Flowchart

- Question
- Definition of sample population
- Experiments and data collection
- Descriptive statistics
- Inferential statistics
- Answering the question

CONFIDENCE INTERVAL

CONFIDENCE INTERVALS CONSIST OF A RANGE OF VALUES (INTERVAL) THAT ACT AS GOOD ESTIMATES OF THE UNKNOWN POPULATION PARAMETER.

- The desired level of confidence is set by the researcher (**not determined by data**). Most commonly, the 95% confidence level is used. However, other confidence levels can be used, for example, 90% and 99%.

-WE KNOW THE VARIANCE ($\sigma 2$) OF THE POPULATION (FOR EXAMPLE FROM PREVIOUS SURVEYS)

- WE DO NOT KNOW THE VARIANCE OF THE POPULATION (MOST FREQUENT CASE)

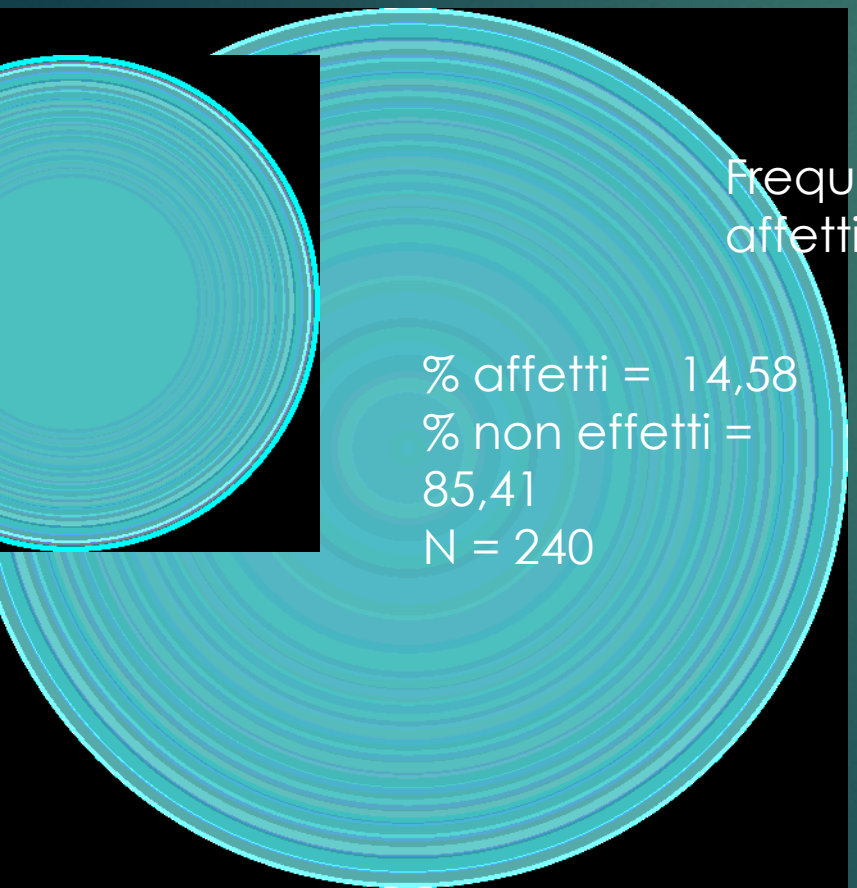
Esempio

CASE 1:

EXCEL: =CONFIDENZA(A; Σ; DIMENSIONI CAMPIONE)

Caso 2

ESAMINO 240 TOPINI E TROVO CHE IL 35 HANNO UNA MUTAZIONE XYZ DI MIO INTERESSE. QUALE SARÀ LA STIMA DELLA PERCENTUALE DI TOPI AFFETTI, CON UN IC DEL 95%?



% affetti = 14,58
% non effetti = 85,41
N = 240

$$\text{Frequenza} +/- 1.96 * \sqrt{\frac{\text{affetti} * (1 - \text{affetti})}{N}}$$

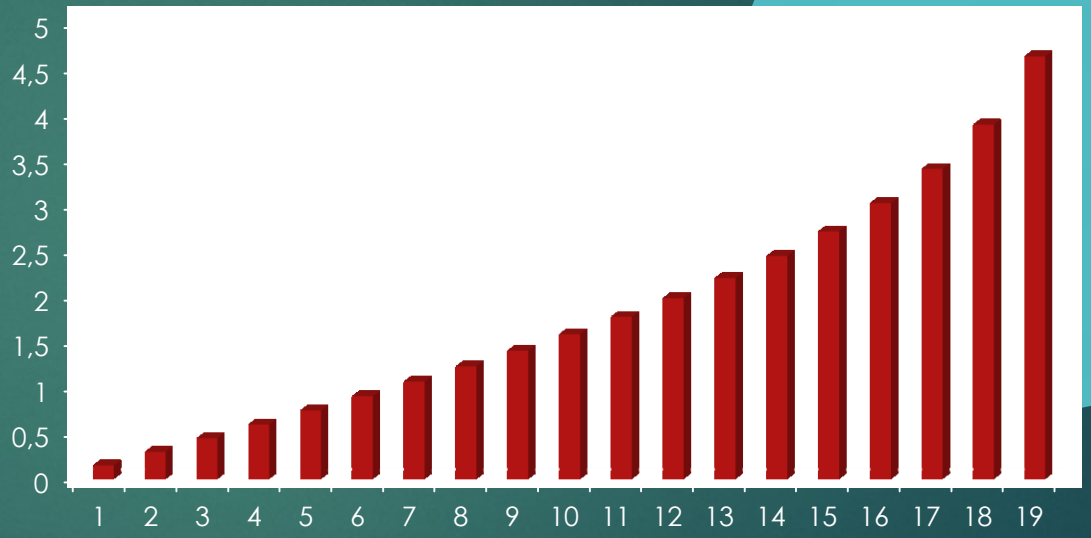
% compresa tra

30,53

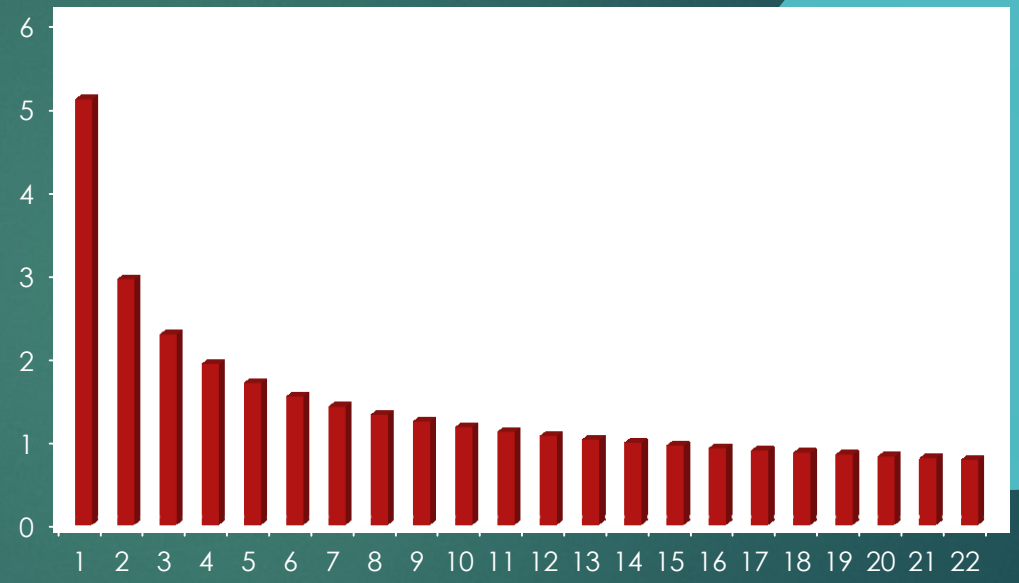
39,46

Some considerations…

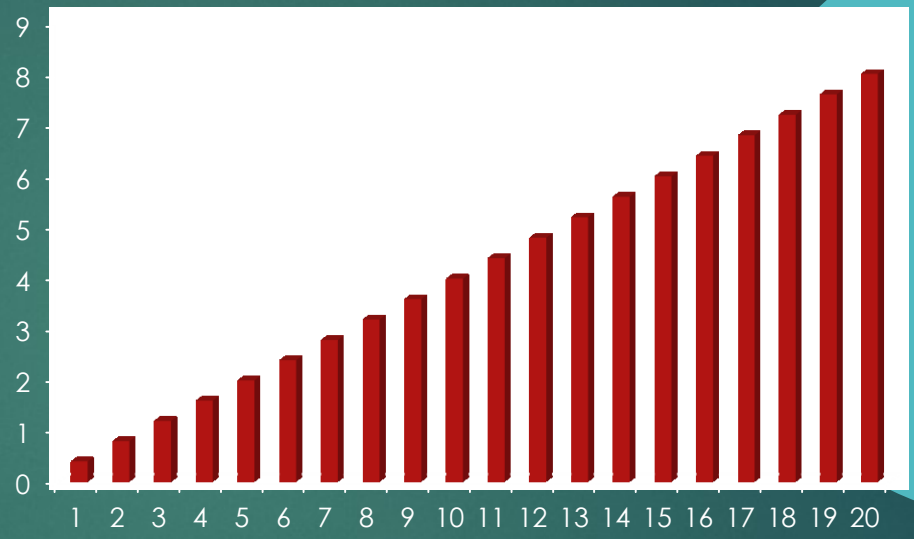
alfa	dev st	numerosità	intervallo
1	5,8	24	
0,95	5,8	24	0,14848
0,9			0,297546
0,85	5,8		0,447802
0,8	5,8		0,599885
0,75	5,8		0,754487
0,7	5,8	24	0,912377
0,65	5,8	24	1,074436
0,6	5,8	24	1,241697
0,55	5,8	24	1,4154
0,5	5,8	24	1,597084
0,45	5,8	24	1,788702
0,4	5,8	24	1,992824
0,35	5,8	24	2,212958
0,3	5,8	24	2,454109
0,25	5,8	24	2,723843
0,2	5,8	24	3,034509
0,15	5,8	24	3,40858
0,1	5,8	24	3,89475
0,05	5,8	24	4,640881
-3,2E-16	5,8	24	



alfa	dev st	numerosità	intervallo confidenza
0,05	5,8	5	5,083831
0,05	5,8	15	2,935151
0,05	5,8	25	2,273558
0,05	5,8	35	1,921507
0,05	5,8	45	1,69461
0,05			1,532833
0,05			1,410001
0,05	5,8		1,312639
0,05	5,8		1,23301
0,05	5,8		1,166311
0,05	5,8		1,109383
0,05	5,8	1	1,060052
0,05	5,8	125	1,016766
0,05	5,8	135	0,978384
0,05	5,8	145	0,944044
0,05	5,8	155	0,913083
0,05	5,8	165	0,884981
0,05	5,8	175	0,859324
0,05	5,8	185	0,835777
0,05	5,8	195	0,814064
0,05	5,8	205	0,793961
0,05	5,8	215	0,775277



alfa	dev st	numerosità	intervallo confidenza
0,05	1	24	0,400076
0,05	2	24	0,800152
0,05	3	24	1,200228
0,05	4	24	1,600304
0,05	5	24	2,00038
0,05	6	24	2,400456
0,05	7	24	2,800532
0,05	8	24	3,200608
0,05	9	24	3,600684
0,05	10	24	4,00076
0,05	11	24	4,400836
0,05	12	24	4,800912
0,05	13	24	5,200988
0,05	14	24	5,601064
0,05	15	24	6,00114
0,05	16	24	6,401216
0,05	17	24	6,801292
0,05	18	24	7,201368
0,05	19	24	7,601443
0,05	20	24	8,001519

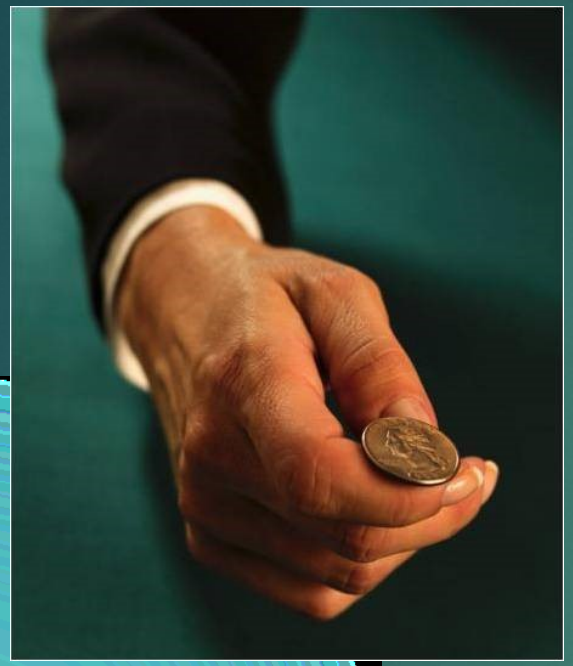


Descriptive statistics

- distribution
- Measure of central tendency
- Measure of varialbility

Yes or not experiments

Binomial distribution



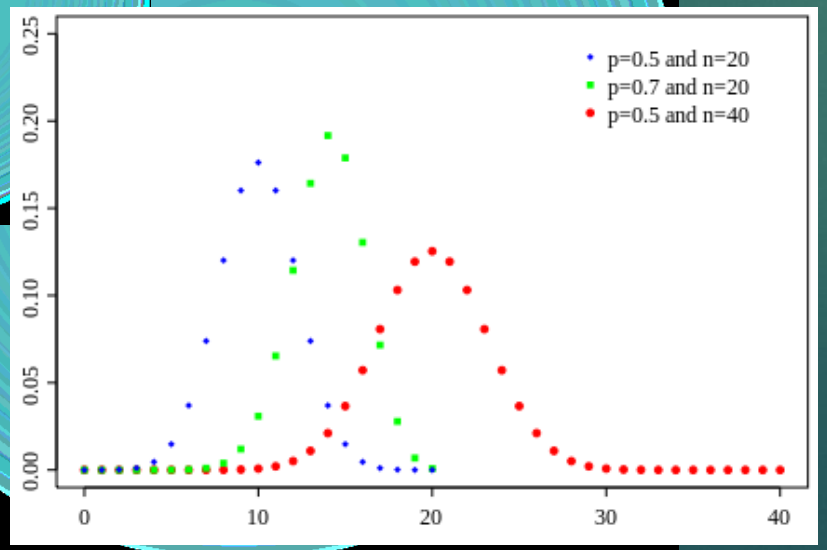
BERNOULLI'S EXPERIMENT: :

1) THE RESULT OF EACH TEST MAY BE ONLY "SUCCESS" OR "FAILURE";

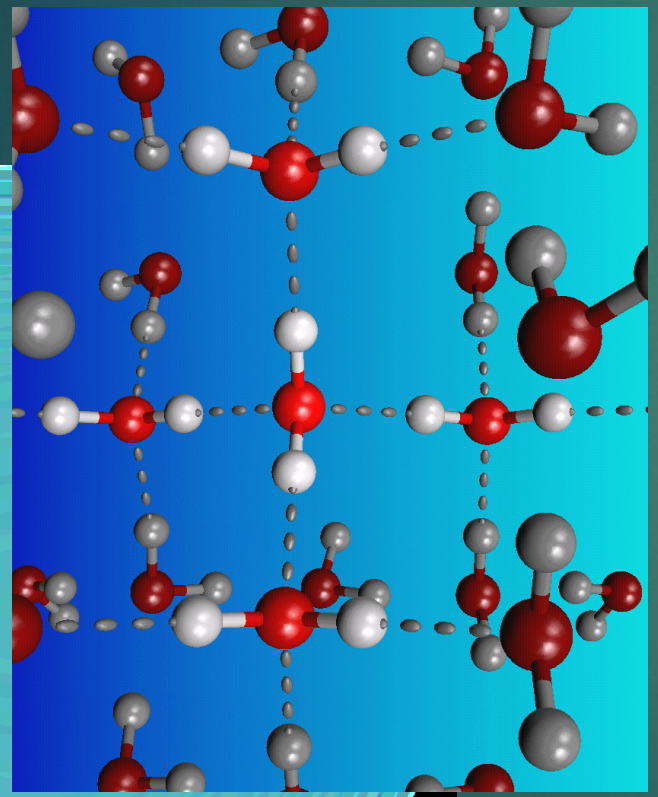
2) THE RESULT OF EACH TEST IS INDEPENDENT OF THE RESULTS OF THE PREVIOUS TESTS;

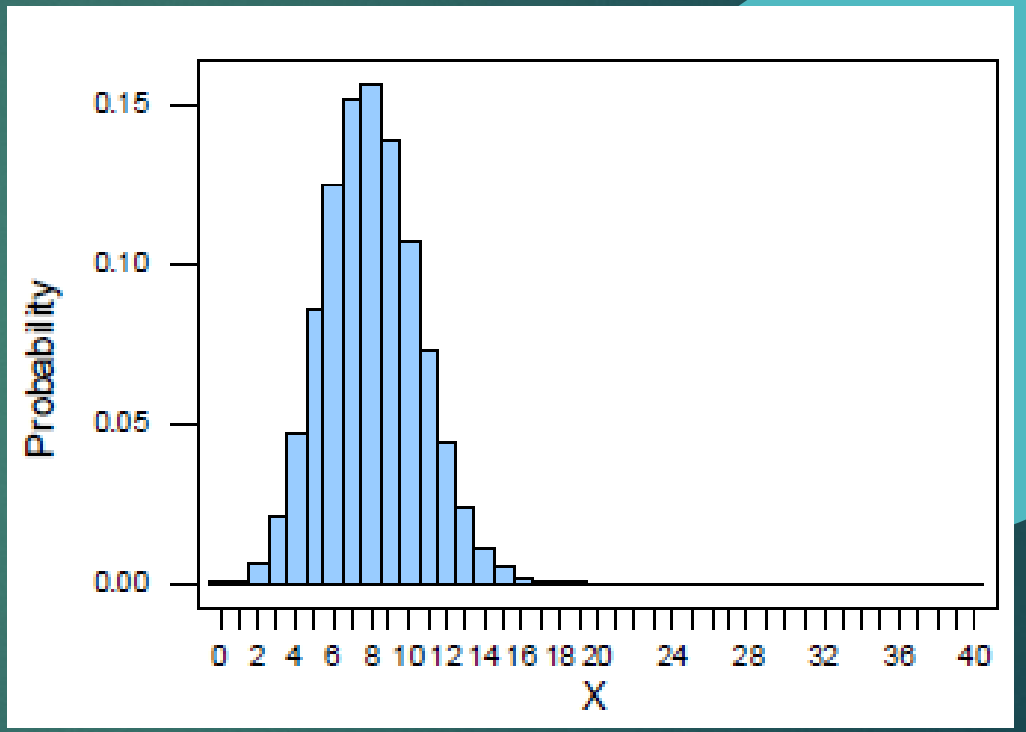
3) THE PROBABILITY P OF "SUCCESS", AND HENCE THE PROBABILITY Q = 1 - P OF "BANKRUPTCY", ARE

CONSTANT IN EACH TRIAL.

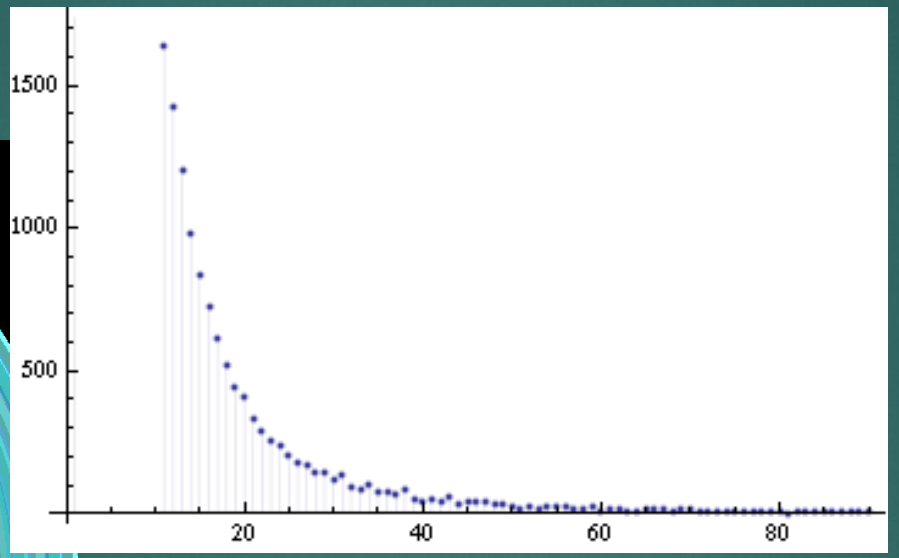


POISSON distribution





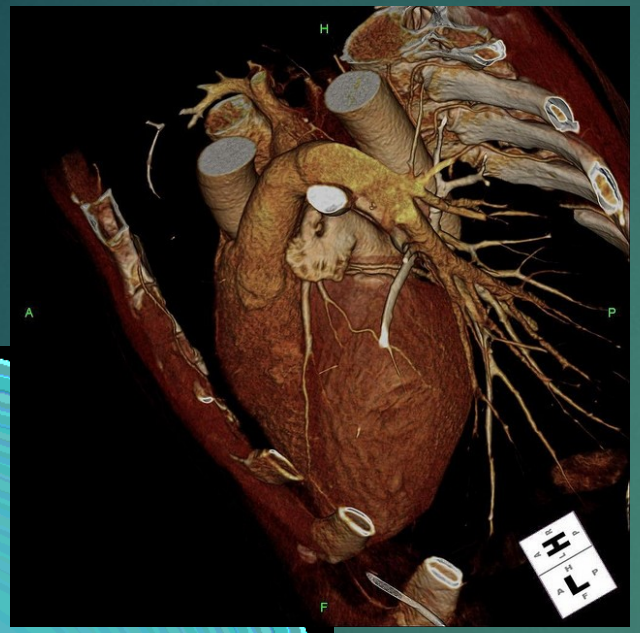
PARETO distribution



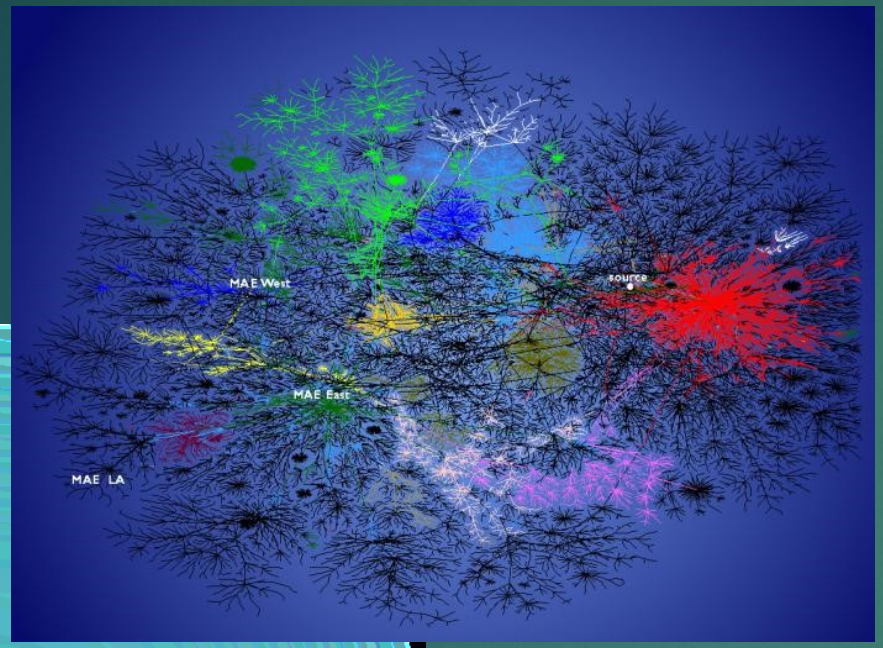
$y = a\,x^{-b}$

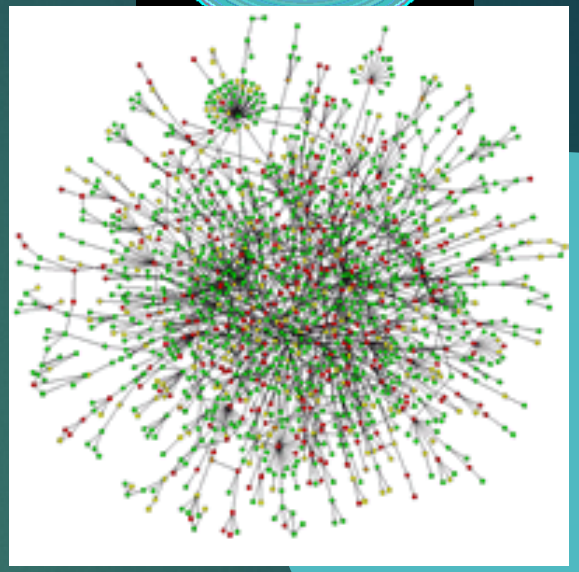
b > 1

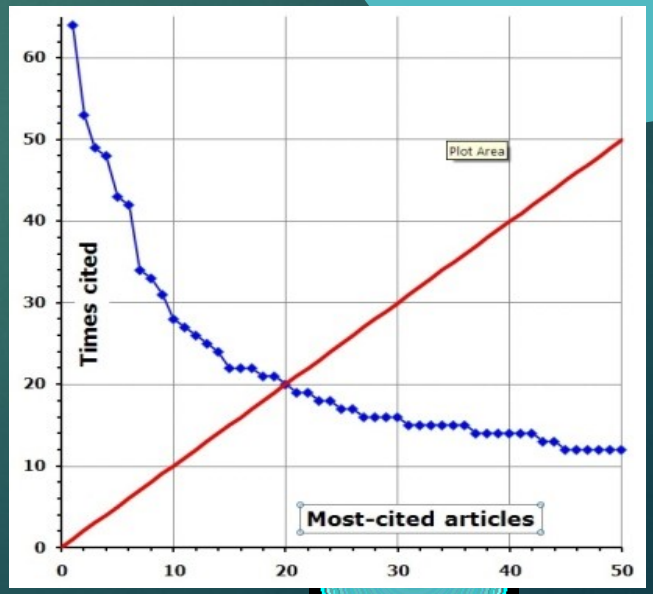
PARETO distribution



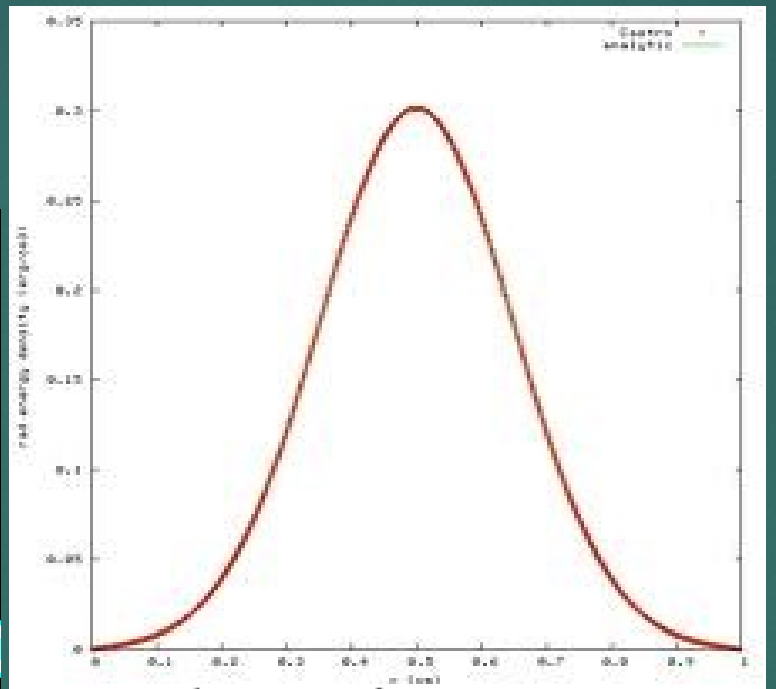
Scale free properties



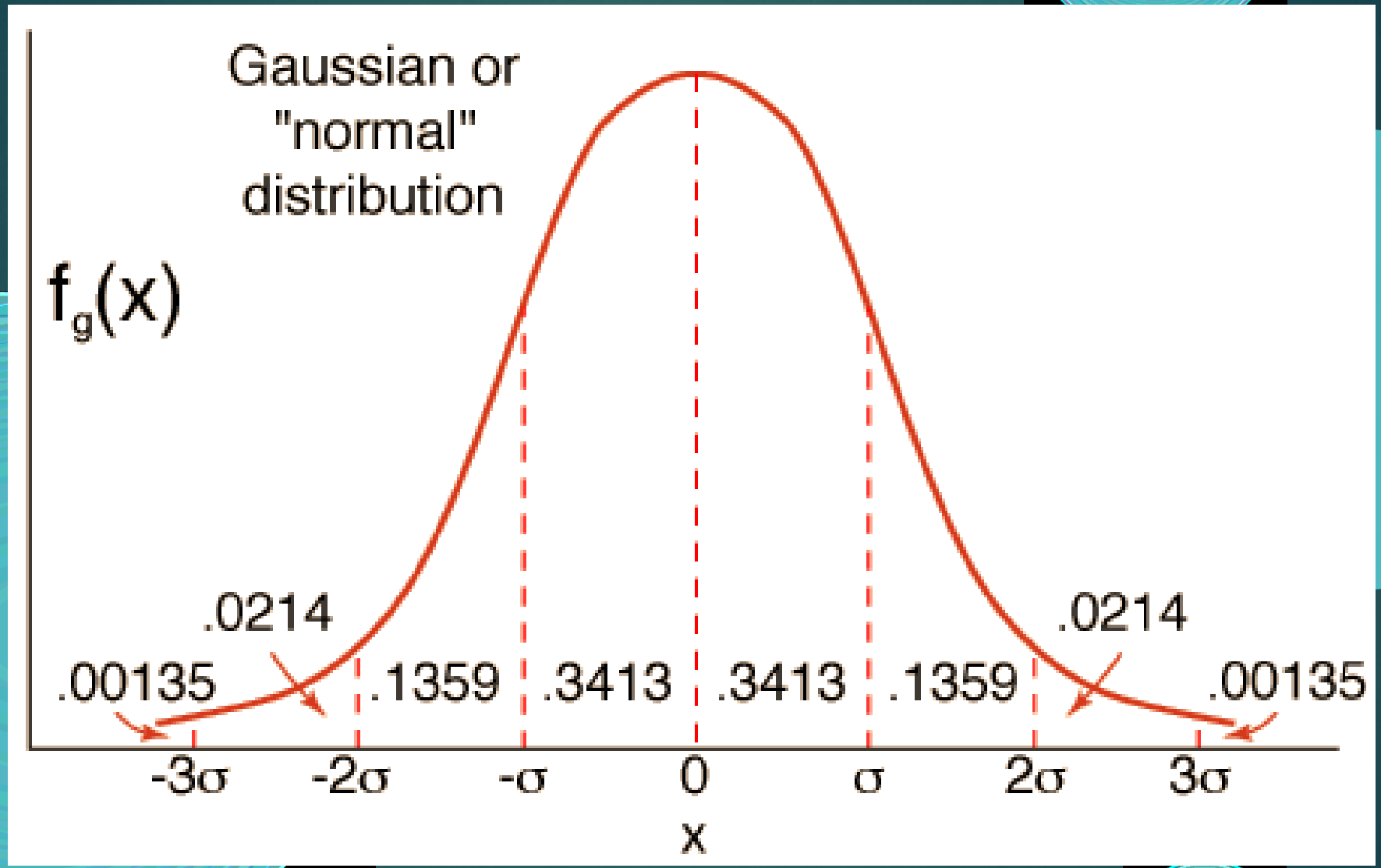




GAUSSIAN distribution

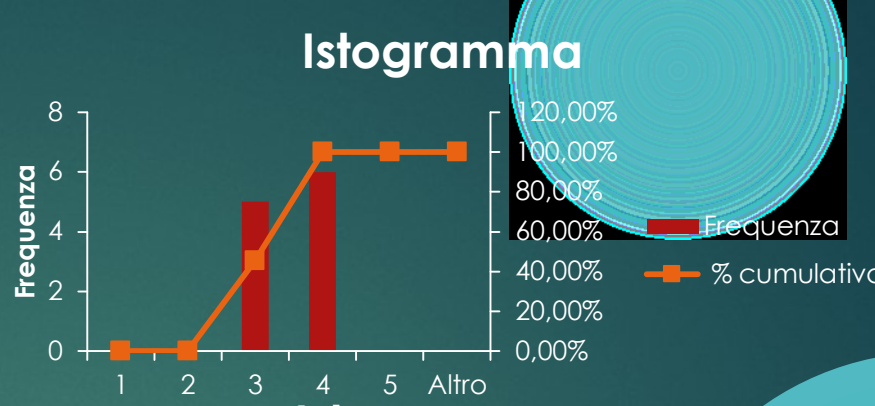


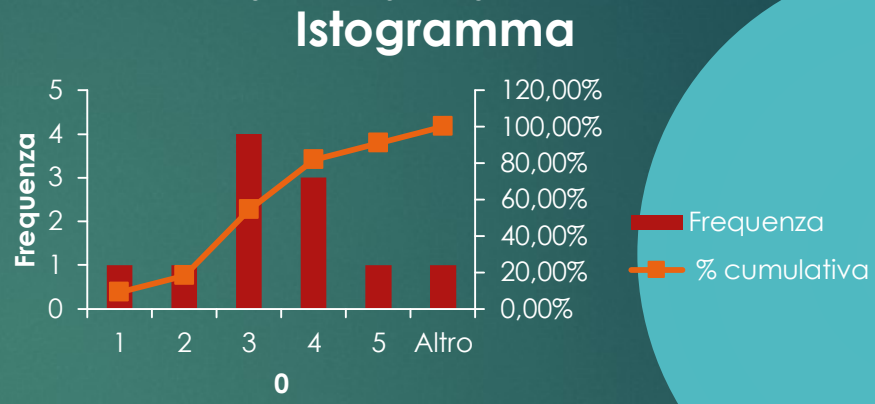
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{con} \quad -\infty < x < \infty$$

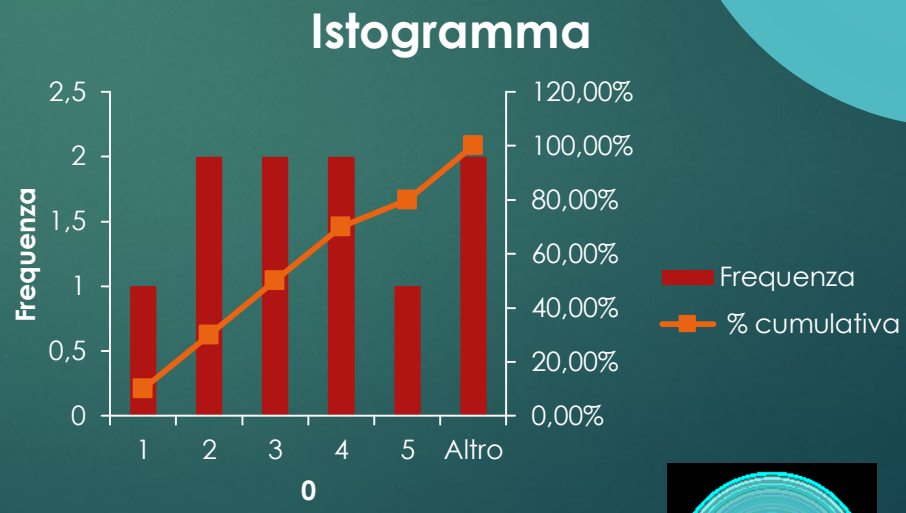

Gaussian or
"normal"
distribution
$f_g(x)$
.0214
.0214
.00135
.1359
.3413
.3413
.1359
.00135
-3σ
-2σ
-σ
0
σ
2σ
3σ
x

	Weight	
A	B	C
3	2	0,5
4	3	1
3,3	4	2
4	3	3
3	3	4
3	4	5
		6
3		7
3,5		2
3,3		3
3,4		4
	1	

media	3,318182	3,333333	3,409091
varianza	0,147636	1,878788	4,140909
dev. standard	0,384235	1,370689	2,034922
err. standard	0,03493	0,114224	0,184993

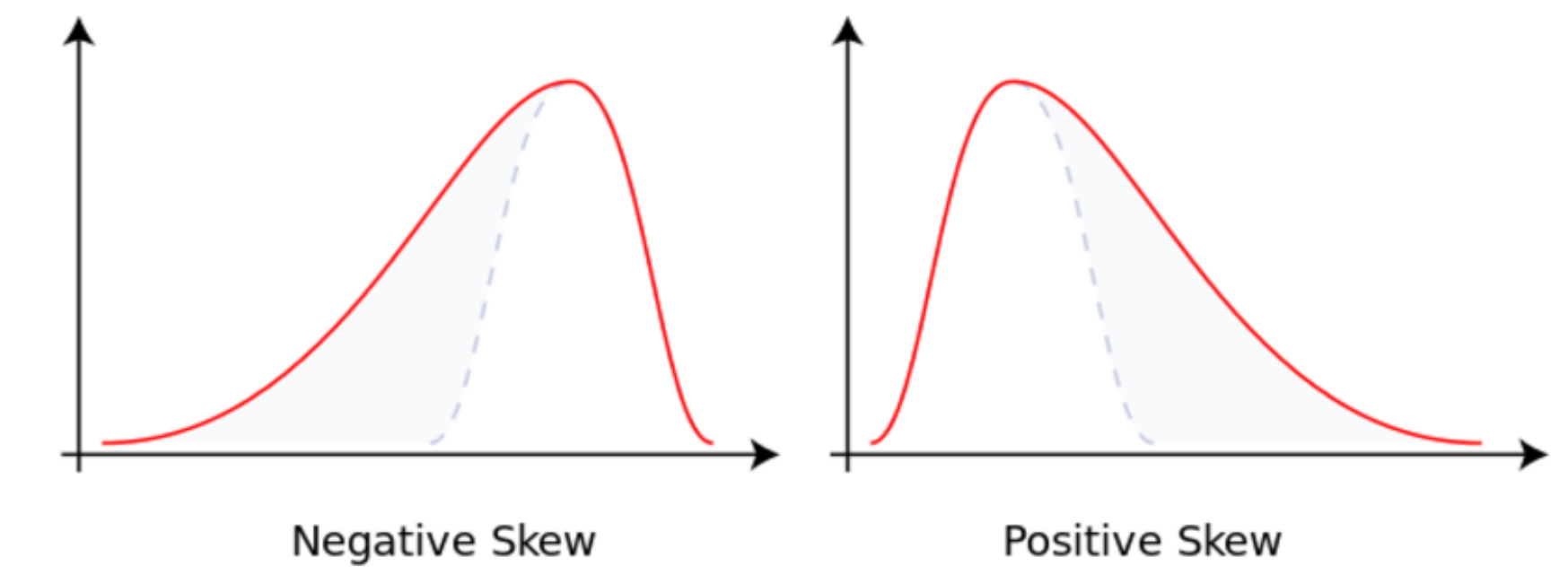




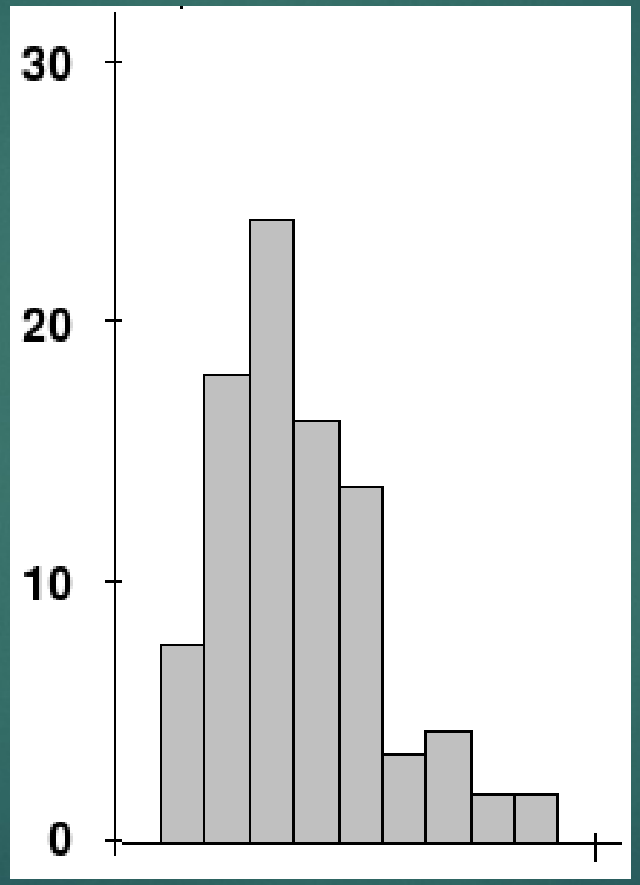


Skewness

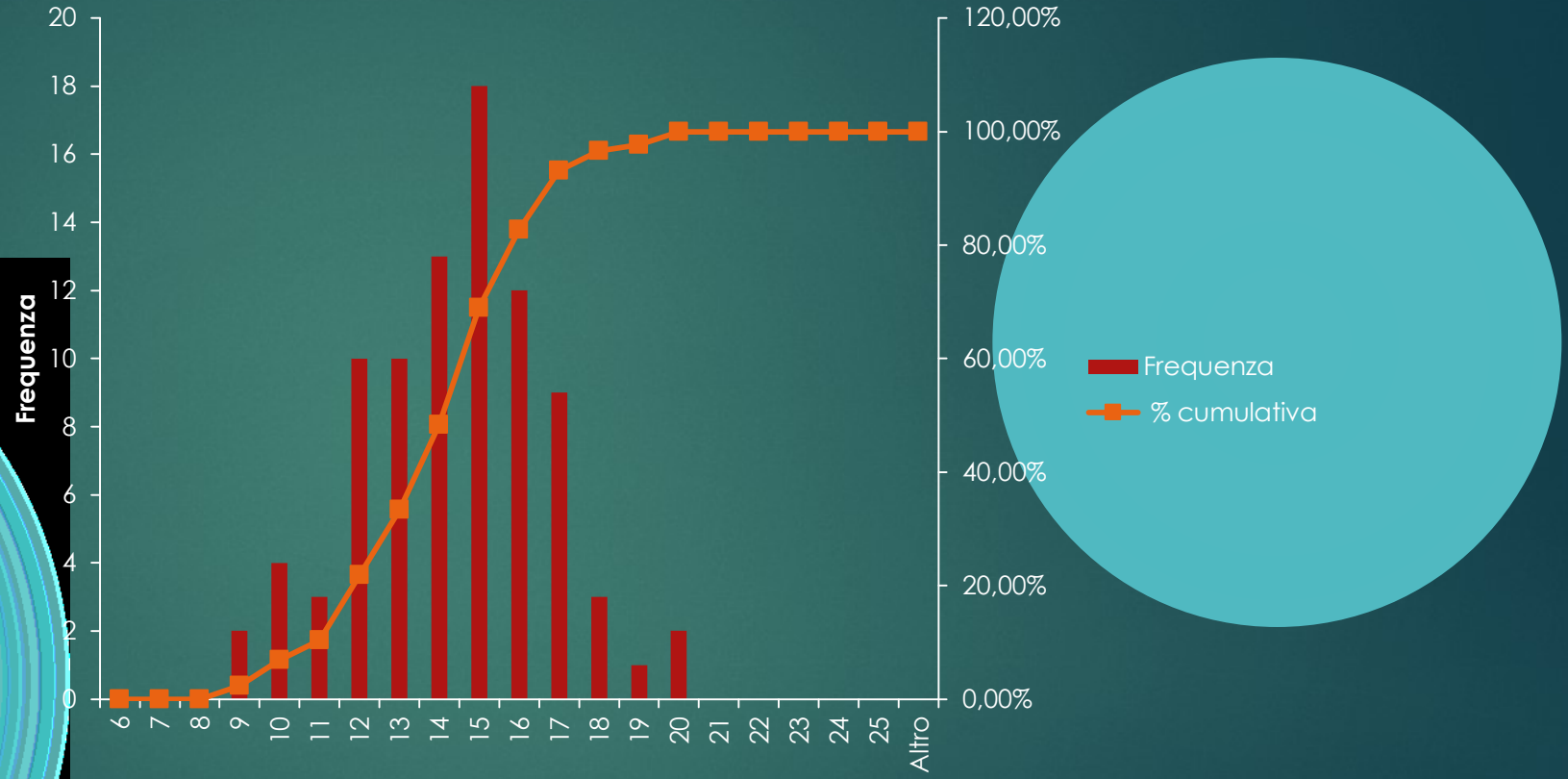
skewnessis a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive or negative, or undefined.



ASIMMETRY



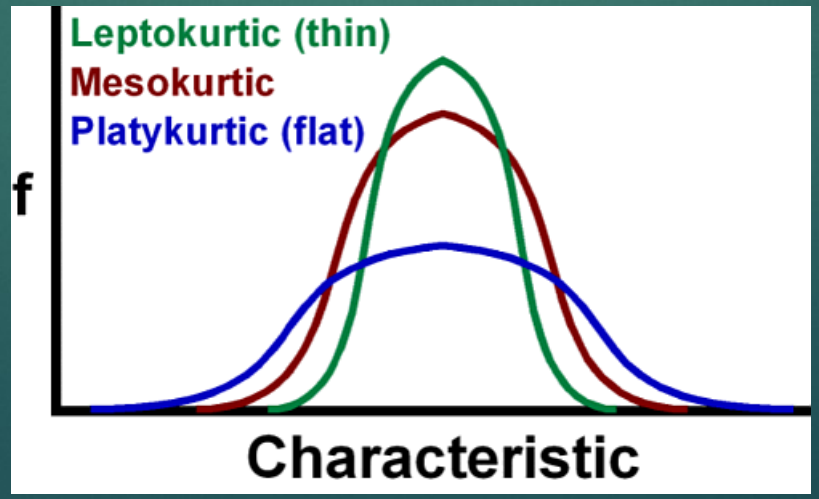
Skewness



asimmetria = -0,08511

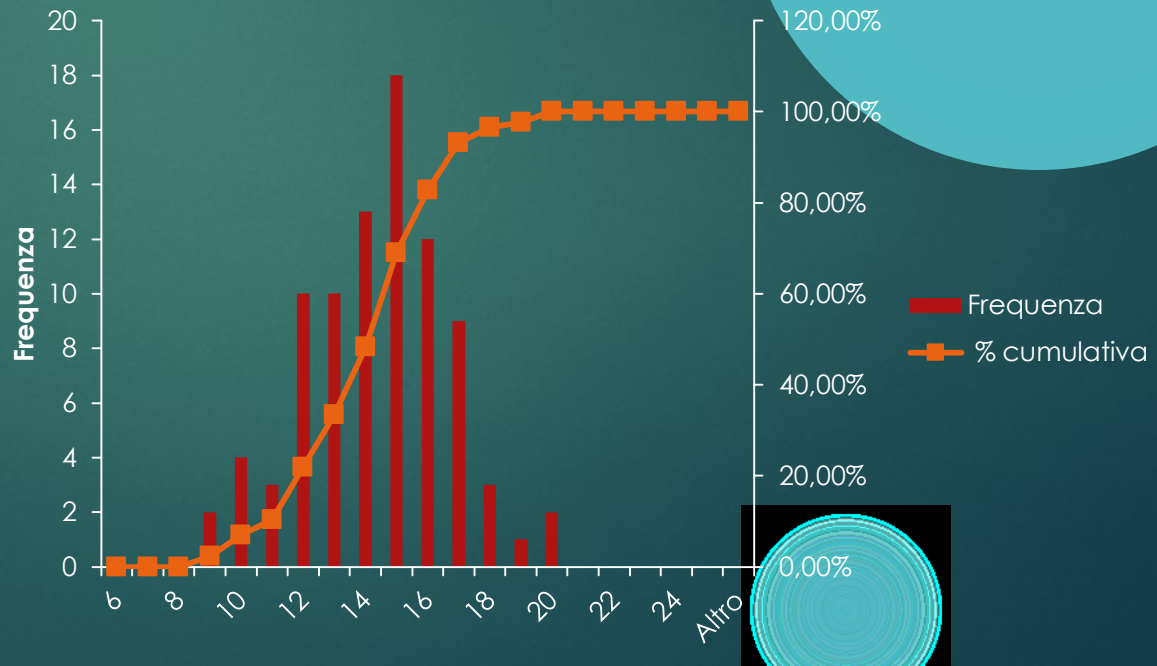
Kurtosis

- La **curtosi** (o **kurtosi**) è un **allontanamento dalla normalità distributiva**, rispetto alla quale si verifica un maggiore *appiattimento* distribuzione platicurtica) o un maggiore *allungamento* (distribuzione leptocurtica).

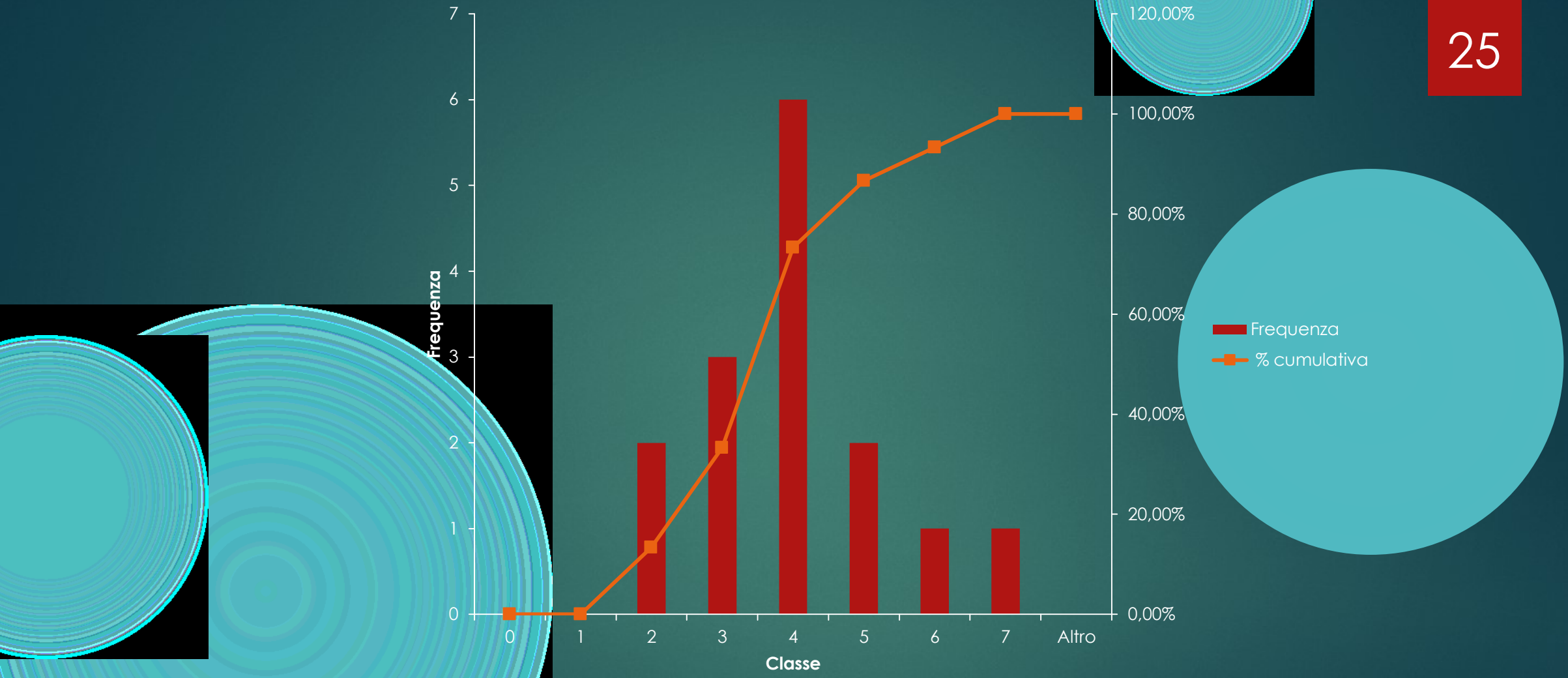


CURTOSI

- is a measure of the "tailedness" of the probability distribution of a real-valued random variable.



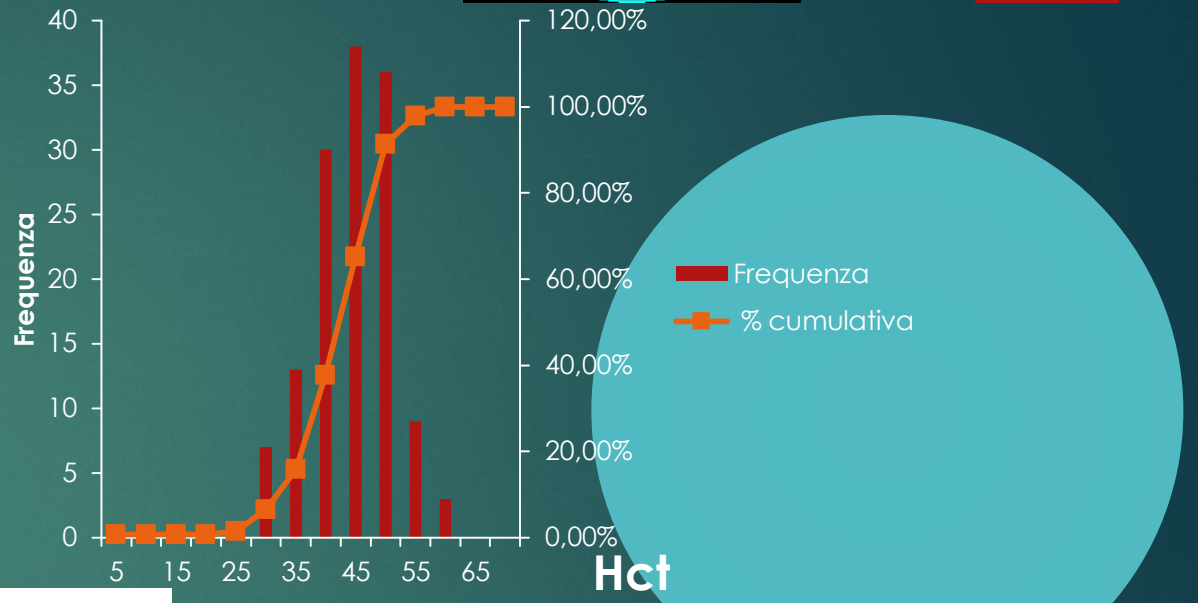
Curtosi = -0,26316

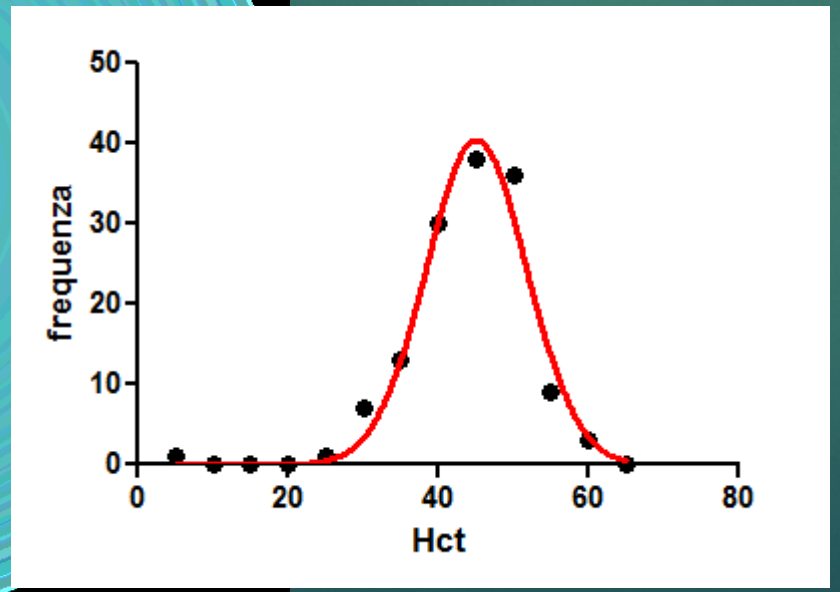


asimmetria	0,800298
curtosi	0,753757

Hct 139 DOGS

	Hct
Media	41,37116
Errore standard	0,626669
Mediana	42,15
Moda	45,4
Deviazione standard	7,361699
Varianza campionaria	54,19461
Curtosi	3,452818
Asimmetria	-0,96729
Intervallo	52,58
Minimo	4,82
Ma	57,4
Somma	5709,22
Conteggio	138
Più grande(1)	57,4
Più piccolo(1)	4,82
Livello di confidenza	1,239196





Goodness of Fit:

$R^2 = 0,9724$

MISURA DELLA TENDENZA CENTRALE

ARITHMETIC MEAN

The arithmetic mean (or simply "mean") of a sample , usualy denoted by $\bar{x}$, is the sum of the sampled values divided by the number of items in the sample

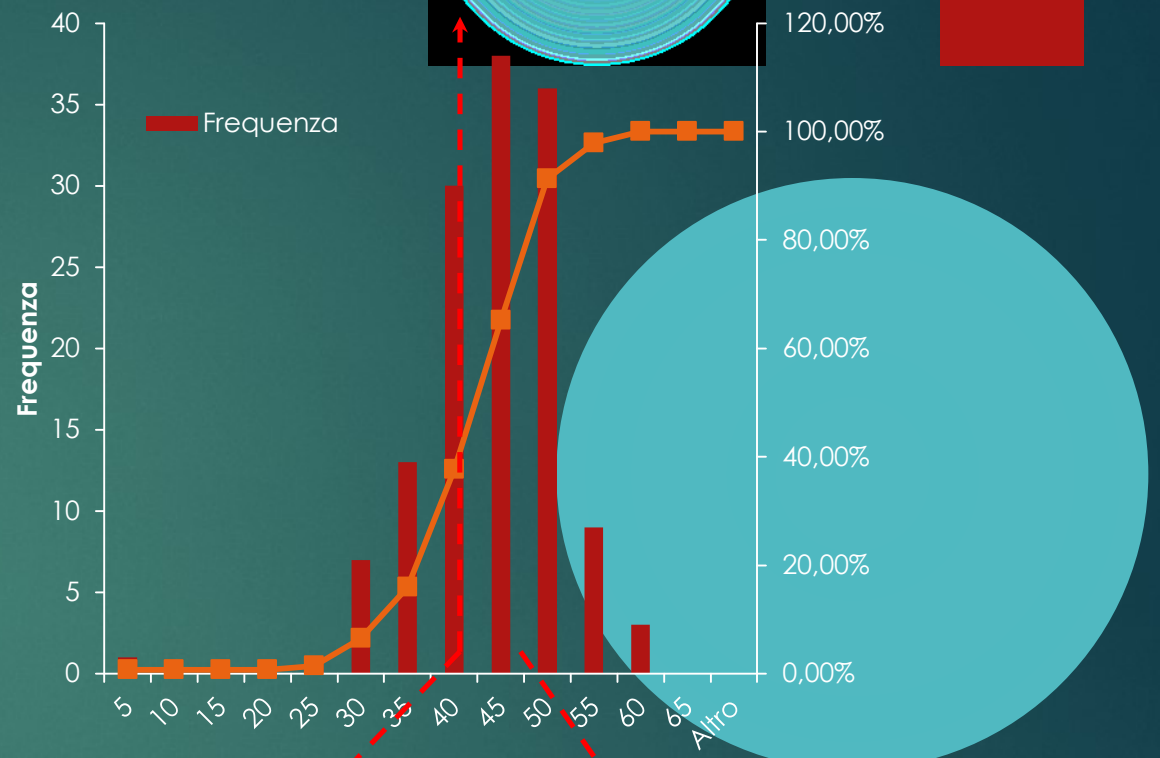
MEDIAN

The **median** is the value separating the higher half of a data sample, a population, or a probability distribution, from the lower half.

MODE

The **mode** of a set of data values is the value that appears most often

Hct	
Media	41,37116
Errore standard	0,626669
Mediana	42,15
Moda	45,4
Deviazione	7,361699
Varianza camp	54,19461
Curtosi	3,452818
Asimmetria	-0,96729
Intervallo	52,58
Minimo	4,82
Massimo	57,4
Somma	5709,22
Conteggio	138
Più grande(1)	57,4
Più piccolo(1)	4,82
Livello di confidenza(95,0%)	1,239196



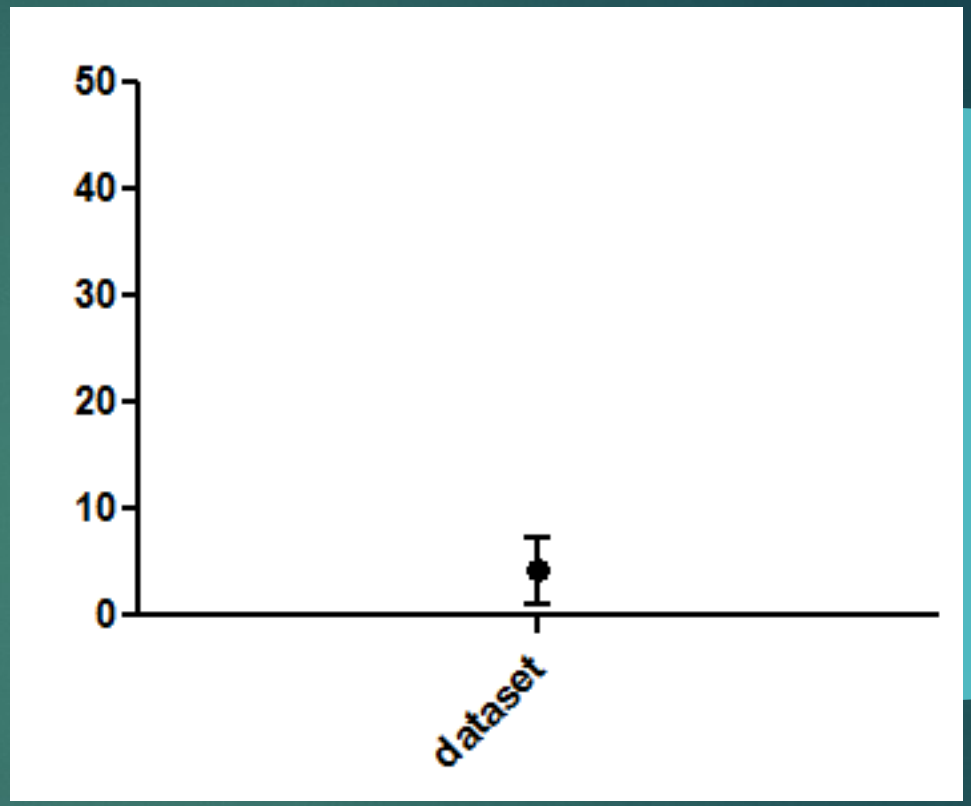
mean

median

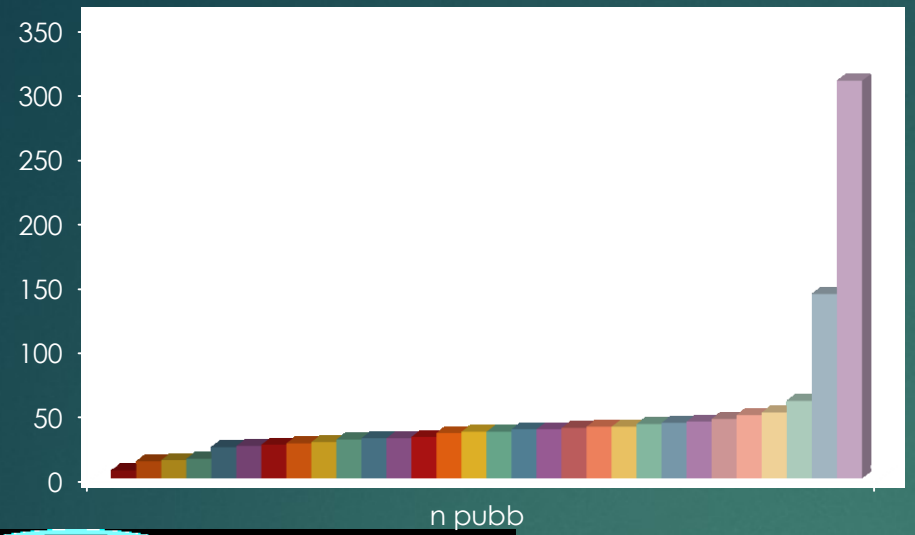
mode

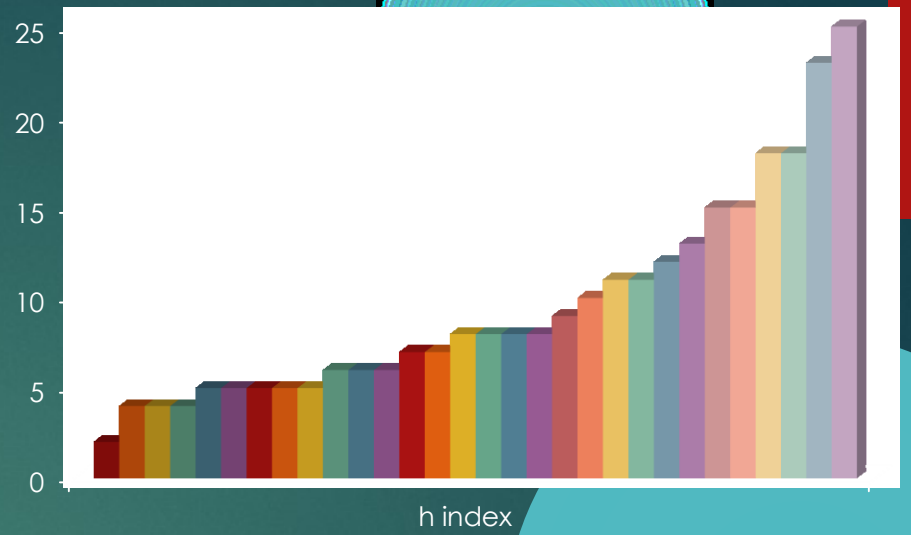
dataset
1
1
1
1
1
1
1
1
45

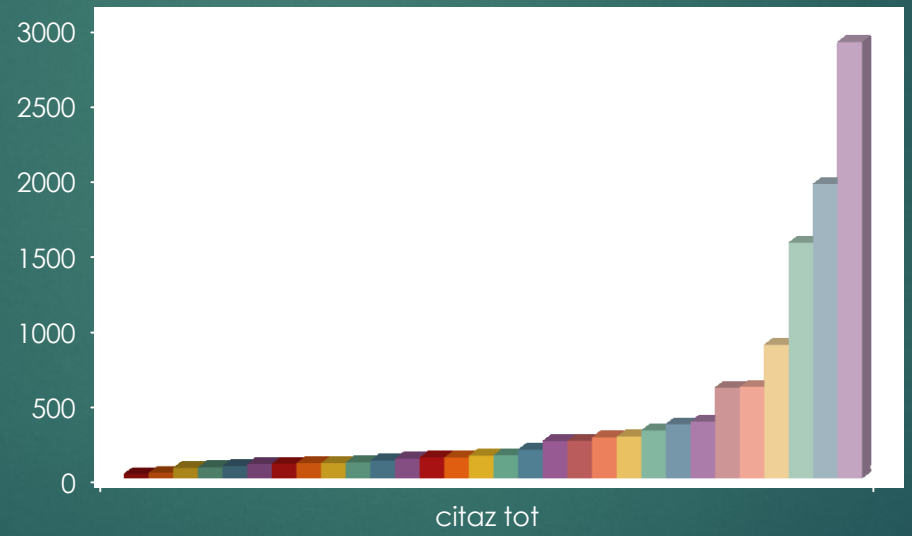
4,214286	media
11,74196	ds
1	mediana
1	10°percentile
1,7	90°percentile



Autore	n pubb	h index	citaz tot
1	36	5	117
2	36	6	96
3	49	12	317
4	25	5	130
5	27	10	247
6	51	18	1565
7	60	18	888
8	143	23	1955
9	42	11	271
10	30	8	153
11	31	7	149
12	46	8	189
13	308	25	2895
14	39	5	100
15	6	2	67
16	24	4	28
17	14	6	72
18	13	4	37
19	40	8	249
20	15	4	277
21	32	11	377
22	43	15	609
23	38	8	137
24	28	5	81
25	44	13	603
26	31	9	105
27	26	6	94
28	40	15	359
29	38	7	137
30	35	5	100







	n pubb	h index	citaz tot
mediana	36,0	8,0	151,0
media	46,3	9,4	413,5

MISURA DELLA VARIABILITA'

VARIANZA $\dfrac{\Sigma (x - \bar{x})^2}{(n - 1)} = \sigma^2$

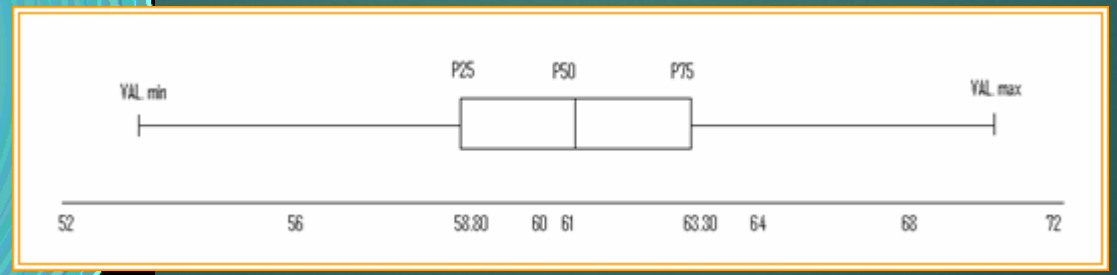
DEVIAZIONE STANDARD $\sqrt{\dfrac{\Sigma (x - \bar{x})^2}{(n - 1)}} = \sigma$

ERRORE STANDARD $\sigma / \sqrt{n}$

PERCENTILE

Percentiles

box plot

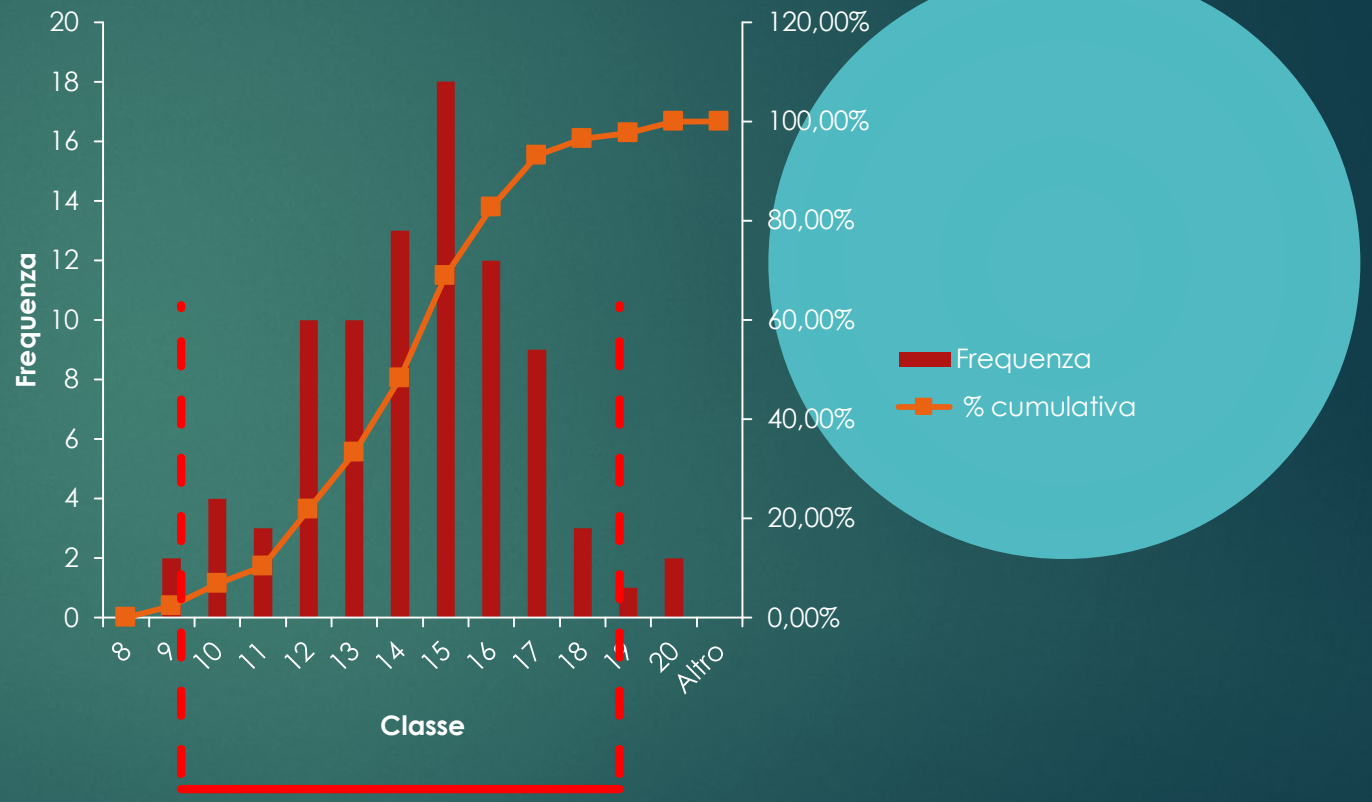


REFERENCE INTERVALS

Reference ranges represent the range of values that includes 95% of the observed results in a healthy subjects control group.

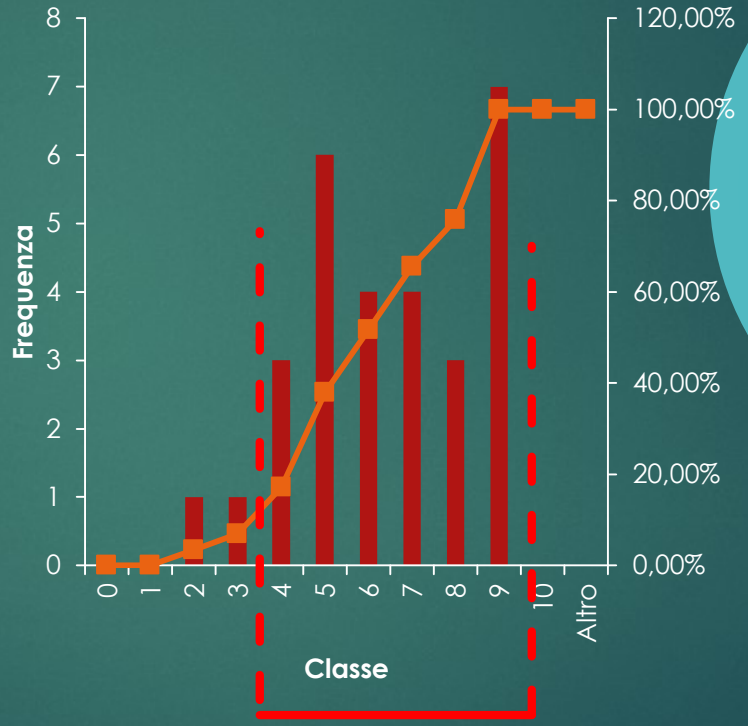
Hgb

media	14,01724
ds	2,298443
min	9,42036
max	18,61



REFERNECE INTERVAL

curtosi	-0,88313
asimmetria	-0,23046
2,5° percentile	3,4
97,5° percentile	9



REFERENCE INTERVAL

A healthy subject undergoing a laboratory analysis has 95% chance that their value will be within the reference ranges;

A healthy subject udergoing 20 laboratory analyzes has a probability of $(0.95)^{20} = 0.358486$ that all results fall within the reference intervals. That is, up to 100 healthy subjects performing each of 20 analyzes, only around 36% will have all values within the reference intervals.