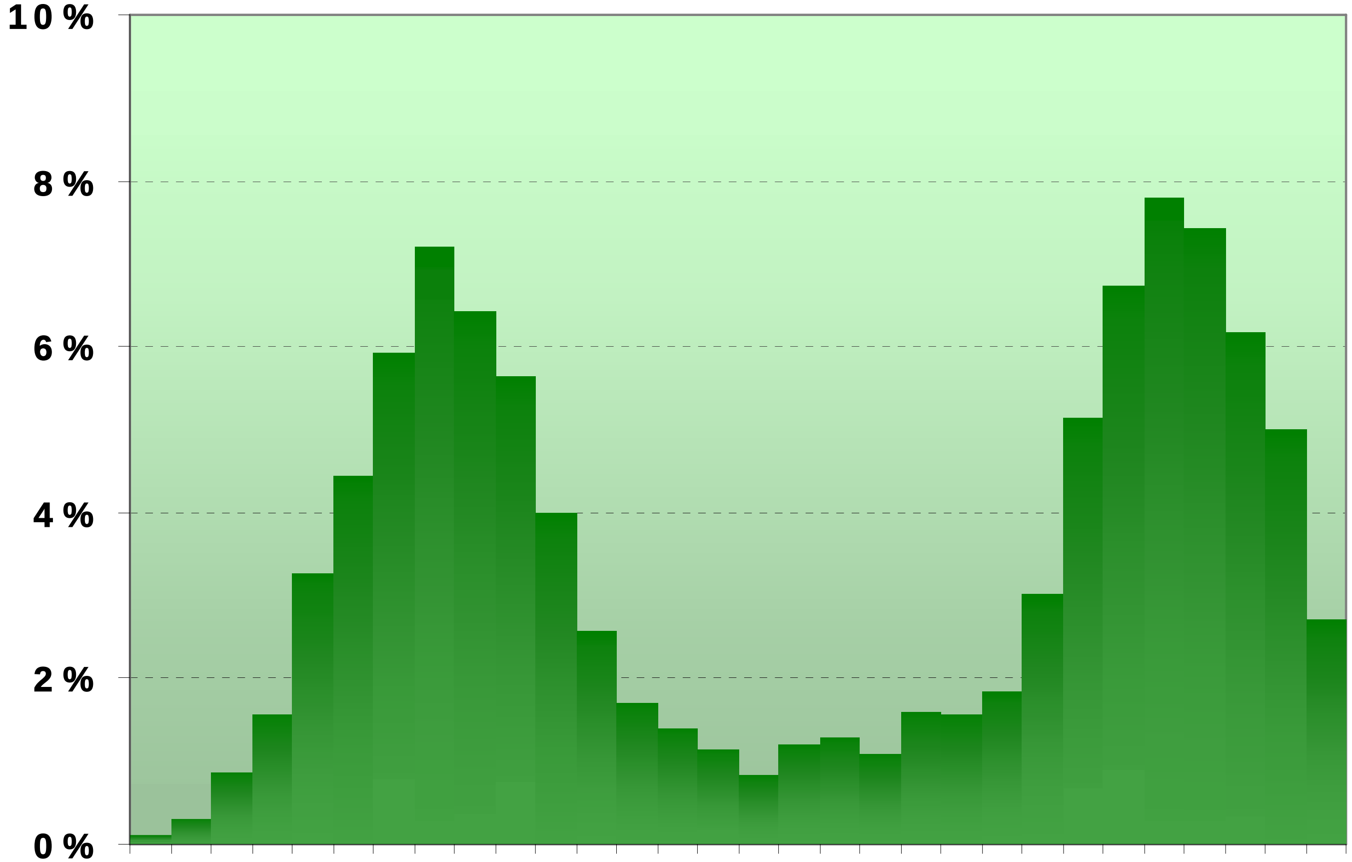
Esempio: Distribuzione con outlier



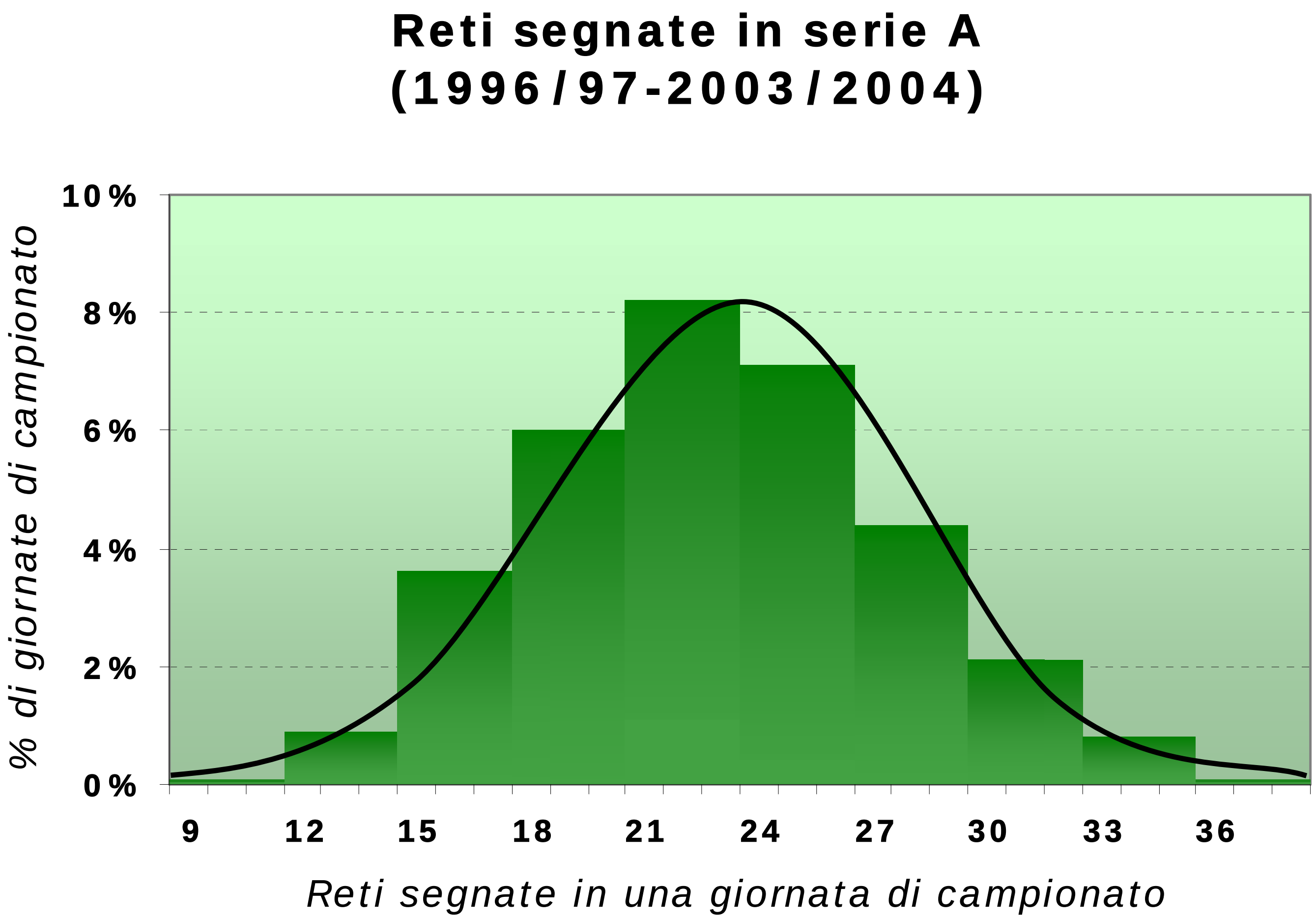
Esempio: Asimmetria a destra



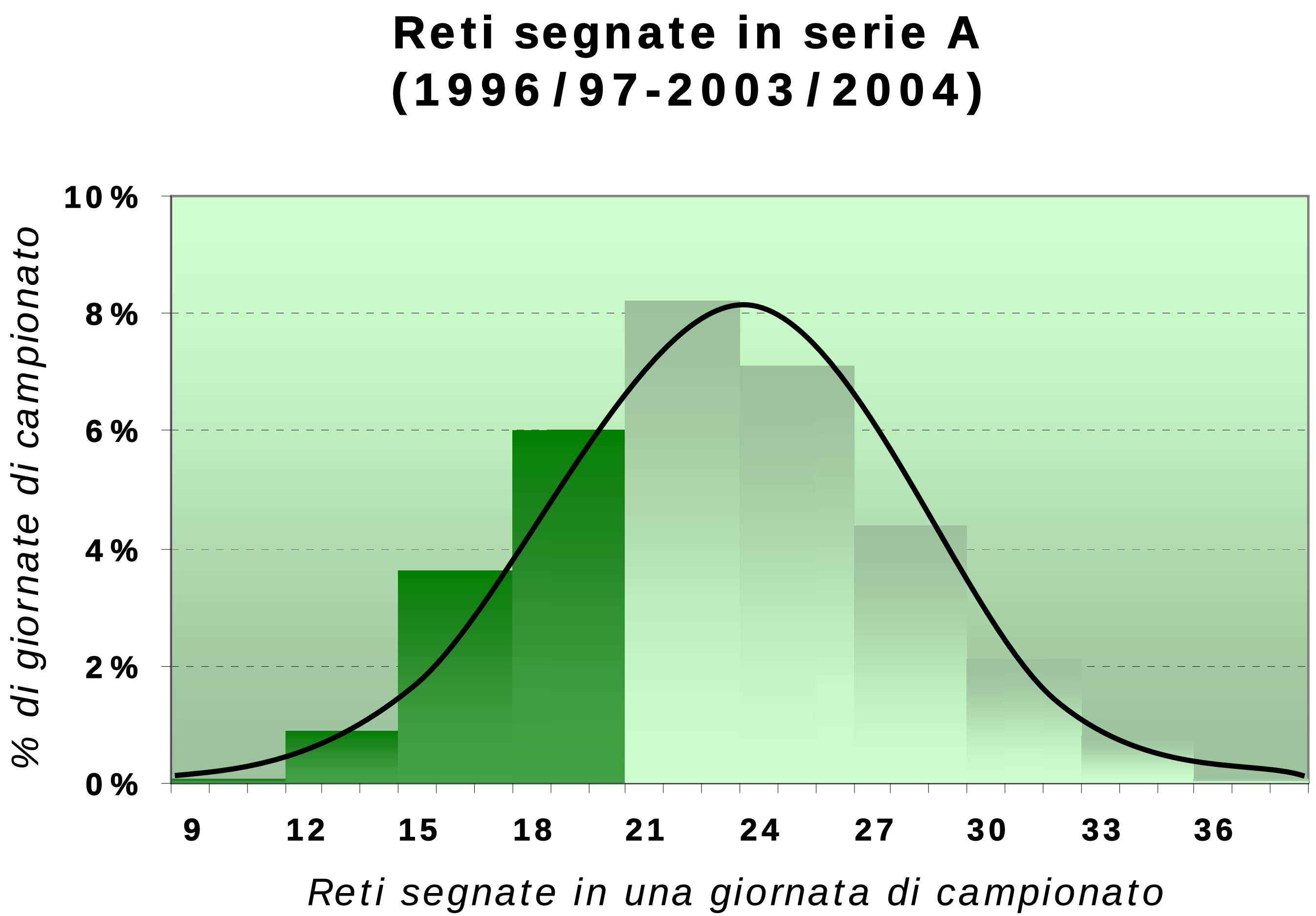
Esempio: Distribuzione bimodale



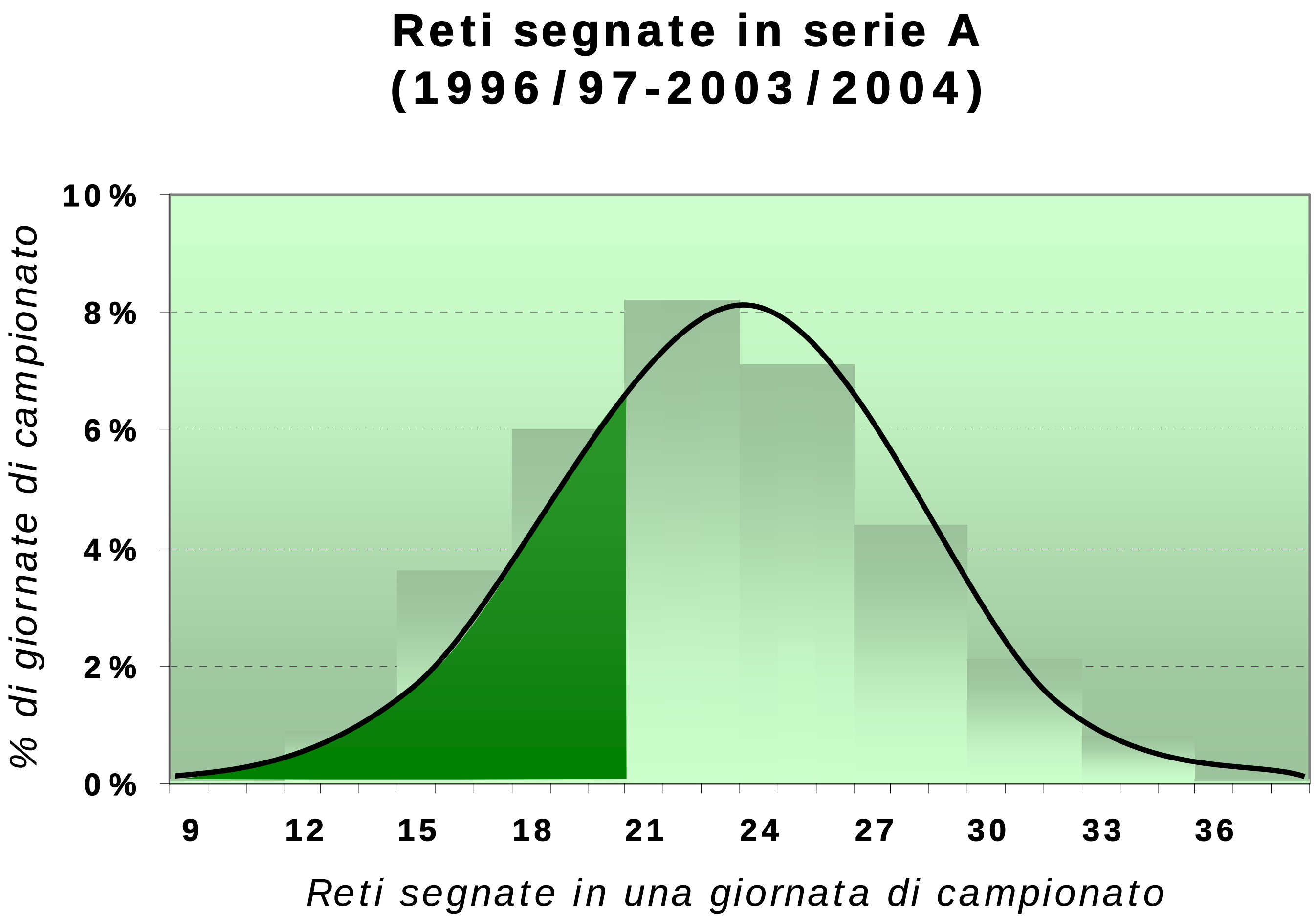
Esempio


Reti segnate in serie A
(1996/97-2003/2004)
10%
8%
6%
4%
2%
0%
% di giornate di campionato
9
12
15
18
21
24
27
30
33
36
Reti segnate in una giornata di campionato

Esempio


Reti segnate in serie A
(1996/97-2003/2004)
10%
8%
6%
4%
2%
0%
% di giornate di campionato
9
12
15
18
21
24
27
30
33
36
Reti segnate in una giornata di campionato

Esempio


Reti segnate in serie A
(1996/97-2003/2004)
10%
8%
6%
4%
2%
0%
% di giornate di campionato
9
12
15
18
21
24
27
30
33
36
Reti segnate in una giornata di campionato

Curva di densità

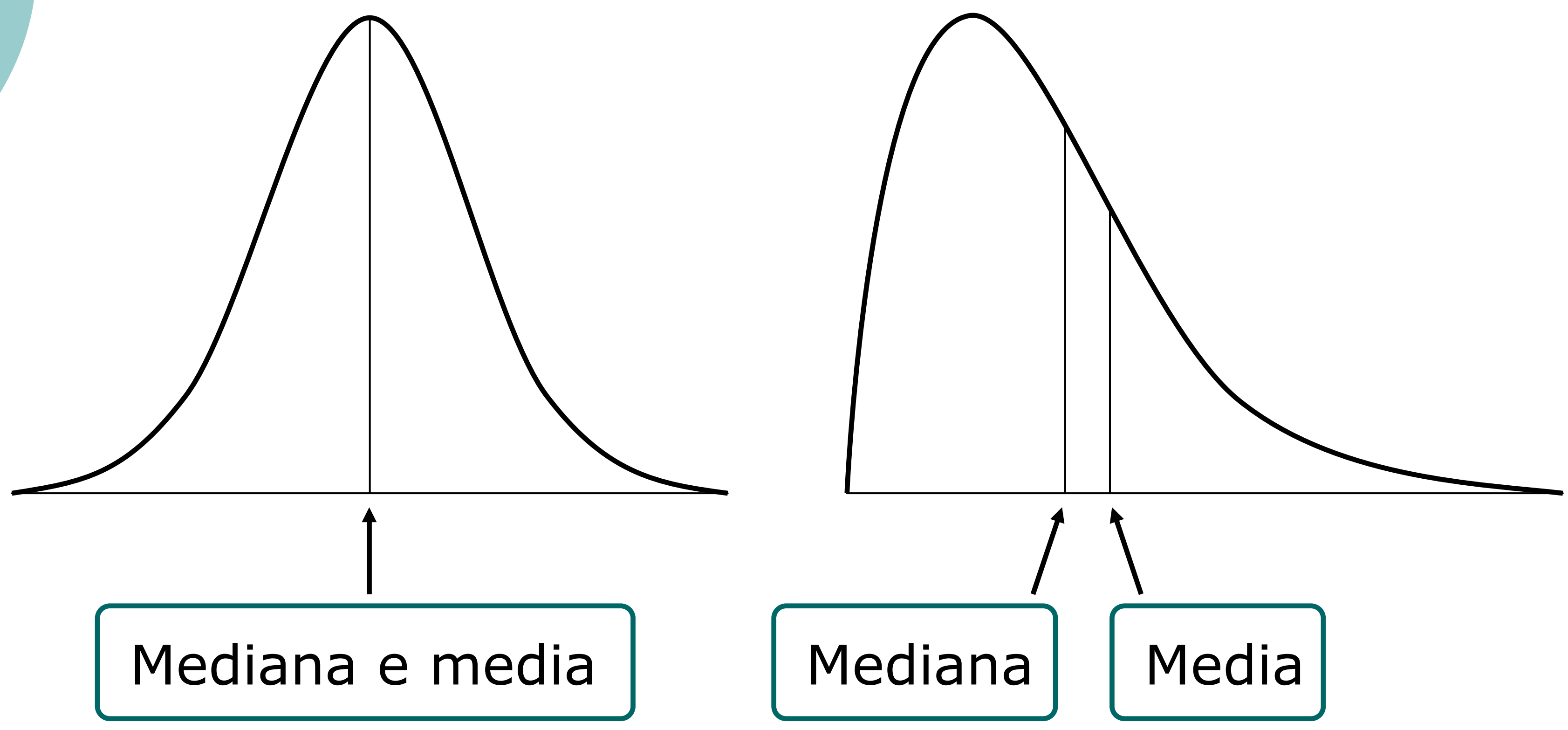
Una curva di densità è una curva tale che:

- Si trova sempre sopra o sull'asse orizzontale
- L'area sotto di essa è esattamente pari a 1

Una curva di densità rappresenta il modello complessivo di una distribuzione.

L'area che sta sotto la curva relativamente a un certo intervallo rappresenta la proporzione di tutte le osservazioni che cadono in quell'intervallo

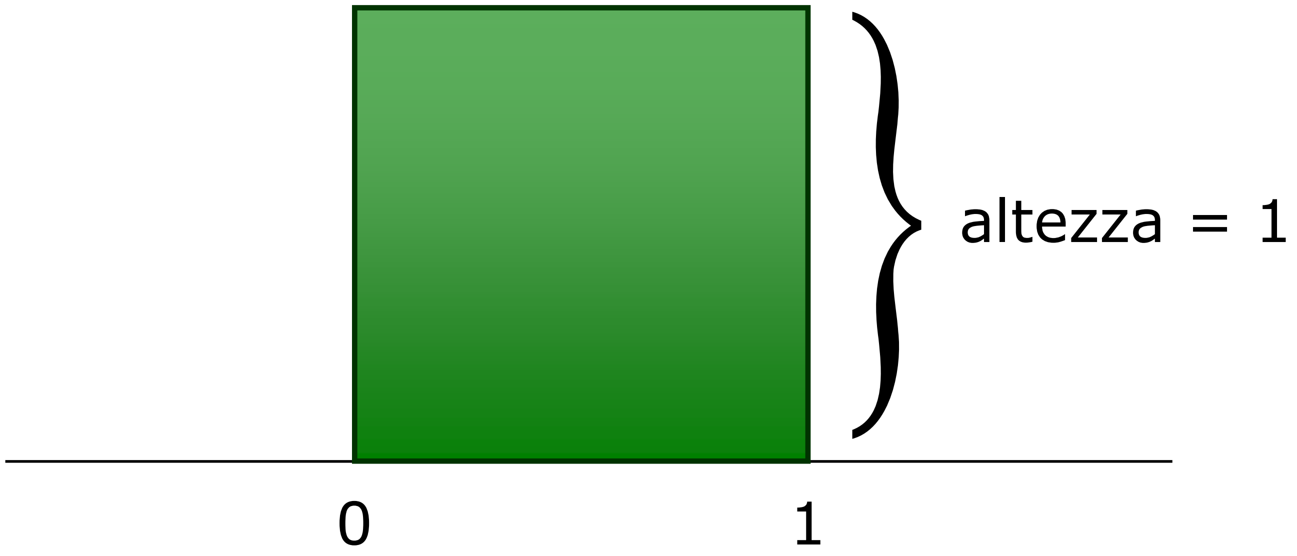
Esempi di curve di densità



Curva di densità: Media e mediana

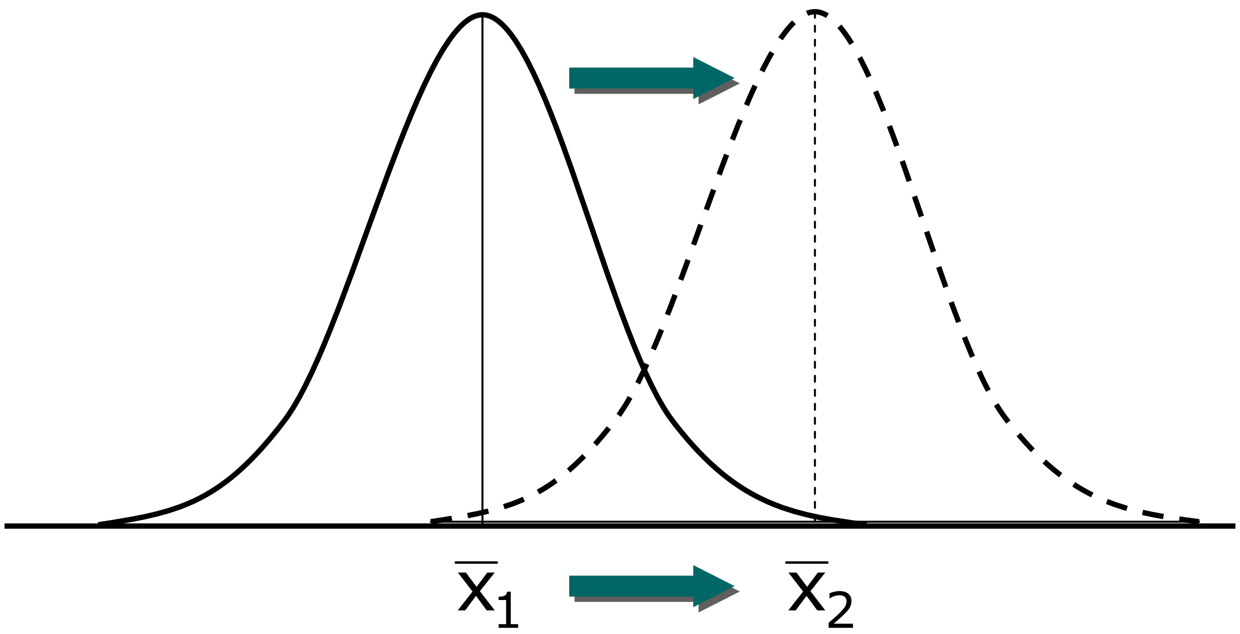
- La mediana di una curva di densità è il punto che divide l’area sotto la curva esattamente a metà
- La media di una curva di densità è il punto in cui, se la curva fosse di materiale solido, essa rimarrebbe in equilibrio

Distribuzione uniforme

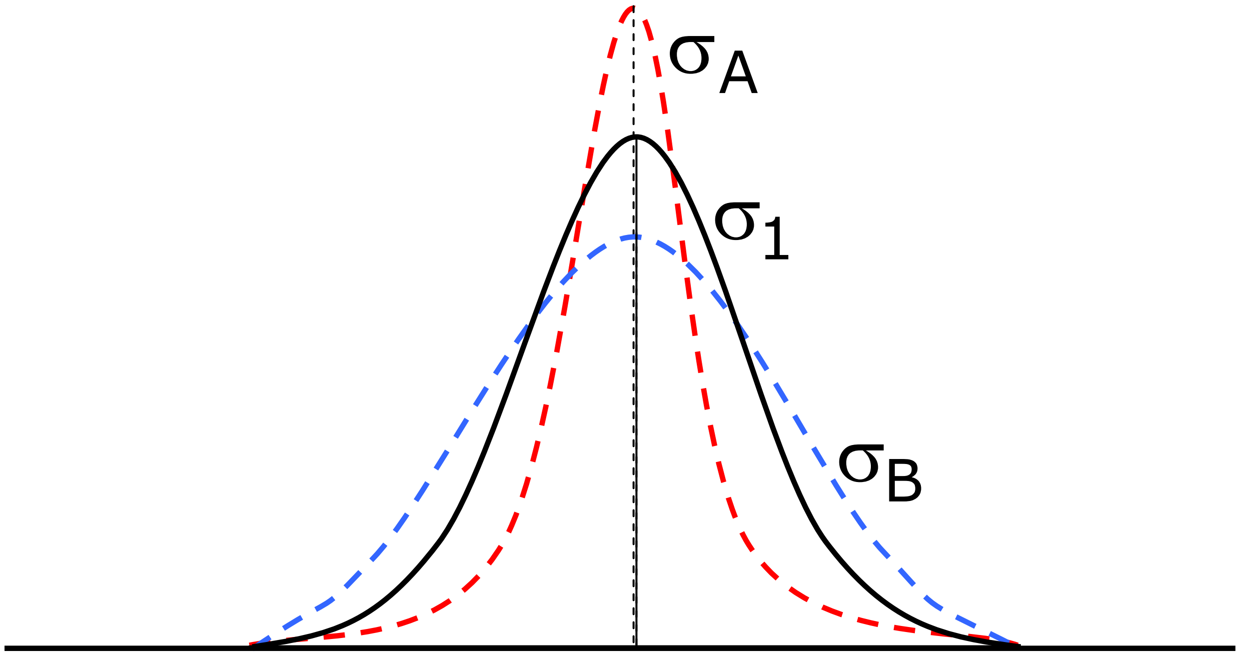


- Quanto vale la superficie totale sotto questa curva?
- Quale percentuale di osservazioni è al di sopra di 0,8?
- Quale percentuale di osservazioni è al di sotto di 0,6?
- Quale percentuale di osservazioni è fra 0,25 e 0,75?

Distribuzione normale

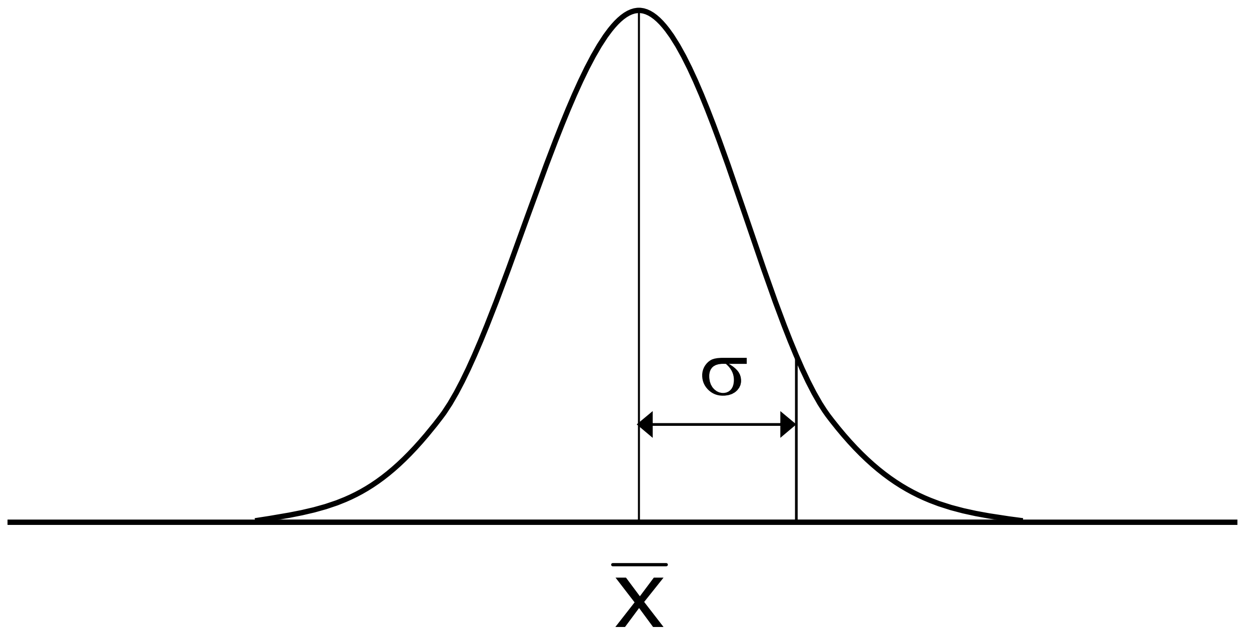


Distribuzione normale



$$\sigma_A < \sigma_1 < \sigma_B$$

Distribuzione normale



Importanza della Normale

- Le distribuzioni normali sono ottime rappresentazioni per alcune distribuzioni di dati reali
- Le distribuzioni normali sono ottime rappresentazioni dei risultati casuali
- Molte elaborazioni dell’inferenza statistica sono basate sulle distribuzioni normali

La regola della Normale

Nella distribuzione Normale con media $\bar{x}$ e deviazione standard σ:

- il 68% delle osservazioni è compreso nell'intervallo $[\bar{x}-\sigma, \bar{x}+\sigma]$
- il 95% delle osservazioni è compreso nell'intervallo $[\bar{x}-2\sigma, \bar{x}+2\sigma]$
- il 99,7% delle osservazioni è compreso nell'intervallo $[\bar{x}-3\sigma, \bar{x}+3\sigma]$

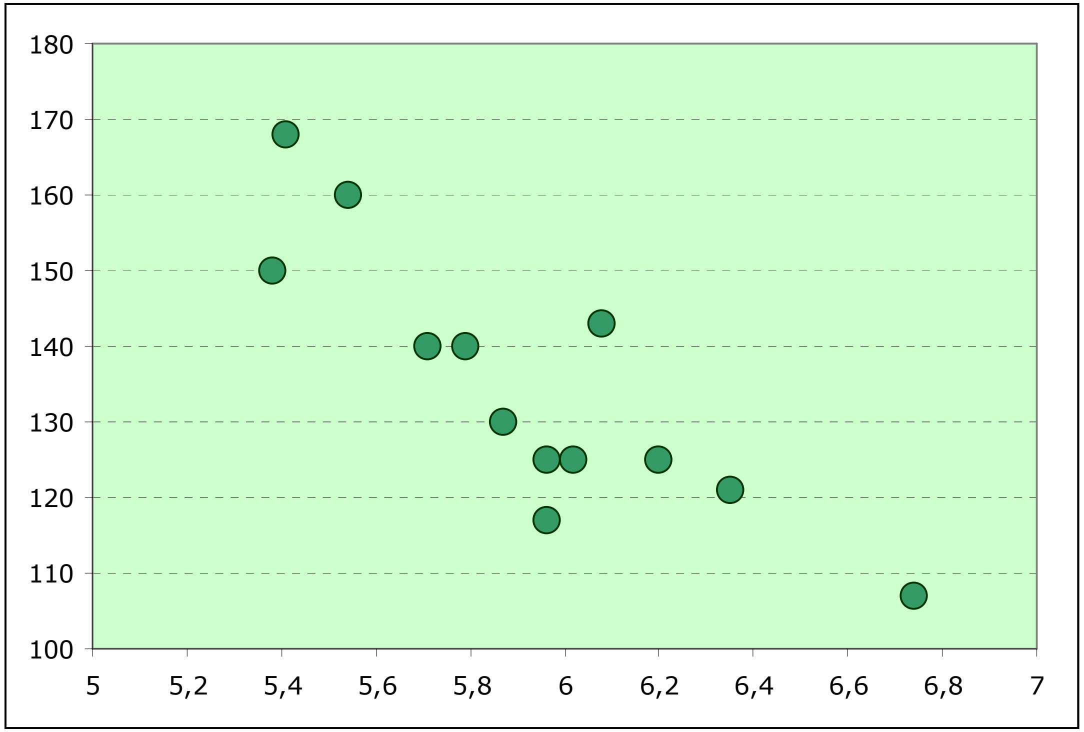
Standardizzazione

Se x è un'osservazione da una distribuzione con media $\overline{x}$ e deviazione standard σ, il valore standardizzato di x è:

$$\boxed{z = \frac{x - \overline{x}}{\sigma}}$$

Relazione tra caratteri quantitativi

X	Y
6,35	121
6,74	107
6,20	125
5,96	125
5,38	150
5,71	140
5,41	168
5,54	160
6,08	143
5,96	117
5,79	140
5,87	130
6,02	125

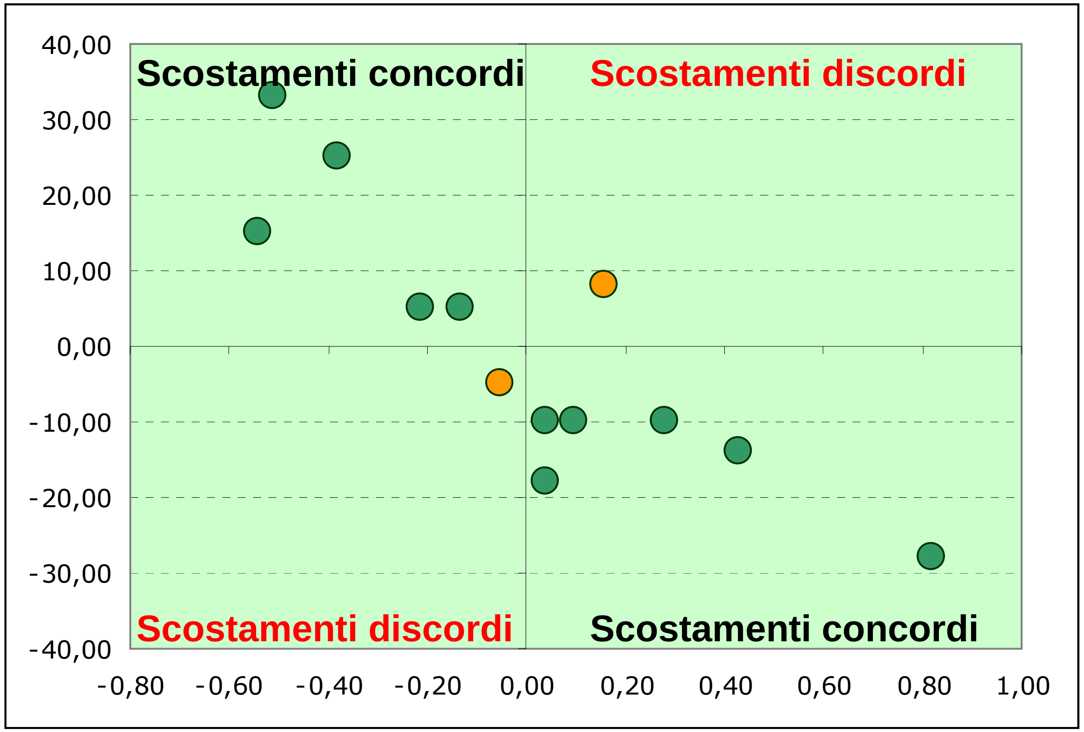


Relazione tra caratteri quantitativi

X	Y
0,43	-13,69
0,82	-27,69
0,28	-9,69
0,04	-9,69
-0,54	15,31
-0,21	5,31
-0,51	33,31
-0,38	25,31
0,16	8,31
0,04	-17,69
-0,13	5,31
-0,05	-4,69
0,10	-9,69

$$\bar{x} = 5,92$$

$$\bar{y} = 134,69$$



Concordanza e Discordanza

- Due caratteri quantitativi presentano concordanza se la maggior parte degli scostamenti sono concordi
- Al contrario, sussiste discordanza se la maggior parte degli scostamenti sono discordi
- Un indice simmetrico per misurare la concordanza o la discordanza è la covarianza

La covarianza

$$\sigma_{xy} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})(x_i - \overline{x})$$

$$\sigma_{xy} = \mathrm{Media}(xy) - \overline{y}\,\overline{x}$$

Il numeratore della covarianza è la codevianza

Calcolo della covarianza

X	Y
6,35	121
6,74	107
6,20	125
5,96	125
5,38	150
5,71	140
5,41	168
5,54	160
6,08	143
5,96	117
5,79	140
5,87	130
6,02	125

$$\sigma_{xy} = \text{Media}(xy) - \bar{y}\bar{x}$$

$$\bar{x} = 5{,}92$$

$$\bar{y} = 134{,}69$$

$$\sigma_{xy} = 792{,}685 - 5{,}92 \cdot 134{,}69$$

$$\sigma_{xy} = -5{,}268$$

Coefficiente di correlazione lineare

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

$$r = \frac{Codev_{xy}}{\sqrt{Dev_x Dev_y}}$$

Calcolo di r

$$\boxed{r = \frac{\sigma_{xy}}{\sigma_x \sigma_y}}$$

$$\sigma_{xy} = -5{,}268$$

$$\sigma_x = 0{,}3652$$

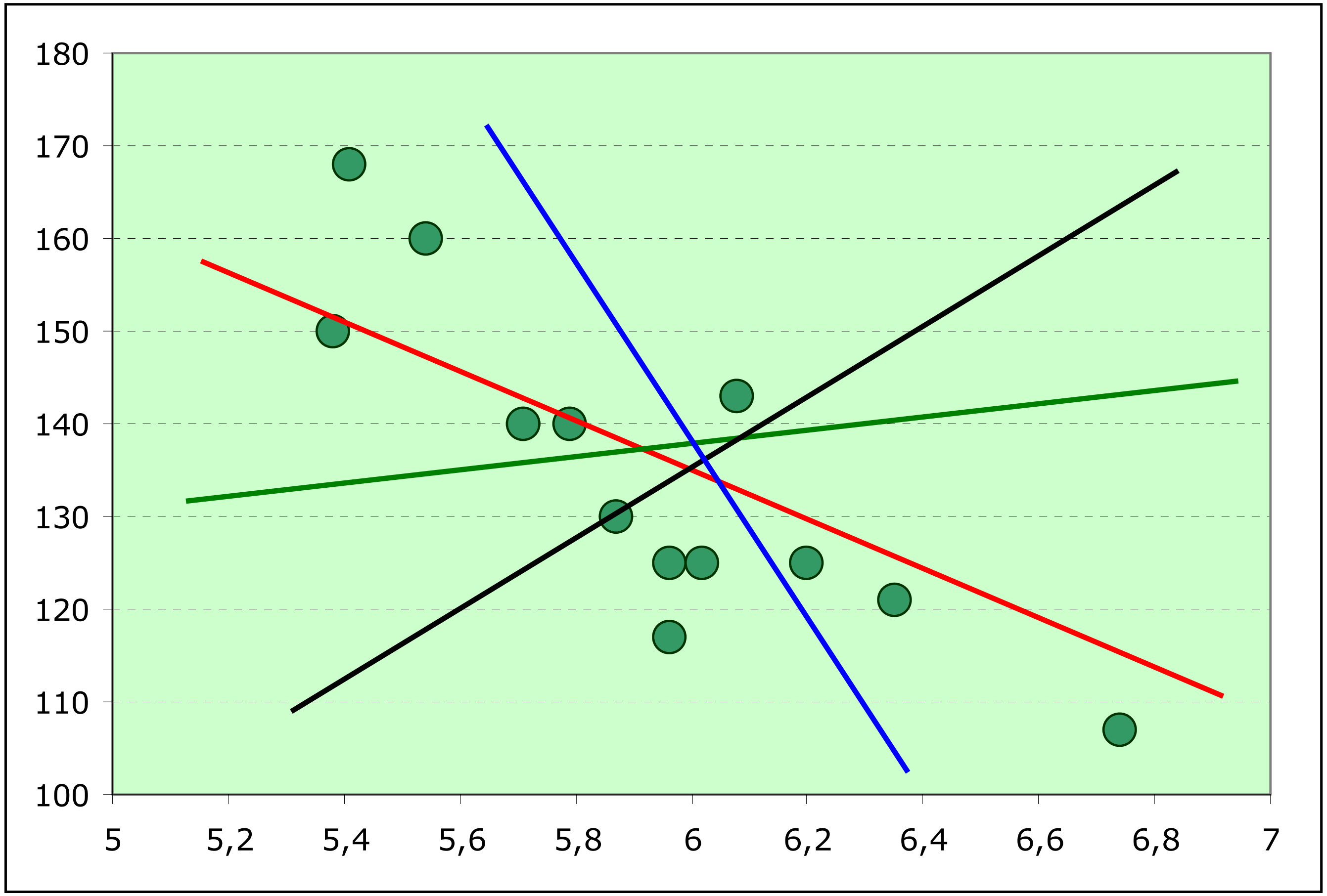
$$\sigma_{xy} = 16{,}8267$$

$$r = \frac{-5{,}268}{0{,}3652 \cdot 16{,}8267} = -0{,}8573$$

Proprietà di r

- $-1 \leq r \leq 1$
- r = 1 se tra X e Y sussiste un perfetto legame lineare e i due caratteri sono concordi
- r = -1 se tra X e Y sussiste un perfetto legame lineare e i due caratteri sono discordi
- r = 0 se i due caratteri sono linearmente indipendenti

Regressione



Parametri della retta di regressione

$$Y = b_0 + b_1 X$$

$$b_0 = \overline{y} - b_1 \overline{x}$$

$$b_1 = \frac{Codev_{xy}}{Dev_x}$$

Calcolo dei parametri della retta

$$\sigma_{xy} = -5{,}268 \qquad \bar{x} = 5{,}92$$

$$Dev_x = 1{,}734 \qquad \bar{y} = 134{,}69$$

$$Codev_{xy} = n \cdot \sigma_{xy} = 13 \cdot (-5{,}268) = -68{,}4846$$

$$b_1 = \frac{Codev_{xy}}{Dev_x} = \frac{-68{,}4846}{1{,}734} = \boxed{-39{,}495}$$

$$b_0 = \bar{y} - b_1\bar{x} = 134{,}69 - (-39{,}495) \cdot 5{,}92 = 368{,}5$$

La retta di regressione