Statistica medica - parte 3

Fisica e Statistica medica

cds in Medicina – cds in Odontoiatria

A.A. 2022/23

I° anno – I° semestre

2 crediti / 24 ore

Prof. Lucio Torelli

Dipartimento Clinico di Scienze mediche, chirurgiche e della salute

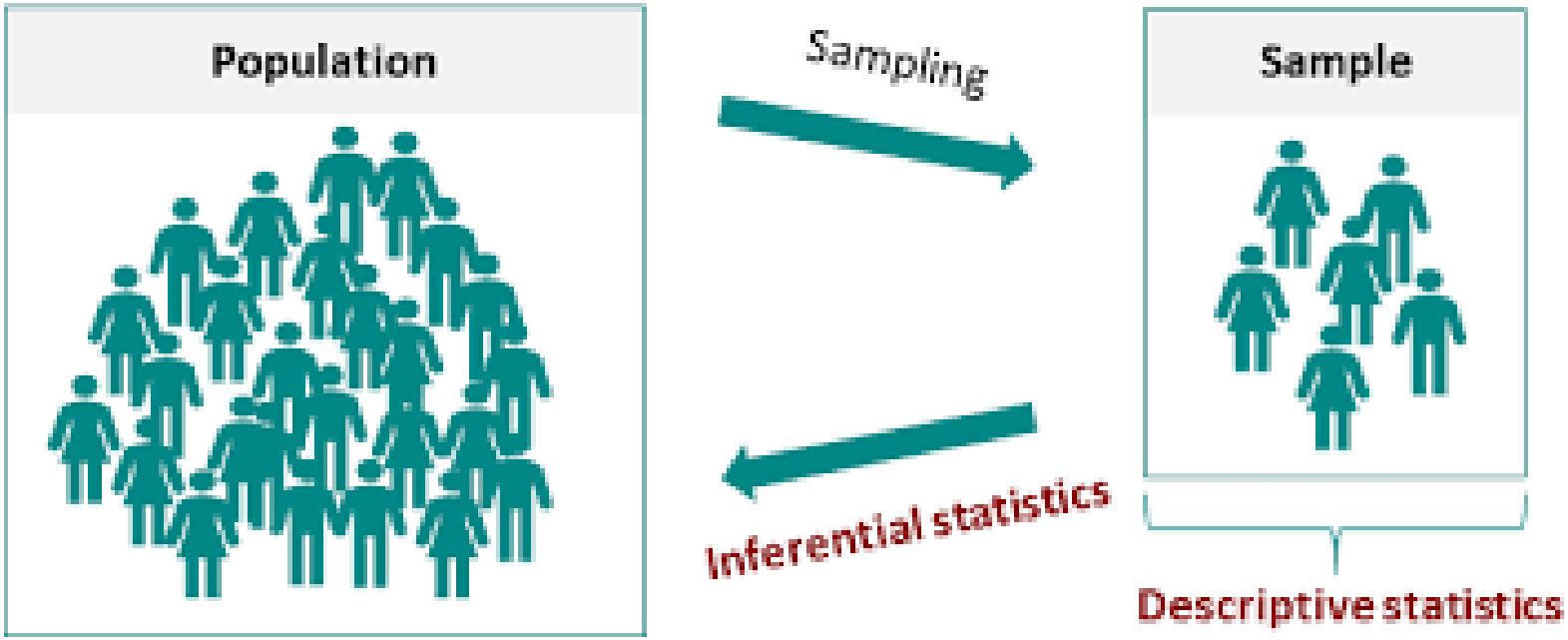
Università degli Studi di Trieste

torelli@units.it

Attenzione: queste slide sono solo alcune note per le lezioni, non sono pertanto un riassunto del seminario

28/11/2022

Inferenza statistica



il passaggio dal campione alla popolazione…

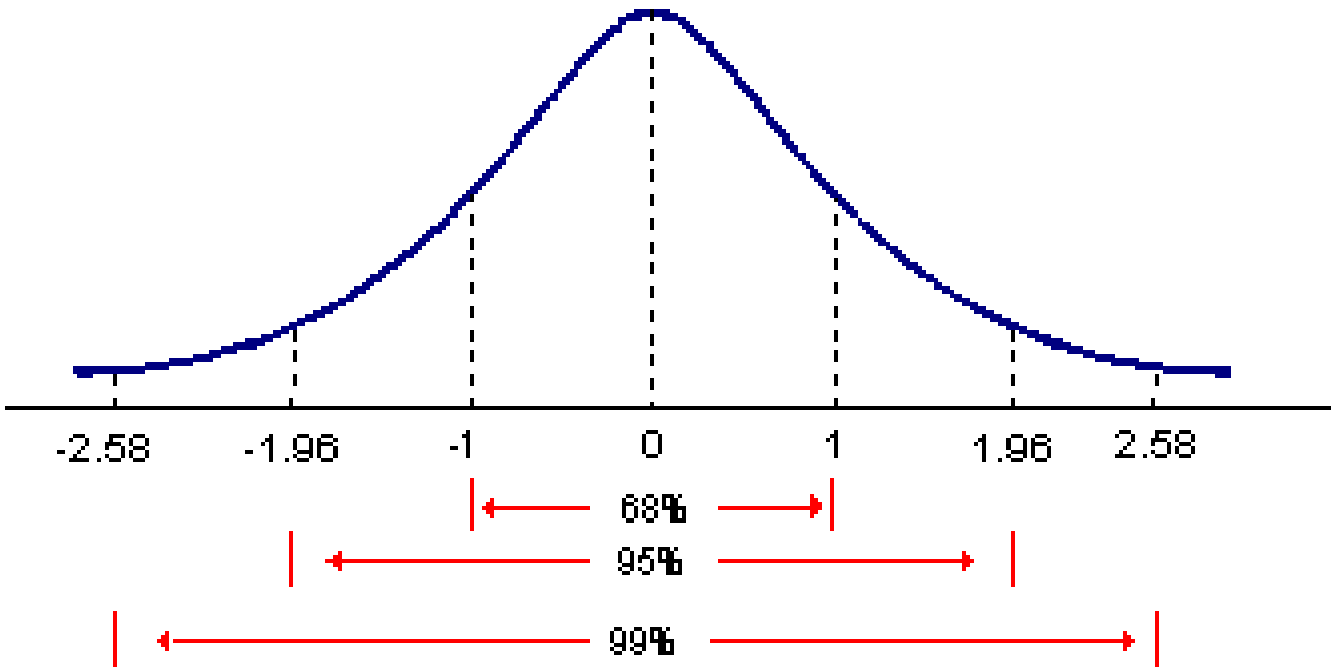
- Obiettivi dello studio
- Disegno dello studio
- Materiali e metodi,
  criteri di inclusione/esclusione (**popolazione** oggetto dello studio)
- Scelta del **campione**
- Valutazione e descrizione del campione
- **Inferenza** sulla popolazione
- Discussione dei risultati

stime

- Stime puntuali
- Stime intervallari
- *Confidence intervals*

Cosa significa se leggiamo:

c.i. per μ al 95%: (26,4; 30,4) ?


-2.58
-1.96
-1
0
1
1.96
2.58
68%
95%
99%

Odds ratio – Confidence Interval

Use this calculator to determine a confidence interval for your odds ratio. An odds ratio is a measure of association between the presence or absence of two properties. For example, it could provide a measure of association between customers who are either older or younger than 25 and either have or have not claimed on their car insurance, in order to determine whether age is associated with the propensity to claim. The value of the odds ratio tells you how much more likely someone under 25 might be to make a claim, for example, and the associated confidence interval indicates the degree of uncertainty associated with that ratio.

Calculator

Contingency table

		Property B: Presence	Property B: Absence
Property A	Presence	15	185
	Absence	14	286

What confidence level do you need? 95 %
Typical choices are 90%, 95%, or 99%

Your odds ratio is 1.66

Your confidence interval is (0.78 , 3.51)

Calculator

Contingency table

		Property B: Presence	Property B: Absence
Property A	Presence	150	1850
	Absence	140	2860

What confidence level do you need? 95 %
Typical choices are 90%, 95%, or 99%

Your odds ratio is 1.66

Your confidence interval is (1.31 , 2.1)

Stime (vedi es precedente)

Formula

This calculator uses the following formulae to calculate the odds ratio (or) and its confidence interval (ci). or = a*d / b*c, where:

- a is the number of times both A and B are present,
- b is the number of times A is present, but B is absent,
- c is the number of times A is absent, but B is present, and
- d is the number of times both A and B are negative.

To calculate the confidence interval, we use the log odds ratio, log(or) = log(a*d/b*c), and calculate its standard error:

$se(\log(or)) = \sqrt{1/a + 1/b + 1/c + 1/d}$

The confidence interval, ci, is calculated as:

$ci = \exp(\log(or) \pm Z\alpha/2 * \sqrt{1/a + 1/b + 1/c + 1/d})$,

where $Z\alpha/2$ is the critical value of the Normal distribution at $\alpha/2$ (e.g. for a confidence level of 95%, α is 0.05 and the critical value is 1.96).

Note: The logarithms included in the formulae above are natural logarithms, i.e., log base e, sometimes denoted ln().

Stime (per una proporzione)

Confidence interval for a proportion

Estimate the proportion with a dichotomous result or finding in a single sample. This calculator gives both binomial and normal approximation to the proportion.

Instructions: Enter parameters in the green cells. Answers will appear in the blue box below.

N =	400		Sample size
x =	20		Number in the sample with the result or finding in question
CL =	95	%	Confidence level

Calculate

Cosa succede
con N=40 e x=2??

e se CL=99% o 80%

$$\hat{p} \pm z\sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}}$$

1. Binomial "exact" calculation

Proportion of positive results = P = x/N = 0.0500
Lower bound = 0.0308
Upper bound = 0.0762

stime

- Trovare il c.i. al *95%*per la stima della media di una popolazione gaussiana $N(\mu,\sigma)$, avendo a disposizione i dati di un campione, opportunamente scelto, di numerosità n, media campionaria m e dev.st. campionaria s.

Risulta: $\left(m - \frac{q \cdot s}{\sqrt{n}} ; m + \frac{q \cdot s}{\sqrt{n}} \right)$ q è il quantile …

Cosa succede al variare di n, e cosa al variare di s?

stime

• **<u>Calcolo a priori della numerosità</u>** *n* del campione per la stima della media di una $N(\mu,\sigma)$, data la lunghezza massima *k* del c.i.

Poiché la lunghezza del c.i. è $\frac{2qs}{\sqrt{n}}$, deve essere: $\frac{2qs}{\sqrt{n}} \leq k,$

da cui risulta: $n \geq \left(\frac{2qs}{k}\right)^2$

Ad es. se vogliamo un c.i. al 95% ($q \approx 2$) di lunghezza non superiore a *1* e se pensiamo che $s \approx 5$, risulta $n \geq 400$.
Ma se ad es. mi basta $k \leq 2$, allora $n \geq 100$.

Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study – *March 11, 2020 –The Lancet*

191 patients (135 from Jinyintan Hospital and 56 from Wuhan Pulmonary Hospital) were included in this study, of whom 137 were discharged and **54 died** in hospital. 91 (48%) patients had a comorbidity, with hypertension being the most common (58 [30%] patients), followed by diabetes (36 [19%] patients) and coronary heart disease (15 [8%] patients). Multivariable regression showed increasing odds of in-hospital death associated with older age (**odds ratio 1·10, 95% CI 1·03–1·17**, per year increase), higher Sequential Organ Failure Assessment (SOFA) score (**5·65, 2·61–12·23**; p<0·0001), and d-dimer greater than 1 μg/mL (**18·42, 2·64–128·55**; p=0·0033) on admission. …
Current smoker (vs non-smoker): **2·23, 0·65–7·63** …
Female sex (vs male): **0·61, 0·31–1·20** …

cos'è un **test di ipotesi**

- è una "procedura di calcolo"
- permette di **rifiutare** un'**ipotesi H_0**
- input
  - dati di una campione
- output
  - un consuntivo
- decisione
  - in base al consuntivo **si rifiuta** oppure **non si rifiuta** l'ipotesi
  - **Semplificando le cose**:

"ipotesi nulla"

$$p < \alpha, \text{ respingo } H_0,$$
$$p > \alpha, \text{ non respingo } H_0$$

ipotesi nulla, ***H0***:

il nuovo farmaco non è efficace

	vero	falso
accetto	decisione corretta	errore (2° tipo)
rifiuto	errore (1° tipo)	decisione corretta

cos'è un test statistico

(***H0***) il nuovo farmaco **non** è efficace

	vero	falso
accetto	1 - α	P(accetto | falsa) β, beta
rifiuto	P(rifiuto | vera) α, alfa	1 - β

α è detta <u>significatività</u> del test; **1-β** <u>potenza</u> del test

Esistono diversi tipi di test …
noi intanto ne vedremo 4 *(altri poi ne vedremo al 3° anno)*:

Test del χ^2 di indipendenza

test esatto di Fisher di indipendenza,

test t di Student per il confronto tra due gruppi,

test t di Student per dati appaiati

Problema … il casco serve??

tavola osservata

	casco sì	casco no	
tr. sì	17	218	235
tr. no	130	428	558
	147	646	793

Da *Biostatica*, Pagano-Gauvreau

Problema … il casco serve??

tavola attesa

	casco sì	casco no	
tr. sì	44	191	235
tr. no	104	455	558
	147	646	793

tavola osservata

	casco sì	casco no	
tr. sì	17	218	235
tr. no	130	428	558
	147	646	793

Da *Biostatica*, Pagano-Gauvreau

Problema … il casco serve??

tavola attesa

	casco sì	casco no	
tr. sì	44	191	235
tr. no	104	455	558
	147	646	793

tavola osservata

	casco sì	casco no	
tr. sì	17	218	235
tr. no	130	428	558
	147	646	793

H0= indipendenza

P < 0,001 …

cosa significa??

Da *Biostatica*, Pagano-Gauvreau

Problema … il casco serve??

tavola attesa

	casco sì	casco no	
tr. sì	44	191	235
tr. no	104	455	558
	147	646	793

tavola osservata

	casco sì	casco no	
tr. sì	17	218	235
tr. no	130	428	558
	147	646	793

P < 0,001 - ***OR = 0,26 c.i. (0,15; 0,44)***

Problema … il trattamento A ha dato effetti significativamente diversi rispetto al trattamento B?

A	B
23	22
21	18
24	17
25	22
25	24
23	25
…	…
…	…
…	…
27	30
26	22
21	
24	
24	

Se dati gaussiani $N(\mu_1,\sigma_1)$, $N(\mu_2,\sigma_2)$, con $\sigma_1 \approx \sigma_2$

Test t di Student, **$H_0 = \{\mu_A \approx \mu_B\}$**

p = 33,4%

Cosa significa??

Pensando ad es. $\alpha = 5\%$ …

$p > \alpha$ non respingo H_0 … cioè …

Problema … la dieta ha effetto?

Prima	dopo
96	90
92	90
84	87
85	82
97	89
88	85
…	…
…	…
…	…
87	84
96	92

Se dati gaussiani $N(\mu_p,\sigma_p)$, $N(\mu_d,\sigma_d)$, con $\sigma_p \approx \sigma_d$

Test t di Student *'per dati accoppiati'*,
$H_0 = \{\mu_p \approx \mu_d\}$

$$p = 0{,}26\%$$

Cosa significa??

Pensando ad es. $\alpha = 5\%$ …

$p < \alpha$ respingo H_0 … cioè …

Esempi:

Leggiamo in un articolo:

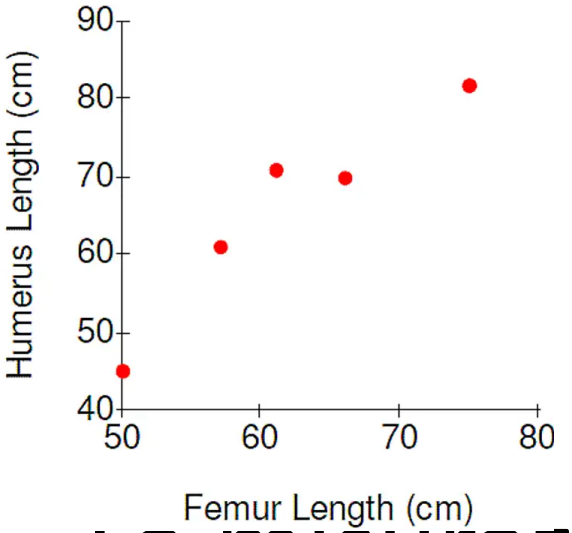
… χ^2, p=48,2%. Cosa significa?

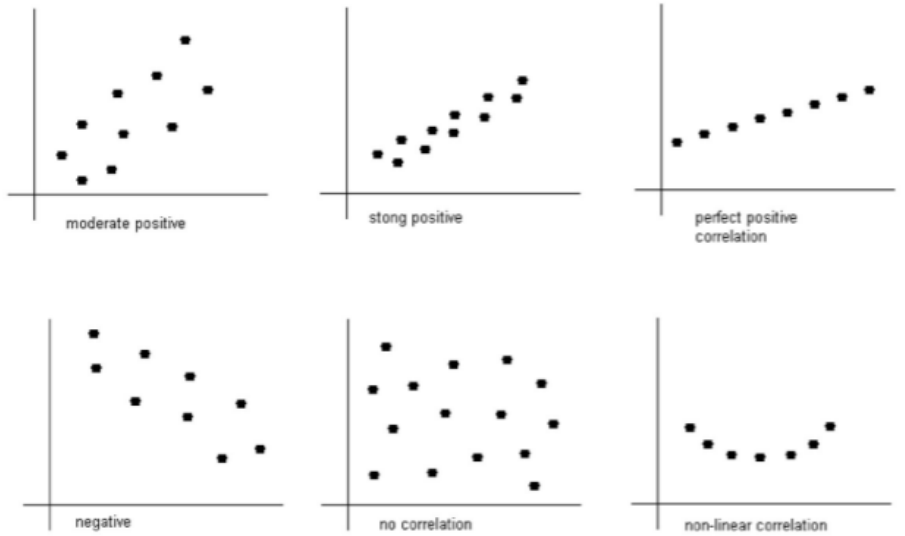
… t test, p=0,8%. Cosa significa?

… t test per dati accoppiati, p=36,2%. Cosa significa?

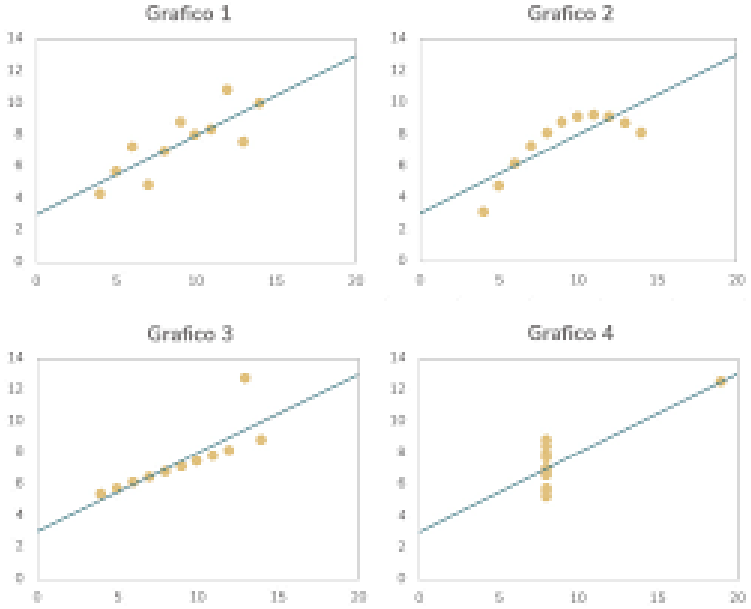
… t test, p=4,8%. Cosa significa?

… χ^2, p=3,2%. Cosa significa?





La misurazione della correlazione (primi cenni)



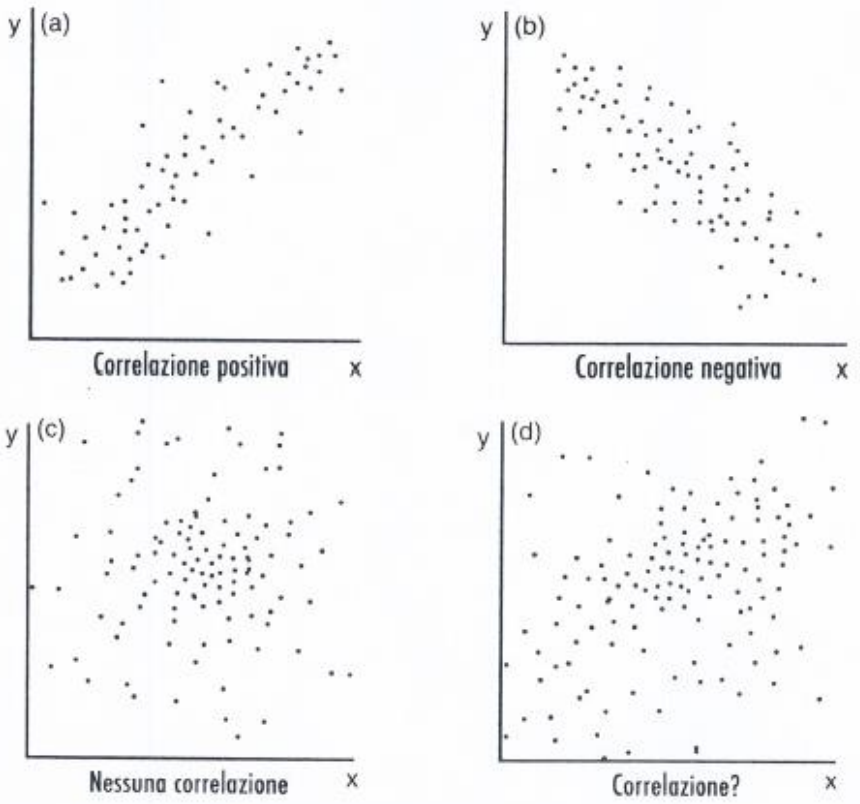


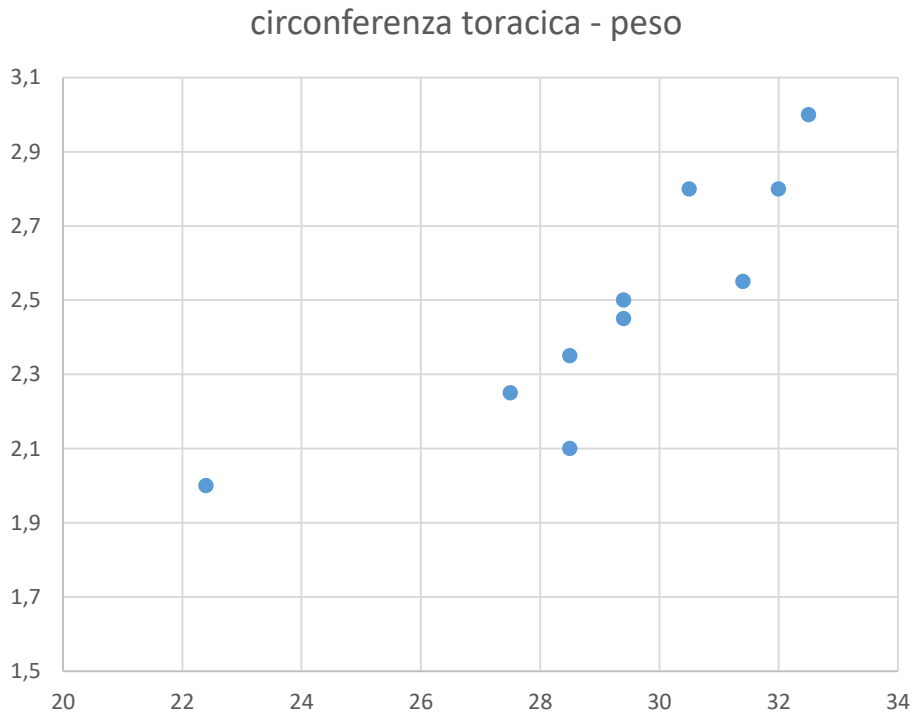
Figura 13.1 Diagramma di dispersione di dati bivariati

da Statistica per le professioni sanitarie – *J.Fowler, P.Jarvis, M.Chevannes* - EdiSES

«… La valutazione soggettiva di un diagramma di dispersione può essere sostituita da una tecnica statistica in grado di fornire indicazioni più precise e obiettive.
La statistica che fornisce un indice di accordo su come le due variabili siano in relazione tra di loro è il **coefficiente di correlazione**.
La statistica è calcolata a partire dai dati campionari, e fornisce una stima del corrispondente parametro nella popolazione.
Il valore numerico del coefficiente di correlazione, ***r***, è compreso tra -1 (perfetta correlazione negativa)
e +1 (perfetta correlazione positiva) …»

«… I coefficienti di correlazione possono essere calcolati attraverso metodi parametrici o non parametrici
Un coefficiente parametrico è il **coefficiente di correlazione di Pearson**.
Dei diversi coefficienti non parametrici, il **coefficiente di correlazione per ranghi di Spearman** è il più comunemente utilizzato …»

circ toracica	peso
22,4	2
27,5	2,25
28,5	2,1
28,5	2,35
29,4	2,45
29,4	2,5
30,5	2,8
32	2,8
31,4	2,55
32,5	3



Excel: PEARSON()

r=	0,86

«… Il quadrato del coefficiente di Pearson, r^2, viene detto **coefficiente di determinazione**.
E' una misura della percentuale della variabilità di una variabile spiegata dalla variabilità dell'altra
Nel nostro esempio, $r^2 = 0,80^2 = 0,74$:
il valore *100-74 =* ***26%*** della variazione del peso non è spiegata dalla variazione della circonferenza toracica.
Il peso del neonato può dipendere anche da fattori diversi dalla misura della circonferenza toracica …»

… è molto utile calcolare anche il *confidence interval* del coefficiente *r* …
… il **valore p** che compare spesso dopo un'analisi di questo tipo dice se i dati respingono o meno l'ipotesi nulla *r=0* …

«…**Attenzione** che due variabili tra loro correlate non implicano necessariamente che una sia la causa …»

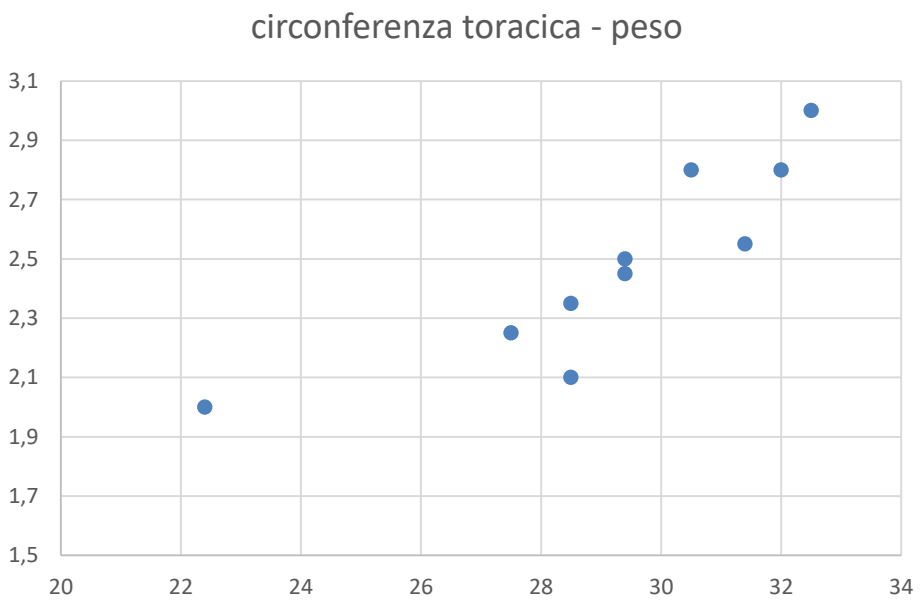
circ toracica	peso
22,4	2
27,5	2,25
28,5	2,1
28,5	2,35
29,4	2,45
29,4	2,5
30,5	2,8
32	2,8
31,4	2,55
32,5	3

socscistatistics

r=	0,8604	r^2=	0,740288	p=	0,0014

ci *r*=**vassarstats**

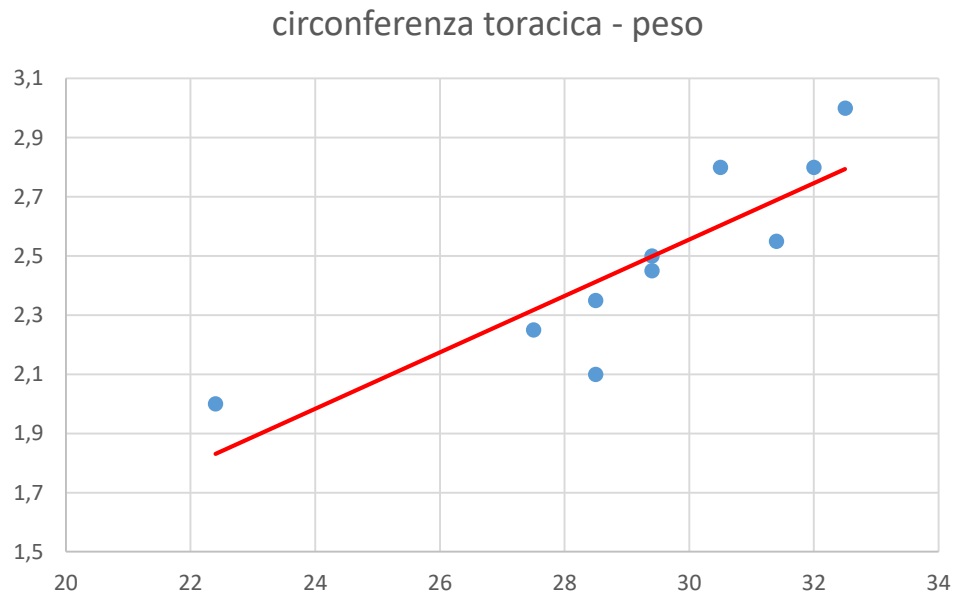
0,504	0,966



Excel:PEARSON()

r=	0,86

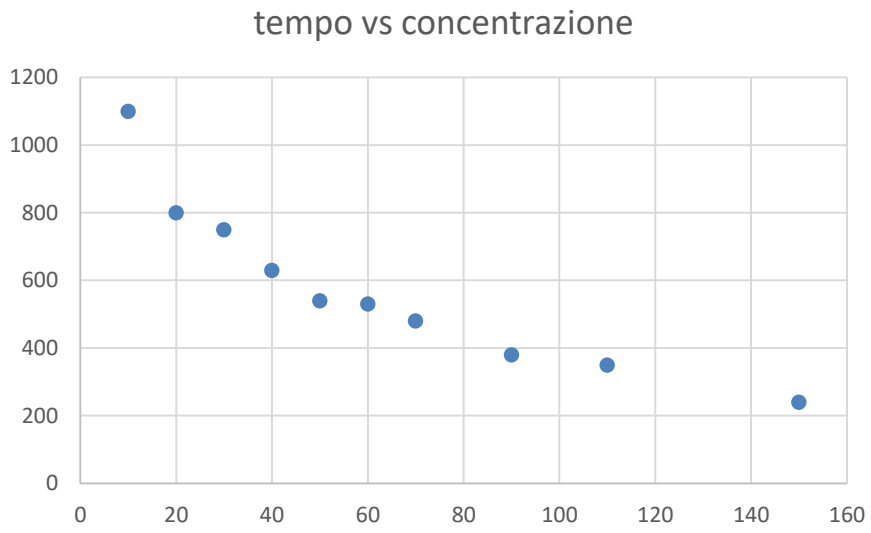
«… Nel presentare un diagramma a dispersione, è talvolta utile tracciare una linea retta all'interno dell'agglomerato di punti per illustrarne la relazione media.
Tale retta, viene chiamata **retta di regressione** …»



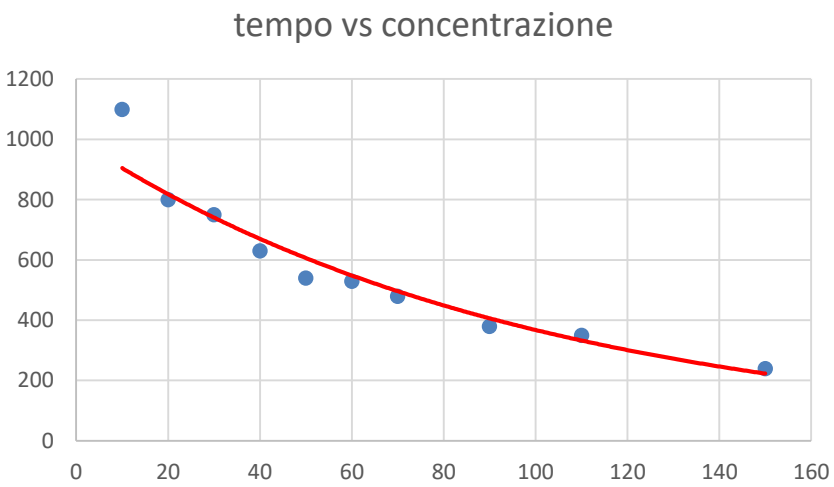
«… Molte relazioni in medicina/riabilitazione non sono però lineari…»

Come fare?

tempo	concetrazione del farmaco nel plasma
10	1100
20	800
30	750
40	630
50	540
60	530
70	480
90	380
110	350
150	240



tempo	concetrazione del farmaco nel plasma
10	1100
20	800
30	750
40	630
50	540
60	530
70	480
90	380
110	350
150	240



tempo	concetrazione del farmaco nel plasma	**Log della concentrazione**
10	1100	**3,04**
20	800	**2,90**
30	750	**2,88**
40	630	**2,80**
50	540	**2,73**
60	530	**2,72**
70	480	**2,68**
90	380	**2,58**
110	350	**2,54**
150	240	**2,38**