

# STATISTICA PER LA RICERCA APPLICATA

Dott.ssa Marta Di Nicola

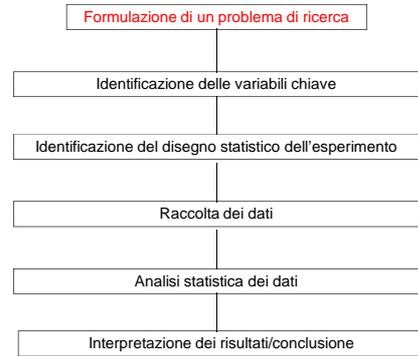
N.P.D. 3° Blocco 2° piano

0871-3554007

m.dinicola@unich.it

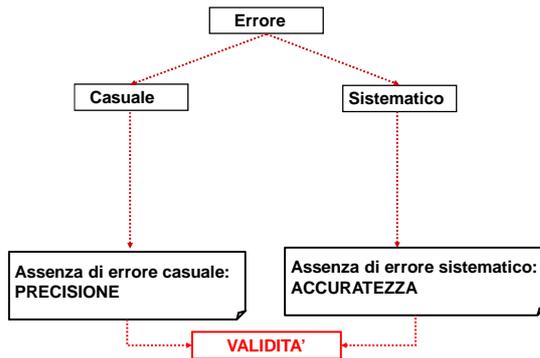
<http://www.biostatistica.unich.it>

1



2

## Gli errori nella ricerca clinica



## Formulazione di un problema di ricerca

Definire chiaramente la domanda cui vogliamo dare risposta e darne una adeguata giustificazione

Definire la domanda di studio in modo che possa essere analizzabile in maniera corretta e non ambigua

- Trasformare concetti vaghi (es. fumare di meno, migliorare la prognosi) in definizioni operative che descrivono le misure che si possono fare e che saranno poi valutate
- Tradurre la domanda di studio in affermazioni relative ad attributi (ad es. la media) della popolazione

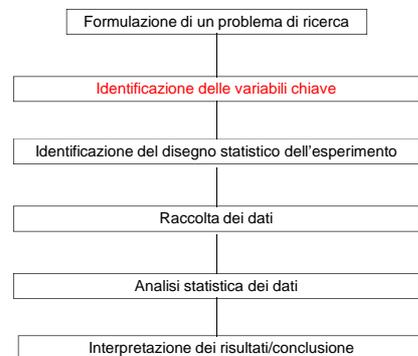
## LA STATISTICA

• La Statistica ha come scopo la conoscenza quantitativa dei fenomeni collettivi. L'analisi statistica mira ad individuare modelli di interpretazione della realtà, attraverso canoni e tecniche che sono astrazioni, semplificazioni di una moltitudine di aspetti e di manifestazioni del reale.

• E' costituita da un insieme dei metodi che consentono di raccogliere, ordinare, riassumere, presentare ed analizzare dati e informazioni, trarne valide conclusioni e prendere decisioni sulla base di tali analisi e risultati.

Dott.ssa Marta Di Nicola

5



6

## GLOSSARIO

**POPOLAZIONE:** L'insieme di **tutti** gli elementi (unità statistiche) oggetto dell'osservazione che hanno una o più caratteristiche comuni. Gli elementi possono essere soggetti, oggetti o eventi.

Esistono popolazioni **FINITE** e popolazioni **INFINITE**.

- Tutti i bambini italiani con problemi di carie nella prima infanzia
- Tutti gli studenti iscritti ad un corso di Laurea nel 2013
- Tutte le compresse di vitamine di un lotto di produzione

**CAMPIONE:** La parte delle unità statistiche sottoposte all'osservazione.

**CARATTERE (O VARIABILE):** Ogni caratteristica di un unità statistica che può essere misurata.

Sex	Status of Oral Hygiene	Level of Post-Surgery Pain
F = female	P = poor	N = no pain
M = male	F = fair	M = mild pain
	G = good	S = severe pain
		E = extremely severe pain

Subject No.	Age (yrs.)	BP (mm Hg)	Pocket Depth (mm)	Cholesterol (mg/dl)
1	56	121/76	6.0	167
2	43	142/95	5.5	180
...	...	...	...	...
115	68	175/124	6.5	243

Note: BP, blood pressure.

Se una variabile può assumere più di un valore con determinate probabilità questa è detta **RANDOM** o **ALEATORIA**.

**MODALITÀ:** Ogni diversa presentazione del carattere o variabile osservata su ciascuna unità statistica.

## Esempi di variabili statistiche

Le **variabili sesso, età, peso, pressione arteriosa, etc** (di pazienti inclusi in uno studio) hanno come **modalità**:

- **maschio** o **femmina** per la variabile "sesso";
- **anni**, per la variabile "età";
- **Kg**, per il "peso corporeo",
- **mmHg**, per la "pressione arteriosa"
- **A, AB, B, O** per il "gruppo sanguigno",
- **elementare, media inferiore, media superiore, università**, per la variabile "titolo di studio"

I dati sperimentali (variabili) si presentano sotto differenti forme, essi possono essere sia di tipo quantitativo sia di tipo qualitativo, ed essere espressi con scale continue o con scale discrete. In particolare:

### VARIABILI QUALITATIVE

### VARIABILI QUANTITATIVE

#### NOMINALI

Date due qualsiasi modalità, è possibile solo affermare se esse sono uguali o diverse.

Sesso; professione; diagnosi medica; ...

#### ORDINALI O PER RANGHI

Esiste un criterio predeterminato per ordinare le modalità

ordine di nascita; giorni della settimana; indice di severità di una malattia; ...

#### DISCRETO

L'insieme delle modalità assumibili può essere messo in " corrisp. biunivoca " con un sottoinsieme dei numeri naturali.

Num. componenti famiglia; num. di figli; num. di denti; num. colonie batteriche in una piastra; ...

#### CONTINUO

la variabile può assumere qualsiasi valore all'interno di intervalli di numeri reali.

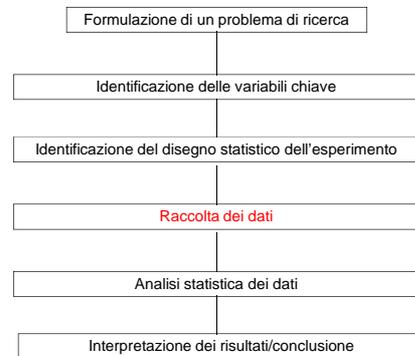
statura; peso; glicemia; PAS; ...



### Descrivere e Riassumere i dati

**Esempio.** Su un campione di pazienti si rilevano le caratteristiche: sesso, età, altezza, peso, pressione arteriosa sistolica (PAS), tasso glicemico.

nome: Rossi Amerigo	nome: Bianchi Paolo	nome: Valenzi Alberica	nome: Alinori Alfonso
sesso: maschio	sesso: maschio	sesso: femmina	sesso: maschio
età: 32	età: 47	età: 45	età: 27
altezza: 172 cm	altezza: 170 cm	altezza: 168 cm	altezza: 183 cm
peso: 64 Kg	peso: 80 Kg	peso: 51 Kg	peso: 85 Kg
PAS: 140 mm/Hg	PAS: 148 mm/Hg	PAS: 125 mm/Hg	PAS: 138 mm/Hg
glicemia: 190 mg/100cc	glicemia: 180 mg/100cc	glicemia: 150 mg/100cc	glicemia: 170 mg/100cc



Le informazioni raccolte per essere "trattate" da un computer devono essere organizzate in strutture chiamate comunemente **Data Base o File Dati**.

Le informazioni vengono, comunemente, organizzate per riga, cioè su ogni riga, consecutivamente, vengono elencati i dati relativi ad un soggetto.

N.	NOME	SESSO	ETA'	ALTEZZA	PESO	PAS	GLIC.
1	Rossi Amerigo	M	32	172	64	140	190
2	Bianchi Paolo	M	47	170	80	148	180
3	ValenziAlberica	F	45	168	51	125	150
4	Alinori Alfonso	M	27	183	85	130	170
5	...	...	...	...	...	...	...
6	...	...	...	...	...	...	...

Dott.ssa Marta Di Nicola 13

### DISTRIBUZIONI SEMPLICI DI FREQUENZE

I dati (cioè le informazioni raccolte) spesso sono di non immediata lettura.

Per questo si procede ad una sistematizzazione e sintesi delle informazioni raccolte, cioè alla loro **tabulazione**. Per ogni variabile si calcolano le **frequenze assolute (f.a.)** che rappresentano il numero di u.s. che presentano una stessa modalità del carattere.

Variabile nominale "Tipologia di corona"

Type of Crown	Number of Crowns
Gold crown	843
Metal ceramic crown	972

↑  
Frequenze assolute

Variabile ordinale "Grado di accordo"

Response Category	Number of Individuals
Strongly disagree	24
Disagree	43
Neutral	49
Agree	33
Strongly agree	30

↑  
Frequenze assolute

Dott.ssa Marta Di Nicola 14

Variabile quantitativa "conta delle colonie batteriche"

120	160	172	176	180	184	184	184	185	
188	190	192	196	196	200	200	200	206	207
210	213	220	236	250	254	272	272	272	275
275	280	280	282	284	286	287	295	296	299
300	300	300	301	302	304	304	304	304	304
306	306	308	315	320	320	325	330	335	346
346	354	356	358	360	364	365	366	380	380
385	386	390	396	396	396	400	408	410	412
412	416	418	424	438	440	448	476	500	588

Il numero degli intervalli potrebbe essere pari a  $\sqrt{n}$

Interval	Frequency	Interval	Frequency
112.5-162.5	2	362.5-412.5	16
162.5-212.5	19	412.5-462.5	6
212.5-262.5	5	462.5-512.5	2
262.5-312.5	27	512.5-562.5	0
312.5-362.5	12	562.5-612.5	1

↑  
Frequenze assolute delle classi

Dott.ssa Marta Di Nicola 15

**Esempio.** Distribuzione doppia di frequenze assolute

Carie	Fumatori	Non fumatori	Totale
SI	160	100	260
NO	120	70	190
Totale	280	170	450



Dott.ssa Marta Di Nicola 16

Ci accorgiamo che il confronto **non** può essere effettuato solo con le f.a. in quanto esse si riferiscono a collettivi di numerosità diversa.

Se vogliamo confrontare le frequenze le dobbiamo "depurare" dalla numerosità del collettivo; ciò lo si fa dividendo le f.a. per la numerosità (N) della popolazione e moltiplicando per 100 (cioè facendo riferimento ad una ipotetica popolazione di 100 unità).

Le frequenze così calcolate sono le **frequenze percentuali (f.%)**

Dott.ssa Marta Di Nicola 17

Carie	Fumatori		Non fumatori	
	f.a.	f.a.%	f.a.	f.a.%
SI	160	57.1	100	58.8
NO	120	42.8	70	41.2
Totale	280	100	170	100

Dott.ssa Marta Di Nicola 18

Tornando all'esempio precedente

$$\left(\frac{19}{90}\right) \cdot 100 = 21\%$$

Interval	Frequency	Relative Frequency (%)	Cumulative Relative Frequency (%)
112.5-162.5	2	2	2
162.5-212.5	19	21	23
212.5-262.5	5	6	29
262.5-312.5	27	30	59
312.5-362.5	12	13	72
362.5-412.5	16	18	90
412.5-462.5	6	7	97
462.5-512.5	2	2	99
512.5-562.5	0	0	99
562.5-612.5	1	1	100
Total	90	100	

Le frequenze cumulate indicano quante u.s. si presentano **fino a** quella modalità. Ha senso calcolare le f.cum solamente per le variabili quantitative o qualitative ordinabili.

29=23+6

Dott.ssa Marta Di Nicola 19

### I GRAFICI STATISTICI

Scopo dei grafici è quello di rendere l'informazione contenuta in una serie di dati:

- ✓ di più facile comprensione;
- ✓ di più diretta lettura.

Pertanto un grafico deve fornire al lettore una informazione sintetica e facile da interpretarsi.

Dott.ssa Marta Di Nicola 20

**ORTOGRAMMA:** Usato per variabili qualitative l'altezza delle barre rappresenta frequenze assolute o percentuale

NUMERO DI CARIE	Frequenza assoluta	Frequenza cumulata
0	21	21
1	15	36
2	6	42
3	3	45
4	2	47
5	1	48
8	1	49
12	1	50
Totale	50	

Dott.ssa Marta Di Nicola 21

PESO CORPOREO	Freq. assoluta	Freq. relativa	Freq. cumulata
40-50	3	0.14	0.14
51-60	5	0.23	0.36
61-70	9	0.41	0.77
71-80	4	0.18	0.95
> 80	1	0.05	1
Totale	22	1	

**ISTOGRAMMA** Indicato per rappresentare distribuzioni in classi (variabili quantitative continue). Costituiti da una serie di barre rettangolari contigue ognuna in rappresentanza di una classe e con area proporzionata alla rispettiva frequenza.

Dott.ssa Marta Di Nicola 22

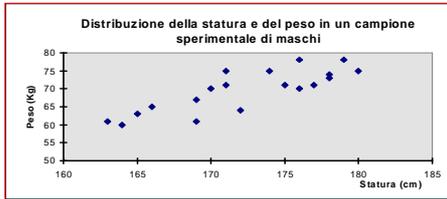
**DIAGRAMMI A SETTORI CIRCOLARI (TORTE)** Indicati per variabili qualitative allo scopo di evidenziare le frequenze % delle singole modalità. L'area di un cerchio viene suddivisa in settori proporzionali alle frequenze %

Dott.ssa Marta Di Nicola 23

**GRAFICI PER SPEZZATE** Si ottengono dai grafici per punti congiungendo i vari punti. Indicati per evidenziare una continuità tra valori come ad es. nella rappresentazione delle serie temporali.

Dott.ssa Marta Di Nicola 24

**SCATTER DI PUNTI** Consente di mettere in relazione due variabili quantitative



Dott.ssa Marta Di Nicola

## Numerosità campionaria (Sample size)

### Perché calcolare la numerosità campionaria?

Per assicurare che il numero di soggetti/animali (unità statistiche) coinvolti nello studio permetta di rispondere adeguatamente al quesito di interesse considerando che:

- uno studio di *dimensioni limitate* avrà una probabilità elevata di non riuscire a riconoscere un trattamento promettente;
- uno studio di *dimensioni eccessive* è economicamente oneroso e rischia di sottoporre un eccessivo numero di soggetti ad un trattamento non efficace.

### Come si calcola la numerosità campionaria?

1. **Obiettivo dello studio** (superiorità, non inferiorità, equivalenza);
2. **Disegno dello studio** (gruppi paralleli, cross-over, etc)
3. **End point primario** (percentuale di "successo", livello medio di piastrene, etc);
4. **Parametri clinici** (entità dell'effetto (effect size), percentuale di drop-out, variabilità del fenomeno);
5. **Parametri statistici** (tipo di test, livello di significatività  $\alpha$ , potenza dello studio  $1-\beta$ ).

Variabili quantitative

$$n = \frac{2\sigma^2}{(\mu_1 - \mu_2)^2} f(\alpha, \beta)$$

Variabili qualitative

$$n = \frac{p_1(100-p_1) + p_2(100-p_2)}{(p_1 - p_2)^2} f(\alpha, \beta)$$

$n$  è il numero di pazienti richiesto per ogni braccio di trattamento  
 $p_1$  è la proporzione di successi attesi nel gruppo di controllo  
 $\sigma^2$  è la varianza della variabile oggetto di studio

$p_1 - p_2$  o  $\mu_1 - \mu_2$  è la differenza fra i trattamenti che si vuole riconoscere

### La scelta del $\Delta$

La scelta della differenza ( $\Delta$ ) che si vuole essere in grado di riconoscere è cruciale ed influenza la dimensione del campione molto più degli altri fattori.

La differenza deve avere senso sul piano clinico:

- Studi di superiorità → minima differenza clinicamente rilevante
- Studi di non-inferiorità → massimo svantaggio clinicamente tollerabile
- Studi di equivalenza → intervallo di equivalenza clinica

$\alpha$  = livello di significatività o errore di primo tipo è la probabilità di concludere che un trattamento è superiore/diverso quando in realtà non lo è.

$1-\beta$  = (1 - errore di secondo tipo) è la probabilità di riconoscere un trattamento che è realmente superiore/diverso.

Errore di I tipo ( $\alpha$ )	Errore di II tipo ( $\beta$ )			
	0.05	0.1	0.2	0.5
0.10	10.8	8.6	6.2	2.7
0.05	13.0	10.5	7.9	3.8
0.01	17.8	14.9	11.7	6.6

### Errori di I e II tipo

		Differenza reale	
		No	Si
Conclusione	I trattamenti non sono diversi		Errore di II° tipo ( $\beta$ )
	I trattamenti <u>sono</u> differenti	Errore di I° tipo ( $\alpha$ )	Potenza dello studio ( $1-\beta$ )

#### Crisis telephone consultation for deliberate self-harm patients: effects on repetition

M.O. Evans, H.G. Morgan, A. Hayward and D.J. Gunnell

Deaths of study subjects were obtained from the local coroner's office and from mortality statistics collected by Avon Health Authority.

##### Statistical analysis

Sample size calculations were based on the 54% reduction in repeat DSH shown in the smaller study (Morgan et al, 1993). On the basis of the expectation that its findings would be applicable to all DSH patients and not merely first-timers, a sample size of 700 was required to detect this reduction with 80% power and a two-sided 5% significance level. Existing in-patient admission rates indicated a target recruitment period of 21 months.

##### Main analyses

All analyses were conducted on an intention-to-treat basis. The effect on DSH repetition of provision of a green card were measured in terms of the odds ratio comparing the odds of repeat

In this trial the researchers know what size of effect they might expect from their intervention.

They are therefore in a position to calculate sample size...

... for stated levels of power and probability of making a Type I error

### Esempio

Il farmaco di riferimento riduce la pressione sistolica di 25 mmHg, il nuovo farmaco per essere competitivo dovrebbe ridurre la pressione sistolica di almeno 30 mmHg (ovvero 5 mmHg in più). La deviazione standard della riduzione della pressione viene stimata in 10 mmHg da studi precedenti. Si adotta un alfa del 5% e una potenza del 90%, pertanto  $f(\alpha, \beta) = 10,5$

$$n = \frac{2(10)^2}{(5)^2} 10,5 = 84$$

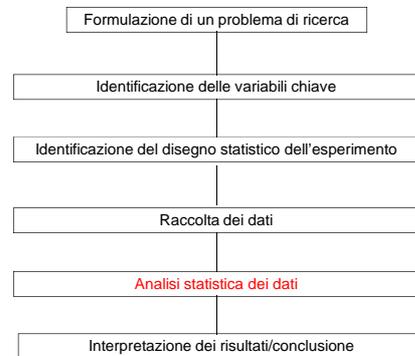
Occorrono almeno 84 soggetti per gruppo.

### Esempio

Nei pazienti affetti da tumore X in stadio avanzato, la sopravvivenza a 5 anni è del 30% con il trattamento standard. Dati preliminari suggeriscono che nei pazienti sottoposti ad un nuovo trattamento la sopravvivenza salga al 40%. Si adotta un alfa del 5% e una potenza dell' 80%, pertanto  $f(\alpha, \beta) = 7,9$

$$n = \frac{30(100-30) + 40(100-40)}{(10)^2} 7,9 = 355,5$$

Occorrono almeno 356 soggetti per gruppo.



**OBIETTIVO:** Individuare un indice che rappresenti significativamente un insieme di dati statistici.

Altezza (cm)

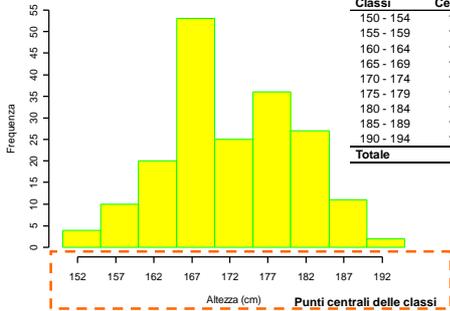


Tabella dei dati

Classi	Punto Centrale	n	%
150 - 154	152	4	2.1
155 - 159	157	10	5.3
160 - 164	162	20	10.6
165 - 169	167	53	28.2
170 - 174	172	25	13.3
175 - 179	177	36	19.1
180 - 184	182	27	14.4
185 - 189	187	11	5.9
190 - 194	192	2	1.1
<b>Totale</b>		<b>188</b>	<b>100.0</b>

### LA MEDIA ARITMETICA

**DEFINIZIONE:** La media aritmetica è quel valore che avrebbero tutte le osservazioni se non ci fosse la variabilità (casuale o sistematica).

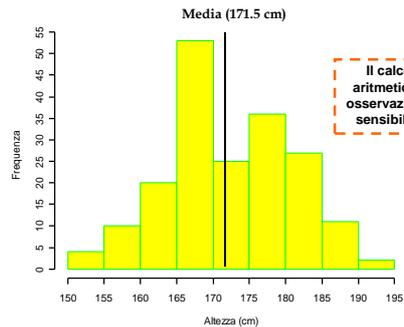
Più precisamente, è quel valore che sostituito a ciascun degli  $n$  dati ne fa rimanere costante la somma.

dato un insieme di  $n$  elementi  $\{x_1, x_2, \dots, x_n\}$

Si dice **media aritmetica semplice** di  $n$  numeri il numero che si ottiene dividendo la loro somma per  $n$ .

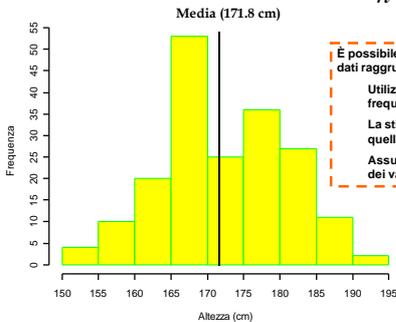
$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

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



Il calcolo della media aritmetica utilizza tutte le osservazioni, risulta quindi sensibile ai valori atipici

$$\mu = \frac{1}{n} \sum_{j=1}^k f_j \cdot x_{(mid)_j}$$



È possibile calcolare la media anche dai dati raggruppati in classi:

Utilizzo i punti centrali e le frequenze di classe

La stima della media approssima quella ottenuta con i singoli dati

Assume la distribuzione uniforme dei valori all'interno delle classi

### Le proprietà della media aritmetica

1. compresa tra il minimo dei dati e il massimo dei dati;
2.  $\sum (x_i - \bar{x}) f_i = 0$  la somma degli scarti dalla media è zero;
3.  $\sum (x_i - z)^2 f_i$  assume valore minimo per  $z =$  media aritmetica;
4. la media dei valori:  $k \cdot \bar{x}$  è pari a la media aritmetica  $\cdot k$  (dove  $k$  è un numero reale qualsiasi)
5. la media dei valori:  $\bar{x} \pm h$  è pari a: media aritmetica  $\pm h$  (dove  $h$  è un numero reale qualsiasi).

La **media aritmetica** è la misura di posizione più usata ma. A volte, altre misure come la **mediana** e la **moda** si dimostrano utili.

Si consideri un campione di valori di VES (velocità di eritrosedimentazione, mm/ora) misurati in 7 pazienti

{8, 5, 7, 6, 35, 5, 4}

In questo caso, la media che è = 10 mm/ora non è un valore tipico della distribuzione: soltanto un valore su 7 è superiore alla media!



**Limite della media aritmetica:**  
è notevolmente influenzata dai valori estremi della distribuzione.

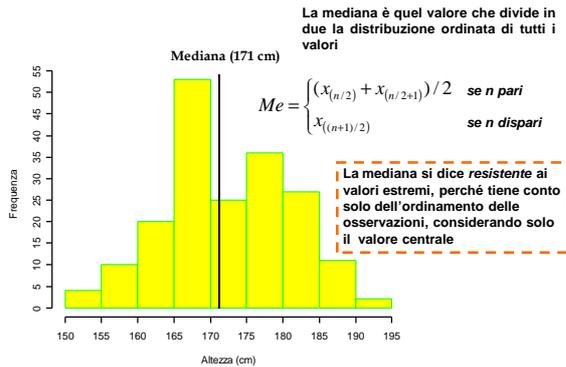
**LA MEDIANA**

**DEFINIZIONE:** La mediana (Me) è quell'osservazione che bipartisce la distribuzione in modo tale da lasciare al "di sotto" lo stesso numero di termini che lascia al "di sopra".

L'idea che è alla base della mediana è di cercare un numero che sia più grande di un 50% delle osservazioni e più piccolo del restante 50%.



per dati almeno ordinali

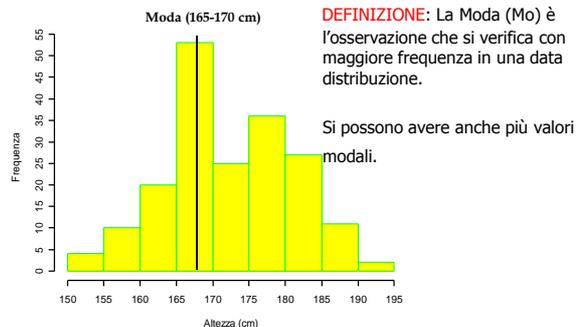


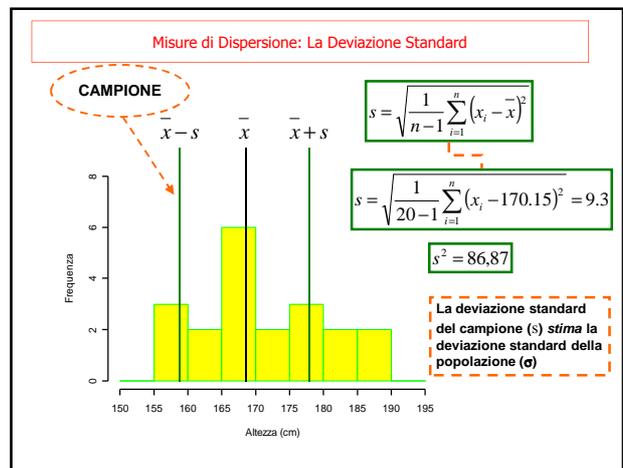
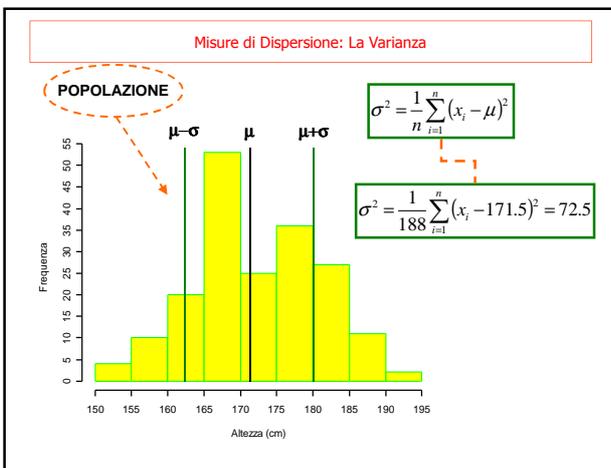
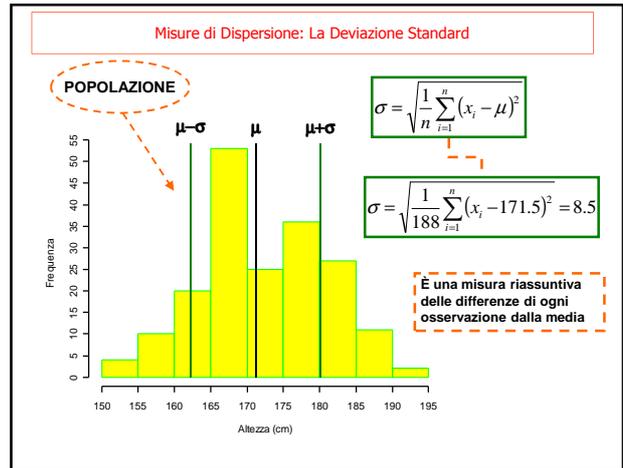
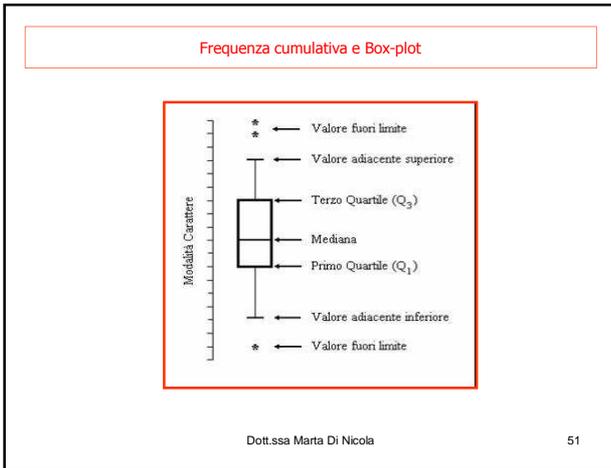
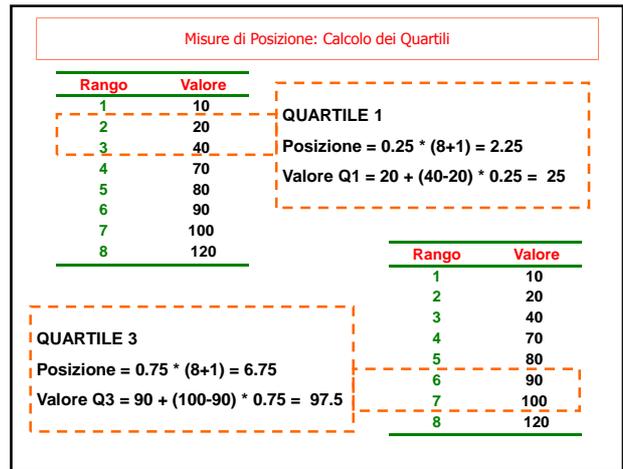
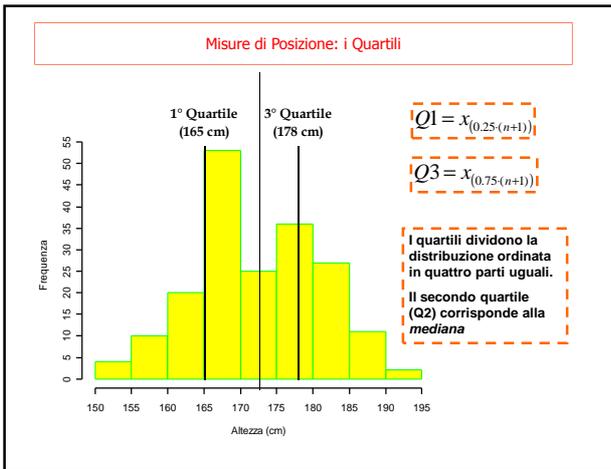
Voti ordinati ( $x_i$ )	Frequenze ( $f_i$ )	Freq. Cum. ( $F_i$ )	Freq.Cum. ( $F_i\%$ )
18	2 (10.5)	2	10.5
20	4 (21.0)	2+4 = 6	31.5
<b>22</b>	<b>8 (42.1)</b>	<b>6+8 = 14</b>	<b>73.6</b>
24	2 (10.5)	14+2 = 16	84.1
27	2 (10.5)	16+2 = 18	94.6
30	1 (5.4)	18+1 = 19	100
<b>Totale</b>	<b>19</b>		

Le fasi operative per il calcolo della mediana sono le seguenti:

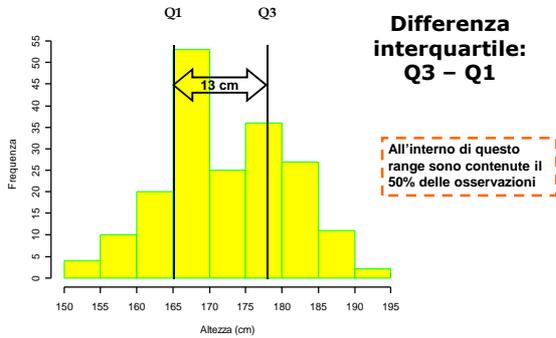
- 1) ordinamento crescente dei dati;
- 2) se il numero di dati n è dispari, la mediana corrisponde al dato che occupa la (n+1)/2 esima posizione;
- 3) se il numero di dati n è pari, la mediana è data dalla media aritmetica dei due dati che occupano la posizione n/2 e quella n/2+1.

**LA MODA**





Misure di Dispersione: La differenza interquartile



Misure di Dispersione: IL COEFFICIENTE DI VARIAZIONE

$$C.V. = \frac{\text{(deviazione standard)}}{\text{(media aritmetica)}}$$

La variabilità guarda alle differenze tra le unità sperimentali. E' però evidente che il significato pratico delle differenze può dipendere dal livello del fenomeno considerato. Può quindi essere interessante disporre di una qualche misura di variabilità *aggiustata* in qualche maniera per tenere conto del livello del fenomeno.

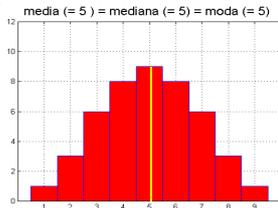
INDICI DI SIMMETRIA

**Distribuzione simmetrica:**

Le osservazioni equidistanti dalla mediana (coincidente in questo caso col massimo centrale) presentano la stessa frequenza relativa

Un esempio importante è fornito dalla **distribuzione normale**

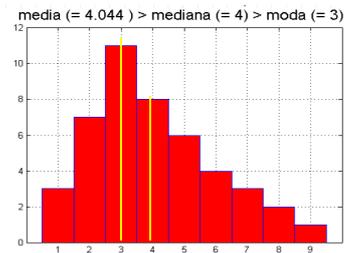
**Media = Mediana = Moda**



**Distribuzione asimmetrica positiva**

La curva di frequenza ha una coda più lunga a destra del massimo centrale

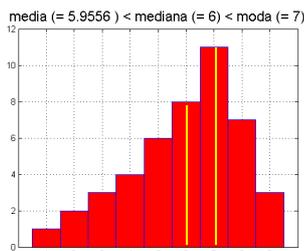
**Media > Mediana > Moda**



**Distribuzione asimmetrica negativa**

La curva di frequenza ha una coda più lunga a sinistra del massimo centrale

**Media < Mediana < Moda**



**JCN Journal of Clinical Nursing**

Journal of Clinical Nursing

ORIGINAL ARTICLE

**Pain experiences and self-management strategies among middle-aged and older adults with arthritis**

Guilan Gong, Jie Li, Xiuyun Li and Jing Mao

**Aims and objectives.** The purposes were (1) to explore pain experiences and the use and perceived effectiveness of pain self-management methods among middle-aged and older adults with osteoarthritis or rheumatoid arthritis in mainland China and (2) to compare those with diagnoses of osteoarthritis and rheumatoid arthritis.

**Background.** Prior research has suggested that pain is a major concern for people with arthritis. However, studies systematically investigating pain experiences and self-management status of arthritis patients are scarce in mainland China.

**Design.** Descriptive survey.

**Methods.** Participants (n = 197) aged 45 and over, diagnosed with either osteoarthritis or rheumatoid arthritis, and experiencing persistent pain were administered three self-report questionnaires: the Demographic Data Questionnaire, the Brief Pain Inventory and the Pain Management Inventory.

**Results.** The mean of the overall pain intensity was 5.6 (SD = 1.3). The median of number of pain sites was 7.0 (QR = 7.0) and the overall pain interference was 6.0 (QR = 2.6). Most participants experienced moderate to severe pain and interference. The current methods used for managing pain were perceived as only moderately effective. The sample used a median of 4.0 (QR = 3.0) self-management methods. Most often used were prescribed medicine, massage, heat and activity pacing. Methods perceived as most helpful included prescribed medicine, over-the-counter medicine, hot herbs and heat. Persons with rheumatoid arthritis had significantly more pain sites, higher pain intensity and greater number of pain management methods used compared to those with osteoarthritis.

**Conclusions.** Pain management is a significant problem in this population. The findings highlight the importance of helping the individual to identify and appropriately use a variety of self-management methods, selecting the appropriate method(s) at any one time.

**Relevance to clinical practice.** Healthcare providers are urged to develop appropriate interventions on pain management tailored to arthritis patients in mainland China.

Key words: arthritis, mainland China, middle-aged and older adults, pain, self-management

60

**JCN**

ORIGINAL ARTICLE  
Pain experiences and older adults  
Guilan Gong, Jie Li

**Table 1** The demographic characteristics of the study

Variables	Median (QR) or Mean $\pm$ SD or n (%) <sup>a</sup>			P-value
	Total (n = 197)	Osteoarthritis (n = 96)	Rheumatoid arthritis (n = 141)	
Age (years)	57.0 (13.0)	64.0 (14.8)	55.0 (13.0)	0.000***
Disease duration (years)	5.0 (10.0)	1.3 (2.2)	7.0 (10.3)	0.000***
BMI	23.0 $\pm$ 2.7	23.3 $\pm$ 2.6	21.4 $\pm$ 2.3	0.954 <sup>b</sup>
Gender				0.495 <sup>c</sup>
Male	33 (16.8%)	11 (19.6%)	22 (15.4%)	
Female	164 (83.2%)	45 (80.4%)	119 (84.4%)	
Marital status				0.060 <sup>d</sup>
Married	166 (84.3%)	43 (76.8%)	123 (87.2%)	
Divorced/widowed	31 (15.7%)	13 (23.2%)	18 (12.7%)	
Education				0.0390 <sup>e</sup>
Below primary school	30 (15.2%)	10 (17.9%)	20 (14.2%)	
Primary school	45 (22.8%)	12 (21.4%)	33 (23.4%)	
Junior high school	56 (28.4%)	9 (16.3%)	47 (33.3%)	
Senior high school or equivalent	49 (24.9%)	16 (28.6%)	33 (23.4%)	
College and above	17 (8.6%)	9 (16.4%)	8 (5.7%)	
Work status				0.200 <sup>f</sup>
Employed	13 (6.6%)	6 (10.7%)	7 (5.0%)	
Unemployed	184 (93.4%)	50 (89.3%)	134 (95.0%)	
Income (year)				0.001 <sup>g</sup>
<1000	58 (29.4%)	10 (17.9%)	48 (34.0%)	
1000-2000	44 (22.3%)	14 (25.0%)	30 (21.3%)	
2000-3000	41 (20.8%)	15 (26.8%)	26 (18.4%)	
>3000	34 (17.3%)	17 (30.4%)	17 (12.1%)	
Residence				0.132 <sup>h</sup>
Living in rural area	87 (44.2%)	20 (35.7%)	67 (47.5%)	
Living in urban area	110 (55.8%)	36 (64.3%)	74 (52.5%)	
Living arrangement				0.229 <sup>i</sup>
Lives alone	14 (7.1%)	6 (10.7%)	8 (5.7%)	
Lives with someone	183 (92.9%)	50 (89.3%)	133 (94.3%)	

<sup>a</sup>p < 0.05, <sup>b</sup>p < 0.01, <sup>c</sup>p < 0.001.  
<sup>d</sup>Median (QR) is given in the variables: age and disease duration; Mean  $\pm$  SD is given in the variable: BMI; n (%) is given for categorical variables.  
<sup>e</sup>Mann-Whitney U test.  
<sup>f</sup>Fisher's exact test.  
<sup>g</sup>Student's t test.  
<sup>h</sup>Chi-squared test.  
<sup>i</sup>BMI, body mass index.

61

**JCN** *Journal of Clinical Nursing*

ORIGINAL ARTICLE  
Pain experiences and self-management strategies among middle-aged and older adults with arthritis  
Guilan Gong, Jie Li

**Data analysis**

All statistical analyses were conducted using SPSS, version 17.0 (SPSS Inc., Chicago, IL, USA). All statistical tests were two-tailed, and significance level was set at 0.05. Normality of data was checked using the Kolmogorov-Smirnov test. Except for the body mass index (BMI) and overall pain intensity score, none of the data satisfied the normal distribution assumption. Descriptive statistics such as mean (M) and standard deviation (SD), median (Md), the 75th percentile and the 25th percentile and interquartile range (QR) were used to indicate the central tendency and variability of the numerical variables. Frequency and percentage were computed for categorical variables. Mann-Whitney test was used to compare demographic characteristics (Chi-squared test or Fisher's exact test was used for categorical data), pain characteristics and the frequency and helpfulness of various methods between persons with OA and RA.

Dott.ssa Marta Di Nicola

62