Test non parametrici

Example (experimental study)

Female remating propensity contingent on sexual cannibalism in sagebrush crickets, *Cyphoderris strepitans*: a mechanism of cryptic female choice

J. Chadwick Johnson, Tracie M. Ivy, Scott K. Sakaluk

Behavioral Ecology, Volume 10, Issue 3, 1 May 1999, Pages 227–233

Binary Exposure :

- *Low nutrients diet (fasting)*

  vs

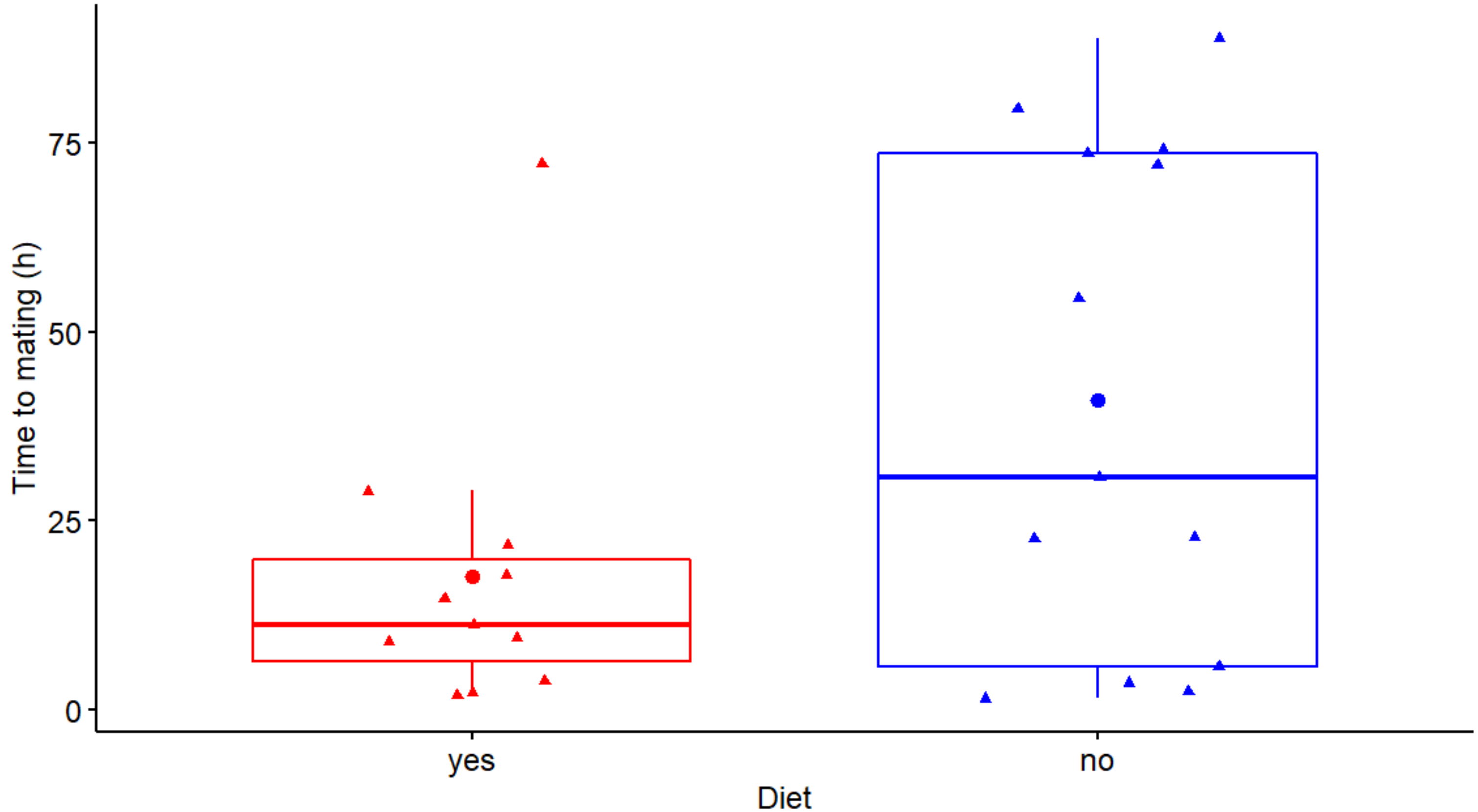
- *High nutrients diet (regular)*

Continuous Outcome :

Time to coupling (hours)

Parametric test – T test

n_E=11 mean (SD) = 17.56 (20.01) n_{NE}=13 mean (SD) = 40.93 (33.61)



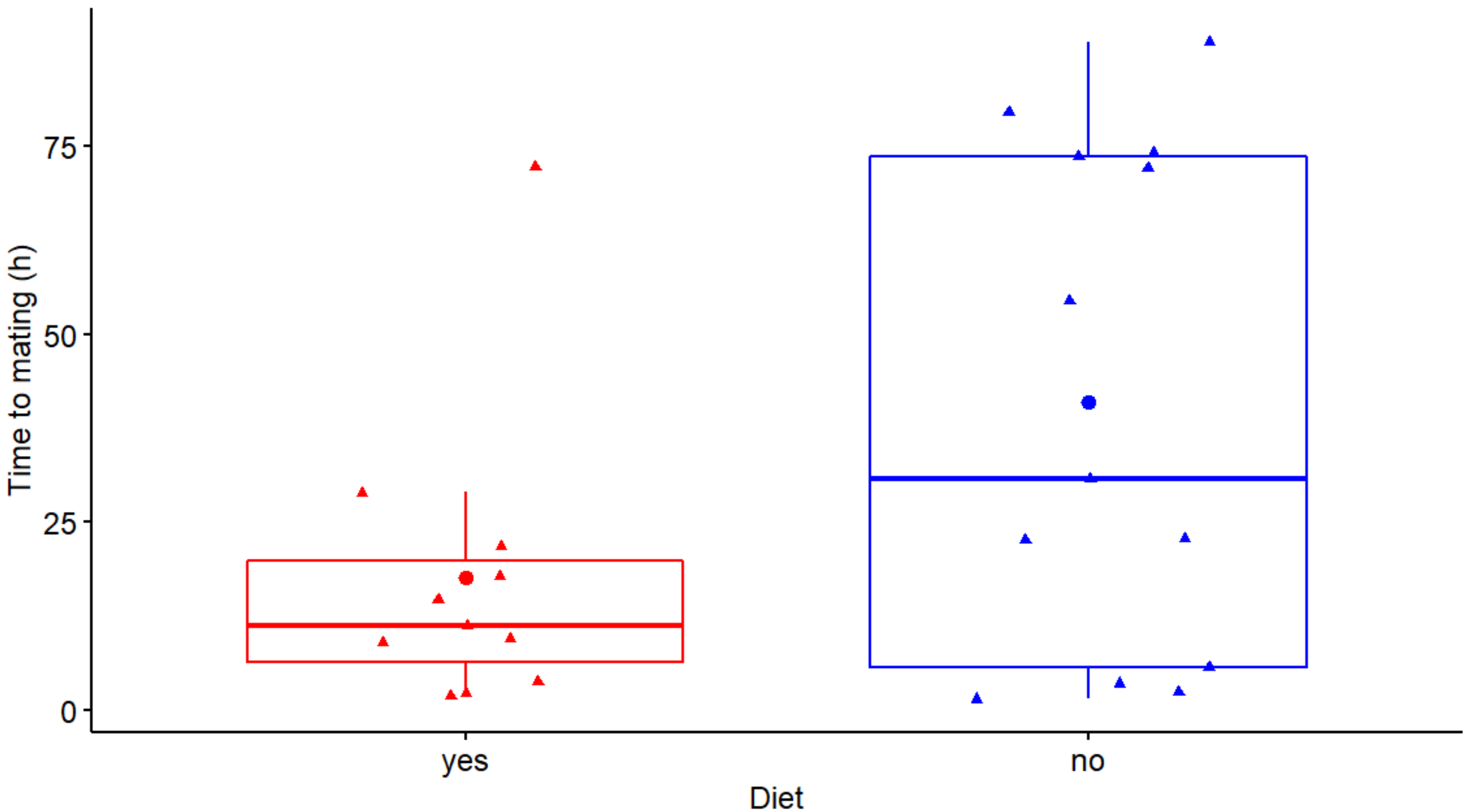
$$t = \frac{(\bar{x}_E - \bar{x}_{NE})}{\sqrt{\left(\frac{s_E^{\ 2}}{n_E} + \frac{s_{NE}^{\ 2}}{n_{NE}}\right)}} = \frac{(17.56 - 40.93)}{\sqrt{\left(\frac{20.01^2}{11} + \frac{33.61^2}{13}\right)}} = -2.10$$

p-value (bilateral test) = 0.048

Non parametric test – Mann Whitney

However, it is not likely that the response in the two populations has a **normal distribution** and the **sample size is small**. Thus, comparing the means using the pararametric T test may not be a valid option. We need a test not requiring any assumption on the distribution of the response variable.

n_E=11 median(IQR)=11.2(6.4-19.8) n_{NE}=13 median(IQR)=30.8(5.7-73.6)



Non parametric test – Mann Whitney

Data are transfomed in ranks, i.e. their position in the ranked sequence of $n_E + n_{NE}$ total data

If two (or more) observations have identical value (ties), than a mean rank is assigned to all these observations.

In our example:

Fasting group		Regular diet group	
Time to coupling	rank	Time to coupling	rank
1.9	2	1.5	1
2.1	3	2.4	4
3.8	6	3.6	5
9	8	5.7	7
9.6	9	22.6	14
13	10	22.8	15
14.7	11	30.8	17
17.9	12	54.4	18
21.7	13	72.1	19
29	16	73.6	21
72.3	20	74.2	22
		79.5	23
		88.9	24

Non parametric test – Mann Whitney

In our example:

$W_E = 110$

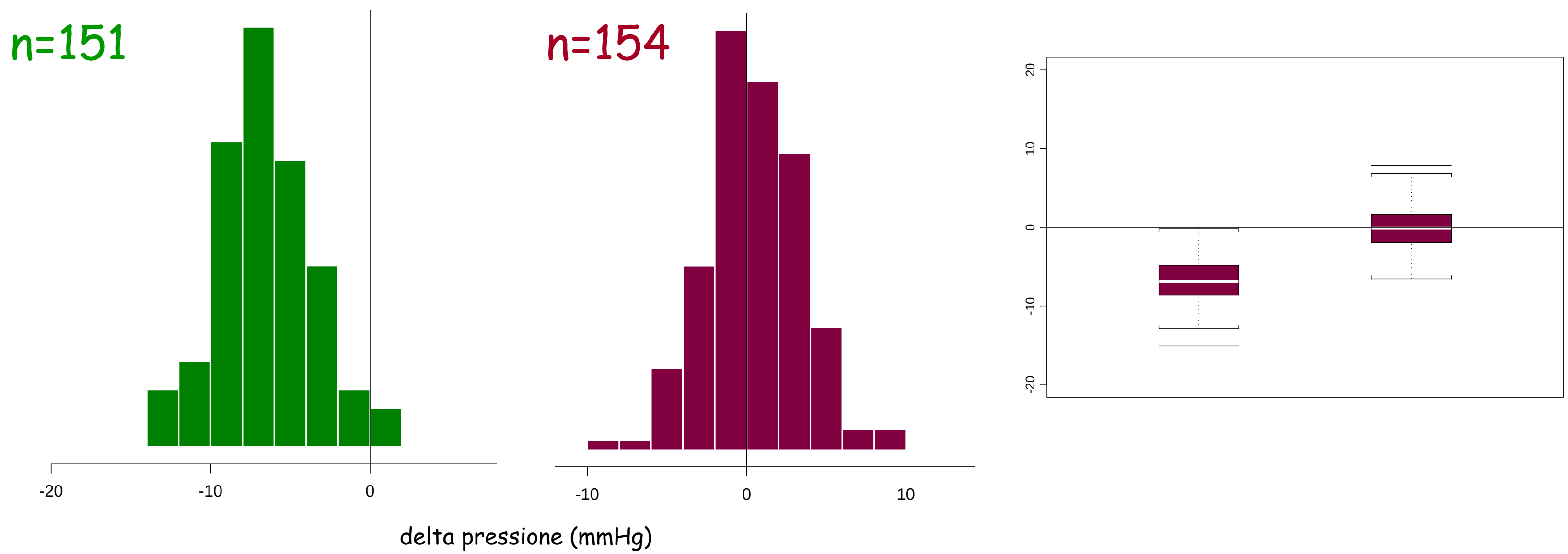
$W_{NE} = 190$

Fasting group		Regular diet group	
Time to coupling	rank	Time to coupling	rank
1.9	2	1.5	1
2.1	3	1.7	4
3.8	6	2.4	5
9	8	3.6	7
9.6	9	5.7	14
13	10	22.6	15
14.7	11	22.8	17
17.9	12	39	18
21.7	13	54.4	19
29	16	72.1	21
72.3	20	73.6	22
		79.5	23
		88.9	24

The p-value (bilateral test) is 0.1191.

Test Parametrici: esempio

Risultati di uno studio che confrontava due farmaci anti-ipertensivi (A e B) nel ridurre la pressione arteriosa (mmHg)



t_{303} = -19.0108, p-value = 0.00001 (two-sided)

95 % CI: [-7.49; -6.08]

Test Parametrici

- Test sulla media μ
  - 1 campione
    - varianza nota: $$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$
    - varianza non nota: $$t_{n-1} = \frac{\bar{X} - \mu_0}{S / \sqrt{n}}$$
  - 2 campioni indipendenti
    - varianza nota: $$Z = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
    - varianza non nota: $$t_{n_1+n_2-2} = \frac{\left(\bar{X}_1 - \bar{X}_2\right)}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$
  - 2 campioni appaiati
    - varianza nota: $$Z = \frac{\bar{d}}{\sqrt{\frac{\sigma_d^2}{n}}}$$
    - varianza non nota: $$t_{n-1} = \frac{\bar{d}}{\sqrt{\frac{s_d^2}{n}}}$$

Ognuno di questi test si basa sulla distanza tra media campionaria e valore atteso sotto H_0 standardizzata per la variabilità campionaria.

Test Parametrici: Assunti

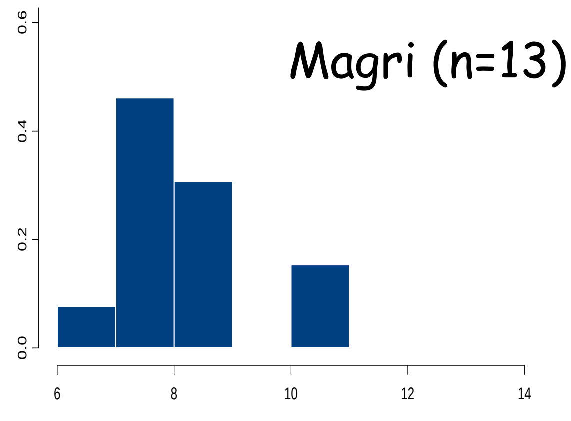
I test parametrici quali il test t di Student e l'analisi della varianza sono basati su alcune assunzioni:

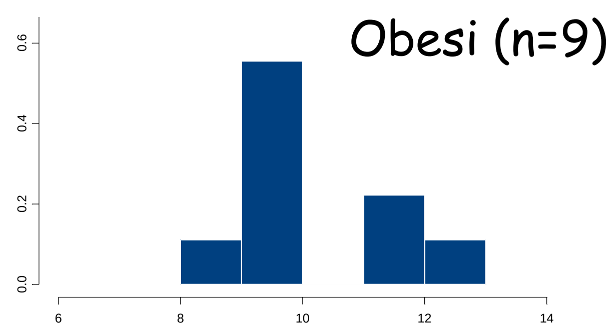
1. Variabile continua misurata senza errore (o con errore indipendente dal valore misurato)
2. Campione estratto casualmente dalla popolazione di riferimento
3. Variabile distribuita in modo (approssimativamente) normale
4. Omogeneità delle varianze (omoschedasticità) tra i gruppi.

Test Parametrici

Dispendio energetico (MJ/giorno) in soggetti magri ed obesi

Magri (n=13)	Obesi (n=9)
6.13	8.79
7.05	9.19
7.48	9.21
7.48	9.68
7.53	9.69
7.58	9.97
7.90	11.51
8.08	11.85
8.09	12.79
8.11	
8.40	
10.15	
10.88	





VIOLAZIONE DELL'ASSUNTO DI NORMALITÀ

Che possibilità abbiamo?

1. Ignorare la violazione degli assunti di normalità
2. Trasformare i dati
3. Utilizzare metodi non parametrici

1) *Ignorare la violazione degli assunti di normalità*

Per n sufficientemente grandi, la distribuzione campionaria delle medie rispetta le assunzioni per l'applicazione del test t grazie al ***teorema del Limite Centrale*** (data una variabile X, con media μ e varianza σ^2, e un campione di numerosità n, la distribuzione della media campionaria ($\overline{X}$) approssima, al crescere di n, una distribuzione gaussiana con media μ e varianza σ^2/n).

Questo teorema è molto generale perché i suoi risultati sono validi qualunque sia la natura della vc X.

Quale n?

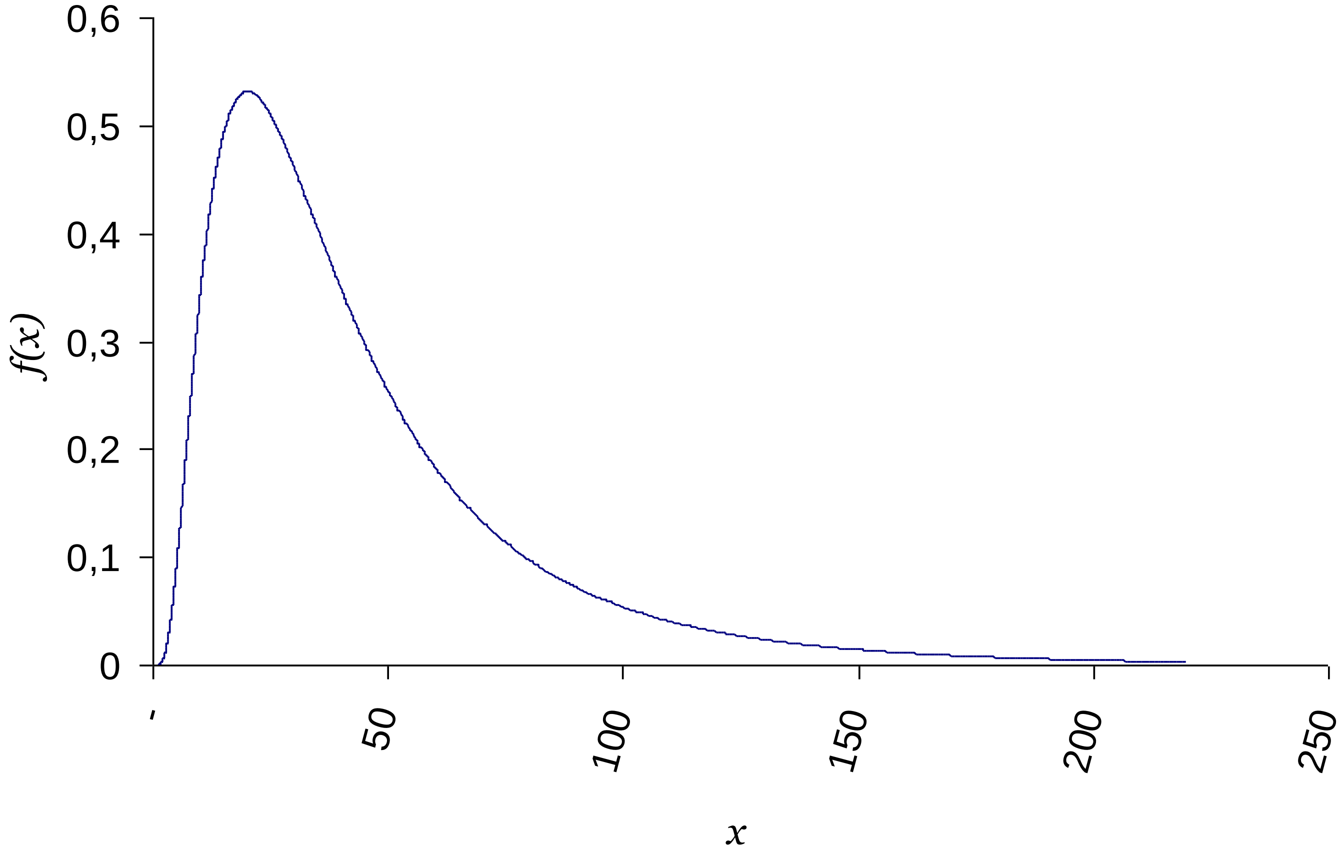
Dipende dalla forma delle distribuzioni:

- Se le distribuzioni hanno la medesima asimmetria può bastare $n \geq 30$
- Se l'asimmetria è diversa $n \geq 500$

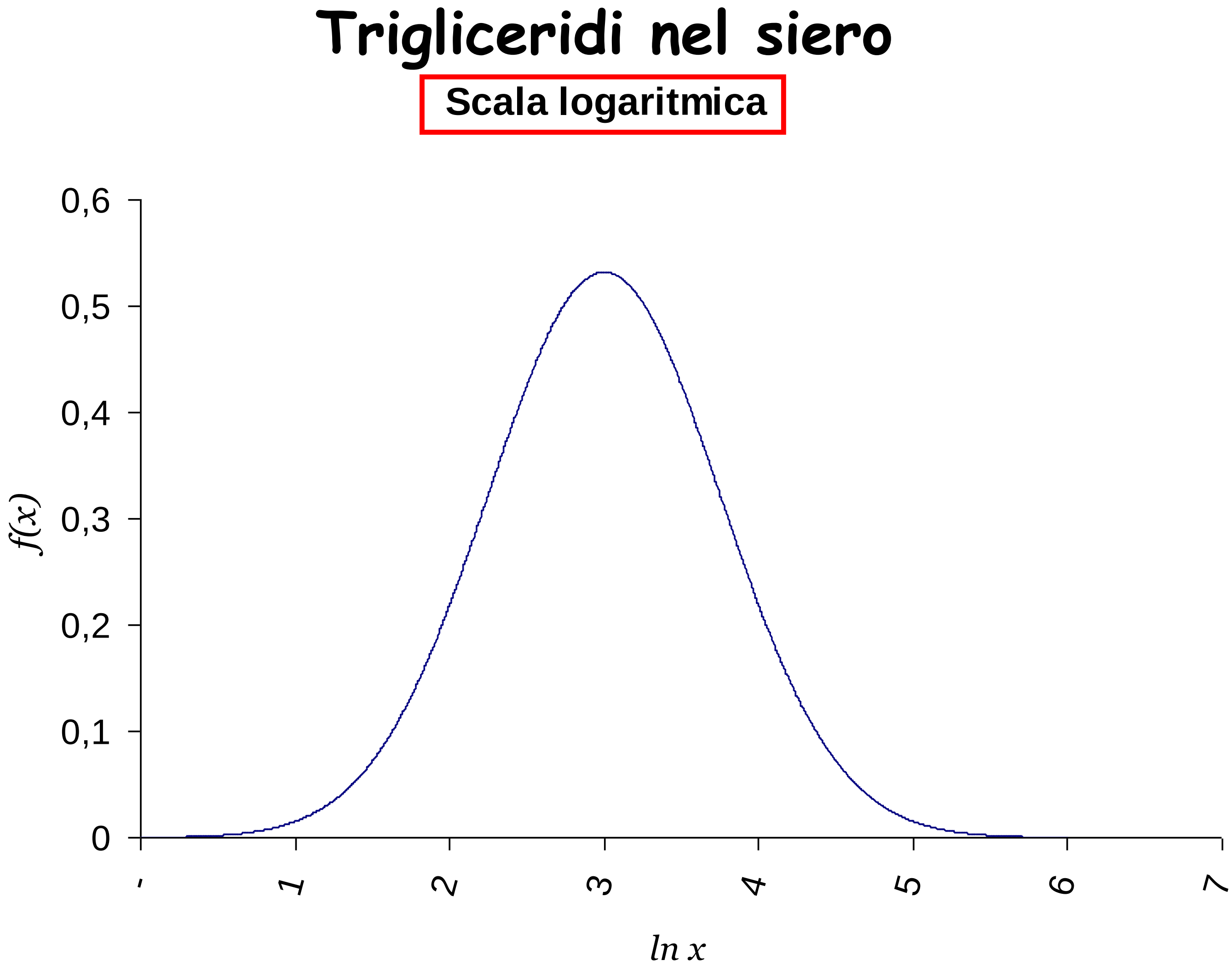
2. Trasformare i dati

Trigliceridi nel siero

Scala naturale



2. Trasformare i dati



Quali trasformate?

- **Logaritmica Y'=ln(Y)** utile quando c'è asimmetria a dx, tutti i valori devono essere >0
- **Arcoseno P'=arcsin[√p]** si applica a dati espressi come proporzioni
- **Radice quadrata Y'=√(Y+1/2)** si usa con i conteggi

3. Utilizzare metodi non parametrici

Parametrico vs Non Parametrico

Metodi PARAMETRICI	Metodi NON PARAMETRICI
Si fanno assunzioni sul meccanismo casuale di generazione dei dati	NON si fanno assunzioni sul meccanismo casuale di generazione dei dati
Esempio: $X \sim N(\mu,\sigma^2)$	$X \sim ?$
	↓
	RANGHI

Quando usare i test non parametrici?

Sono giustificati quando:

1. le variabili hanno evidenti scostamenti dalla normalità (o sono fortemente asimmetriche o presentano più di un picco);
2. quando il campione è troppo piccolo per comprendere se esiste una distribuzione normale dei dati
3. quando le osservazioni sono rappresentate da classifiche ordinali (es. gravità di una malattia da 1 a 4)

TEST DEI SEGNI

Si può usare in alternativa al test t per un campione o per dati appaiati.

Valuta se la mediana di una popolazione sia uguale a un valore specificato dall'ipotesi nulla

Poco potente

TEST DEI SEGNI

Se Ho è vera, ci attendiamo che metà delle misure sia situata al di sopra e metà al di sotto della mediana specificata.

Le misure al di sopra del valore specificato nell'ipotesi nulla vengono classificate con «+», quelle al di sotto con «-»

È un test binomiale in cui il numero di osservazioni sopra o sotto la mediana viene confrontato con quello atteso

$$Ho: P(+)=P(-)=0.5$$

TEST DEI SEGNI

- Il numero di segni+ (o meno) è capitato per «caso» , quando l'ipotesi nulla è vera, o tale numero è da attribuire ad altro? (Ho falsa)

- La statistica test

$$P(k \leq x|n,p) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}$$

- Se P<α/2 allora rifiuto Ho

TEST DEI SEGNI: *esempio*

Un campione casuale di 15 studenti infermieri è stato sottoposto ad un test per valutare il loro livello di autoritarismo con i seguenti risultati.

Studente	Valore
1	75
2	90
3	85
4	110
5	115
6	95
7	132
8	74

Studente	Valore
9	82
10	104
11	88
12	124
13	110
14	76
15	98

Saggiare, con α al 5%, l'ipotesi nulla che il punteggio mediano per la popolazione campionata sia 100.

TEST DEI SEGNI: *esempio*

Un campione casuale di 15 studenti infermieri è stato sottoposto ad un test per valutare il loro livello di autoritarismo con i seguenti risultati.

Studente	Valore
1	75 -
2	90 -
3	85 -
4	110 +
5	115 +
6	95 -
7	132 +
8	74 -

Studente	Valore
9	82 -
10	104 +
11	88 -
12	124 +
13	110 +
14	76 -
15	98 -

9- e 6+

È una distribuzione bernoulliana con p=0.5 se Ho è vera

TEST DEI SEGNI: esempio

Sistema d'ipotesi

$$\begin{cases} H_0: P(+)=P(-) \\ H_1: P(+)\neq P(-) \end{cases}$$

$$P[X \leq 6] = \sum_{i=0}^{6} \binom{15}{i} (0.5)^i (0.5)^{15-i}$$

$P[X \leq 6]=$
$=P[X=0]+ P[X=1]+ P[X=2]+ P[X=3]+ P[X=4]+ P[X=5]+ P[X=6]=$
$=0.3036$
P-value=2*0.3036=0.6072

P-value >> 0.05
Non rifiuto Ho, gli studenti infermieri hanno un livello di autoritarismo mediano che non differisce da 100

TEST DEI SEGNI: dati appaiati

- Si testa l'ipotesi nulla che la mediana della differenza sia 0

- Per ogni coppia di valori
  - si calcola la differenza X-Y (o viceversa)
  - si pone il segno – se X<Y,
  - si pone il segno + se X>Y

Se la mediana è 0 ci aspettiamo che

$$Ho: P(+)=P(-)=0.5$$

Poi si procede come per il test ad un solo campione

TEST DEI SEGNI: dati appaiati

Un gruppo di ricercatori in ambito odontoiatrico voleva sapere se insegnare alla gente come pulire i denti potesse migliorare la loro igiene orale.

Sono state selezionate 12 coppie, ad un solo membro della coppia sono state date istruzioni su come pulire i denti. Sei mesi dopo i 24 soggetti sono stati valutati da un igienista dentale che ha dato loro un punteggio. Un valore basso indica alto livello di igiene

Coppia	Con istruzioni	Senza istruzioni
1	1.5	2.0
2	2.0	2.0
3	3.5	4.0
4	3.0	2.5
5	3.5	4.0
6	2.5	3.0
7	2.0	3.5
8	1.5	3.0
9	1.5	2.5
10	2.0	2.5
11	3.0	2.5
12	2.0	2.5

TEST DEI SEGNI: dati appaiati

Se consideriamo la differenza X-Y

Ho: P(+)=P(-) (mediana=0)

H1: P(+)<P(-) (mediana negativa)

Alpha=0.05

Otteniamo 2+ e 9- (lo 0 viene escluso dalle analisi)

$$P(k \leq 2|11{,}0.5) = \sum_{k=0}^{2} \binom{11}{k} 0.5^k (0.5)^{11-k}$$

P=0.0327<0.05, allora rifiuto Ho

Coppia	(X)	(Y)	(X-Y)
1	1.5	2.0	-
2	2.0	2.0	0
3	3.5	4.0	-
4	3.0	2.5	+
5	3.5	4.0	-
6	2.5	3.0	-
7	2.0	3.5	-
8	1.5	3.0	-
9	1.5	2.5	-
10	2.0	2.5	-
11	3.0	2.5	+
12	2.0	2.5	-

TEST DELLA MEDIANA

Si saggia se <u>due campioni indipendenti </u>provengano da popolazioni con al stessa mediana

Ho: M1=M2

H1: M1≠M2

- Si calcola la mediana «pooled»
- Si costruisce una tabella in cui i soggetti vengono raggruppati in due gruppi (sotto e sopra la mediana)
- Si esegue un test chi quadro con 1 gdl
- Se la statistica test <3.841 non si rifiuta Ho, i due campioni sono stati estratti da popolazioni con al medesima mediana

TEST DELLA MEDIANA: esempio

C'è differenza nel livello di salute mentale tra gli studenti maschi di scuole medie inferiori rurali ed urbane?

Sono stati sottoposti a test 12 bambini di una scuola rurale e 16 di una scuola urbana. (alpha=0.05)

Ho: Mu=Mr

H1: Mu≠Mr

Scuola urbana															
35	26	27	21	27	38	23	25	25	27	45	46	33	26	46	41
Scuola rurale															
29	50	43	22	42	47	42	32	50	37	34	31				

Me=33.5 (33+34)/2

TEST DELLA MEDIANA: *esempio*

Me=33.5 (33+34)/2

	Urbana	Rurale	Totale
Valori sopra la mediana	6	8	14
Valori sotto la mediana	10	4	14
Totale	16	12	28

X^2=2.33

Poiché 2.33<3.841, i due campioni possono essere stati estratti da popolazioni con la stessa mediana

I Ranghi

Il rango di X_i è la sua **posizione** in una sequenza ordinata

$R_i = R(X_i) =$ # di osservazioni campionarie $\leq X_i$

<u>Esempio</u>: X=[15, 3, 7, 18]   R=[3, 1, 2, 4]

Alle <u>osservazioni coincidenti</u> ("ties") si assegna il rango medio osservato

<u>Esempio</u>: X=[15, 7, 7, 18]   R=[3, 1.5, 1.5, 4]

Test basati sui Ranghi

<u>Idea di base</u>: costruire statistiche test che siano funzione dei ranghi

$$T = f(R_1, \ldots, R_i, \ldots, R_n)$$

I test basati sui ranghi contengono solo una parte delle informazioni fornite dal campione.

La perdita di informazioni implica, in generale, una perdita di efficienza che risulta però trascurabile (asintoticamente).

Test basati sui Ranghi

Se l'ipotesi nulla è vera, i diversi ranghi saranno distribuiti circa nello stesso modo nei due gruppi

Sono utili quando i dati contengono outlier perché non influenzano i ranghi

Test di Wilcoxon per un campione o dati appaiati

(Wilcoxon Signed-Rank test)

Valuta la simmetria di una distribuzione rispetto ad un valore ipotizzato ed è un test sul valore mediano.

Si valutano gli scostamenti dalla mediana considerandone il segno ma anche l’entità.

Ci si aspetta, sotto l’ipotesi nulla, che le somme dei ranghi positivi e negativi si equivalgano.

Test di Wilcoxon
(Wilcoxon Signed-Rank test)

Si ordinano gli scarti in valore assoluto dalla mediana e vi si assegna il corrispondente rango.

Se alcune osservazioni compaiono più di una volta, a ciascuna viene attribuito un rango pari alla media aritmetica dei ranghi che tali a tali osservazioni sarebbero stati attribuiti se esse fossero state diverse.

Si assegna a ciascun rango il segno della deviazione.

Si sommano i ranghi positivi (Sp) e quelli negativi (Sn), tenendo presente che la somma di tutti i ranghi è pari a $n(n+1)/2$

Si sceglie il minore tra Sp e Sn $T=\min(Sp, Sn)$

Test di Wilcoxon (Wilcoxon Signed-Rank test)

Nel caso di piccoli campioni si confronta T con il valore di soglia critica fornito da tabelle opportunamente predisposte.

Per campioni sufficientemente grandi la statistica test T ha distribuzione approssimativamente normale di parametri:

$\mu = n \cdot (n + 1)/4$
$\sigma^2 = n \cdot (n + 1)(2 n + 1)/24$

i valori di tali parametri vengono utilizzati per calcolare una statistica test che segue una distribuzione normale standardizzata.

Test di Wilcoxon
(Wilcoxon Signed-Rank test)

Un docente sottopone ad un campione casuale di 15 studenti un test che è stato predisposto nell'ipotesi che il punteggio mediano (in centesimi) sia 80.

Gli studenti riportano i seguenti voti:

65 92 86 84 60 90 82 95 100 83 70 88 82 96 85

I dati sono conformi alle attese del docente?

Si controlli scegliendo un valore α=0.05

Xi	Xi-80	segno	rango positivo	rango negativo
82	2	pos	1.5	
82	2	pos	1.5	
83	3	pos	3	
84	4	pos	4	
85	5	pos	5	
86	6	pos	6	
88	8	pos	7	
90	10	pos	8.5	
70	10			8.5
92	12	pos	10	
65	15			11.5
95	15	pos	11.5	
96	16	pos	13	
60	20			14.5
100	20	pos	14.5	
Somma			85.5	34.5

Il valore soglia è pari a 25. Non si rifiuta l'ipotesi nulla

p-value = 0.1553

Confronto di due campioni: dati non appaiati

Test di Wilcoxon-Mann-Whitney
(Wilcoxon-Rank Sum test)

<u>Assunzioni</u>: 2 campioni casuali e indipendenti

$X=(X_1, \ldots, X_i, \ldots, X_{n1})$ n_1 VC iid con $F(X)$

$Y=(Y_1, \ldots, Y_j, \ldots, Y_{n2})$ n_2 VC iid con $F(Y)$

F è comune ma non nota

$$\begin{cases} H_0: F(X) = F(Y) \\ H_1: F(X) = F(Y+\Delta) \end{cases}$$

alternativa di posizione
(non di forma)
Δ parametro di shift

Confronto di due campioni: dati non appaiati

1. Si costruisce il campione globale ("pooled") $n=n_1+n_2$
2. Si calcolano i ranghi sulle n osservazioni
   R^X_i =rango di X nel campione globale $(i=1,\ldots,n_1)$
   R^Y_j =rango di Y nel campione globale $(j=1,\ldots,n_2)$
3. $$W^X = \sum_{i=1}^{n_1} R_i^X$$

   o equivalentemente

   $$W^Y = \sum_{j=1}^{n_2} R_j^Y$$

   N.B. Si consiglia di costruire W sul campione meno numeroso

4. Si rifiuta H_0 per valori molto grandi o molto piccoli

Confronto di due campioni: dati non appaiati

Per n piccoli, i valori critici per rifiutare Ho si trovano nella tavola statistica apposita. L'ipotesi nulla viene rifiutata se U_i è maggiore o uguale al valore critico di U

Per n_1 e n_2 >10 la trasformazione della statistica U data da

$$Z = \frac{2U - n_1 n_2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/3}}$$

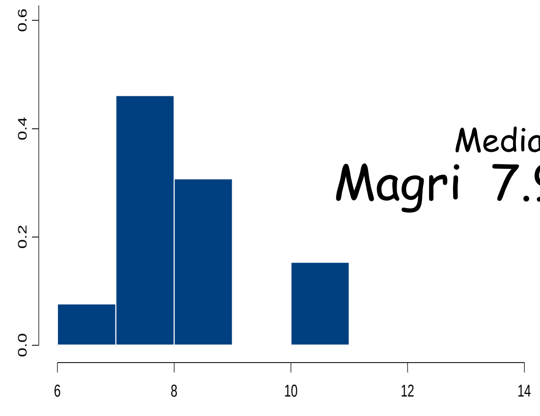
ha una distribuzione campionaria, se l'ipotesi nulla è vera, ben approssimata dalla distribuzione normale.

Confronto di due campioni: dati non appaiati

Dispendio energetico (MJ/giorno) in soggetti magri ed obesi

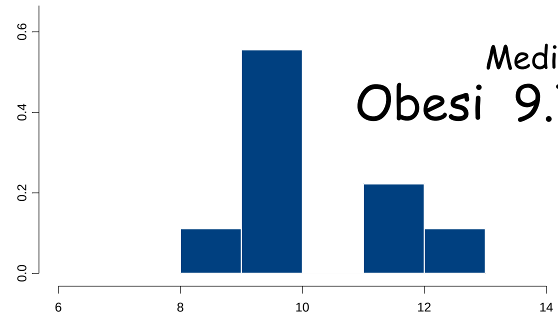
Magri (n=13)		Obesi (n=9)	
1	6.13		
2	7.05		
3.5	7.48		
3.5	7.48		
5	7.53		
6	7.58		
7	7.90		
8	8.08		
9	8.09		
10	8.11		
11	8.40		
		8.79	12
		9.19	13
		9.21	14
		9.68	15
		9.69	16
		9.97	17
18	10.15		
19	10.88		
		11.51	20
		11.85	21
		12.79	22

$W_M=103$ $W_O=150$



Mediana 1°-3°Quartile

Magri 7.9 7.48-8.11



Mediana 1°-3°Quartile

Obesi 9.7 9.2-11.5

Confronto di due campioni: dati non appaiati

Dispendio energetico (MJ/giorno) in soggetti magri ed obesi

Classified by Variable group

group	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
Magri	13	103.0	149.50	14.970751	7.923077
Obesi	9	150.0	103.50	14.970751	16.666667

Wilcoxon Two-Sample Test

Statistic (S)	150.0000

Normal Approximation

Z	3.0727
One-Sided Pr > Z	0.0011
Two-Sided Pr > |Z|	0.0021

Exact Test

One-Sided Pr >= S	5.287E-04
Two-Sided Pr >= |S - Mean|	0.0010

Alcuni test non parametrici

Test per la verifica di omogeneità fra:

i) 2 campioni indipendenti
- Test di Wilcoxon-Mann-Whitney (Wilcoxon-Rank Sum test)
- Test sulla mediana

ii) 2 campioni appaiati
- Test di Wilcoxon per dati appaiati (Wilcoxon signed rank test)

iii) >2 campioni indipendenti
- Test di Kruskal-Wallis

Test Non Parametrici

PRO

- Assunti meno restrittivi ⇨ maggior applicabilità
- Flessibilità rispetto alla natura dei dati (scala almeno ordinale)
- Possono essere impiegati anche in presenza di campioni di numerosità esigua

CONTRA

- Sono meno potenti rispetto agli analoghi parametrici, ma solo qualora siano verificati gli assunti.
- Sono utilizzabili quasi esclusivamente come strumento di verifica di ipotesi, difficilmente sono utili nel processo di stima di quantità/parametri che abbiano un significato clinico (in ambito parametrico è invece possibile mettere in relazione l'esito del test con l'IC per arricchire il processo inferenziale)