



TECNICHE DI ANALISI DEI DATI

AA 2017/2018

PROF. V.P. SENESE

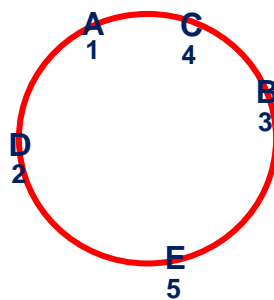
Questi materiali sono disponibili per tutti gli studenti al seguente indirizzo:

<https://goo.gl/hxL9zG>

Seconda Università di Napoli (SUN) – Dipartimento di Psicologia – TECNICHE DI ANALISI DEI DATI – © Prof. V.P. Senese

SCALA NOMINALE

Quando nel sistema empirico è possibile esclusivamente riconoscere una suddivisione in categorie distinte e mutualmente escludentesi allora numeri uguali vengono assegnati ad elementi uguali, mentre numeri diversi ad elementi diversi (*f*).



SCALA NOMINALE

PROPRIETÀ FORMALI:

equivalenza simmetrica:

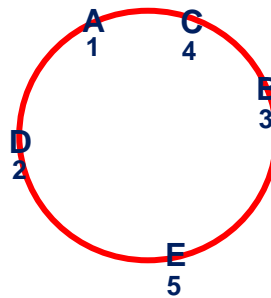
$A=B \triangleright B=A$;

equivalenza transitiva:

$A=B$ e $B=C \triangleright A=C$;

non equivalenza simmetrica:

$A \neq B \triangleright B \neq A$



SCALA NOMINALE

OPERAZIONI MATEMATICHE:

nessuna, solo **conteggio delle frequenze** per ogni categoria.

ANALISI STATISTICHE APPLICABILI:

analisi descrittive (frequenze) e test non parametrici che si basano sul conteggio delle frequenze.

SCALA NOMINALE

Sia \mathcal{A}_1 una classe di bambini **Italiani** e **Immigrati** ($N = 8$).

$f \Rightarrow$ Italiani = 1; Immigrati = 2

$\mathcal{B}_1 = \{1, 1, 2, 2, 1, 2, 1, 1\} \Rightarrow$ CITTAD. = 11... 11?!?

CITTAD.	Italiani	Immigrati
CODICE	1	2
FREQUENZA	5	3
FREQ. %	63%	37%

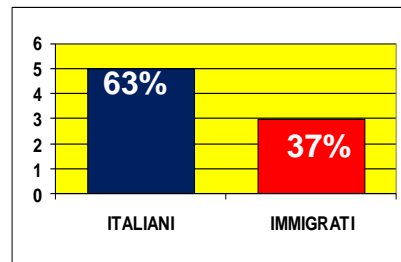


Grafico a barre variabile cittadinanza

SCALA ORDINALE

Quando nel sistema empirico tutti gli elementi godono di una proprietà ma in quantità o in grado diverso e sono **ordinabili** rispetto a tale grado, allora i numeri vengono assegnati in base all'ordine. In questo sistema numeri uguali indicano quantità uguali, mentre numeri diversi indicano esclusivamente una **graduatoria** (f).



SCALA ORDINALE

PROPRIETÀ FORMALI:

equivalenza (nominale)

relazione d'ordine asimmetrica:

$A < B \triangleright B > A$ e non $B < A$

relazione d'ordine transitiva:

$A < B$ e $B < C \triangleright A < C$



SCALA ORDINALE

OPERAZIONI MATEMATICHE:

nessuna, solo **conteggio delle frequenze** per ogni categoria.

ANALISI STATISTICHE APPLICABILI:

analisi descrittive (frequenze) e test non parametrici che si basano sul conteggio delle frequenze e sugli indici di posizione.

SCALA ORDINALE

Sia \mathcal{A}_1 una Scuola Media ($N = 70$).

$f \Rightarrow$ I MEDIA = 1; II MEDIA = 2; III MEDIA = 3

$\mathcal{B}_1 = \{1, 1, 2, 2, 1, 2, 1, 1, 3, 3, 2, 2, 1, 1, 2, 2, 2, 3, \dots, 1, 3, 2, 2, 1, \dots\} \Rightarrow$ CLASSE = 132... 132?!?

CLASSE	I MED.	II MED.	III MED.
CODICE	1	2	3
FREQUENZA	23	32	15
FREQ. %	33%	46%	21%
FREQ. CUM.	23	55	70
FREQ. C.%	33%	79%	100%

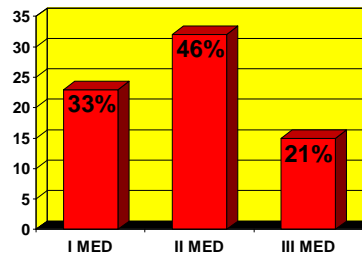


Grafico a barre variabile classe

SCALE NON PARAMETRICHE

Le variabili misurate su scala Nominale e Ordinale sono dette:

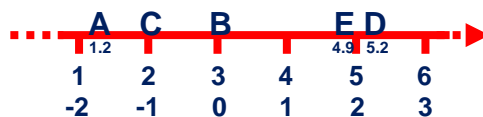
VARIABILI QUALITATIVE o NON PARAMETRICHE

VARIABILI **DICOTOMICHE**
con due categorie
(es. Si/No; M/F; V/F; ecc.)

VARIABILI **POLICOTOMICHE**
con più di due categorie
(es. B/M/A; I/II/III; ecc.)

SCALA INTERVALLI

Quando nel sistema empirico tutti gli elementi godono di una proprietà ma in quantità o in grado diverso ed è possibile stabilire un'**unità di misura** della proprietà, allora i numeri vengono assegnati in base alla differenza di intensità della proprietà. In questo sistema numeri uguali indicano quantità uguali, mentre numeri diversi indicano il **grado di presenza dell'unità stabilita** (*f*).



SCALA INTERVALLI

PROPRIETÀ FORMALI:

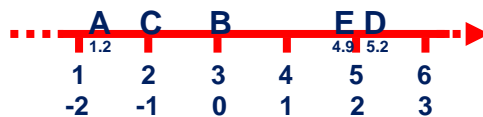
equivalenza (NOMINALE);

relazione d'ordine (ORDINALE);

costanza del rapporto tra intervalli:

tra 1 e 2 c'è la stessa distanza che tra 3 e 4

$3=2+1$ o $4=3+1$



SCALA INTERVALLI

OPERAZIONI MATEMATICHE:

Addizione, sottrazione e moltiplicazione, ma non è possibile stabilire rapporti diretti tra le misure (es., non si può dire che “4” è il doppio di “2”).

ANALISI STATISTICHE APPLICABILI:

analisi descrittive, test non parametrici e test parametrici.

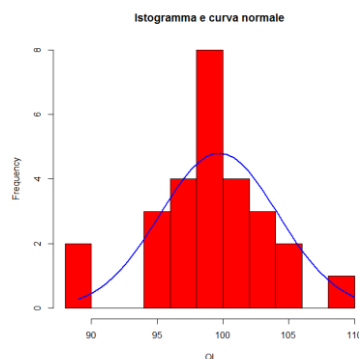
SCALA INTERVALLI

Sia \mathcal{A}_1 un gruppo di studenti universitari ($N = 27$).

$f \Rightarrow$ punteggio ad un test d'intelligenza

$\mathcal{B}_1 = \{89, 89, 95, 95, 95, 98, 98, 98, 98, 100, 100, 100, 100, 100, 100, 100, 100, 100, 102, 102, 102, 102, 103, 103, 103, 105, 105, 110\} \Rightarrow QI = 2692... 2692!$

QI	FREQ	FREQ%
89	2	7
95	3	11
98	4	14
100	8	29
102	4	14
103	3	11
105	2	7
110	1	4
112	1	4
TOT	27	100



SCALA A RAPPORTI

Quando nel sistema empirico tutti gli elementi godono di una proprietà ma in quantità o in grado diverso ed è possibile stabilire sia un'unità di misura della proprietà, sia un elemento di intensità nulla, allora i numeri vengono assegnati in base alla differenza di **intensità della proprietà** (f) e il sistema numerico adottato godrà di tutte le proprietà dei numeri reali.



SCALA A RAPPORTI

PROPRIETÀ FORMALI:

equivalenza (NOMINALE);

relazione d'ordine (ORDINALE);

costanza del rapporto tra intervalli (INTERVALLI);

costanza del rapporto tra valori:

ad es. 4 è il doppio di 2

$4/2=2$ e $8/4=2$

non può assumere valori negativi



SCALA A RAPPORTI

OPERAZIONI MATEMATICHE:

tutte anche la divisione.

ANALISI STATISTICHE APPLICABILI:

analisi descrittive, test non parametrici e test parametrici.

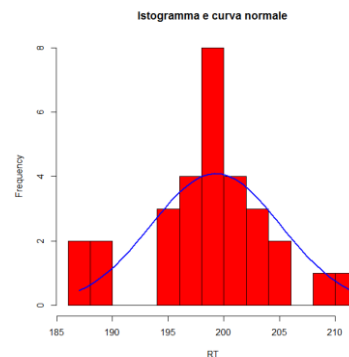
SCALA A RAPPORTI

Sia A_1 un gruppo di adulti ($N = 30$).

$f \Rightarrow$ Tempi di reazione ad uno stimolo (RT)

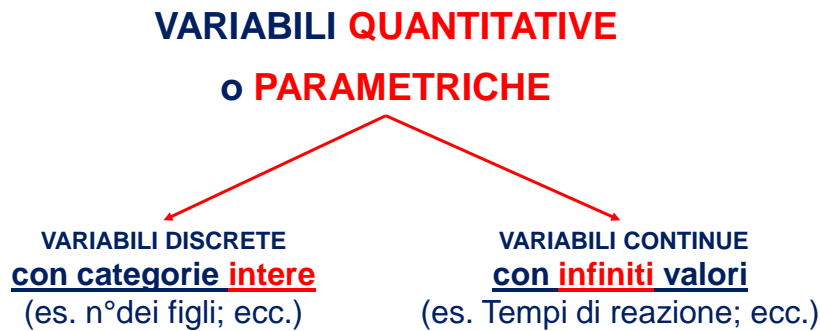
$B_1 = \{187, 187, 189, 189, 195, 195, 195, 198, 198, 198, 198, 200, 200, 200, 200, 200, 200, 200, 200, 200, 202, 202, 202, 202, 202, 202, 203, 203, 203, 205, 205, 210, 212\} \Rightarrow RT = 5978 \dots 5978!$

RT	FREQ	FREQ%	F % C
187	2	7	7
189	2	7	14
195	3	10	24
198	4	13	37
200	8	27	64
202	4	13	77
203	3	10	87
205	2	7	94
210	1	3	97
212	1	3	100
TOT	30	100	--



SCALE PARAMETRICHE

Le variabili misurate su scala ad **Intervalli** e a **Rapporto** sono dette:



SCALE DI MISURA

- 1 Ogni scala possiede le caratteristiche di quella che la precede.
- 2 È possibile **trasformare** una misurazione ottenuta su una scala di livello superiore in una di livello inferiore, mentre non è vero il contrario.
- 3 I **test statistici** applicabili alle scale di livello inferiore sono utilizzabili anche per scale di livello superiore, mentre non è vero il contrario.
- 4 La scala di misura utilizzata per una variabile può dipendere dalla **definizione operativa** della variabile e dal suo uso all'interno della ricerca.

SCALE DI MISURA

In alcuni casi si rende necessario trasformare una variabile misurata ad un livello di scala superiore in un livello di scala inferiore. Sebbene questa trasformazione sia possibile bisogna ricordarsi che c'è una perdita di informazioni.

Scala a
ORDINALE

SOGG.	LATENZA	
10	34	
04	30	= 1
07	28	= 1
06	28	
03	25	◀
09	21	= 2
08	15	
12	15	= 2
11	14	
05	14	= 3
01	13	
02	12	

MISURAZIONE

Il Signor M. ha 35 anni. È alto m 1.82 e pesa 75 kg. È biondo. Ha gli occhi azzurri. È laureato. Risiede in una città del nord Italia. Attualmente lavora come pilota automobilistico. È sposato. Ha un QI pari a 120.



Il Signor M. ha 35 anni. È alto m 1.82 e pesa 75 kg. È biondo. Ha gli occhi azzurri. È laureato. M. risiede in una città del nord Italia. Attualmente lavora come pilota automobilistico. È sposato. Ha un QI pari a 120.

Come possiamo trasformare questi dati in una miscela empirica in una MISURA UNA FUNZIONE



MISURAZIONE

Il Signor M. ha 35 anni. È alto m 1.82 e pesa 75 kg. È biondo. Ha gli occhi azzurri. È laureato. M. risiede in una città del nord Italia. Attualmente lavora come pilota automobilistico. È sposato. Ha un QI pari a 120.

Definiamo le **scale di misura** e le f

SESSO	N - (1 = M; 2 = F)
ETÀ	R - (x = ANNI)
ALTEZZA	R - (x = CM)
PESO	R - (x = KG)
COL. CAPELLI	N - (1 = CHI.; 2 = SCU.)
COL. OCCHI	N - (1 = CHI.; 2 = SCU.)
TIT. STUDIO	O - (0 = N; 1 = M; 2 = D; 3 = L)
RESIDENZA	N - (1 = N; 2 = C; 3 = S; 4 = I)
LAVORO	N - (0 = NO; 1 = SI)
STATO CIVILE	N - (0 = C/N; 1 = S; 2=S/D)
QI	I - (x = PUNT. QI)

ID	001
S	1
E	35
A	1.82
P	75
CC	1
CO	1
TS	3
R	2
L	1
SC	1
QI	120

MISURAZIONE

Il Signor R. ha 28 anni. È alto m 1.74 e pesa 80 kg. È castano. Ha gli occhi castani. È diplomato. R. risiede in una città del sud Italia. Attualmente è disoccupato. Non è mai stato sposato. Ha un QI pari a 122.

SESSO	(1 = M; 2 = F)
ETÀ	(x = ANNI)
ALTEZZA	(x = CM)
PESO	(x = KG)
COL. CAPELLI	(1 = CHI.; 2 = SCU.)
COL. OCCHI	(1 = CHI.; 2 = SCU.)
TIT. STUDIO	(0 = N; 1 = M; 2 = D; 3 = L)
RESIDENZA	(1 = N; 2 = C; 3 = S; 4 = I)
LAVORO	(0 = NO; 1 = SI)
STATO CIVILE	(0 = C/N; 1 = S)
QI	(x = PUNT. QI)

ID	001	002	...	n
S	1	1
E	35	28
A	1.82	1.74
P	75	80
CC	1	2
CO	1	2
TS	3	2
R	2	3
L	1	0
SC	1	0
QI	120	122

MATRICE **VARIABILI**x**SOGGETTI**

MISURAZIONE

	A	B	C	Y	Z
1	ID	Age	Sex	PARQ_M1	PARQ_M2
2	1	25	1	2	1
3	2	24	0	2	1
4	3	22	0	1	1
5	4	21	0	1	1
6	5	25	0	1	1
7	6	24	0	2	1
8	7	25	0	1	1
9	8	25	0	2	1
10	9	25	0	2	1

Db **SOGGETTI**x**VARIABILI**

Formato **standard** per la memorizzazione dei dati da analizzare statisticamente (**Db**).

File: xlsx; csv; txt, ecc.

	A	B
1	ID	Identification number
2	Age	Age (Years)
3	Sex	Gender: 1=M; 0=F
25	PARQ_M1	Item 1 - PARQ Mother: 1="Qua
26	PARQ_M2
27	PARQ_M3	
28	PARQ_M4	
29	PARQ_M5	
30	PARQ_M6	
31	PARQ_M7	

Legenda

MISURAZIONE

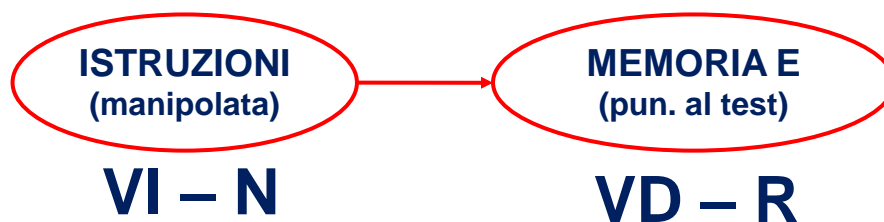
La **misurazione** non deve essere fine a se stessa, ma deve essere guidata in base a degli specifici **obiettivi** regolati in base a delle considerazioni **teoriche** che hanno portato alla formulazione di **ipotesi scientifiche**.

IMPORTANTE:

- posso analizzare solo ciò che ho **misurato**;
- il **tipo di analisi** che posso fare dipende da **ciò** che ho misurato e da **come** ho raccolto i dati;
- per ciascun quesito teorico devo trovare la **giusta tecnica di analisi**.

ESEMPIO #1

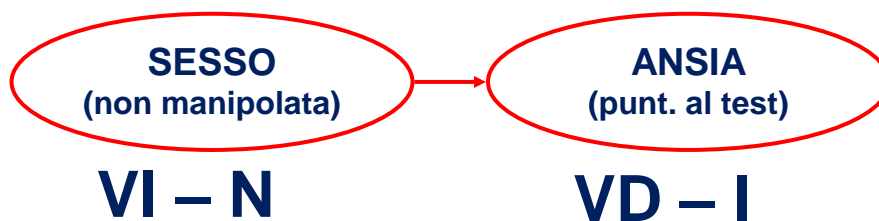
40 adolescenti vengono convocati per un esperimento di memoria episodica (ME – lista di parole). Vengono divisi in modo **casuale** in due gruppi da 20. Al primo gruppo, prima di iniziare la prova, viene detto che si tratta di un compito **molto difficile**, mentre al secondo gruppo non viene data alcuna indicazione. Viene valutato il numero di parole ricordate correttamente.



ESEMPIO #2

Allo scopo di valutare l'effetto del sesso sull'ansia di tratto ad un gruppo di **50 soggetti** (25 M e 25 F) è stato somministrato un questionario di valutazione dell'ansia di tratto.

Quali sono le variabili considerate?

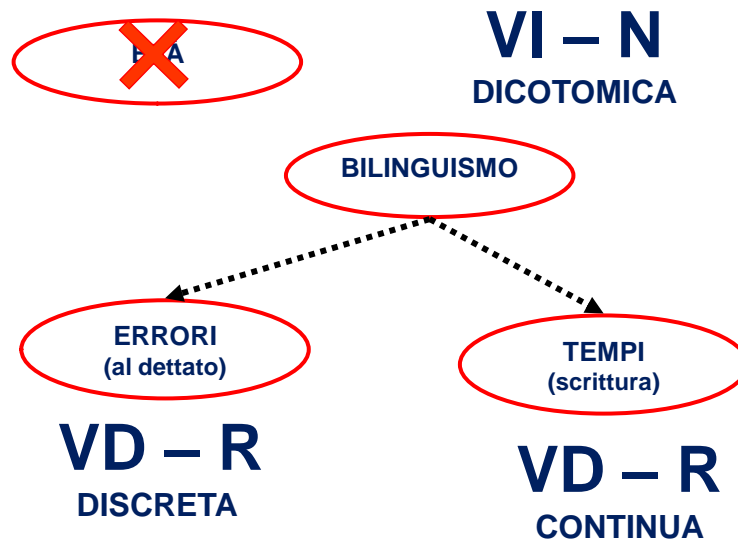


ESEMPIO #3

Ad un gruppo di **30 bambini** italiani di **10 anni**, 15 dei quali bilingue, viene somministrata una **prova di scrittura in italiano** (vengono dettate 35 parole) e si valutano gli errori commessi e i tempi impiegati nello scrivere ciascuna parola.

Indicare le variabili oggetto di studio, la relazione e il livello di misura.

ESEMPIO #3



ESEMPIO #3

Dopo la prova risulta che uno dei bambini ha commesso 22 errori...

...22?

INDICI DI TENDENZA CENTRALE

Una misura di **tendenza centrale** è:

un indice o statistica che rappresenta un insieme di misure mediante un unico valore.

LA MODA LA MEDIANA LA MEDIA

LA MODA

La **MODA** è:

*un indice o statistica che indica la categoria numerica o le categorie che si presentano con **maggior frequenza**.*

APPLICAZIONE

È una statistica significativa per **tutte** le scale di misura:

- Nominali
- Ordinali
- Intervalli
- Rapporti

LA MODA

In un campione di **30 studenti** universitari è stata rilevata la variabile "**Facoltà**" il cui dominio è: **Giurisprudenza** (G=1); **Medicina** (M=2); **Lettere** (L=3); **Psicologia** (P=4)

$\mathcal{B}_1 = \{1, 1, 1, 2, 2, 4, 4, 3, 3, 4, 4, 4, 3, 1, 1, 3, 3, 3, 1, 1, 2, 2, 2, 1, 1, 1, 1, 1, 2, 1\}$

TABELLA DI FREQUENZE

FAC.	FREQ	FREQ%
1	13	43%
2	6	20%
3	6	20%
4	5	17%
TOT	30	100%

Moda = 1

Moda = Giurisprudenza

LA MEDIANA

La **MEDIANA** è:

*un indice o statistica che indica la categoria numerica che divide una distribuzione (ordinata per valore) a **metà**.*

APPLICAZIONE

È una statistica **significante** per le scale di misura:

- Ordinali
- Intervalli
- Rapporti

Non è significativa per le scale

- Nominali

LA MEDIANA

Se il numero delle osservazioni (n) è **dispari**:

$$Me = \frac{n + 1}{2}$$

Se il numero delle osservazioni (n) è **pari**:

$$Me = \frac{n}{2} < i < \frac{n}{2} + 1$$

dati non raggruppati

$$i_{Me} = \frac{n}{2} + \left(\frac{n}{2} - fc_{inf} \right) \cdot a$$

$$Me = l_{inf} + \frac{f_{i_{Me}}}{f_{i_{Me}}}$$

dati raggruppati

LA MEDIANA

In un campione di **70 ragazzi** è stata rilevata la variabile “Classe” il cui dominio è: I Media (=1); II Media (=2); III Media (=3).

$\mathcal{B}_1 = \{1, 1, 2, 2, 1, 2, 1, 1, 3, 3, 2, 2, 1, 1, 2, 2, 2, 3, \dots, 1, 3, 2, 2, 1, \dots\} \Rightarrow$ **ORDINARE** \Rightarrow
 $\mathcal{B}_1 = \{1, 1, 1, 1, 1, 1, 1, 1, 1, \dots, 2, 2, 2, 2, 2, 2, 2, 2, \dots, 3, 3, 3, 3, \dots\}$

$N = 70$

TABELLA DI FREQUENZE

CLASSE	FREQ	FREQ%	FREQ C	FREQ C%
1	23	33%	23	33%
2	32	46%	55	79%
3	15	21%	70	100%
TOT	70	100%	-	

LA MEDIANA

TABELLA DI FREQUENZE

CLASSE	FREQ	FREQ%	FREQ C
1 (.5 - 1.5)	23	33%	23
2 (1.5 - 2.5)	32	46%	55
3 (2.5 - 3.5)	15	21%	70
TOT	70	100%	

$$i_{Me} = \frac{n}{2} \quad Me = l_{inf} + \frac{\left(\frac{n}{2} - fc_{inf}\right) \cdot a}{f_{i_{Me}}}$$

$$N = 70$$

$$i_{Me} = 35$$

$$l_{inf} = 1.5$$

$$fc_{inf} = 23$$

$$f_{i_{Me}} = 32$$

$$a = 1.5 - 0.5$$

$$i_{Me} = \frac{70}{2} = 35$$

$$Me = 1.5 + \frac{\left(\frac{70}{2} - 23\right) \cdot (2.5 - 1.5)}{32} =$$

$$= 1.5 + \frac{(12) \cdot (1)}{32} = 1.9 \approx 2$$

$$\mathbf{Me = 2}$$

LA MEDIA

La **MEDIA** è:

un indice o statistica che corrisponde alla **somma** di tutti i numeri di una distribuzione **diviso** il numero di osservazioni.

APPLICAZIONE

È una statistica **significante** per le scale di misura:

- Intervalli
- Rapporti

Non è significativa per le scale di misura:

- Nominali
- Ordinali

LA MEDIA

MEDIA = \bar{x} o μ (*mu*)

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

dati non raggruppati

N = osservazioni
 x_i = valori osservati
 Σ = sommatoria

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

dati raggruppati

x_i = valori osservati
 f_i = frequenze del valore x_i
 Σ = sommatoria

LA MEDIA

Ad un campione di **30 ragazzi** è stata presentata uno stimolo (la parola "Classe") e sono stati registrati e tempi di reazione (**RT**) in un compito di produzione.

B_1 = {187, 187, 189, 189, 195,195, 195, 198, 198, 198, 198, 200, 200, 200, 200, 200, 200, 200, 202, 202, 202, 202, 203, 203, 203, 205, 205, 210}

RT	FREQ	FREQ _c	$f x_i$
187	2	2	374
189	2	4	378
195	3	7	585
198	4	11	792
200	8	19	1600
202	4	23	808
203	3	26	609
205	2	28	410
210	1	29	210
212	1	30	212
Σ	30		5978

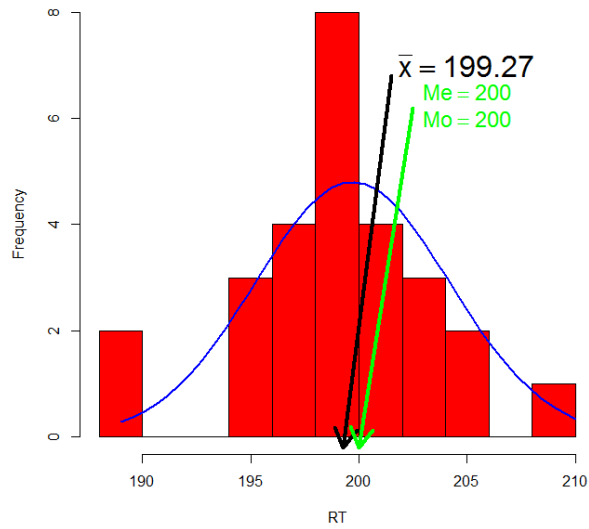
$$\bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} = \frac{\sum_{i=1}^n x_i}{n} = \frac{5978}{30} = 199.27$$

Moda = 200
Me = 200

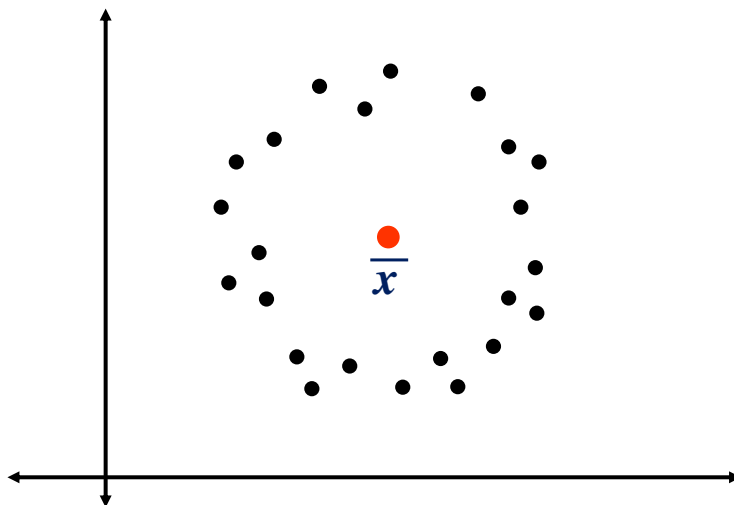
$\bar{x} = 199.27$

LA MEDIA

Indici di tendenza centrale



INDICI DI TENDENZA CENTRALE



INDICI DI DISPERSIONE

Una misura di **dispersione** è:

*un indice o statistica che rappresenta la **variabilità** delle osservazioni campionate.*

- **NdE**
- **Gamma**
- **Scostamento S.**
- **Scarto quadratico**
- **σ^2 (S²) e σ (S)**

NdE

Corrisponde al **numero di classi** (quindi di valori) con cui si è manifestato il fenomeno indagato.

REPARTO	A	B	C	D	E	LUNEDÌ
INGRESSI	12	1	3	7	4	NdE = 5
REPARTO	A	B	C	D	E	GIOVEDÌ
INGRESSI	12	-	3	7	-	NdE = 3

GAMMA

Corrisponde alla grandezza dell'**intervallo** nel quale sono contenuti i dati (**quantitativi**) di una distribuzione.

$$\text{Gamma} = X_{max} - X_{min}$$

SCOSTAMENTO SEMPLICE

Corrisponde alla sommatoria degli scarti dalla media.

$$\text{Scostament o S.} = \frac{\sum_{i=1}^n |x_i - \bar{X}|}{n}$$

SCARTO QUADRATICO DEVIANZA

Corrisponde alla sommatoria degli **scarti dalla media**, al quadrato.

$$\text{Scarto quadratico} = \sum_{i=1}^n (x_i - \bar{X})^2$$

VARIANZA

Corrisponde allo scarto quadratico **medio**.

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

s^2 = varianza campionari a

σ^2 (*sigma*) = varianza della popolazione e

DEVIAZIONE STANDARD

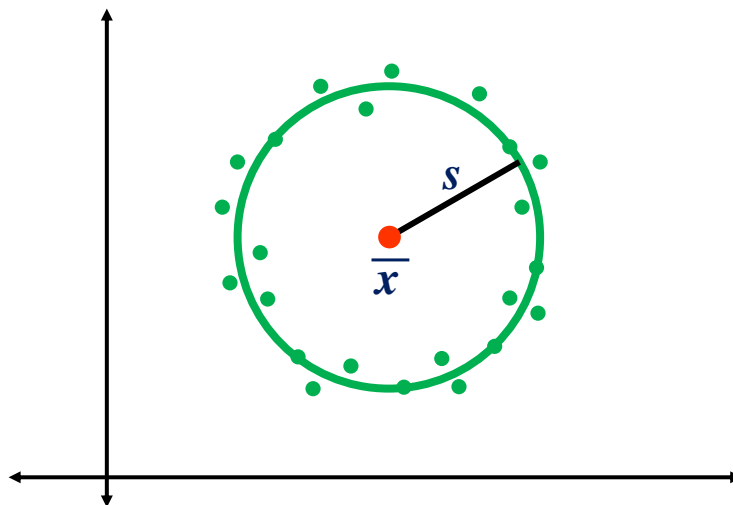
Corrisponde alla radice quadrata dello scarto quadratico medio.

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

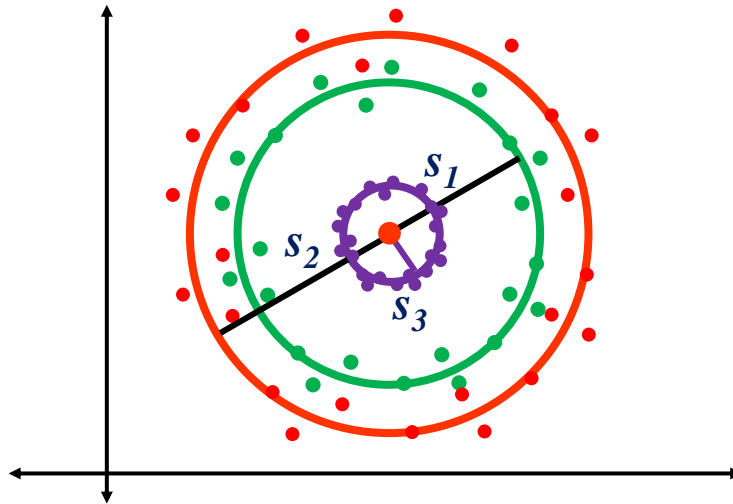
s = dev. st. campionari a

σ (*sigma*) = dev. st. della popolazione e

DEVIAZIONE STANDARD



DEVIAZIONE STANDARD



INDICI DI POSIZIONE

Una misura di **POSIZIONE** è:

un indice o statistica che rappresenta la posizione di una data osservazione all'interno di una distribuzione ordinata.

Consentono di definire la **percentuale** di valori che si trova al di sopra o al di sotto del valore di interesse (**quantili**).

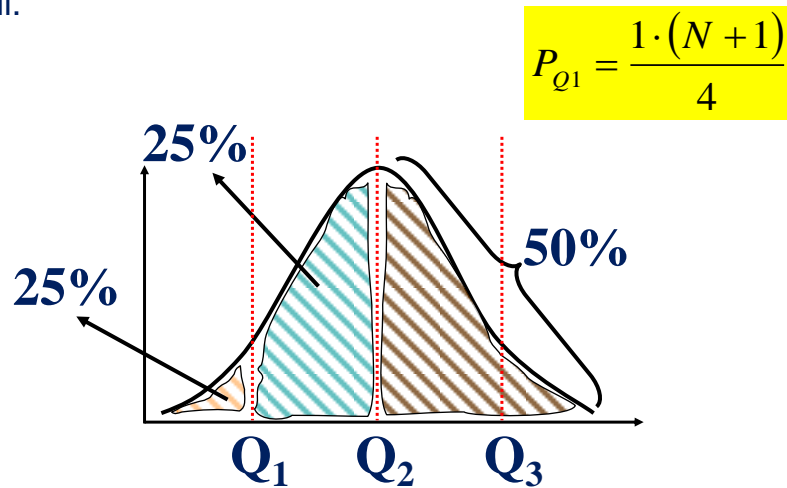
- **QUARTILI**
- **PERCENTILI**
- **DECILI**
- ...

$$P_q = \frac{\text{Quantile} \cdot (N + 1)}{n \cdot \text{Quantili}}$$

$$P_x = \frac{\text{Posizione } (x) \cdot n \cdot \text{Quantili}}{(N + 1)}$$

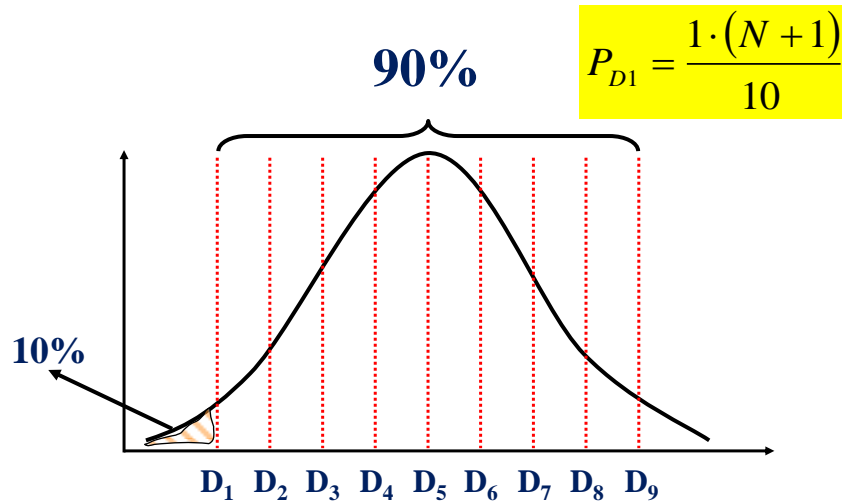
QUARTILI

Dividono una distribuzione **ordinata** in **quattro** parti uguali.



DECILI

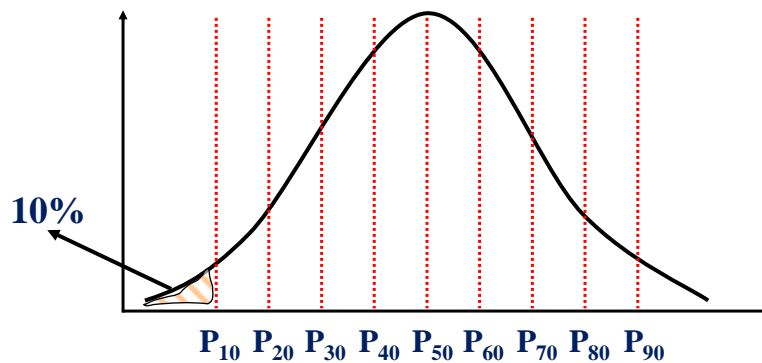
Dividono una distribuzione **ordinata** in **dieci** parti uguali.



DECILI

Dividono una distribuzione **ordinata** in **cento** parti uguali.

$$P_{P_{10}} = \frac{10 \cdot (N + 1)}{100}$$

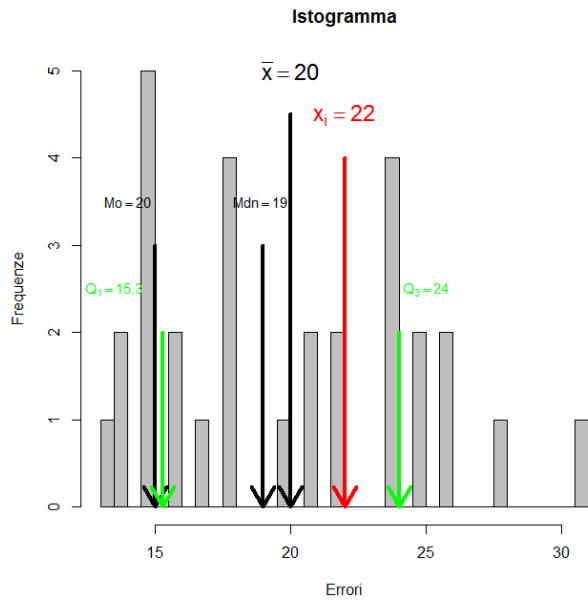


ESEMPIO #3

x	f	fc	$fc\%$
13	1	1	3.3
14	2	3	10.0
15	5	8	26.7
16	2	10	33.3
17	1	11	36.7
18	4	15	50.0
19	0	15	50.0
20	1	16	53.3
21	2	18	60.0
22	2	20	66.7
23	0	20	66.7
24	4	24	80.0
25	2	26	86.7
26	2	28	93.3
27	0	28	93.3
28	1	29	96.7
29	0	29	96.7
30	0	29	96.7
31	1	30	100.0

- $N = 30$
- $x_i = 22$
- $X_{Min} = 13$
- $X_{Max} = 31$
- $\text{Gamma} = 18$
- $NdE = 14$
- **Media = 20**
- $Mdn = 19$
- $\text{Moda} = 15$
- $Q1 = 15.3$
- $Q3 = 24$
- $P_{x=22} = 65^\circ$
- $s^2 = 25$
- $s = 5$

ESEMPIO #3



- $N = 30$
- $x_i = 22$
- $X_{Min} = 13$
- $X_{Max} = 31$
- $Gamma = 18$
- $NdE = 14$
- **Media = 20**
- $Mdn = 19$
- $Moda = 15$
- $Q1 = 15.3$
- $Q3 = 24$
- $P_{x=22} = 65^\circ$
- $s^2 = 25$
- $s = 5$