

# COSTRUIRE UN PICCOLO “SET” DI DATI



Individuazione degli obiettivi dello studio



Individuazione delle variabili che possono aiutare il raggiungimento degli obiettivi dello studio



Preparazione degli strumenti di rilevazione

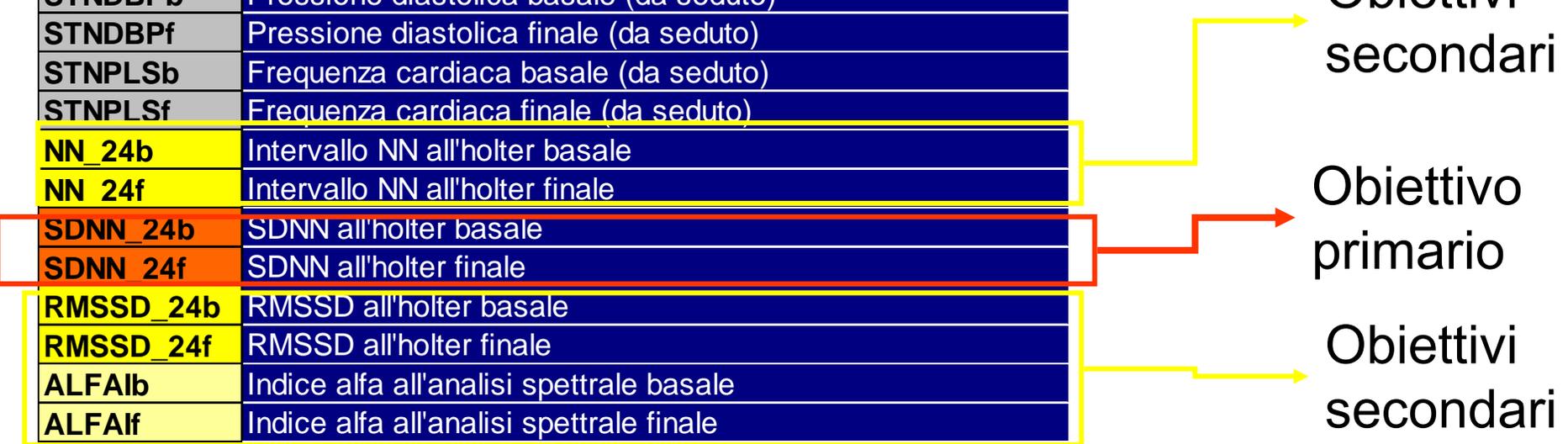
PATNO	Numero di selezione del paziente
SEX	Sesso del paziente
AGE	Età del paziente (criterio inclusione 18-80)
Treat	Trattamenti in studio: 1 vs 2 (si vuole testare la superiorità del trattamento 1 sul trattamento 2 per i parametri holter)
AMI	Cardiopatía infartuale
IPERT	Cardiopatía ipertensiva
DILAT	Cardiopatía dilatativa primitiva
FE	Frazione d'iezione (criterio di inclusione FE<40%)
NYHAb	Classe funzionale NYHA (Criterio inclusione NYHA>1)
NYHAf	Classe funzionale NYHA finale
WGT1Nb	Peso basale
WGT1Nf	Peso finale
STNSBPb	Pressione sistolica basale (da seduto)
STNSBPf	Pressione sistolica finale (da seduto)
STNDBPb	Pressione diastolica basale (da seduto)
STNDBPf	Pressione diastolica finale (da seduto)
STNPLSb	Frequenza cardiaca basale (da seduto)
STNPLSf	Frequenza cardiaca finale (da seduto)
NN_24b	Intervallo NN all'holter basale
NN_24f	Intervallo NN all'holter finale
SDNN_24b	SDNN all'holter basale
SDNN_24f	SDNN all'holter finale
RMSSD_24b	RMSSD all'holter basale
RMSSD_24f	RMSSD all'holter finale
ALFAIb	Indice alfa all'analisi spettrale basale
ALFAIf	Indice alfa all'analisi spettrale finale

Valutazione del trattamento con ACE inibitore vs Inibitore dell'angiotensina II sui parametri holter nei pazienti con scompenso cardiaco

Obiettivi secondari

Obiettivo primario

Obiettivi secondari



# PREPARARE UN "SET" DI DATI

Microsoft Access - [SCOMPENSO]

File Modifica Visualizza Inserisci Formato Record Strumenti Finestra ?

Arial 9 G C S

PAZIENTE	0
SESSO	
ETA'	0
TRATTAMENTO (1 / 2)	0
CARDIOPATIA INFARTUALE	0
...ALTRE VARIABILI...	
CLASSE NYHA DI INGRESSO	
CLASSE NYHA FINALE	
...ALTRE VARIABILI...	
INTERVALLO NN HOLTER INGRESSO	0
INTERVALLO NN HOLTER DOPO TRATTAMENTO	0
DEVIATIONE STANDARD CICLO CARDIACO 24 ORE INGRESSO	0
DEVIATIONE STANDARD CICLO CARDIACO 24 ORE DOPO TRATTAMENTO	0
...ALTRE VARIABILI...	

Record: 1 di 1

Visualizzazione Maschera

MA NUM

# STRUTTURA DI UN "SET" DATI

	A	B	C	D	E	F	G	H
1	patno	sex	age	treat	ami	sdnn_24b	sdnn_24f	
2								
3								
4								
5								
6								
7								
8								

Ogni riga  
rappresenta  
un paziente

Ogni  
colonna  
rappresenta  
una  
variabile

Misure ripetute nel  
tempo possono essere  
poste in colonne  
affiancate

# IL "SET" DI DATI COMPLETO

Microsoft Excel + Analyse-it - Dati

File Modifica Visualizza Inserisci Formato Strumenti Dati Analyse Finestra ?

100% 12 € % ,00 +,00

E5 = 0

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	S
1	PATNO	SEX	AGE	Treat	AMI	IPERT	DILAT	FE	NYHAb	NYHaf	WGT1Nb	WGT1Nf	STNSBPb	STNSBPf	S
2	1	F	61	2	0	1	0	39	2	2	108	108,5	160	175	
3	2	M	71	1	0	1	0	39	2		54		145		
4	3	F	50	2	0	0	1	34	2		49		105		
5	4	M	59	1	0	1	0	25	2	2	105	107	140	150	
6	5	M	69	1	1	0	0	39	2	2	70	72,5	110	110	
7	6	F	72	2	0	0	1	38	3	3	68	74	120	145	
8	7	F	78	1	0	1	0	36	3	3	62	61	130	110	
9	8	M	65	2	0	1	0	36	3	2	78	78	160	130	
10	9	F	75	1	0	1	0	22	2	3	43	45	120	120	
11	10	M	60	1	0	0	1	19	2	2	60	63,5	140	150	
12	11	M	51	2	0	0	1	30	2	2	70	69	120	110	
13	12	M	54	2	1	0	0	20	3	2	79	78	130	105	
14	13	M	47	1	1	0	0	35	2	2	68	72	130	105	
15	14	M	62	1	0	1	0	24	3	3	68	65	140	100	
16	15	M	74	2	1	0	0	27	2	2	60	63	150	145	
17	16	M	39	1	0	0	1	28	2	2	60	61	120	105	
18	17	F	63	2	0	0	1	25	2	2	73	74	120	120	
19	18	M	66	1	0	1	0	28	2	2	62	62	150	120	
20	19	M	71	2	1	0	0	31	2	2	66,5	65	130	115	
21	20	M	41	2	0	0	1	35	2	2	77	72,5	155	135	
22	21	F	73	2	0	1	0	35	3	2	71	64	165	120	

Scompenso / Struttura1 / IMA / Struttura2

Pronto NUM

**VARIABILE:** qualunque caratteristica che possa assumere valore diverso in tempi, spazi, persone differenti

Se misurabile ed espressa con un valore numerico

## **QUANTITATIVA**

**Continua:** misurabile su una scala continua, il valore numerico dipende dalle caratteristiche dello strumento di misura adottato

**Discreta:** assume solo valori interi

Se espressa con un aggettivo o sostantivo

## **QUALITATIVA**

**Nominale:** il valore della variabile è espresso da un aggettivo o da un sostantivo

**Ordinale:** il valore della variabile è espresso da aggettivi o sostantivi in cui è possibile riconoscere un criterio di ordinamento

# GUARDARE I DATI: Come sono le variabili in studio?

## Variabili quantitative continue

- ✿ Peso (kg)
- ✿ Pressione arteriosa (mmHg)
- ✿ Durata media del ciclo cardiaco nelle 24 ore (ms)
- ✿ Deviazione Standard della durata ciclo cardiaco nelle 24 ore (ms)

## Variabili quantitative discrete

- ✿ Frequenza cardiaca (batt./min)

## Variabili qualitative

- ✿ Presenza / assenza di cardiopatia dilatativa (dicotomica 1/0)
- ✿ Classe funzionale NYHA

# SCALE DI MISURA

- **Nominale:** la caratteristica viene espressa da un attributo, senza una effettiva misurazione (gruppo sanguigno, sesso...)
- **Ordinale:** la caratteristica pur esprimendo una qualità consente un ordinamento dei risultati (indici dello stato di salute)
- **Rango:** gli elementi possono essere ordinati in base alla grandezza delle osservazioni, per cui è possibile costruire una graduatoria (scala dell'ansia etc...)
- **Quantitativa discreta:** la misura è effettuata solo per numeri interi
- **Quantitativa continua:** la misura è effettuata per tutti i possibili valori in un intervallo continuo

## **GUARDARE I DATI**

**Verificare la validità dei dati inseriti:**

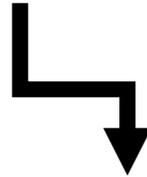
**Presenza di valori troppo bassi (outlyer)**

**Presenza di valori troppo alti (outlyer)**

**Presenza di valori mancanti (missing data: celle vuote, codice 9 o 99 o 999, simbolo \*, simbolo “-”, simbolo “.”)**

**Concentrazione di osservazioni su pochi valori**

**Etc...**



**Studio della distribuzione di frequenza della  
variabile sia quantitativa che qualitativa**



**Tabella di frequenza**

N	
Test entrata	Totale
0	161
0,09	2
0,11	1
0,13	4
0,16	1
0,18	2
0,2	6
0,23	4
0,25	1
0,27	5
0,29	1
0,32	3
0,36	1
0,39	1
0,41	1
0,43	1
0,45	1
0,46	1
0,48	2
0,52	4
0,55	1
0,64	6
0,66	2
0,69	3
0,75	1
0,78	2
0,85	1
0,96	1
1,01	3
1,05	1
1,35	1
1,38	1
1,51	1
1,54	1
1,74	1
Totale complessivo	229

Si vuole valutare il livello di alcol ematico mediante il test all'etilometro. Di lato si osservano i singoli valori osservati.

Per sintetizzare ed esporre i risultati con una tabella conviene renderla più piccola, determinando delle "classi di alcolemia".

Per determinare il numero di classi (K) può essere utile la formula di Sturges:  
 $K=1+3,322(\text{Log } n)$

Dove n è la numerosità campionaria.

L'ampiezza (w) della classe sarà data da:

$$W=R/K$$

Dove R è il range ossia la differenza tra il valore più grande e quello più piccolo presenti nei dati

Per determinare il numero di classi (K) si usa la formula di Sturges:

$$K=1+3,322(\text{Log } n)=1+3,322(\text{Log } 229)=8,8$$

dove n è la numerosità campionaria.

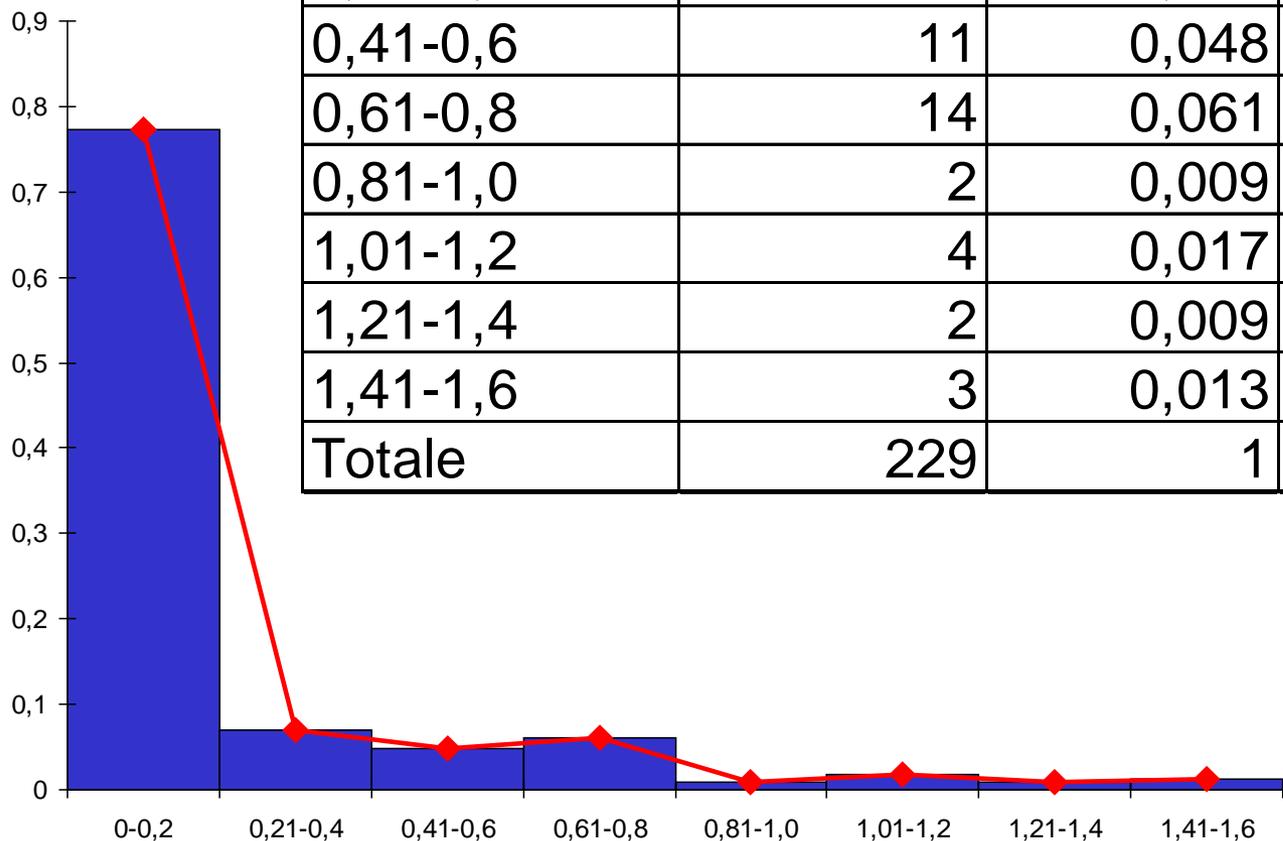
l'ampiezza (w) della classe sarà data da:

$$W= R / K=(1.74-0)/8=0.2$$

dove R è il range ossia la differenza tra il valore più grande e quello più piccolo

Il risultato della formula di Sturges relativo all'esempio in questione è mostrato in tabella e nel grafico sottostante

Livello di alcol	Frequenza assoluta	Frequenza relativa	Frequenza cumulativa assoluta	Frequenza cumulativa relativa
0-0,2	177	0,773	177	0,773
0,21-0,4	16	0,070	193	0,843
0,41-0,6	11	0,048	204	0,891
0,61-0,8	14	0,061	218	0,952
0,81-1,0	2	0,009	220	0,961
1,01-1,2	4	0,017	224	0,978
1,21-1,4	2	0,009	226	0,987
1,41-1,6	3	0,013	229	1,000
Totale	229	1		



# DESCRIVERE I DATI

INDICI SINTETICI:



Tendenza centrale:

media

mediana

moda

Variabilità:

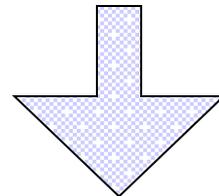
varianza - deviazione standard

coefficiente di variazione

range

differenza interquartile (differenza tra 25° e 75°

percentile)



I valori degli indici possono essere determinati per strato (qualora vi fossero variabili per le quali abbia senso stratificare) e presentati per mezzo di tabelle

# INDICI DI TENDENZA CENTRALE



Media

Aritmetica:  
utilizzabile sempre, in particolare quando i dati seguono una distribuzione normale

$$\bar{x} = \frac{\sum x_i}{N}$$

Geometrica:  
utilizzata per dati con distribuzione log-normale

$$\bar{x}_g = \frac{\sum \log x_i}{N}$$

Armonica: indicata quando i dati sono relativi alla misura tempo

$$m_g = \sqrt[n]{\prod x_i}$$

$$\bar{x}_a = \frac{N}{\sum \frac{1}{x_i}}$$

Moda:  
valore più frequente

Mediana: valore che bipartisce un'insieme di dati ordinati

# MISURE DI DISPERSIONE



**Range**

Differenza tra valore massimo e minimo

**Differenza interquartile**

Differenza tra valore del 75° e del 25° percentile

## **Varianza**

Misura media dei quadrati dello scostamento dei singoli valori dalla media aritmetica

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{N - 1} = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N - 1}$$

**Deviazione standard**

$$\sqrt{S^2}$$

**Coefficiente di variazione = C.V. = (Deviazione standard/media)x100**



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

Gradi di libertà

Ovvero il numero di determinazioni che posso scegliere liberamente meno i vincoli



Valore 1	10		25
Valore 2	5		15
Valore 3	25		18
Valore 4	16		22
Valore 5	44	?	
<b>Totale</b>	<b>100</b>		<b>100</b>

20!!!!!!

Vincolo:  $\sum (x_i - \bar{x}) = 0$

$$\sum (x_i - \bar{x}) = 0$$

$$\sum (x_i - \bar{x}) = \sum x_i - \sum \bar{x} = \sum x_i - n\bar{x} =$$

$$= \sum x_i - n \frac{\sum x_i}{n} = 0$$

# IL BOX-PLOT

Il grafico a scatola (box-plot) è una particolare rappresentazione di una distribuzione: gli elementi utilizzati per costruire la scatola sono i quantili e gli estremi della Distribuzione.

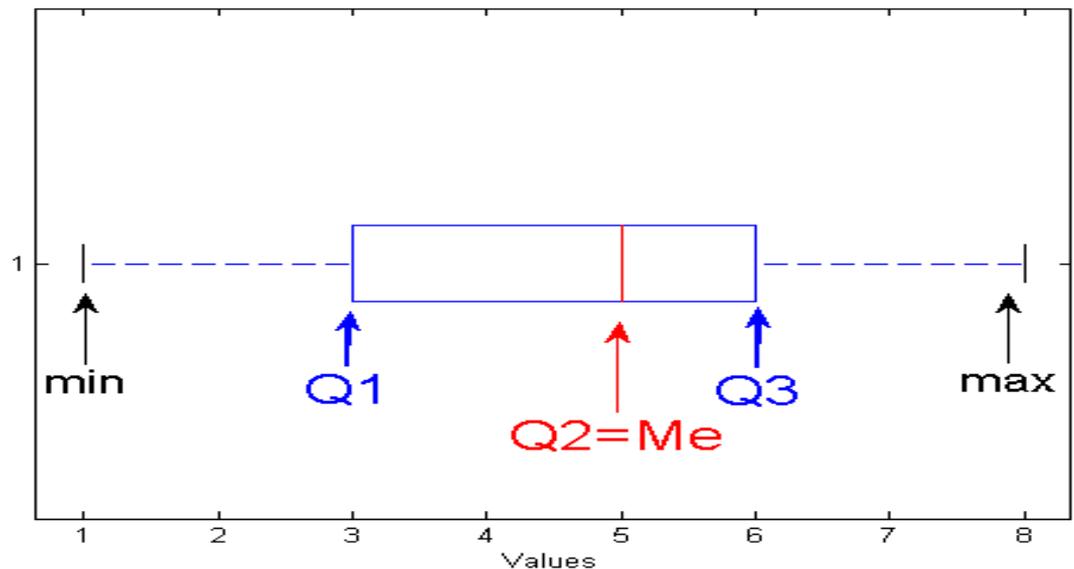
I valori anomali, che quindi non partecipano alla costruzione della scatola, vengono determinati dal confronto con il campo di variazione interquartile. In particolare vengono considerate due soglie:

$$Q1 - 1,5 \times (Q3 - Q1)$$

rappresenta il valore al di sotto del quale una modalità considerata outlier

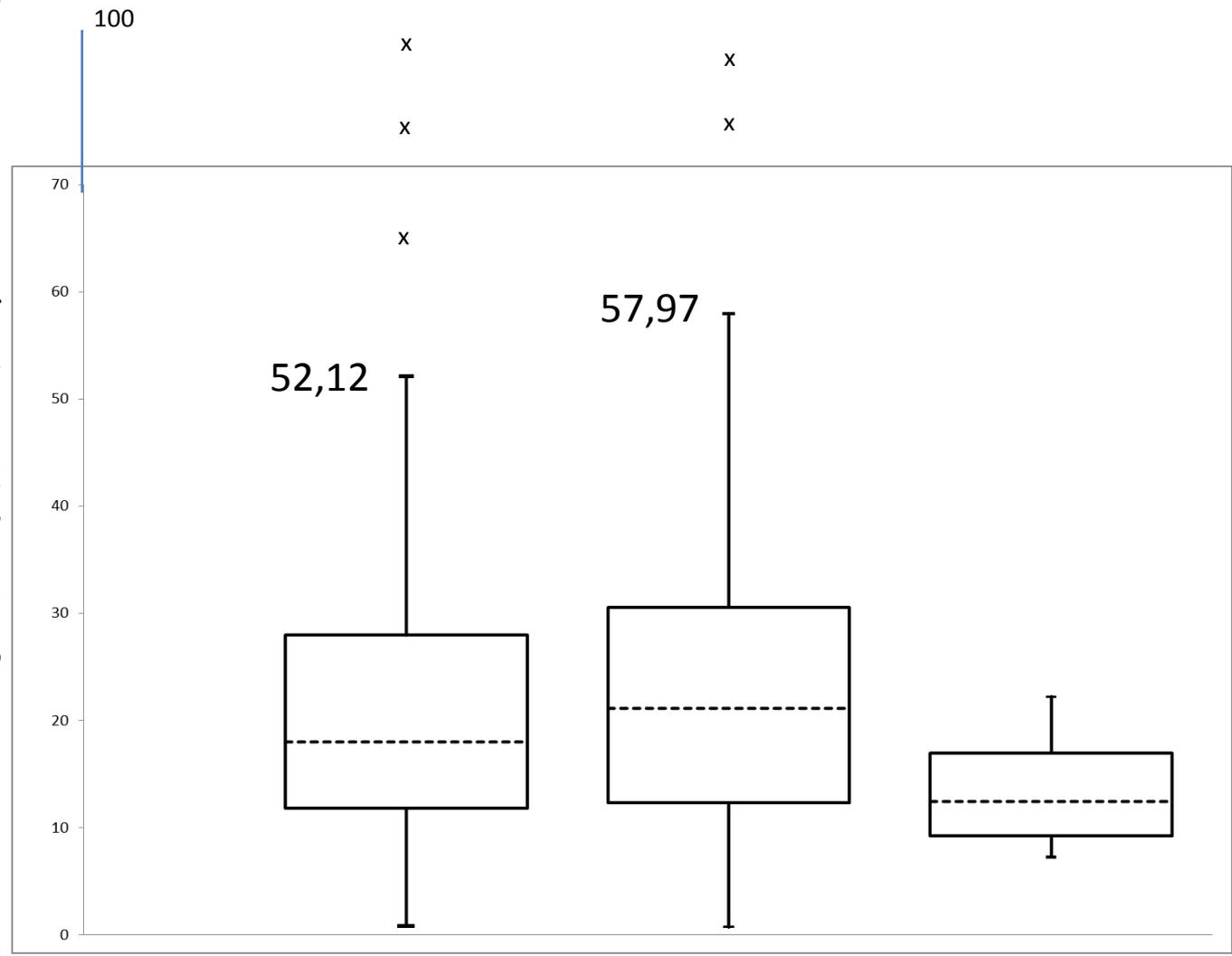
$$Q3 + 1,5 \times (Q3 - Q1)$$

rappresenta il valore al di sopra del quale una modalità considerata outlier



min	0,8	0,73
v1	0,9	2,69
v2	2,6	
v3	4,7	
v4		
v5		
5perc	5,9	6,04
10perc	7,54	6,72
q1	11,82	12,3
mediana	17,99	21,12
q3	27,94	30,57
90perc	44,2	57,3
95perc	60,9	75,8
v6		
v7		
v8		
v9	64	
v10	86,5	85,5
max	99,48	96,54

TGF B – in diverse classi di Hcc pts



# DESCRIVERE I DATI



## RAPPRESENTAZIONI CON GRAFICI

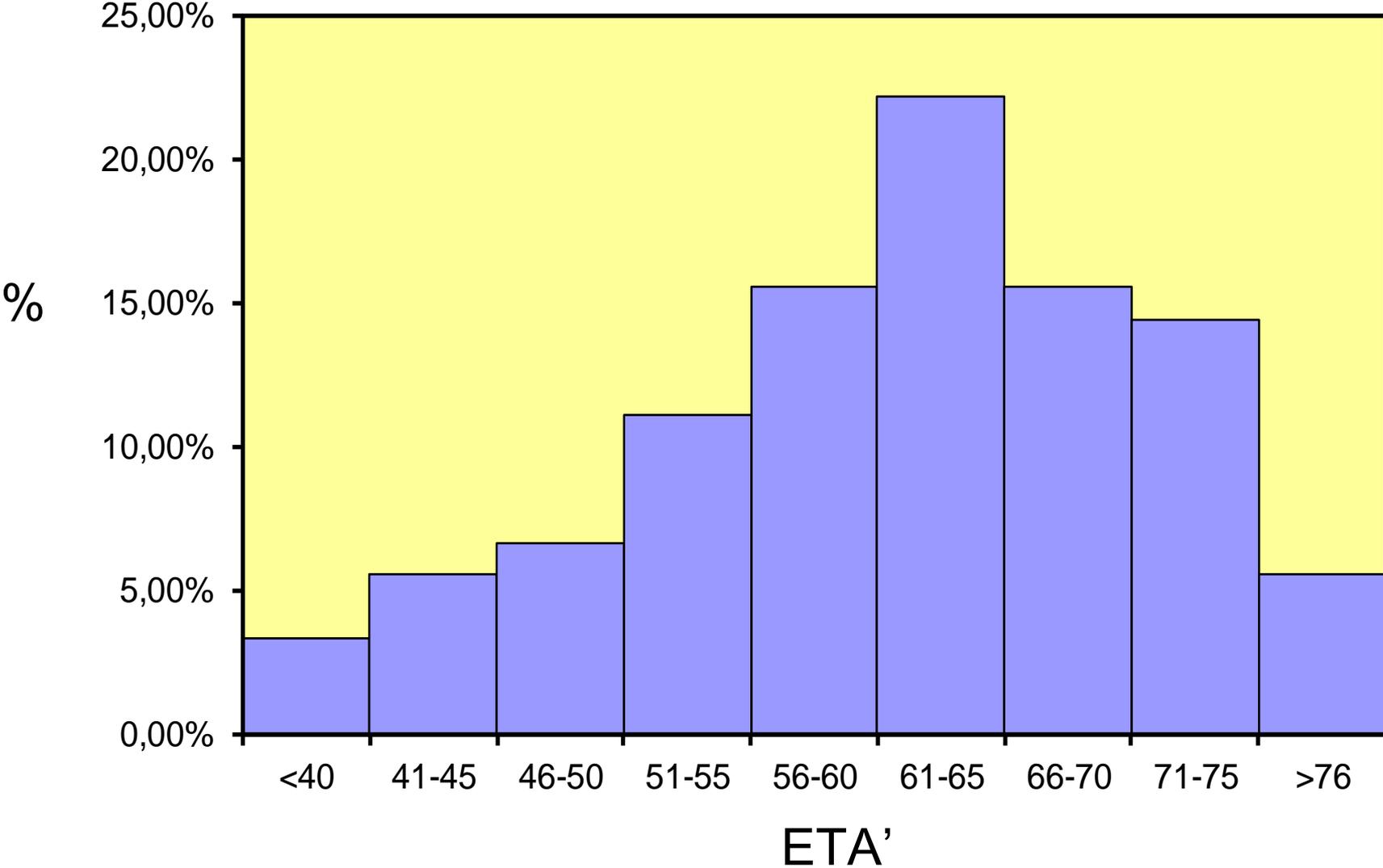
**ISTOGRAMMI:** sull'asse x c'è la variabile continua suddivisa in classi e sull'asse y la frequenza (più correttamente quella "relativa", cioè la percentuale) con cui quella classe si presenta

Età

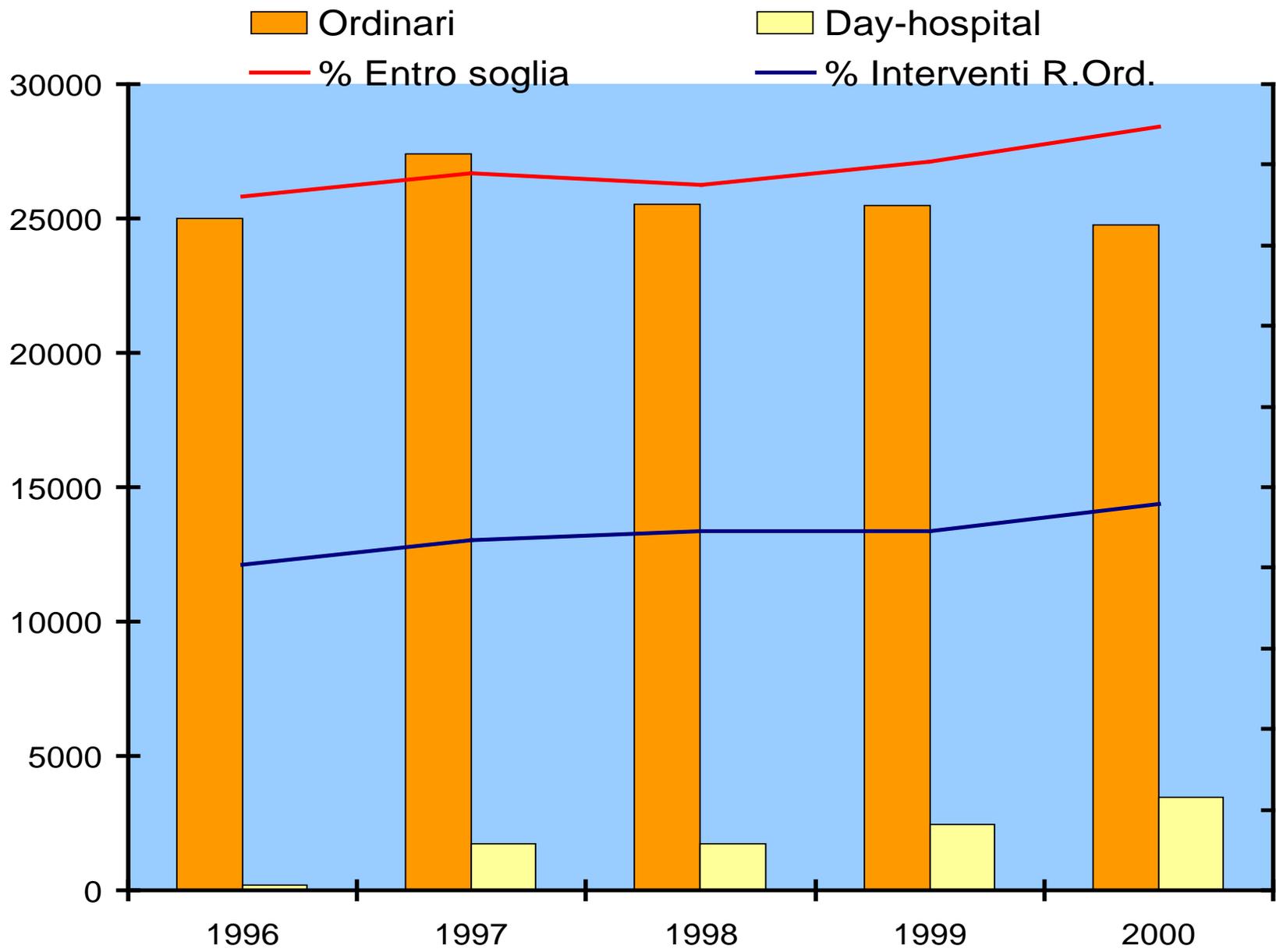
Peso

Frequenza cardiaca

# DISTRIBUZIONE DEI PAZIENTI PER CLASSI DI ETA'.



# Distribuzione dei ricoveri nel quinquennio 96-00 e percentuale di ricoveri con DRG chirurgico



# DESCRIVERE I DATI



## RAPPRESENTAZIONI CON GRAFICI

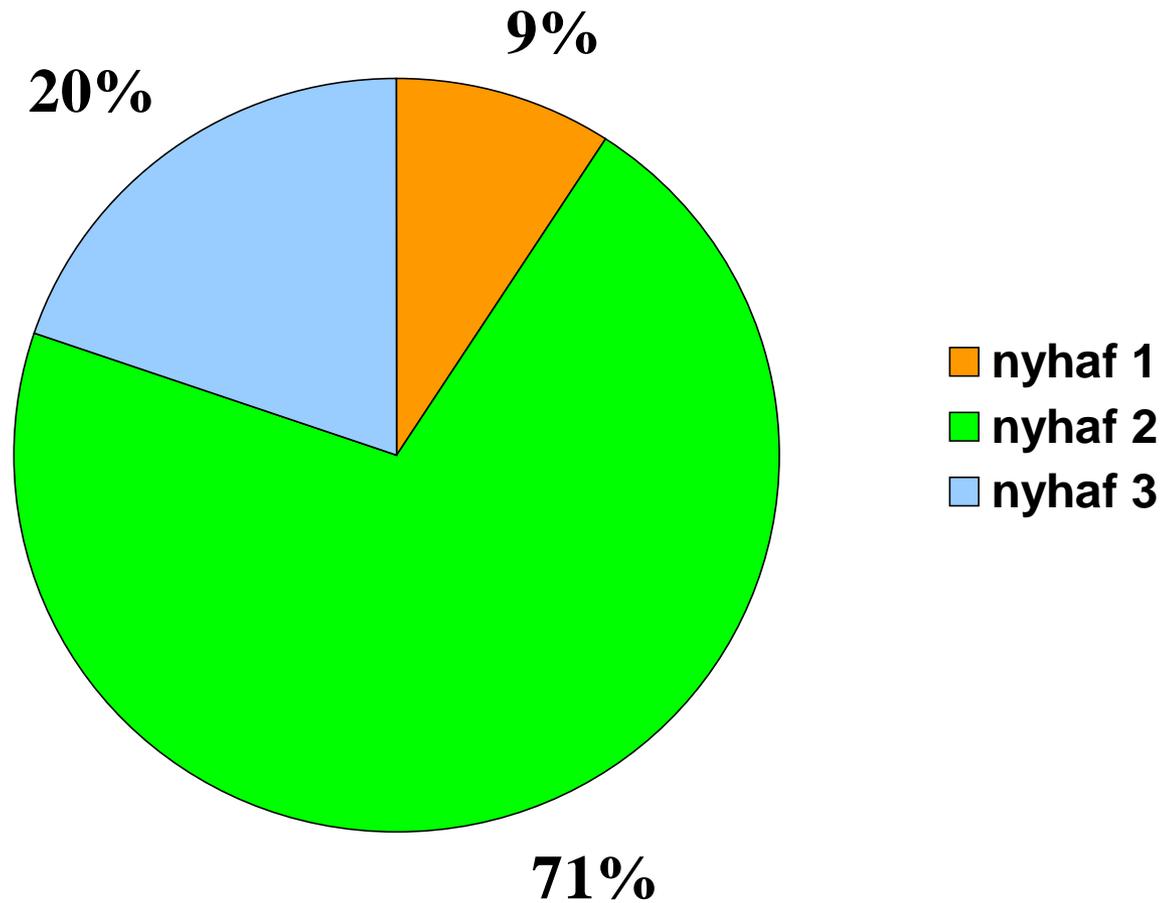
DIAGRAMMI A TORTA: l'intera circonferenza rappresenta il 100%, ciascuno spicchio indica la percentuale con cui si presenta un carattere. E' indicato per le variabili qualitative

Sesso

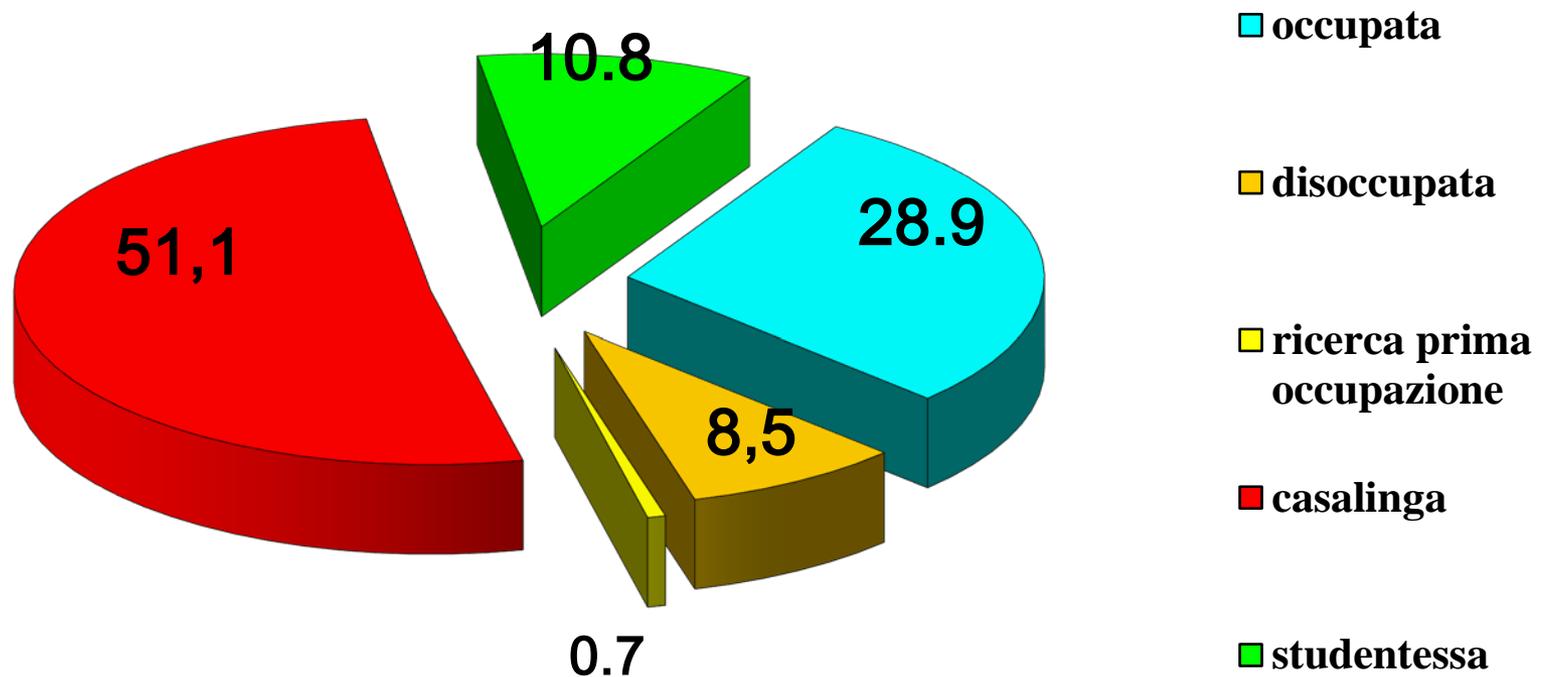
Trattamento

Classe NYHA...

# DISTRIBUZIONE DEI PAZIENTI PER CLASSE NYHA ALLA FINE DELLO STUDIO.



# IVG e condizione professionale anno 2001



# DESCRIVERE I DATI

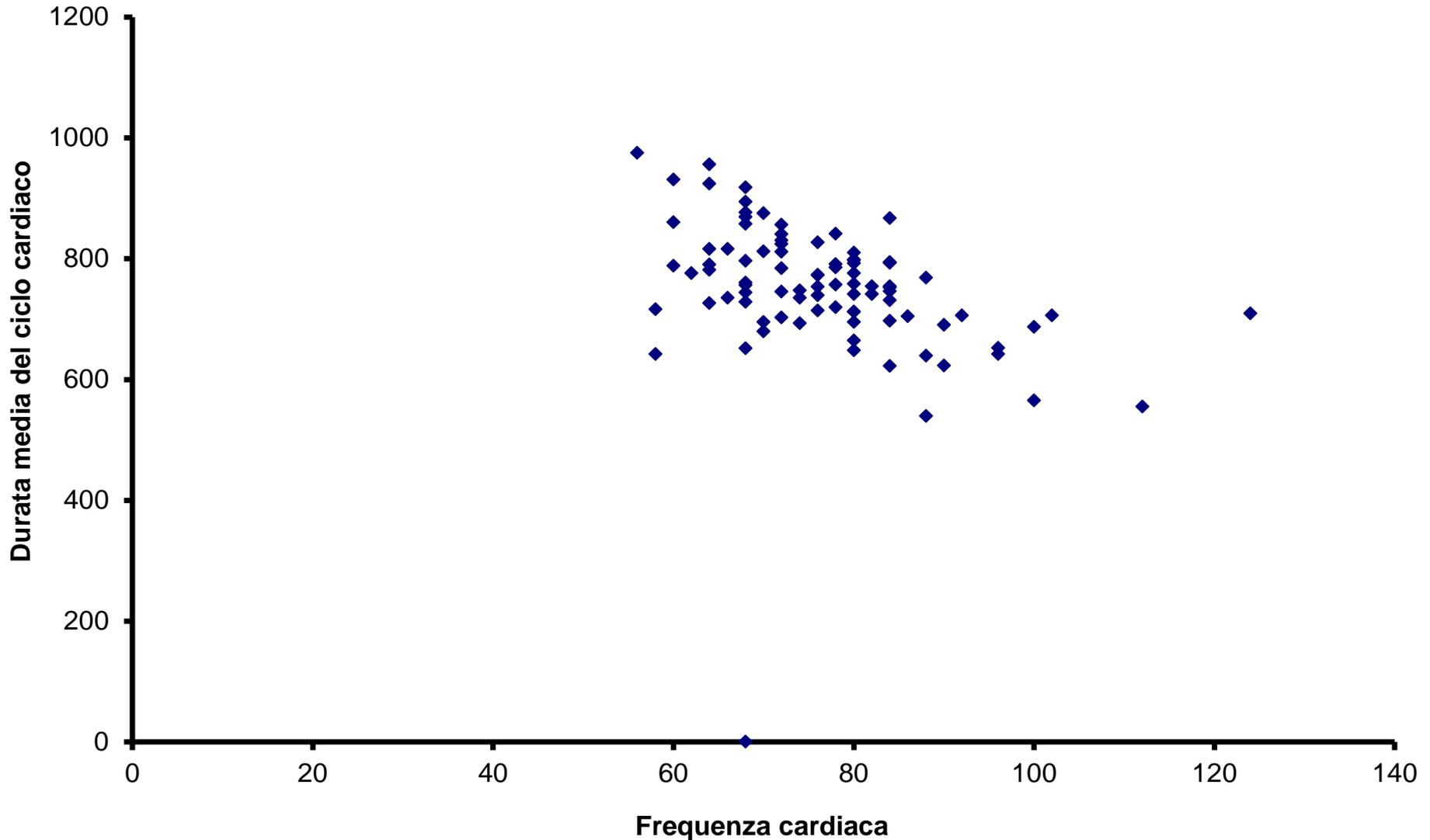


## RAPPRESENTAZIONI CON GRAFICI

**DIAGRAMMI A DISPERSIONE:** utili per valutare le relazioni tra variabili quantitative; sull'asse x e sull'asse y ci sono le due variabili di cui si vuole studiare la relazione.

Il grafico mostra come ciascuna osservazione si colloca nel piano cartesiano in relazione ai valori delle due variabili

# DIAGRAMMA A DISPERSIONE PER L'ANALISI DELLA RELAZIONE TRA FREQUENZA CARDIACA E DURATA MEDIA DEL CICLO CARDIACO.

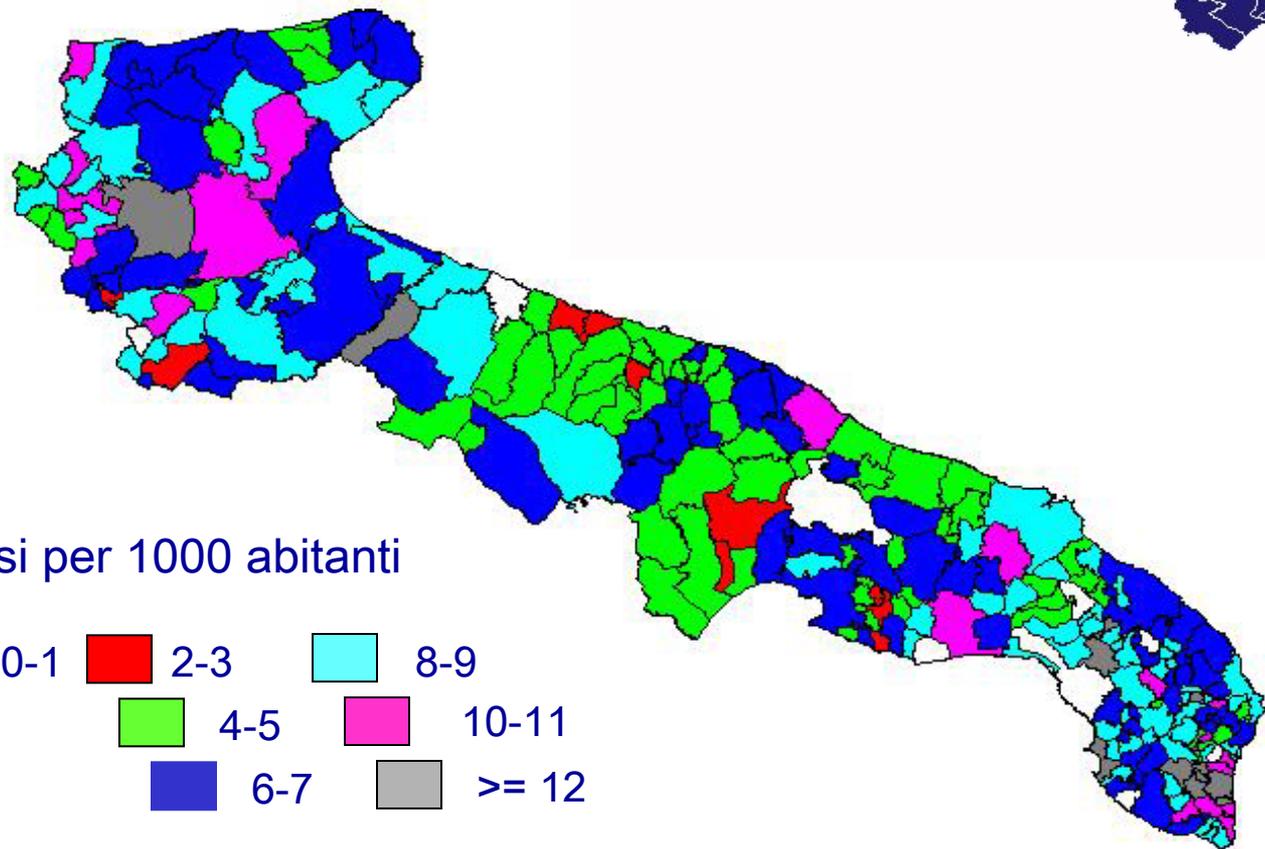


# DESCRIVERE I DATI



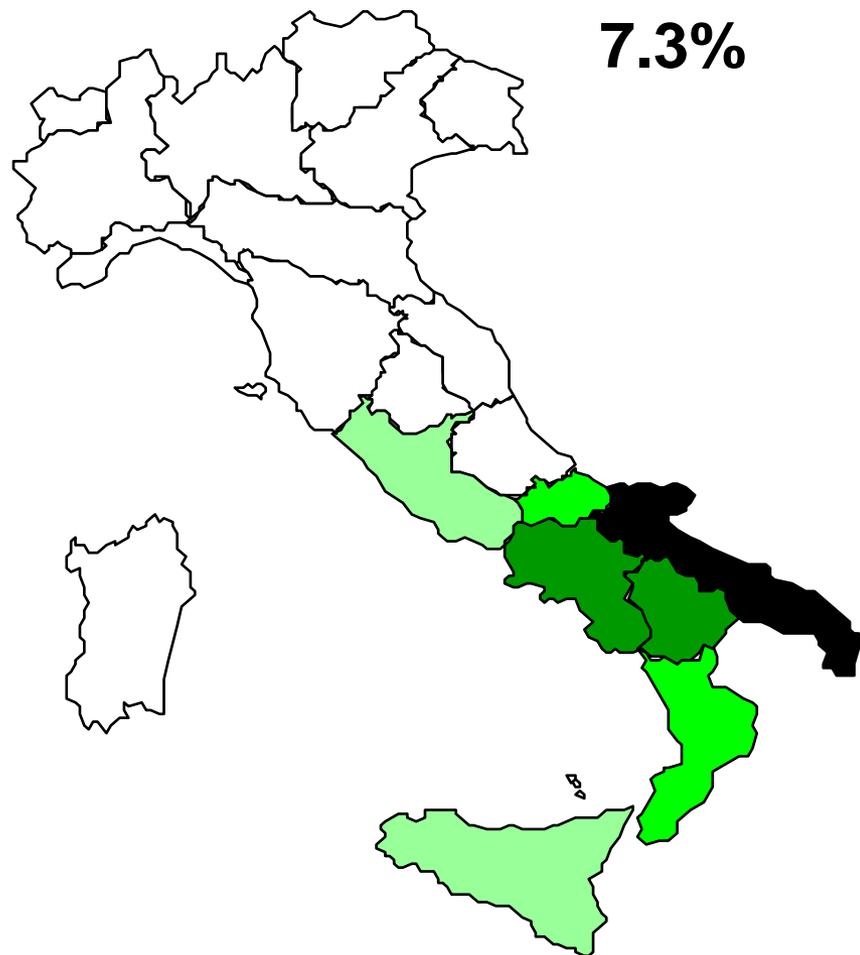
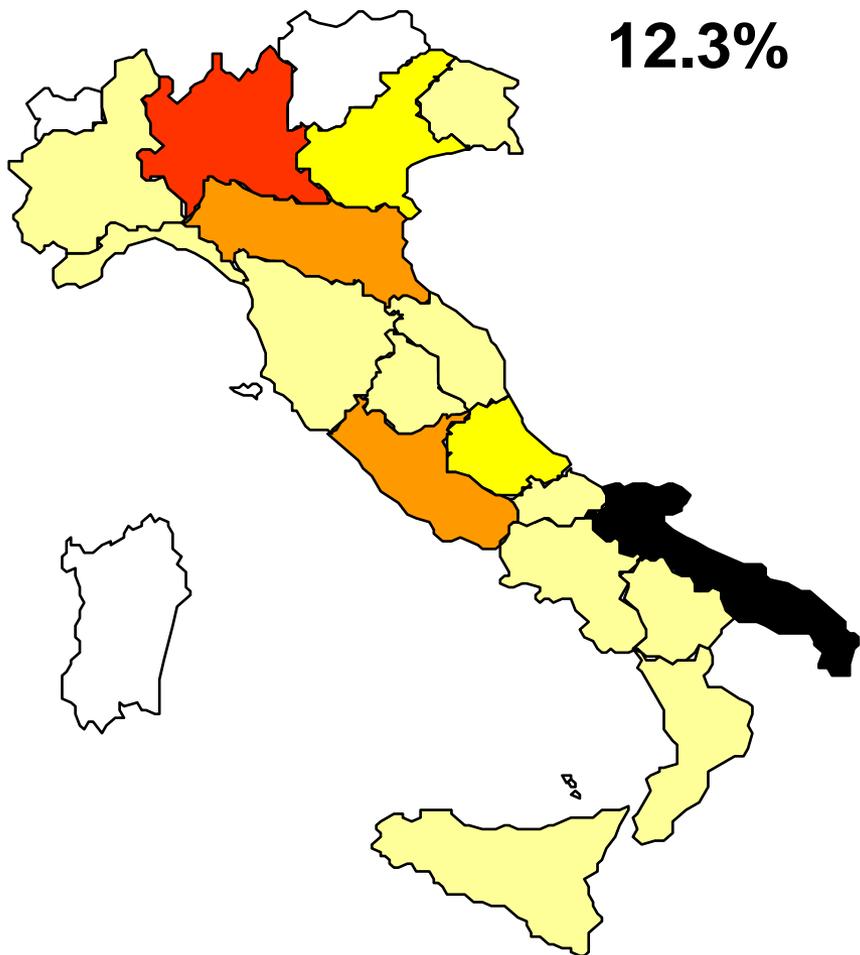
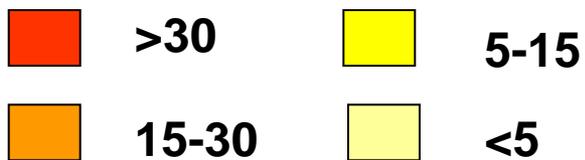
RAPPRESENTAZIONI CON CARTOGRAMMI

# Tassi di ospedalizzazione per 1000 abitanti e distribuzione geografica dei posti letto



Tassi per 1000 abitanti

# Distribuzione geografica dei ricoveri esportati ed importati anno 2002



Città	Valore	Città	Valore
1	68	29	57
2	63	30	51
3	42	31	12
4	27	32	32
5	30	33	49
6	36	34	38
7	28	35	42
8	32	36	27
9	79	37	31
10	27	38	50
11	22	39	38
12	23	40	21
13	24	41	16
14	25	42	24
15	44	43	69
16	65	44	47
17	43	45	23
18	25	46	22
19	74	47	43
20	51	48	27
21	36	49	49
22	42	50	28
23	28	51	23
24	31	52	19
25	28	53	46
26	25	54	30
27	45	55	43
28	12	56	49
		57	12

Si dispone dei dati di inquinamento atmosferico di 57 città. Il valore è riferito a ossidi di azoto in  $\text{microgr./m}^3$  di aria.

Si costruisca la distribuzione di frequenza applicando la regola di Sturges.

Si costruisca l'istogramma (con il poligono di frequenza).

Per determinare il numero di classi (K) si usa la formula di Sturges:

$$K=1+3,322(\text{Log } 57)=1+3,322*1,7559=7$$

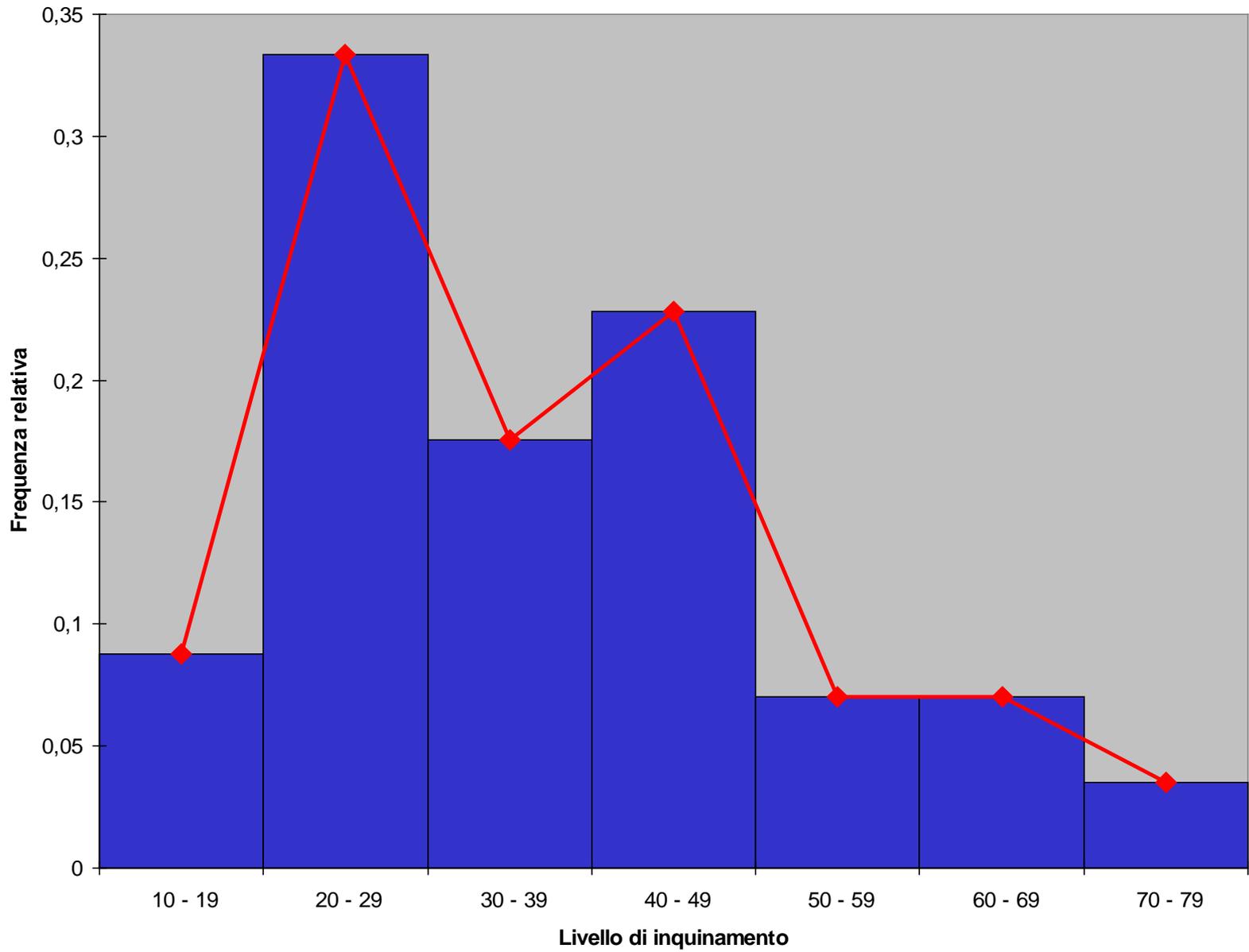
Dove n è la numerosità campionaria 57.

L'ampiezza (w) della classe sarà data da:

$$W= R / K = (79-12) / 7 = 67 / 7 = 9,6$$

Dove R è il range ossia la differenza tra il valore più grande 79 e quello più piccolo 12

Livello di inquinamento	Frequenza assoluta	Frequenza cumulativa assoluta	Frequenza relativa	Frequenza cumulativa relativa	Frequenza relativa
10 - 19	5	5	0,09	0,09	8,77
20 - 29	19	24	0,33	0,42	33,33
30 - 39	10	34	0,18	0,60	17,54
40 - 49	13	47	0,23	0,82	22,81
50 - 59	4	51	0,07	0,89	7,02
60 - 69	4	55	0,07	0,96	7,02
70 - 79	2	57	0,04	1,00	3,51
	57		1,00		100,00



Città	Valore	F.Cum ass	F. cum rel.	Città	Valore	F.Cum a	F. cum re
28	12	1	2%	32	32	30	53%
31	12	2	4%	6	36	31	54%
57	12	3	5%	21	36	32	56%
41	16	4	7%	34	38	33	58%
52	19	5	9%	39	38	34	60%
40	21	6	11%	3	42	35	61%
11	22	7	12%	22	42	36	63%
46	22	8	14%	35	42	37	65%
12	23	9	16%	17	43	38	67%
45	23	10	18%	47	43	39	68%
51	23	11	19%	55	43	40	70%
13	24	12	21%	15	44	41	72%
42	24	13	23%	27	45	42	74%
14	25	14	25%	53	46	43	75%
18	25	15	26%	44	47	44	77%
26	25	16	28%	33	49	45	79%
4	27	17	30%	49	49	46	81%
10	27	18	32%	56	49	47	82%
36	27	19	33%	38	50	48	84%
48	27	20	35%	20	51	49	86%
7	28	21	37%	30	51	50	88%
23	28	22	39%	29	57	51	89%
25	28	23	40%	2	63	52	91%
50	28	24	42%	16	65	53	93%
5	30	25	44%	1	68	54	95%
54	30	26	46%	43	69	55	96%
24	31	27	47%	19	74	56	98%
37	31	28	49%	9	79	57	100%
8	32	29	51%				

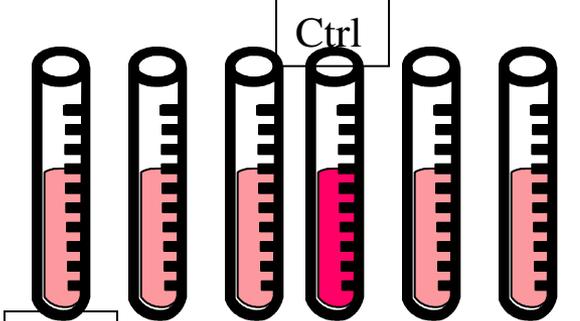
MEDIA

$$\bar{x} = \frac{\sum x_i}{n} = \frac{2093}{57} = 36,72$$

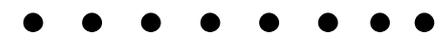
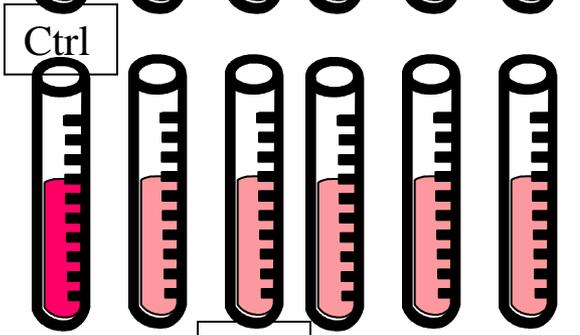
VARIANZA

$$\begin{aligned} S^2 &= \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1} = \frac{90949 - \frac{4380649}{57}}{56} = \\ &= \frac{90949 - 76853,49}{56} = \frac{14095,51}{56} = 251,71 \end{aligned}$$

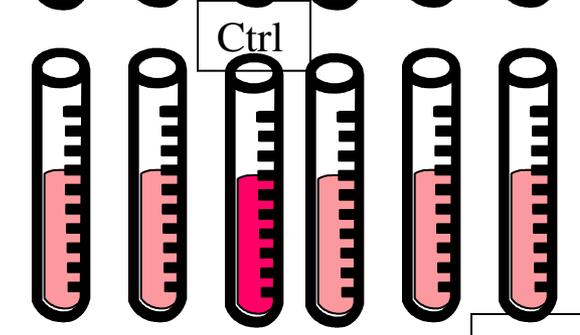
Giorno 1



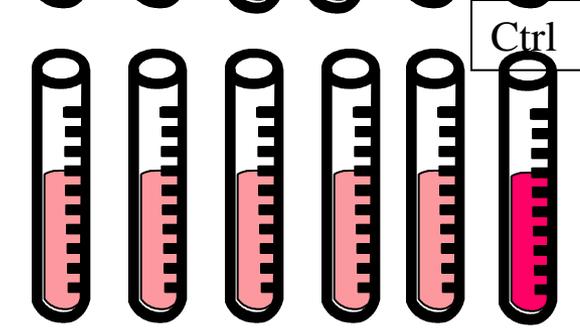
Giorno 2



Giorno 3

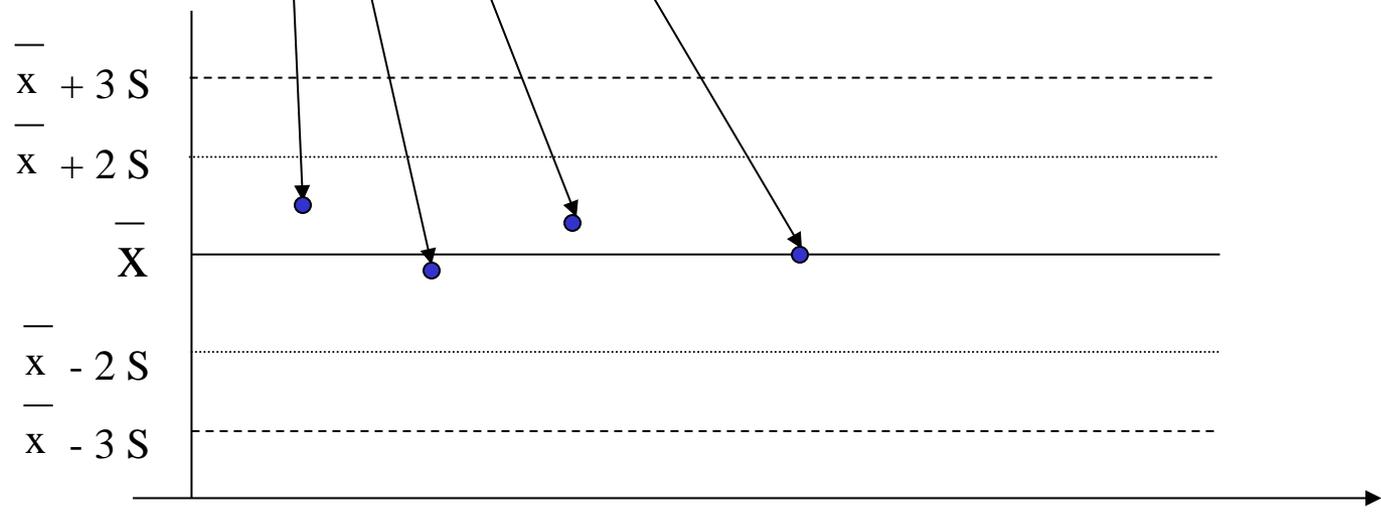
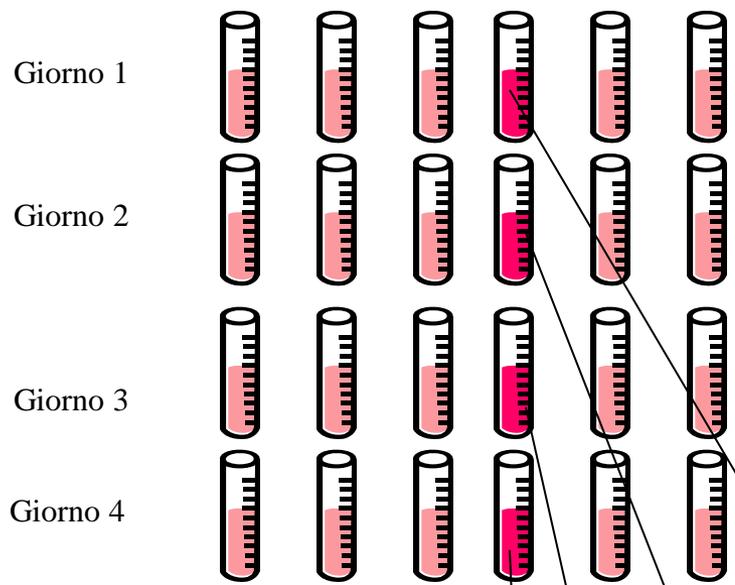


Giorno 4



Determino media e deviazione standard dei sieri di controllo





Giorni

Alcuni autori hanno condotto uno studio in cui hanno misurato i livelli di istamina nel liquido del lavaggio broncoalveolare in soggetti con rinite allergica, soggetti con asma e soggetti normali. Una delle misure ottenuta è stata la proteina totale ( $\mu\text{g/ml}$ ). Alcuni risultati sono riportati nella seguente tabella.

76	114	73	128	82	55
77	74	47	88	62	35
149	77	73	58	61	72
54	57	62	84	106	66
55	91	67	44	61	59
51	71	44	53	63	95
57	61	57	72	54	
88	100	78	59	83	
86	51	85	59	79	
54	72	41	67	153	
95	62	69	109	70	

MEDIA

$$\bar{x} = \frac{\sum x_i}{n} = \frac{4445}{61} = 72,87$$

VARIANZA

$$S^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1} = \frac{356911 - \frac{19758025}{61}}{60} = 550.15$$

DEVIAZIONE STANDARD

$$S = 23,45$$

$$K = 1 + 3,322(\text{Log } 61) = 1 + 3,322(\text{Log } 61) = 6,93$$

$$R = 153 - 35 = 118$$

$$W = R / K = (118) / 7 = 17,02$$

classi	Freq assoluta	Freq cumulata	Freq relativa	Freq rel. cumulata	Freq %
35-51	7	7	0.11	0.11	11%
52-68	23	30	0.38	0.49	38%
69-85	18	48	0.30	0.78	30%
86-102	7	55	0.11	0.90	11%
103-119	3	58	0.05	0.95	5%
120-136	1	59	0.02	0.96	2%
137-153	2	61	0.03	1.00	3%

**61**

**1.00**

**100%**

