Test di Permutazione

Aldo Solari

Scuola di Alta Formazione Statistica
Inferenza Statistica in Biologia e Scienze Umane
Asti, 8 Luglio 2016

Test di permutazione

- Introdotti da R. A. Fisher nel 1935
- Noti come *permutation tests* o *randomization tests*
- Molti dei test non parametrici classici sono casi particolari di test di permutazione (e.g. Wilcoxon-Mann-Whitney, Ranghi con segno di Wilcoxon, Esatto di Fisher, McNemar, etc.)
- Utilizzo limitato nel passato perchè sono computazionalmente intensivi

Due punti di vista

Disegno randomizzato

Gli individui disponibili vengono assegnati casualmente ai trattamenti X e Y

Campionamento casuale

Gli individui vengono pescati a caso da due popolazioni X e Y

Disegno randomizzato

Ipotesi nulla

H_0: il livello del colesterolo è indipendente dalla dieta

Controfattuale

Un certo individuo che è stato assegnato dalla randomizzazione alla "dieta a base di pesce", ha un livello di colesterolo pari a 7.11.

Supponiamo che H_0 sia vera.

Allora, se l'esito della randomizzazione fosse stato "dieta a base di carne", il livello di colesterolo misurato sarebbe stato lo stesso.

Randomizzazioni potenziali

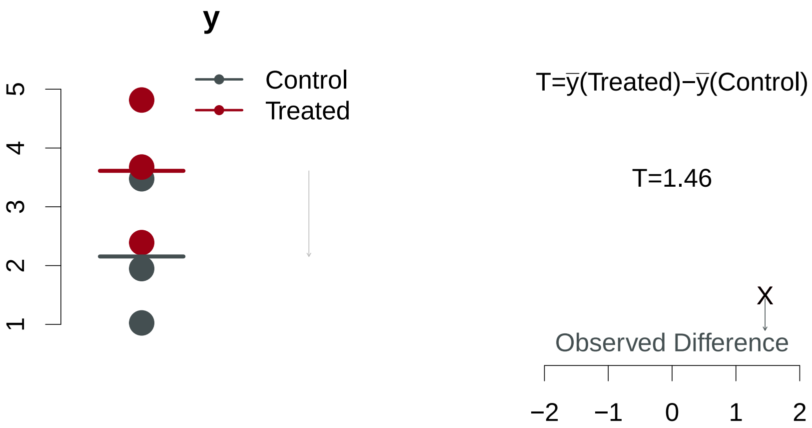
		Osservata	Potenziale 1	Potenziale 2	...
1	7.11	fish	meat	fish	...
2	6.16	fish	fish	meat	...
3	7	fish	fish	fish	...
4	6.8	fish	fish	meat	...
5	6.51	meat	meat	meat	...
6	5.86	fish	meat	meat	...
7	7.84	meat	fish	fish	...
8	6.55	fish	fish	fish	...
9	7.56	meat	meat	meat	...
10	7.61	meat	meat	fish	...
11	11.5	meat	fish	fish	...
12	5.42	fish	fish	fish	...
Statistica t		2.3451	0.6353	-1.1132	...

Distribuzione di permutazione

- Randomizzazioni possibili: $\binom{n_x + n_y}{n_x} = \binom{7+5}{5} = 792$
- Le randomizzazioni sono equiprobabili
- Ogni randomizzazione fornisce un dataset generato sotto l'ipotesi nulla
- Possiamo ottenere la "distribuzione di permutazione" calcolando il valore della statistica test t_r per l'r-sima randomizzazione, $r = 1, \ldots, \binom{n_x+n_y}{n_x}$
- p-value $= \dfrac{\sum_{r=1}^{\binom{n_x+n_y}{n_x}} I\{t_r \geq t_{oss}\}}{\binom{n_x+n_y}{n_x}} = \dfrac{6}{792} = 0.0076$

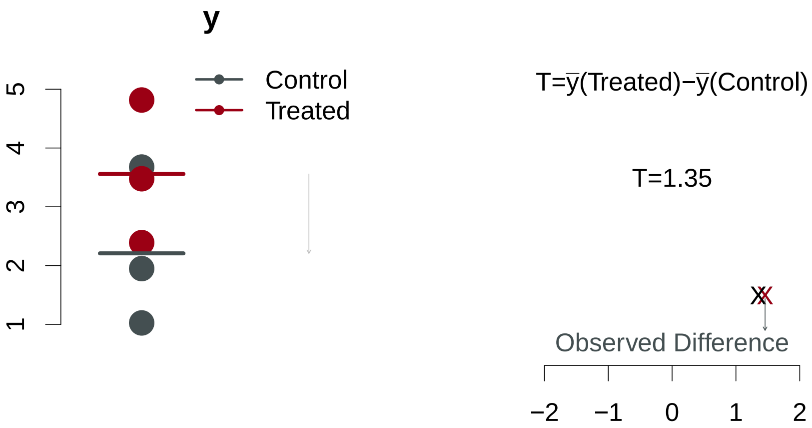
Distribuzione di permutatione della statistica t di Student

1

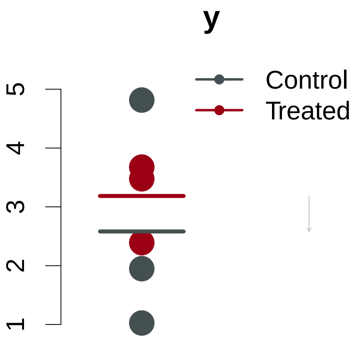


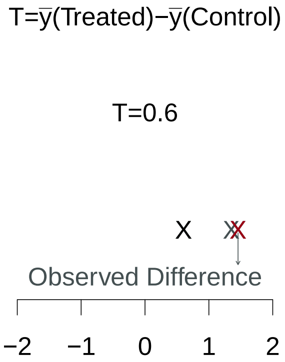
[1]fonte: Livio Finos

2



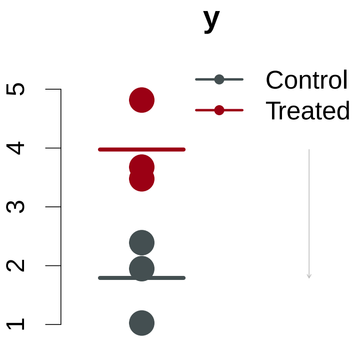
[2] fonte: Livio Finos

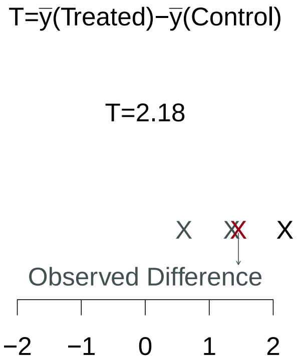




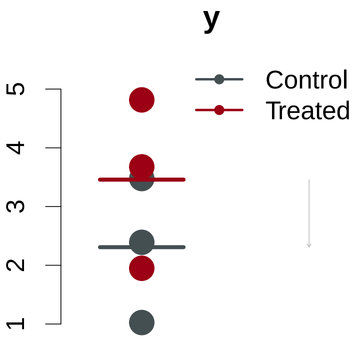
[3] fonte: Livio Finos

4



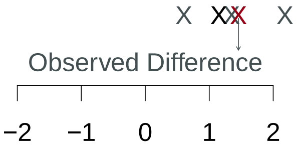


[4] fonte: Livio Finos



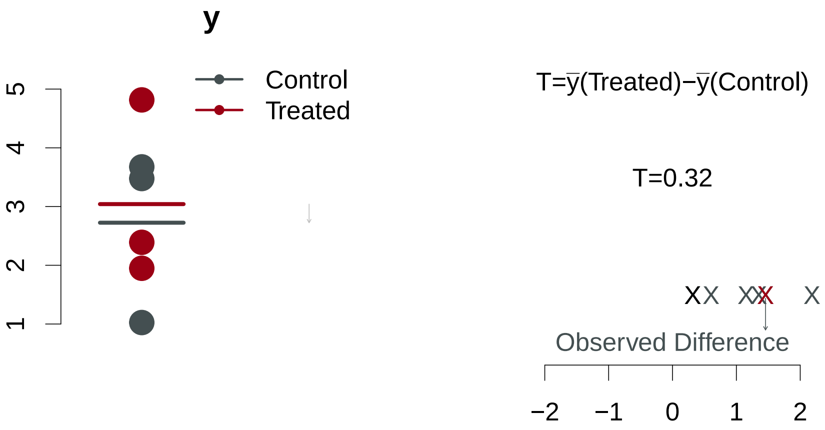
$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$

$T = 1.15$



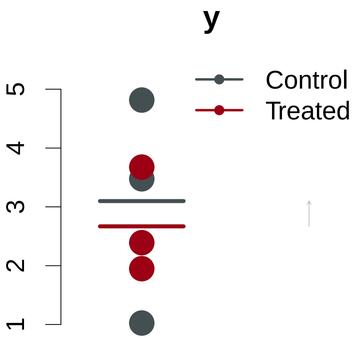
[5] fonte: Livio Finos

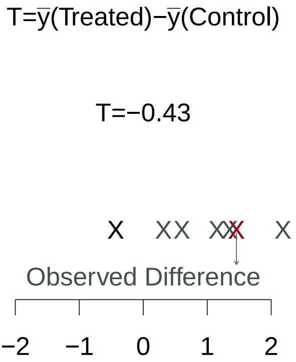
[6]



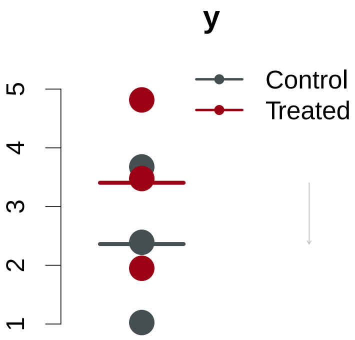
[6]fonte: Livio Finos

[7]



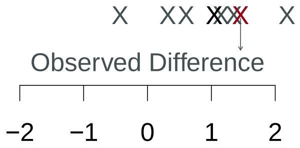


[7] fonte: Livio Finos



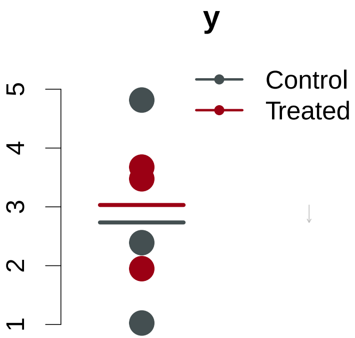
T=$\bar{y}$(Treated)−$\bar{y}$(Control)

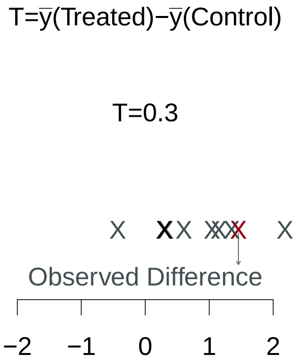
T=1.04



[8] fonte: Livio Finos

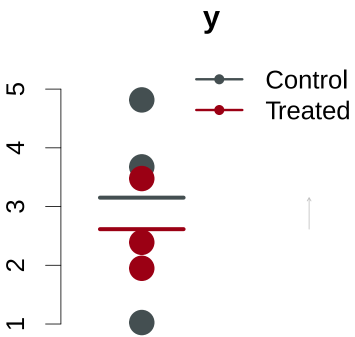
9

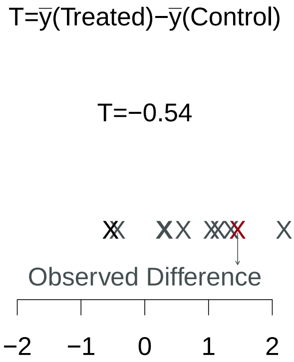




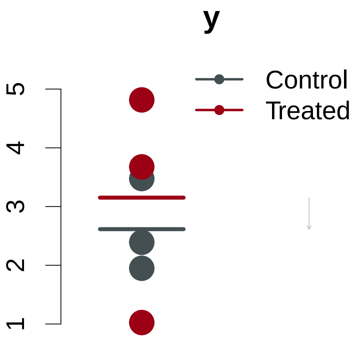
[9] fonte: Livio Finos

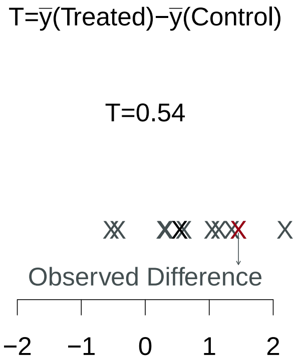
10



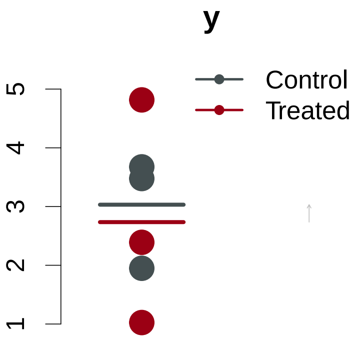


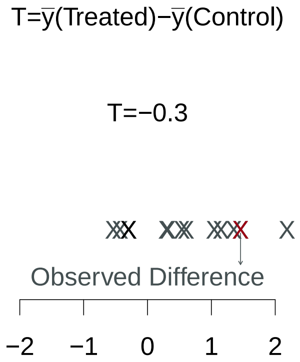
[10] fonte: Livio Finos





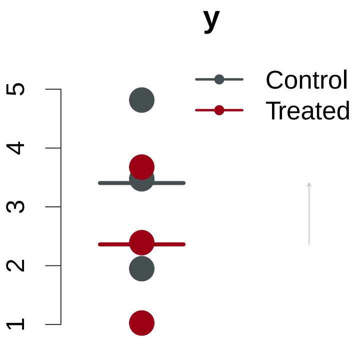
[11] fonte: Livio Finos

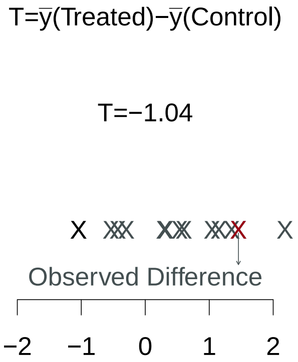




[12] fonte: Livio Finos

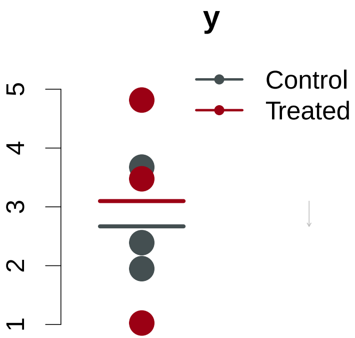
13

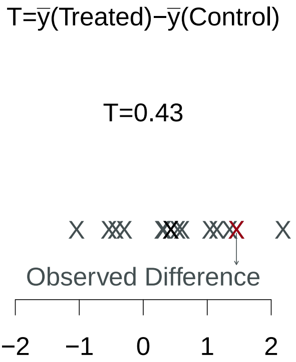




[13]fonte: Livio Finos

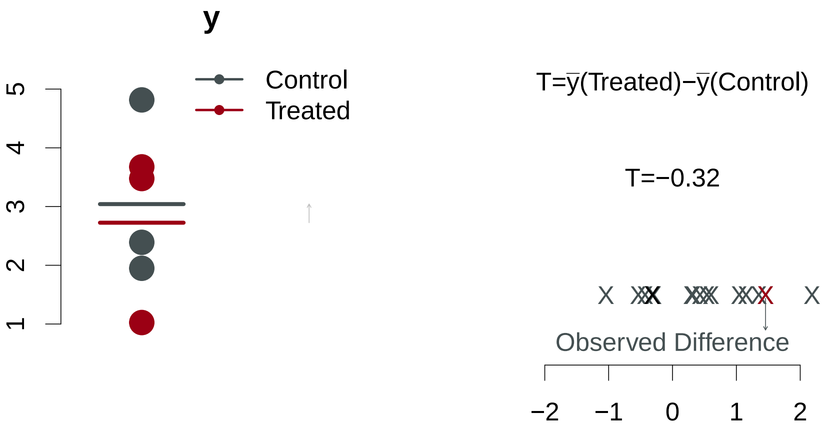
[14]



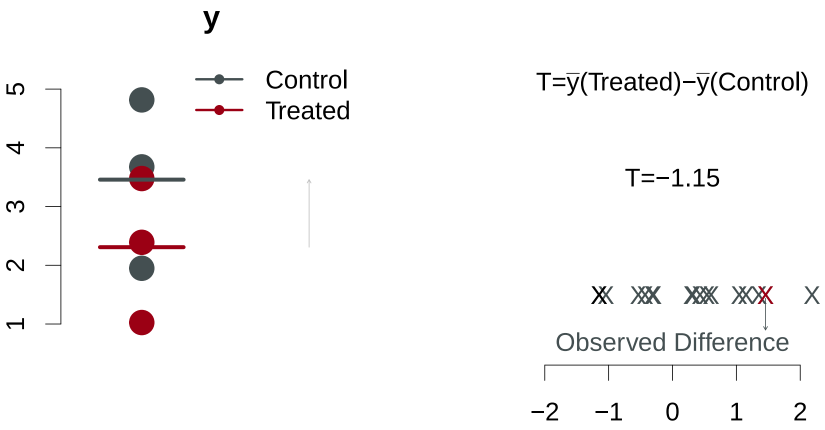


[14]fonte: Livio Finos

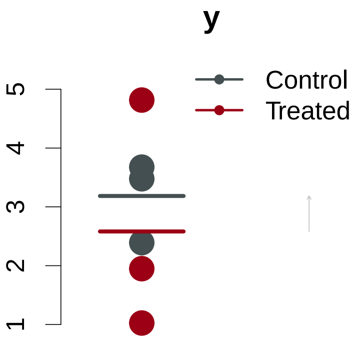
15

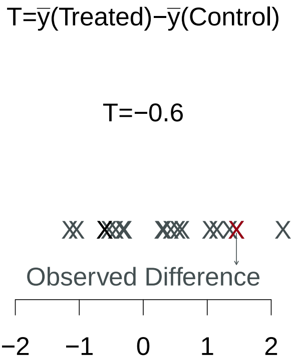


[15] fonte: Livio Finos

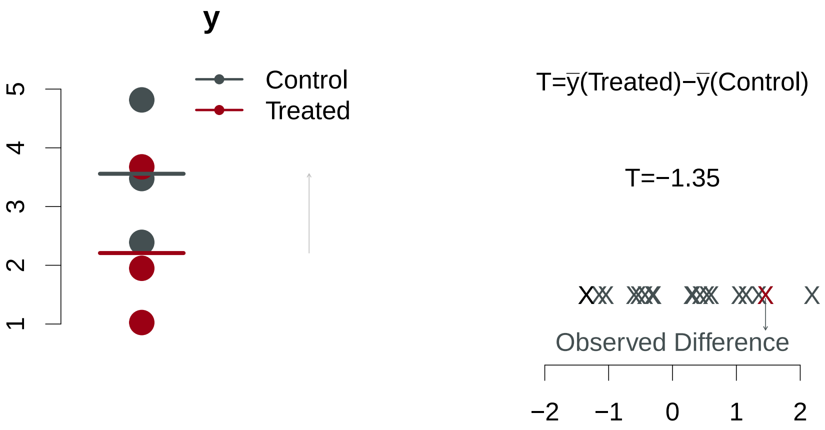


[16] fonte: Livio Finos

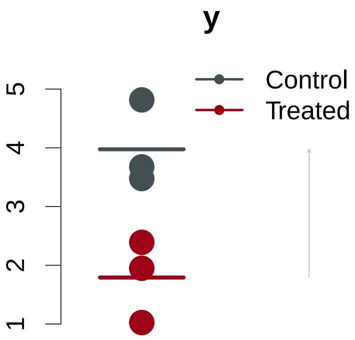




[17] fonte: Livio Finos

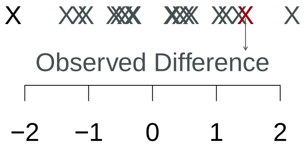


[18]fonte: Livio Finos

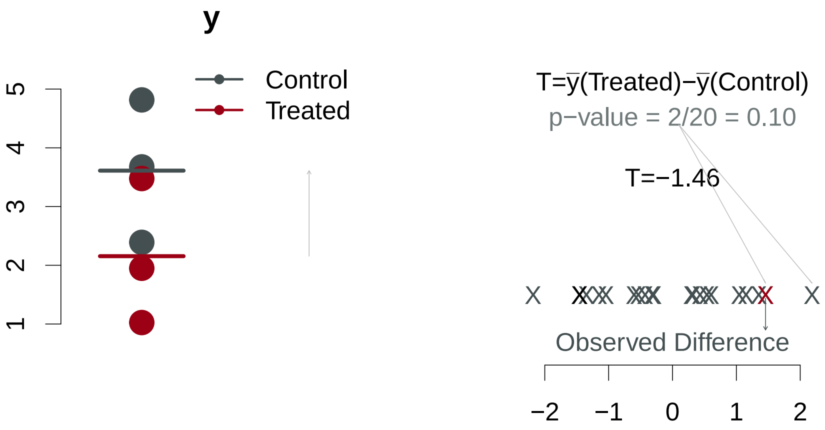


$T = \bar{y}(\text{Treated}) - \bar{y}(\text{Control})$

$T = -2.18$



19 fonte: Livio Finos



[20] fonte: Livio Finos

Campionamento casuale

- $X_1, \ldots, X_{n_x}$ i.i.d. X con f.d.r. $\mathcal{F}_x$
- $Y_1, \ldots, Y_{n_y}$ i.i.d. Y con f.d.r. $\mathcal{F}_y$
- $Z_1, \ldots, Z_{n_x+n_y} = X_1, \ldots, X_{n_x}, Y_1, \ldots, Y_{n_y}$ campione *pooled*
- Se è vera $H_0 : \mathcal{F}_x = \mathcal{F}_y$, allora $Z_1, \ldots, Z_{n_x+n_y}$ i.i.d. $Z \Rightarrow$

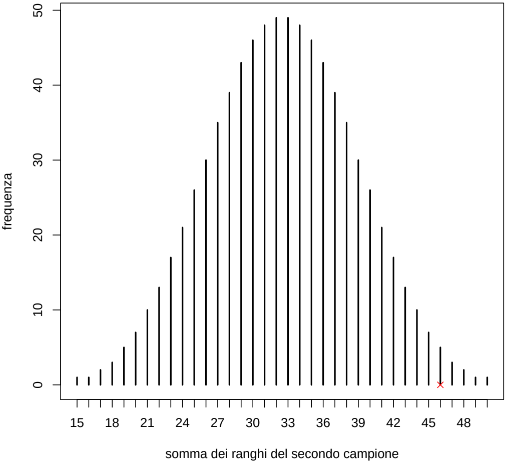
  $$(Z_1, \ldots, Z_{n_x+n_y}) \stackrel{\mathrm{d}}{=} (Z_{\pi(1)}, \ldots, Z_{\pi(n_x+n_y)})$$

  dove $\{\pi(1), \ldots, \pi(n_x + n_y)\}$ è una permutazione di $\{1, \ldots, n_x + n_y\}$
- Si possono calcolare $(n_x + n_y)!$ datasets equiprobabili

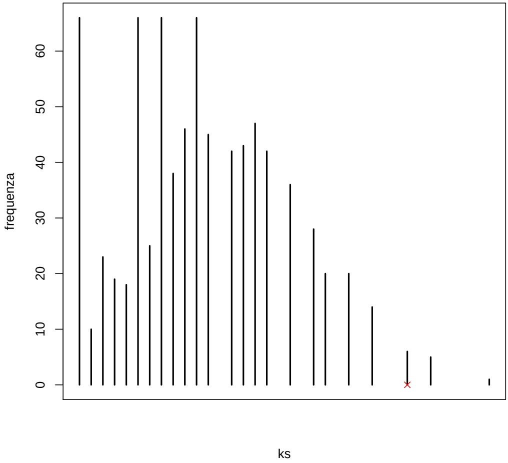
Test di Wilcoxon-Mann-Whitney

	Z_i	R_i	oss.	π_1	π_2	...
1	7.11	8	fish	meat	fish	...
2	6.16	3	fish	fish	meat	...
3	7	7	fish	fish	fish	...
4	6.8	6	fish	fish	meat	...
5	6.51	4	meat	meat	meat	...
6	5.86	2	fish	meat	meat	...
7	7.84	11	meat	fish	fish	...
8	6.55	5	fish	fish	fish	...
9	7.56	9	meat	meat	meat	...
10	7.61	10	meat	meat	fish	...
11	11.5	12	meat	fish	fish	...
12	5.42	1	fish	fish	fish	...
Statistica W			46	33	24	...

Distribuzione di permutazione della statistica W di Wilcoxon



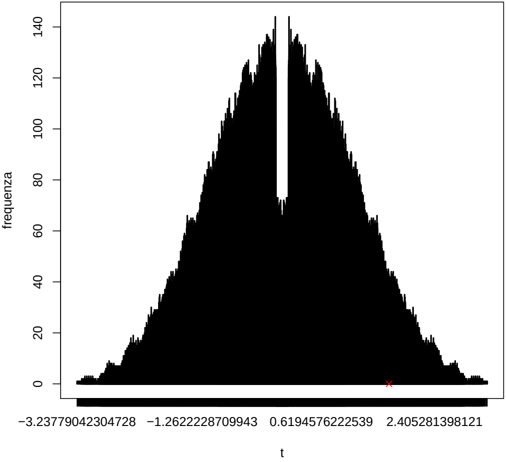
Distribuzione di permutazione della statistica KS di Kolmogorov Smirnov



Dati appaiati

			osservato			potenziale 1			...
1	0.72	0.82	SO2	Air	-0.10	Air	SO2	0.10	...
2	1.05	0.86	SO2	Air	0.19	SO2	Air	0.19	...
3	1.40	1.86	SO2	Air	-0.46	SO2	Air	-0.46	...
4	2.30	1.64	SO2	Air	0.66	Air	SO2	-0.66	...
5	13.49	12.57	SO2	Air	0.92	SO2	Air	0.92	...
...	...	...	...	...	...	...	...	...	...
t				2.07			-0.10		...

Distribuzione di permutazione della statistica t di Student per dati appaiati



Distribuzione di permutazione della statistica W dei ranghi con segno