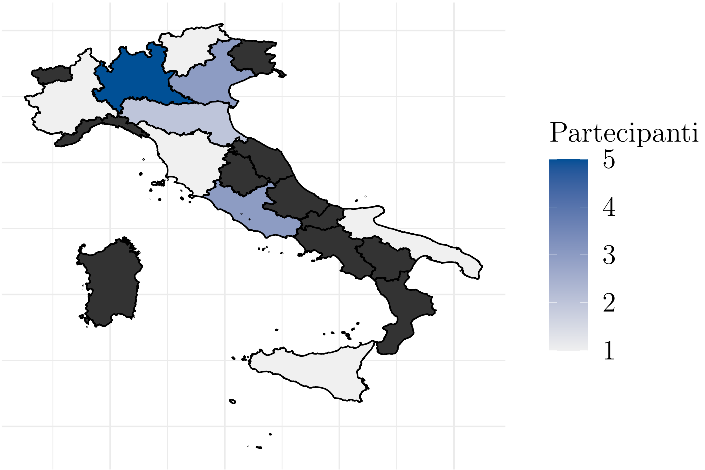
1. Inferenza statistica

Tutto quello che sappiamo, non sappiamo, crediamo di sapere...

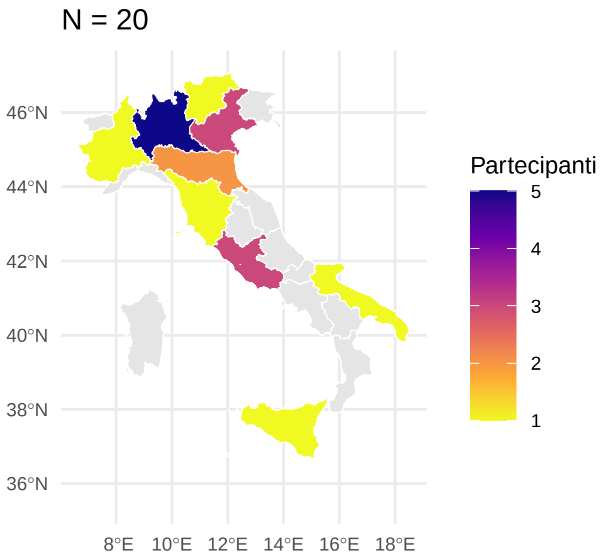
Massimiliano Pastore
Università di Padova

Da dove veniamo

$N = 20$



Da dove veniamo



Dove andiamo?

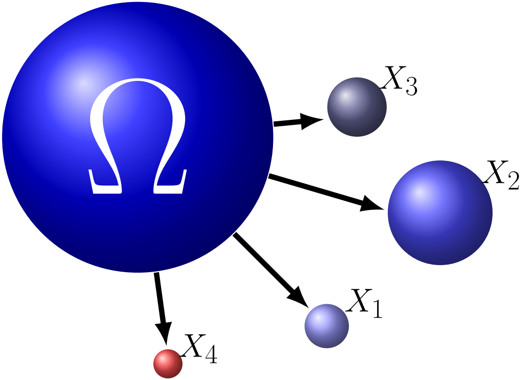
Contenuti

Popolazione e campione

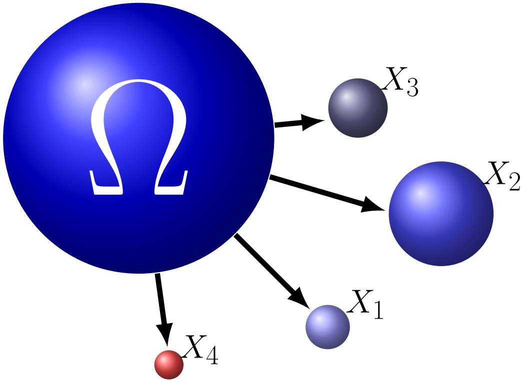
- A **population** can be defined as including all people or items with the characteristic one wishes to understand. It can be a group of existing objects or a hypothetical and potentially infinite group of objects conceived as a generalization from experience ('Glossary of statistical terms: Population'. `Statistics.com`).
- A **sample** is a set of individuals or objects collected or selected from a statistical population by a defined procedure (Peck & al. 2008).

Peck, R., Olsen, C., & Devore, J. L. (2008). Introduction to statistics and data analysis (3rd ed.), Belmont, Cal.: Thomson Brooks/Cole.

Popolazione e campione

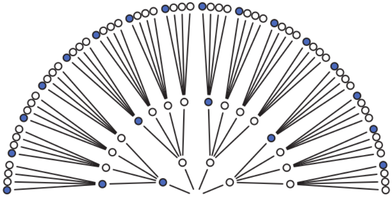


Popolazione e campione



I campioni che si estraggono da una popolazione possono essere più o meno simili alla popolazione stessa.

Il giardino dei sentieri biforcati



- Nel racconto di Jorge Louis Borges, *Il giardino dei sentieri biforcati*, si fa riferimento ad un libro ed un labirinto in cui si cerca di descrivere tutti i possibili risultati di un evento, ognuno dei quali produce una ulteriore proliferazione di conseguenze possibili.

Thanks to Richard Mc Elreath.

In pratica

- Noi vorremmo conoscere la popolazione, ma possiamo solo osservare uno o (nei casi più fortunati) più campioni.
- Pertanto, conoscere le proprietà delle statistiche ottenute sui campioni ci serve per avere informazioni su come potrebbe essere la popolazione.

Parametri e statistiche

- Un **parametro** è una caratteristica della popolazione espressa con un certo valore (solitamente numerico).
- Una **statistica** è un valore che, per mezzo di una funzione, viene associato ad una caratteristica di un qualsiasi campione di ampiezza n, appartenente ad una data popolazione.
- Generalmente i parametri sono valori incogniti, mentre le statistiche valori rilevati empiricamente che vengono utilizzati come stimatori dei parametri incogniti.

Parametri e statistiche

- Un **parametro** è una caratteristica della popolazione espressa con un certo valore (solitamente numerico).
- Una **statistica** è un valore che, per mezzo di una funzione, viene associato ad una caratteristica di un qualsiasi campione di ampiezza n, appartenente ad una data popolazione.
- Generalmente i parametri sono valori incogniti, mentre le statistiche valori rilevati empiricamente che vengono utilizzati come stimatori dei parametri incogniti.

NOTA BENE

Le statistiche si indicano con lettere latine (es. $\overline{x}$, s^2) mentre i parametri si indicano con lettere greche (es. μ, σ^2).

Inferenza

- L'**inferenza** è un processo logico per il quale, data una o più premesse, è possibile trarre una conclusione.
- Ne possiamo distinguere due tipi:

Inferenza

- L'**inferenza** è un processo logico per il quale, data una o più premesse, è possibile trarre una conclusione.
- Ne possiamo distinguere due tipi:

Inferenza deduttiva

Quando parte da premesse certe ed arriva a conclusioni certe.

Esempio

- Ogni triangolo rettangolo ha un angolo di 90°.
- Il triangolo T è rettangolo.

Inferenza

- L'**inferenza** è un processo logico per il quale, data una o più premesse, è possibile trarre una conclusione.
- Ne possiamo distinguere due tipi:

Inferenza deduttiva

Quando parte da premesse certe ed arriva a conclusioni certe.

Esempio

- Ogni triangolo rettangolo ha un angolo di 90°.
- Il triangolo T è rettangolo.
- Il triangolo T ha un angolo di 90°.

Inferenza

- L'**inferenza** è un processo logico per il quale, data una o più premesse, è possibile trarre una conclusione.
- Ne possiamo distinguere due tipi:

Inferenza deduttiva

Quando parte da premesse certe ed arriva a conclusioni certe.

Esempio

- Ogni triangolo rettangolo ha un angolo di 90°.
- Il triangolo T è rettangolo.
- Il triangolo T ha un angolo di 90°.

Inferenza induttiva

Quando parte da premesse certe ed arriva a conclusioni probabili.

Esempio

- Ho visto un corvo ed era nero.
- Ho visto un secondo corvo ed era nero.
- ....

Inferenza

- L'**inferenza** è un processo logico per il quale, data una o più premesse, è possibile trarre una conclusione.
- Ne possiamo distinguere due tipi:

Inferenza deduttiva

Quando parte da premesse certe ed arriva a conclusioni certe.

Esempio

- Ogni triangolo rettangolo ha un angolo di 90°.
- Il triangolo T è rettangolo.
- Il triangolo T ha un angolo di 90°.

Inferenza induttiva

Quando parte da premesse certe ed arriva a conclusioni probabili.

Esempio

- Ho visto un corvo ed era nero.
- Ho visto un secondo corvo ed era nero.
- ....
- Probabilmente tutti i corvi sono neri.

Elezioni presidenziali USA 1936

Elezioni presidenziali USA 1936

- La rivista *Literary Digest* condusse un sondaggio elettorale basato su 2.3 milioni di rispondenti sulla base del quale veniva dato per sicuro vincente Alf Landon.

Elezioni presidenziali USA 1936

- La rivista *Literary Digest* condusse un sondaggio elettorale basato su 2.3 milioni di rispondenti sulla base del quale veniva dato per sicuro vincente Alf Landon.
- In realtà poi stravinse Roosvelt con oltre il 60% dei voti.

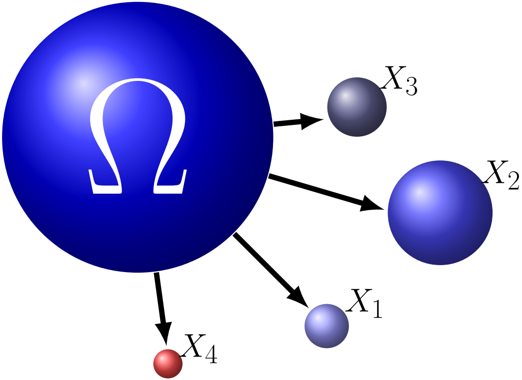
Elezioni presidenziali USA 1936

- La rivista aveva contattato 10 milioni di cittadini dalle liste del registro automobilistico e dall'elenco telefonico.

Elezioni presidenziali USA 1936

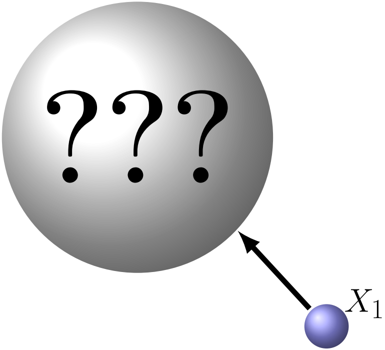
- La rivista aveva contattato 10 milioni di cittadini dalle liste del registro automobilistico e dall'elenco telefonico.
- Contemporaneamente, lo statistico George Gallup aveva predetto la vittoria di Roosevelt utilizzando un campione molto più piccolo (50000 persone). Non solo, ma con un esiguo sottoinsieme di 3000 persone incluse nella lista del *Digest* Gallup replicò i risultati ottenuti dalla rivista.
- Come mai?

Dal campione alla popolazione



I campioni che si estraggono da una popolazione possono essere più o meno simili alla popolazione stessa.

Dal campione alla popolazione



Quindi come facciamo ad avere informazioni sulla popolazione a partire dai dati del campione?

L'idea del campionamento

- Se non sappiamo come è fatta una popolazione, campioniamo ripetutamente da essa.
- Questo approccio diventa cruciale nell'approccio bayesiano; se non conosciamo la forma analitica della *posterior*, la approssimiamo per campionamento.

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

Example

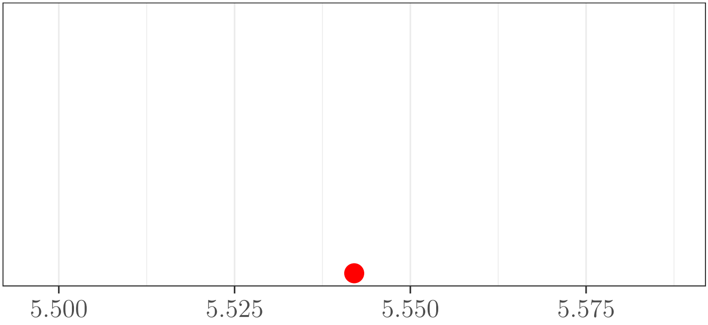
Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 1 )
```

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 1 )
```



Example

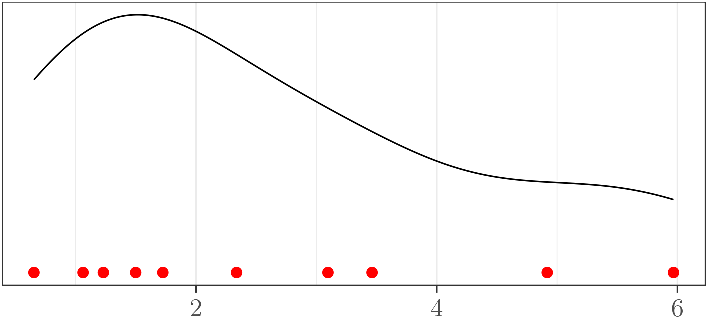
Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 10 )
```

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 10 )
```



Example

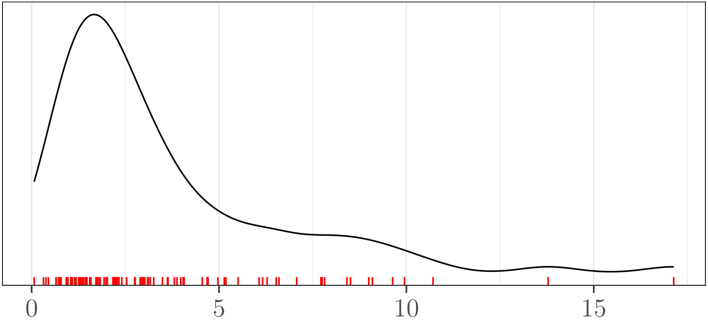
Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 100 )
```

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 100 )
```



Example

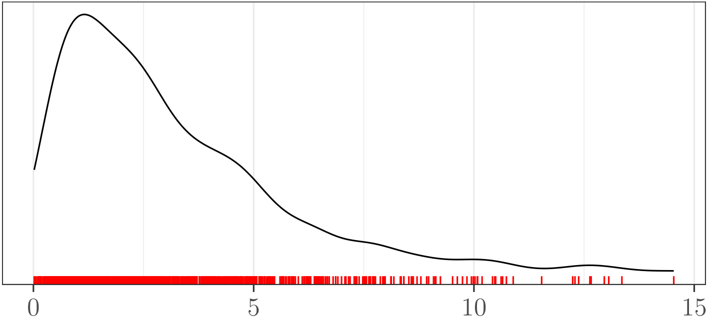
Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 1000 )
```

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 1000 )
```



Example

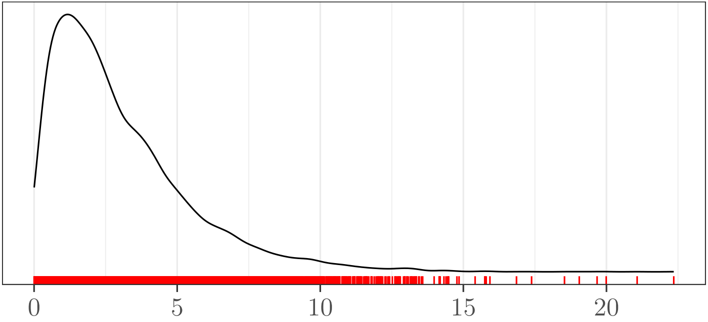
Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 10000 )
```

Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

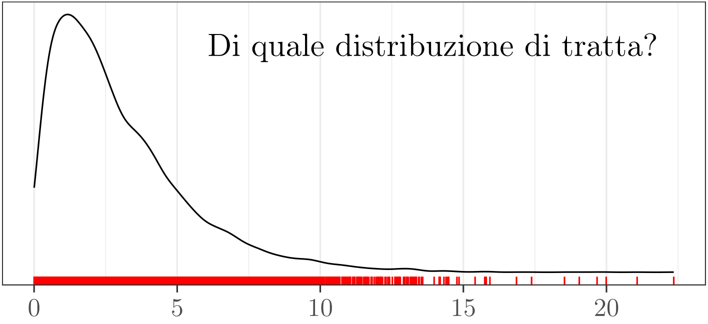
```
> y <- sample( Omega, 10000 )
```



Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

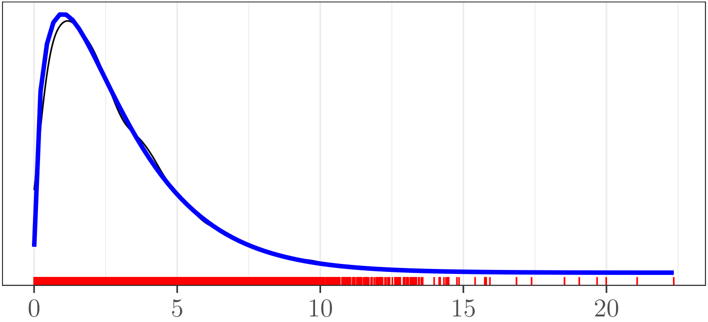
```
> y <- sample( Omega, 10000 )
```



Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

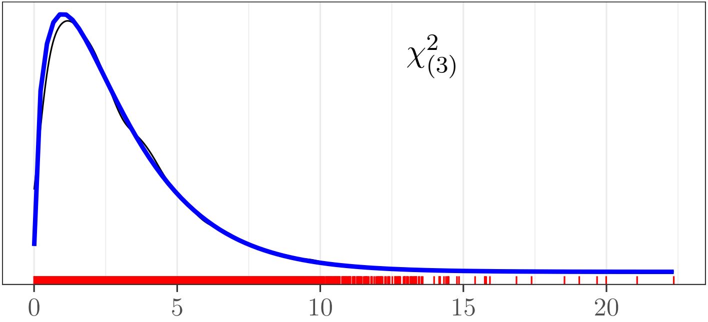
```
> y <- sample( Omega, 10000 )
```



Example

Supponiamo di voler approssimare una distribuzione ignota per campionamento.

```
> y <- sample( Omega, 10000 )
```



Inferenza statistica

- *Statistical inference can be formulated as a set of operations on data that yield estimates and uncertainty statements about predictions and parameters of some underlying process or population* (Gelman, Hill & Vehtari, 2020).
- I due approcci principali all'inferenza sono quello frequentista e quello bayesiano.

Inferenza in senso frequentista

- L'approccio NHST
- Il teorema di Bayes
- I limiti dell'approccio NHST

Problema 1

Vogliamo stimare la proporzione (o la percentuale) di studenti iscritti ai corsi della Scuola di Psicologia a cui piace la Statistica.

Problema 1

Vogliamo stimare la proporzione (o la percentuale) di studenti iscritti ai corsi della Scuola di Psicologia a cui piace la Statistica.

SONDAGGIO

Approccio frequentista

Usiamo i dati di un questionario rilevato all'inizio di un corso di Analisi dei Dati:

- Rispondenti: $n = 46$
- Piace (> 6): $n_{\text{piace}} = 9$

Approccio frequentista

Usiamo i dati di un questionario rilevato all'inizio di un corso di Analisi dei Dati:

- Rispondenti: $n = 46$
- Piace (> 6): $n_{\text{piace}} = 9$
- Proporzione campionaria: $\frac{n_{\text{piace}}}{n} = \frac{9}{46} = 0.2$
- Percentuale campionaria: $\frac{n_{\text{piace}}}{n} \times 100 = \frac{9}{46} \times 100 = 20\%$

```
> p_piace <- 0:100 / 100
```

```
> p_piace <- 0:100 / 100
```

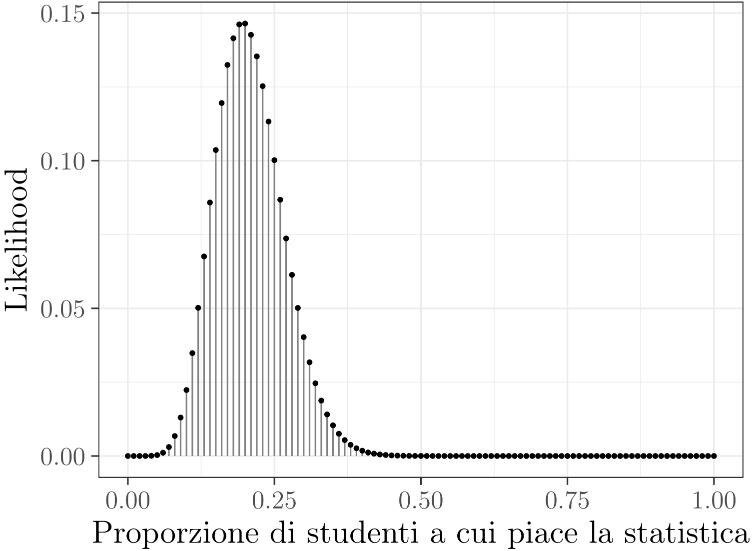
```
> L <- dbinom( 9, 46, prob = p_piace )
```

```
> p_piace <- 0:100 / 100
```

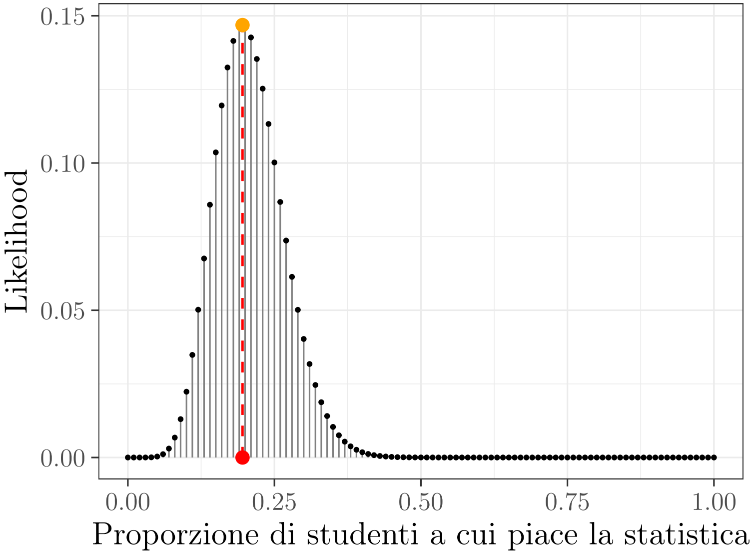
```
> L <- dbinom( 9, 46, prob = p_piace )
```

```
> plot( p_piace, L, type = 'h' )
> points( p_piace, L, pch = 19 )
```

Verosimiglianza



Verosimiglianza



Test di ipotesi

Test di ipotesi

Ipotesi H_0

Test di ipotesi

Ipotesi H_0

La proporzione di studenti a cui piace la statistica è uguale a quella degli studenti a cui non piace.
Formalmente

$$H_0 : \pi_{\text{piace}} = 0.5$$

Test di ipotesi

Ipotesi H_0

La proporzione di studenti a cui piace la statistica è uguale a quella degli studenti a cui non piace.
Formalmente

$$H_0 : \pi_{\text{piace}} = 0.5$$

Ipotesi H_1

Test di ipotesi

Ipotesi H_0

La proporzione di studenti a cui piace la statistica è uguale a quella degli studenti a cui non piace.
Formalmente

$$H_0 : \pi_{\text{piace}} = 0.5$$

Ipotesi H_1

La proporzione di studenti a cui piace la statistica è diversa da quella degli studenti a cui non piace.

Test di ipotesi

Ipotesi H_0

La proporzione di studenti a cui piace la statistica è uguale a quella degli studenti a cui non piace.
Formalmente

$$H_0 : \pi_{\text{piace}} = 0.5$$

Ipotesi H_1

La proporzione di studenti a cui piace la statistica è diversa da quella degli studenti a cui non piace.
Può essere bidirezionale

$$H_1 : \pi_{\text{piace}} \neq 0.5$$

o monodirezionale

$$H_1 : \pi_{\text{piace}} < 0.5 \qquad H_1 : \pi_{\text{piace}} > 0.5$$

```
> S <- 0:46  # spazio campionario
```

```
> S <- 0:46  # spazio campionario
```

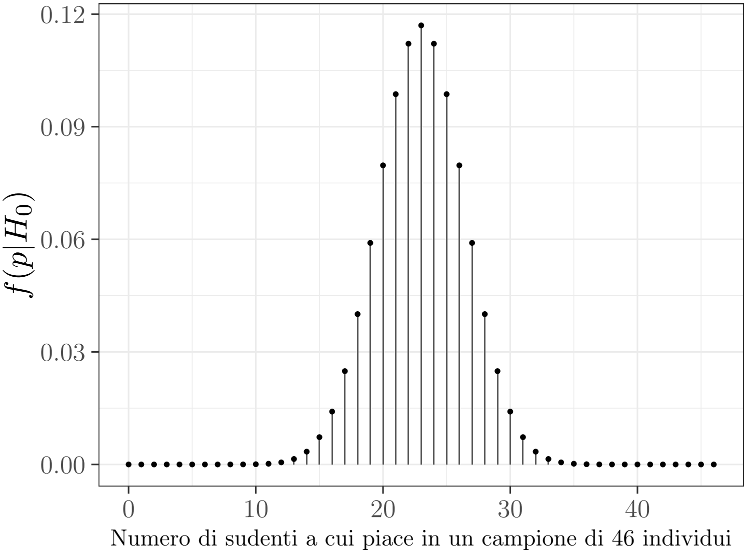
```
> P <- dbinom( S, 46, prob = .5 )
```

```r
> S <- 0:46  # spazio campionario
```

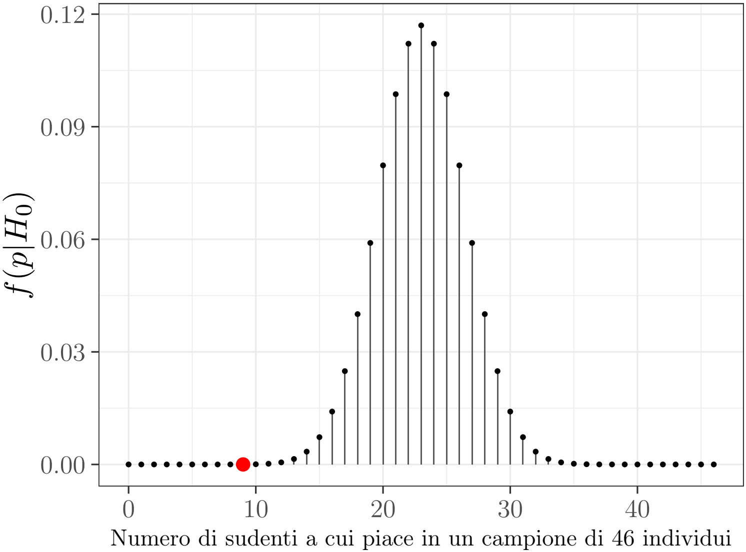
```r
> P <- dbinom( S, 46, prob = .5 )
```

```r
> plot( S, P, type = 'h' )
> points( S, P, pch = 19 )
```

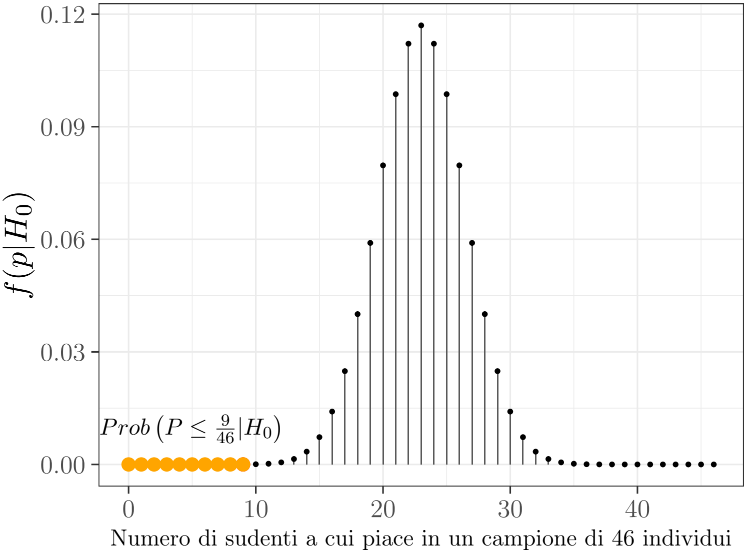
Distribuzione campionaria sotto H_0



Distribuzione campionaria sotto H_0



Distribuzione campionaria sotto H_0



$$Prob\left(P \leq \frac{9}{46} | H_0\right) =$$

$$Prob\left(P \leq \frac{9}{46} | H_0\right) =$$

```
> pbinom( 9, 46, prob = .5 )
```

$$Prob\left(P \leq \frac{9}{46}|H_0\right) =$$

```
> pbinom( 9, 46, prob = .5 )
```

```
[1] 2.028018e-05
```

Test Binomiale

```
> binom.test( 9, 46, alternative = 'less' )
```

Test Binomiale

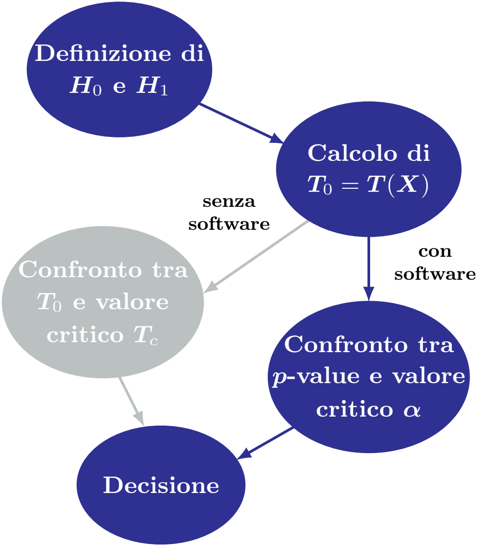
```
> binom.test( 9, 46, alternative = 'less' )
```

```
     Exact binomial test

data:  9 and 46
number of successes = 9, number of trials = 46,
                                    p-value = 2.028e-05
alternative hypothesis:
     true probability of success is less than 0.5
95 percent confidence interval:
 0.0000000 0.3165904
sample estimates:
probability of success
             0.1956522
```

L'approccio NHST

L'approccio NHST



L'approccio NHST

- Lo schema appena descritto entra in un approccio classico che è più corretto chiamare *Null Hypothesis Significance Testing* (**NHST**) piuttosto che *Verifica di ipotesi*.
- Possiamo riassumere tale approccio nel seguente modo:
  1. Dato un campione di osservazioni X, si calcola una determinata statistica test $T_0 = T(X)$.

L'approccio NHST

- Lo schema appena descritto entra in un approccio classico che è più corretto chiamare *Null Hypothesis Significance Testing* (**NHST**) piuttosto che *Verifica di ipotesi*.
- Possiamo riassumere tale approccio nel seguente modo:
  1. Dato un campione di osservazioni X, si calcola una determinata statistica test $T_0 = T(X)$.
  2. Conoscendo $f(T|H_0)$, ossia la distribuzione campionaria di T sotto l'ipotesi H_0 determiniamo $P(T \geq T_0|H_0)$ cioè la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente T_0.

L'approccio NHST

- Lo schema appena descritto entra in un approccio classico che è più corretto chiamare *Null Hypothesis Significance Testing* (**NHST**) piuttosto che *Verifica di ipotesi*.
- Possiamo riassumere tale approccio nel seguente modo:
  1. Dato un campione di osservazioni X, si calcola una determinata statistica test $T_0 = T(X)$.
  2. Conoscendo $f(T|H_0)$, ossia la distribuzione campionaria di T sotto l'ipotesi H_0 determiniamo $P(T \geq T_0|H_0)$ cioè la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente T_0.
  3. Se tale probabilità (che chiameremo *p-value*) risulta essere inferiore ad una soglia prefissata di $\alpha = 0.05$ concluderemo rigettando l'ipotesi H_0 altrimenti no.

L'approccio NHST

- Secondo tale approccio i risultati che otteniamo rispondono alla domanda:

Assumendo che H_0 sia vera, qual è la probabilità di osservare (a caso) valori uguali o più estremi rispetto a quelli empiricamente rilevati?

L'approccio NHST

- Secondo tale approccio i risultati che otteniamo rispondono alla domanda:

Assumendo che H_0 sia vera, qual è la probabilità di osservare (a caso) valori uguali o più estremi rispetto a quelli empiricamente rilevati?

- In realtà noi vorremmo rispondere alla domanda:

Dati questi valori osservati, qual è la probabilità che la mia ipotesi (H_1) sia vera?

L'approccio NHST

- Secondo tale approccio i risultati che otteniamo rispondono alla domanda:

Assumendo che H_0 sia vera, qual è la probabilità di osservare (a caso) valori uguali o più estremi rispetto a quelli empiricamente rilevati?

- In realtà noi vorremmo rispondere alla domanda:

Dati questi valori osservati, qual è la probabilità che la mia ipotesi (H_1) sia vera?

- **Non si tratta esattamente della stessa cosa!!**

$P(R|H_0) \neq P(H_0|R)$

- I test statistici che si rifanno all'approccio NHST stimano

$P(R|H_0)$

$P(R|H_0) \neq P(H_0|R)$

- I test statistici che si rifanno all'approccio NHST stimano

$P(R|H_0)$

Probabilità di ottenere il risultato R condizionata al fatto che sia vera H_0 (verosimiglianza).

$P(R|H_0) \neq P(H_0|R)$

- I test statistici che si rifanno all'approccio NHST stimano

$P(R|H_0)$

Probabilità di ottenere il risultato R condizionata al fatto che sia vera H_0 (verosimiglianza).

- In realtà noi saremmo interessati a stimare

$P(H_0|R)$

$P(R|H_0) \neq P(H_0|R)$

- I test statistici che si rifanno all'approccio NHST stimano

$P(R|H_0)$

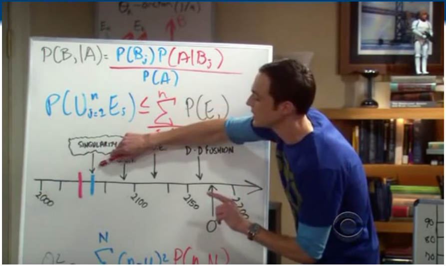
Probabilità di ottenere il risultato R condizionata al fatto che sia vera H_0 (verosimiglianza).

- In realtà noi saremmo interessati a stimare

$P(H_0|R)$

Probabilità che sia vera l'ipotesi H_0 condizionata all'aver osservato il risultato R (probabilità a posteriori).

Teorema di Bayes



Il legame tra le due probabilità è definito dal **Teorema di Bayes**.

$$P(H_0|R) = \frac{P(H_0)P(R|H_0)}{P(H_0)P(R|H_0) + P(H_1)P(R|H_1)}$$

in cui:

- $P(H_0)$ è la probabilità a priori di H_0
- $P(R|H_0)$ è la probabilità del risultato R condizionato ad H_0
- $P(H_0|R)$ è la probabilità a posteriori

Esempio (Cohen, 1994)

L'incidenza della schizofrenia negli adulti è circa del 2%. Supponiamo di avere un test per la diagnosi di tale patologia con il 95% di accuratezza nell' individuazione dei soggetti realmente schizofrenici, e circa il 97% nell' individuazione dei soggetti sani.

Esempio (Cohen, 1994)

L'incidenza della schizofrenia negli adulti è circa del 2%. Supponiamo di avere un test per la diagnosi di tale patologia con il 95% di accuratezza nell' individuazione dei soggetti realmente schizofrenici, e circa il 97% nell' individuazione dei soggetti sani.

Possiamo allora indicare con:

- H_0: il soggetto è normale
- H_1: il soggetto è schizofrenico
- R = risultato del test positivo per la schizofrenia

Pertanto avremo le seguenti probabilità:

$P(H_0) = 0.98$	

Esempio (Cohen, 1994)

L'incidenza della schizofrenia negli adulti è circa del 2%. Supponiamo di avere un test per la diagnosi di tale patologia con il 95% di accuratezza nell' individuazione dei soggetti realmente schizofrenici, e circa il 97% nell' individuazione dei soggetti sani.

Possiamo allora indicare con:

- H_0: il soggetto è normale
- H_1: il soggetto è schizofrenico
- R = risultato del test positivo per la schizofrenia

Pertanto avremo le seguenti probabilità:

$P(H_0) = 0.98$	$P(H_1) = 0.02$

Esempio (Cohen, 1994)

L'incidenza della schizofrenia negli adulti è circa del 2%.
Supponiamo di avere un test per la diagnosi di tale patologia con il 95% di accuratezza nell' individuazione dei soggetti realmente schizofrenici, e circa il 97% nell' individuazione dei soggetti sani.
Possiamo allora indicare con:

- H_0: il soggetto è normale
- H_1: il soggetto è schizofrenico
- R = risultato del test positivo per la schizofrenia

Pertanto avremo le seguenti probabilità:

$P(H_0) = 0.98$	$P(H_1) = 0.02$
$P(R\|H_0) = 0.03$	

Esempio (Cohen, 1994)

L'incidenza della schizofrenia negli adulti è circa del 2%.
Supponiamo di avere un test per la diagnosi di tale patologia con il 95% di accuratezza nell' individuazione dei soggetti realmente schizofrenici, e circa il 97% nell' individuazione dei soggetti sani.
Possiamo allora indicare con:

- H_0: il soggetto è normale
- H_1: il soggetto è schizofrenico
- R = risultato del test positivo per la schizofrenia

Pertanto avremo le seguenti probabilità:

$P(H_0) = 0.98$	$P(H_1) = 0.02$
$P(R\|H_0) = 0.03$	$P(R\|H_1) = 0.95$

Si noti che $P(R|H_0)$ è minore del 5%.

Esempio (Cohen, 1994)

Calcoliamo la probabilità a posteriori utilizzando il teorema di Bayes:

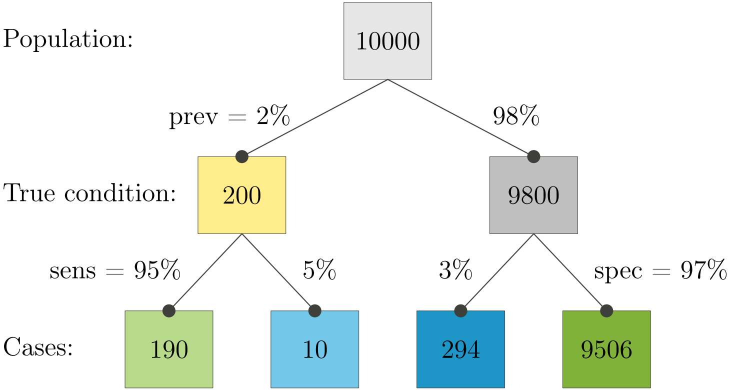
$$P(H_0|R) = \frac{0.98 \times 0.03}{0.98 \times 0.03 + 0.02 \times 0.95} \simeq 0.61$$

Esempio (Cohen, 1994)

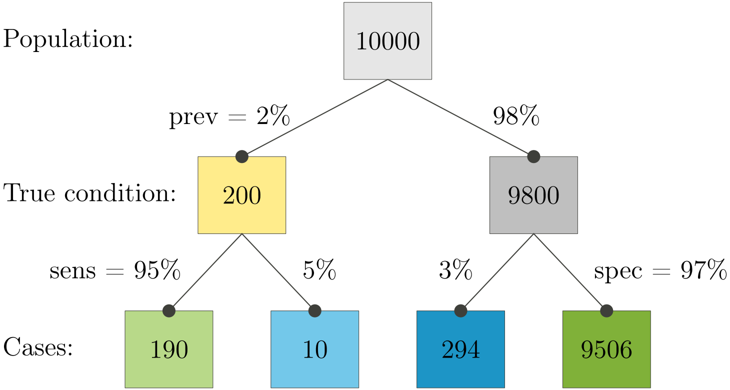
Calcoliamo la probabilità a posteriori utilizzando il teorema di Bayes:

$$P(H_0|R) = \frac{0.98 \times 0.03}{0.98 \times 0.03 + 0.02 \times 0.95} \simeq 0.61$$

Nonostante la verosimiglianza $P(R|H_0) = 0.03$ sia minore di 0.05 abbiamo una probabilità a posteriori molto più alta, $P(H_0|R) = 0.61$.



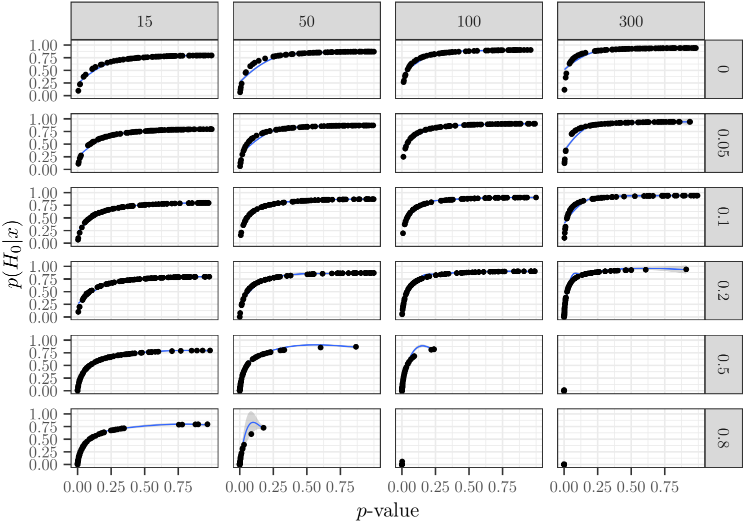
Se consideriamo tutta la popolazione, qual è la proporzione di individui positivi al test che non sono realmente malati?



Se consideriamo tutta la popolazione, qual è la proporzione di individui positivi al test che non sono realmente malati?

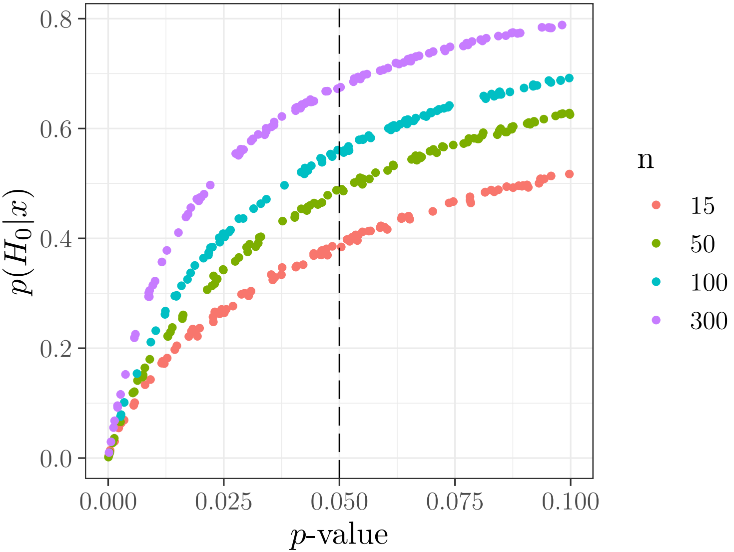
$$\frac{294}{(190 + 294)} = \frac{294}{484} \approx 0.61 = P(H_0|R)$$

Relazione tra p-value e $p(H_0|x)$ (Pastore & Altoé, 2013)[1]



[1] Pastore, M., Altoé, G. (2013). Bayes Factor e p-value: così vicini, così lontani. *Giornale italiano di psicologia*, 40, 175-193.

Relazione tra p-value e $p(H_0|x)$ con H_0 vera!!



The null ritual (Gigerenzer Krauss & Vitouch, 2004)

The null ritual (Gigerenzer Krauss & Vitouch, 2004)

- The null ritual is an invention of statistical textbook writers in the social sciences.
- The null ritual does not exist in statistic proper. What does exist are conflicting theories of inference, most relevantly those of Fisher and Neyman-Pearson.
- One rarely finds a hint at this controversy in statistics textbooks written by social scientists. As a result, the null ritual is confused with Fisher's theory of null hypothesis testing.

Gigerenzer, G., Krauss, S., & Vitouch, O. (2004). The null ritual. In D. Kaplan (Ed.), *The Sage book of quantitative methodology for the social sciences* (pp. 391–408). Thousand Oaks, CA: Sage.

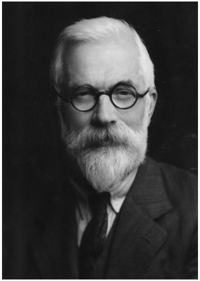
Neyman & Pearson scheme

1. Set up two statistical hypotheses, H_1 and H_2, and decide on α (Type 1 error rate), β (Type 2 error rate), and the sample size before the experiment, based on subjective cost-benefit considerations.
2. If the data fall into the rejection region of H_1, accept H_2; otherwise accept H_1.
3. The usefulness of this procedure is limited among others to situations where there is a disjunction of hypotheses (e.g., either μ_1 or μ_2 is true), where there is repeated sampling, and where you can make meaningful cost-benefit trade-offs for choosing α and β.

Fisher null hypothesis testing

No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.
(Fisher, 1956)

1. Set up a statistical null hypothesis. The null need not be a nil hypothesis (e.g., zero difference).
2. Report the exact level of significance (e.g., $p = .055$ or $.045$). Do not use a conventional 5% level all the time.
3. Use this procedure only if you know little about the problem at hand.



Fisher, R. A. (1956). *Statistical methods and scientific inference*. Edinburgh, UK: Oliver & Boyd.

Limiti dell'approccio NHST

> *Statistical signifcance is not a scientific test. It is a philosophical, qualitative test. It does not ask how much. It asks whether. Existence, the question of whether, is interesting. But it is not scientific.* (Ziliak & McCloskey, 2008)

Nell'utilizzo dell'approccio NHST bisogna assolutamente tenere conto dei seguenti aspetti:

Limiti dell'approccio NHST

Statistical signifcance is not a scientific test. It is a philosophical, qualitative test. It does not ask how much. It asks whether. Existence, the question of whether, is interesting. But it is not scientific. (Ziliak & McCloskey, 2008)

Nell'utilizzo dell'approccio NHST bisogna assolutamente tenere conto dei seguenti aspetti:

- NHST tende a indurre confusione tra la probabilità dell'ipotesi condizionata ai dati (probabilità a posteriori) e probabilità dei dati condizionati all'ipotesi (verosimiglianza).

Limiti dell'approccio NHST

> *Statistical signifcance is not a scientific test. It is a philosophical, qualitative test. It does not ask how much. It asks whether. Existence, the question of whether, is interesting. But it is not scientific.* (Ziliak & McCloskey, 2008)

Nell'utilizzo dell'approccio NHST bisogna assolutamente tenere conto dei seguenti aspetti:

- NHST tende a indurre confusione tra la probabilità dell'ipotesi condizionata ai dati (probabilità a posteriori) e probabilità dei dati condizionati all'ipotesi (verosimiglianza).
- NHST viene erroneamente considerato un metodo per la verifica delle ipotesi. In realtà esso tiene conto solo di H_0 e permette solo la falsificazione di tale ipotesi senza che questo abbia relazione con la veridicità di H_1.

Ziliak, S. T., & McCloskey, D. N. (2008). *The cult of statistical significance.* Ann Arbor, MI: University of Michigan Press.

Limiti dell'approccio NHST

- Il criterio $\alpha = 0.05$ è puramente arbitrario ed è stato più volte messo in discussione (es. Johnson, 2013; Benjamin & al., 2017; Lakens, 2018)

Limiti dell'approccio NHST

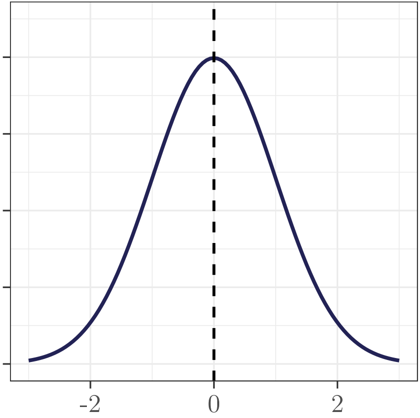
- Il criterio $\alpha = 0.05$ è puramente arbitrario ed è stato più volte messo in discussione (es. Johnson, 2013; Benjamin & al., 2017; Lakens, 2018)
- I test tradizionali tendono a sovrastimare l'evidenza contro H_0: infatti H_0 nei contesti reali non è mai esattamente vera e pertanto aumentando a dovere il numero di osservazioni è quasi sempre possibile rigettarla (Wagenmakers, 2007).

Benjamin, D. J., Berger, J. O., Johannesson, M., Nosek, B. A., Wagenmakers, E. J., Berk, R., . . . & Cesarini, D. (2018). Redefine statistical significance. *Nature Human Behaviour*, 2(1), 6–10.
Johnson, V. E. (2013). Revised standards for statistical evidence. *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 110, 19313–19317.
Lakens, D. (2018). Justify Your Alpha by Decreasing Alpha Levels as a Function of the Sample Size, disponibile on-line:
http://daniellakens.blogspot.com/2018/12/testing-whether-observed-data-should.html
Wagenmakers, E. J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.

Limiti dell'approccio NHST

- L'ipotesi nulla legata ad un unico valore puntuale può portare a conclusioni improprie (Berger & Sellke, 1987; Sellke, Bayarri & Berger, 2001).

  Ad esempio, se:
  $H_0 : \mu = 0$



Berger, J. O., & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association*, 82(397), pp. 112–122.

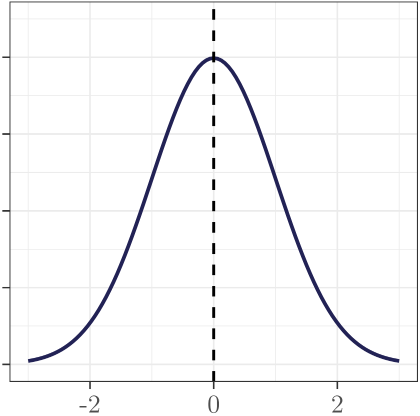
Sellke, T., Bayarri, M., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71.

Limiti dell'approccio NHST

- L'ipotesi nulla legata ad un unico valore puntuale può portare a conclusioni improprie (Berger & Sellke, 1987; Sellke, Bayarri & Berger, 2001).

  Ad esempio, se:
  $H_0 : \mu = 0$

  $$P(\mu = 0) =$$



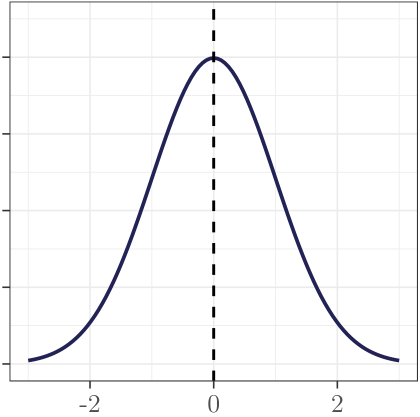
Berger, J. O., & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association*, 82(397), pp. 112–122.
Sellke, T., Bayarri, M., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71.

Limiti dell'approccio NHST

- L'ipotesi nulla legata ad un unico valore puntuale può portare a conclusioni improprie (Berger & Sellke, 1987; Sellke, Bayarri & Berger, 2001).

  Ad esempio, se:
  $H_0 : \mu = 0$

  $$P(\mu = 0) = 0$$



Berger, J. O., & Sellke, T. (1987). Testing a Point Null Hypothesis: The Irreconcilability of P Values and Evidence. *Journal of the American Statistical Association*, 82(397), pp. 112–122.
Sellke, T., Bayarri, M., & Berger, J. O. (2001). Calibration of p values for testing precise null hypotheses. *The American Statistician*, 55(1), 62–71.

Il paradosso del t-test (Rouder & al, 2009)

- Consideriamo un esperimento in cui 100 soggetti rispondono a 100 stimoli in due condizioni diverse. Per ciascun soggetto avremo pertanto 200 tempi di reazione (TR).
- Assumiamo che per ciascun soggetto la media dei TR sia compresa tra 500 e 1000 ms, con una deviazione standard di 300.
- Siano x_{1i} e x_{2i}, $i = \{1, ..., n\}$ le medie dei TR nelle due condizioni per ciascun soggetto.
- Vogliamo confrontare questi valori per stabilire se esista una differenza significativa tra le medie nelle due condizioni.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.

Il paradosso del t-test (Rouder & al, 2009)

Confrontando le medie nelle due condizioni otteniamo il seguente risultato:

```
 Paired t-test

data:  x2 and x1
t = 2.2419, df = 99, p-value = 0.0272
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
  1.320234 21.653010
sample estimates:
mean difference
       11.48662
```

con una differenza tra le medie pari a circa 11ms.

Il paradosso del t-test (Rouder & al, 2009)

Confrontando le medie nelle due condizioni otteniamo il seguente risultato:

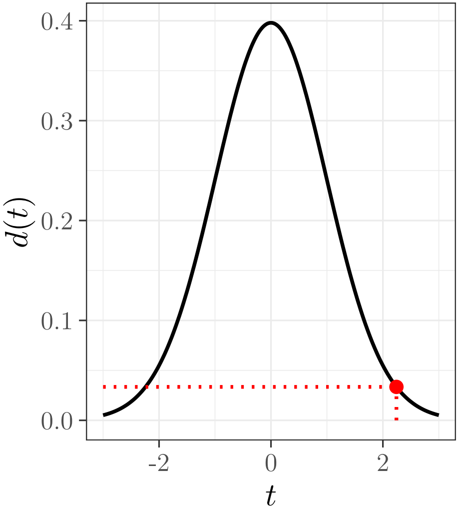
```
 Paired t-test

data:  x2 and x1
t = 2.2419, df = 99, p-value = 0.0272
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
  1.320234 21.653010
sample estimates:
mean difference
       11.48662
```

con una differenza tra le medie pari a circa 11ms.

Un modo per quantificare l'evidenza del risultato consiste nel calcolare la verosimiglianza del valore osservato di t (2.242) sotto varie ipotesi alternative.

Verosimiglianza

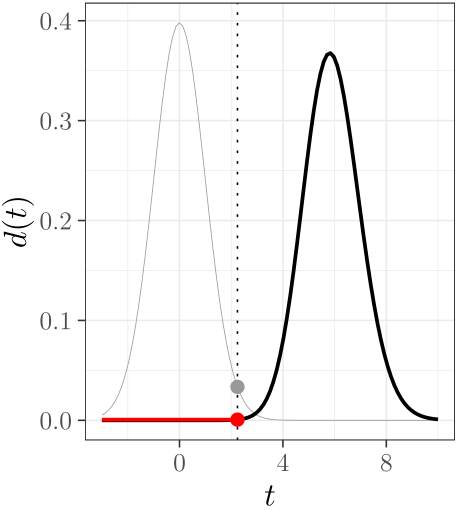


Il valore di verosimiglianza di $t = 2.242$ sotto l'ipotesi H_0 è dato dalla densità della distribuzione t con 99 gradi di libertà, cioè circa 0.033.

Se H_0 è vera vuol dire che non esiste differenza tra i TR nelle due condizioni.

Formalmente
$H_0 : \mu_2 - \mu_1 = 0.$

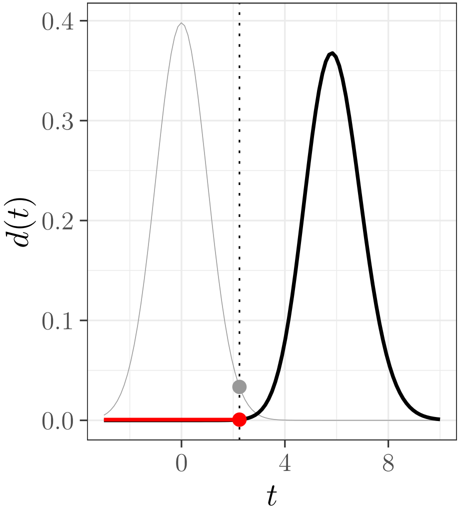
Rapporto di verosimiglianza



Supponiamo che la reale differenza tra i TR sia invece pari a 30ms e calcoliamo il relativo valore di verosimiglianza per $t = 2.2419$ sotto l'ipotesi $H_1 : \mu_2 - \mu_1 = 30$.

Tale valore è dato dalla densità della distribuzione t con 99 gradi di libertà e parametro di noncentralità $\sqrt{N}\frac{\Delta_\mu}{\sigma_{\Delta_\mu}}$ (circa 5.86) ed è inferiore a 0.001.

Rapporto di verosimiglianza

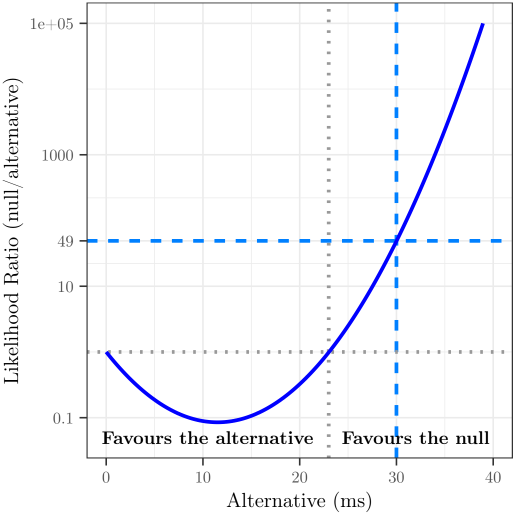


Supponiamo che la reale differenza tra i TR sia invece pari a 30ms e calcoliamo il relativo valore di verosimiglianza per $t = 2.2419$ sotto l'ipotesi $H_1 : \mu_2 - \mu_1 = 30$.

Tale valore è dato dalla densità della distribuzione t con 99 gradi di libertà e parametro di noncentralità $\sqrt{N}\frac{\Delta_\mu}{\sigma_{\Delta_\mu}}$ (circa 5.86) ed è inferiore a 0.001.

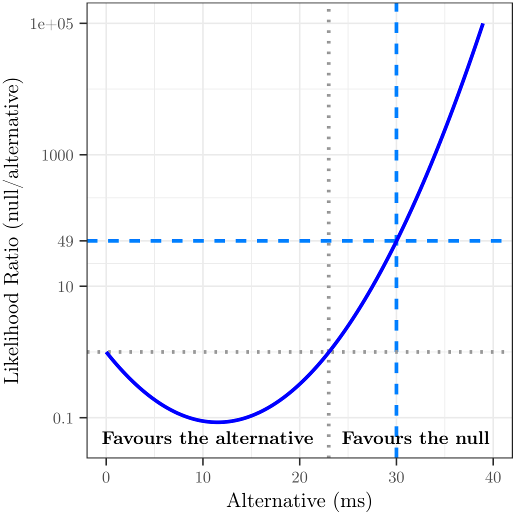
Il rapporto di verosimiglianza tra $L(H_0)$ e $L(H_1)$ risulta di 48 a 1 in favore di H_0, nonostante il test sia risultato statisticamente significativo (!!!!).

Likelihood Ratio



Rappresentiamo graficamente il rapporto di verosimiglianza tra H_0 e diversi valori di ipotesi alternativa, fino a 40ms.

Likelihood Ratio



Rappresentiamo graficamente il rapporto di verosimiglianza tra H_0 e diversi valori di ipotesi alternativa, fino a 40ms.

Si osserva il seguente paradosso:

l'ipotesi H_0 risulta essere più verosimile rispetto all'ipotesi alternativa per differenze superiori a 23ms.

Interpretazione del *p-value*: 7 marzo 2016

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics

THE WORLD OF STATISTICS

THIS is STATISTICS

Accreditation | Community | Members Only |

ABOUT ASA | MEMBERSHIP | AWARDS & RECOGNITION | CAREER CENTER | EDUCATION | PUBLICATION

ASA Headlines

ASA Releases 'Statement on Statistical Significance and *P*-Values'

The ASA "**Statement on Statistical Significance and *P*-Values**" includes six principles underlying the proper use and interpretation of the *p*-value and is intended to improve the conduct and interpretation of quantitative science and inform the growing emphasis on reproducibility of science research. "The p-value was never intended to be a substitute for scientific reasoning," said Ron Wasserstein, the ASA's executive director, in a **press release**. "Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a 'post $p<0.05$ era.'" The statement is published in ***The American Statistician*** along with more than a dozen discussion papers to provide further perspective on this broad and complex topic.

General Article

Failing Grade: 89% of Introduction-to-Psychology Textbooks That Define or Explain Statistical Significance Do So Incorrectly

Scott A. Cassidy, Ralitza Dimova, Benjamin Giguère, Jeffrey R. Spence, and David J. Stanley
Department of Psychology, University of Guelph

(*Advances in Methods and Practices in Psychological Science*, 2019)

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 235–245: Statistical Inference in the 21st Century
https://doi.org/10.1080/00031305.2018.1527253

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Abandon Statistical Significance

Blakeley B. McShane[a], David Gal[b], Andrew Gelman[c], Christian Robert[d], and Jennifer L. Tackett[e]

[a]Department of Marketing, Kellogg School of Management, Northwestern University, Evanston, IL; [b]Department of Managerial Studies, College of Business Administration, University of Illinois at Chicago, Chicago, IL; [c]Department of Statistics and Department of Political Science, Columbia University, New York, NY; [d]Centre de Recherche en Mathématiques de la Décision (CEREMADE), Université Paris-Dauphine, Paris, France; [e]Department of Psychology, Northwestern University, Evanston, IL

ABSTRACT
We discuss problems the null hypothesis significance testing (NHST) paradigm poses for replication and more broadly in the biomedical and social sciences as well as how these problems remain unresolved by proposals involving modified *p*-value thresholds, confidence intervals, and Bayes factors. We then discuss our own proposal, which is to abandon statistical significance. We recommend dropping the NHST paradigm—and the *p*-value thresholds intrinsic to it—as the default statistical paradigm for research, publication, and discovery in the biomedical and social sciences. Specifically, we propose that the *p*-value be demoted from its threshold screening role and instead, treated continuously, be considered along with currently subordinate factors (e.g., related prior evidence, plausibility of mechanism, study design and data quality, real world costs and benefits, novelty of finding, and other factors that vary by research domain) as just one among many pieces of evidence. We have no desire to "ban" *p*-values or other purely statistical measures. Rather, we believe that such measures should not be thresholded and that, thresholded or not, they should not take priority over the currently subordinate factors. We also argue that it seldom makes sense to calibrate evidence as a function of *p*-values or other purely statistical measures. We offer recommendations for how our proposal can be implemented in the scientific publication process as well as in statistical decision making more broadly.

ARTICLE HISTORY
Received October 2017
Revised September 2018

KEYWORDS
Null hypothesis significance testing; *p*-Value; Replication; Sociology of science; Statistical significance

Abandon the idea that the goal of the statistical analysis is to get some sort of certainty. Instead, accept posterior ambiguity: don't try to learn more from the data than you really can.

(Andrew Gelman, 2018)

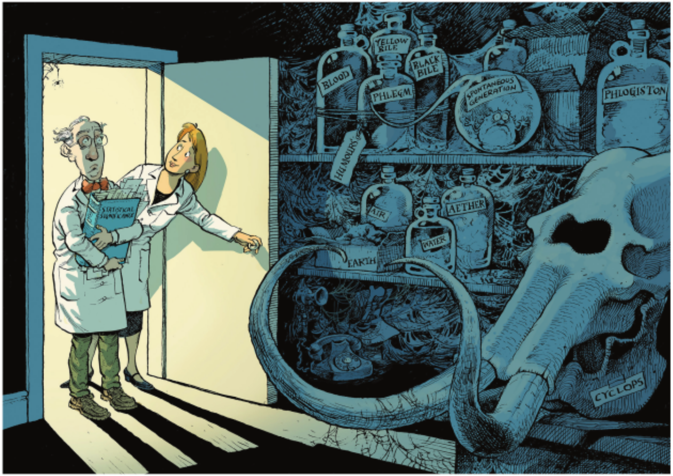
Abandon the idea that the goal of the statistical analysis is to get some sort of certainty. Instead, accept posterior ambiguity: don't try to learn more from the data than you really can.

(Andrew Gelman, 2018)

Accept uncertainty. Be thoughtful, open, and modest. Remember "ATOM".

(Wasserstein, Schirm & Lazar, 2019)

ILLUSTRATION BY DAVID PARKINS



Retire statistical significance

NATURE, 21 March 2019

Adopting More Holistic Approaches

Inferential Statistics as Descriptive Statistics: There Is No Replication Crisis if We Don’t Expect Replication

Valentin Amrhein, David Trafimow & Sander Greenland

Pages 262-270 | Received 19 Mar 2018, Accepted 25 Oct 2018, Published online: 20 Mar 2019

The American Statistician, 2019

Interpretazione del *p-value*

In sostanza:

- *p-value* è la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente se (e solo se) l'ipotesi H_0 è vera.

Interpretazione del *p-value*

In sostanza:

- *p-value* è la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente se (e solo se) l'ipotesi H_0 è vera.
- Non rappresenta in alcun modo la probabilità che sia vera H_0 e pertanto non può essere considerato un misuratore del grado di falsità della stessa.

Interpretazione del *p-value*

In sostanza:

- *p-value* è la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente se (e solo se) l'ipotesi H_0 è vera.
- Non rappresenta in alcun modo la probabilità che sia vera H_0 e pertanto non può essere considerato un misuratore del grado di falsità della stessa.
- *p-value* risente della numerosità campionaria: più aumenta n più tende a diminuire.

Interpretazione del *p-value*

In sostanza:

- *p-value* è la probabilità di ottenere un risultato uguale o più estremo rispetto a quello rilevato empiricamente se (e solo se) l'ipotesi H_0 è vera.
- Non rappresenta in alcun modo la probabilità che sia vera H_0 e pertanto non può essere considerato un misuratore del grado di falsità della stessa.
- *p-value* risente della numerosità campionaria: più aumenta n più tende a diminuire.
- Non è una misura dell'evidenza statistica, va usato solo come criterio decisionale per rigettare o meno H_0.

Interpretazione del *p-value*

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

The Practical Alternative to the *p* Value Is the Correctly Used *p* Value

Daniël Lakens
Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology

Perspectives on Psychological Science
1–10

DOI: 10.1177/1745691620958012
www.psychologicalscience.org/PPS
SAGE

(Lakens, 2021)

Inferenza in senso bayesiano

Inferenza Bayesiana

*A frequentist is a person whose long-run ambition is to be wrong 5% of the time.
A Bayesian is one who, vaguely expecting a horse, and catching a glimpse of a donkey, strongly believes he has seen a mule.*

- Nonostante il teorema di Bayes sia storicamente più antico (risale circa alla metà del '700) la sua affermazione in un approccio statistico alternativo può essere datata verso la metà degli anni '60 ad opera di Lindley.
- La ragione di questo *ritardo* è da imputare a due principali ragioni, una storica ed una tecnologica.

Inferenza Bayesiana

- Negli ultimi anni è cresciuta una *corrente Bayesiana*, anche nell'ambito della psicologia (es. Dienes and Mclatchie, 2018; Etz and Vandekerckhove, 2018; Kruschke and Liddell, 2018).
- Vi sono almeno tre ragioni che possono spiegare questo fatto:
  1. L'approccio Bayesiano permette una valutazione delle ipotesi in termini di evidenza.
  2. Può essere utilizzato anche in casi di elevata complessità.
  3. Molti software statistici hanno ormai implementato la possibiltà di fare analisi in senso Bayesiano.

Dienes, Z., & Mclatchie, N. (2018). Four reasons to prefer Bayesian analyses over significance testing. *Psychonomic Bulletin & Review*, 25(1), 207–218.
Etz, A., & Vandekerckhove, J. (2018). Introduction to Bayesian inference for psychology. *Psychonomic Bulletin & Review*, 25(1), 5–34.
Kruschke, J. K., & Liddell, T. M. (2018b). The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, 25(1), 178–206.

Statistical Modeling, Causal Inference, and Social Science

https://statmodeling.stat.columbia.edu/

Statistical Modeling, Causal Inference, and Social Science
https://statmodeling.stat.columbia.edu/

Bayesians are frequentists.

Posted on January 8, 2024 9:50 AM by Andrew

Statistical Modeling, Causal Inference, and Social Science

`https://statmodeling.stat.columbia.edu/`

Bayesians are frequentists.

Posted on January 8, 2024 9:50 AM by Andrew

Anything you can do with Bayesian inference you can do in other ways. Bayesian inference is a bit like calculus: . . .

Posted on June 5, 2025 9:01 AM

Approccio bayesiano vs NHST

- Nell'approccio NHST i parametri (quasi sempre incogniti) sono considerati come quantità fisse.
- Nell'approccio bayesiano invece, i parametri sono trattati come variabili casuali con una distribuzione di probabilità.

Esempio:

- Supponiamo che μ indichi la media (incognita) di una popolazione.
- Con l'approccio bayesiano possiamo stimare ad esempio $P(\mu < T)$, $P(T_1 < \mu < T_2)$.
- Con l'approccio classico questo non è possibile in quanto la probabilità che la media vera sia inferiore a T è 0 oppure 1.

Sistema di Ipotesi

- Nel sistema di ipotesi bayesiano viene definita una distribuzione a priori per il parametro oggetto del test.
- Questa distribuzione viene combinata con le informazioni derivate dai dati (verosimiglianza) per ottenere una distribuzione a posteriori.
- Formalmente:

$$\text{posterior} \propto \text{prior} \times \text{likelihood}$$

Problema 2

Vogliamo stimare la percentuale (o la proporzione) di studenti iscritti ai corsi della Scuola di Psicologia a cui piace la Statistica, ma questa volta utilizzando una procedura di stima bayesiana.

Definizione della *prior*

- Sia $\theta = p(\text{piace})$.
- Come primo passo dobbiamo definire i nostri *prior beliefs*.
- Secondo NHST siamo abituati a definire solo l'ipotesi $H_0 : \theta = .5$
- In questo caso possiamo fare una operazione molto più raffinata, ovvero individuare più valori possibili di θ ed assegnare a ciascuno di essi una probabilità $p(\theta)$.

Definizione della *prior*

- Sia $\theta = p(\text{piace})$.
- Come primo passo dobbiamo definire i nostri *prior beliefs*.
- Secondo NHST siamo abituati a definire solo l'ipotesi $H_0 : \theta = .5$
- In questo caso possiamo fare una operazione molto più raffinata, ovvero individuare più valori possibili di θ ed assegnare a ciascuno di essi una probabilità $p(\theta)$.
- **Sondaggione**

Definizione della *prior*

MENTIMETER

Definizione della *prior*

```
# valori del parametro
> theta <- c( 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 )
```

Definizione della *prior*

```
# valori del parametro
> theta <- c( 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 )
```

```
# probabilità a priori
> (prior <- c(6, 6, 5, 1, 1, 0, 0, 0, 0) / 19)
```

Definizione della *prior*

```
# valori del parametro
> theta <- c( 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 )
```

```
# probabilità a priori
> (prior <- c(6, 6, 5, 1, 1, 0, 0, 0, 0) / 19)
```

```
[1] 0.32 0.32 0.26 0.05 0.05 0.00 0.00 0.00 0.00
```

Definizione della *prior*

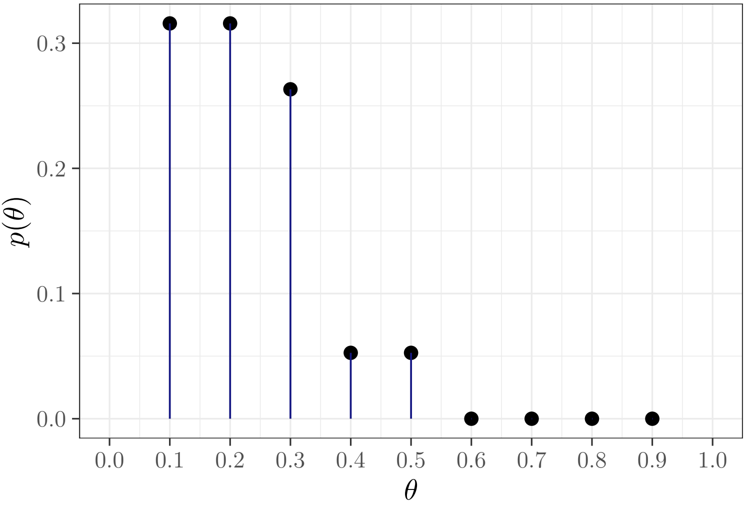
```
# valori del parametro
> theta <- c( 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 )
```

```
# probabilità a priori
> (prior <- c(6, 6, 5, 1, 1, 0, 0, 0, 0) / 19)
```

```
[1] 0.32 0.32 0.26 0.05 0.05 0.00 0.00 0.00 0.00
```

```
> plot( theta, prior, type = "h", lwd = 3,
+         xlim = c( 0, 1 ), ylim = c( 0, 1 ) )
```

Prior



- Abbiamo scelto 9 valori plausibili per θ nella popolazione, a ciascuno dei quali è stata assegnata una probabilità $p(\theta)$.

Definizione della *likelihood*

- A questo punto *prendiamo in esame i dati*, su 46 partecipanti, a 9 piace la statistica, a 37 non piace.
- Sulla base di questo risultato possiamo calcolare i valori di *likelihood* relativi a ciascuno dei 9 valori di θ con la formula seguente:

$$p(D|\theta) = \theta^9(1-\theta)^{37}$$

Definizione della *likelihood*

$$p(D|\theta) = \theta^9(1-\theta)^{37}$$

```
> (L <- theta ^ 9 * ( 1 - theta ) ^ 37)
```

Definizione della *likelihood*

$$p(D|\theta) = \theta^9(1-\theta)^{37}$$

```
> (L <- theta ^ 9 * ( 1 - theta ) ^ 37)
```

```
[1] 2e-11 1.3e-10 3.7e-11 1.6e-12 1.4e-14
[6] 1.4e-14 1.9e-17 1.8e-21 1.8e-27 3.9e-38
```

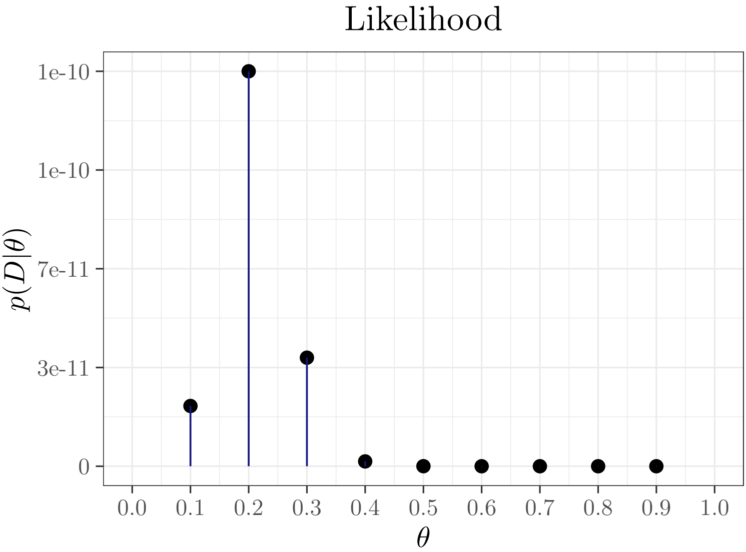
Definizione della *likelihood*

$$p(D|\theta) = \theta^9(1-\theta)^{37}$$

```
> (L <- theta ^ 9 * ( 1 - theta ) ^ 37)
```

```
[1] 2e-11 1.3e-10 3.7e-11 1.6e-12 1.4e-14
[6] 1.4e-14 1.9e-17 1.8e-21 1.8e-27 3.9e-38
```

```
> plot( theta, L, type = "h", lwd = 3, xlim = c( 0, 1 ))
```


Likelihood
1e-10
1e-10
7e-11
3e-11
0
$p(D|\theta)$
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
θ

Calcolo della *posterior*

- Per il calcolo della distribuzione a posteriori utilizziamo infine il teorema di Bayes:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Calcolo della *posterior*

- Per il calcolo della distribuzione a posteriori utilizziamo infine il teorema di Bayes:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

- Nel caso binomiale il calcolo di $p(D)$ è piuttosto semplice ovvero:

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

Calcolo della *posterior*

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

Calcolo della *posterior*

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

```
[1] 5.81e-11
```

Calcolo della *posterior*

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

```
[1] 5.81e-11
```

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

Calcolo della *posterior*

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

```
[1] 5.81e-11
```

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

```
> ( post <- L * prior / pData )
```

Calcolo della *posterior*

$$p(D) = \sum_\theta p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

```
[1] 5.81e-11
```

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

```
> ( post <- L * prior / pData )
```

```
[1] 0.11 0.72 0.17 0.0015 1.3e-05    0    0    0    0
```

Calcolo della *posterior*

$$p(D) = \sum_{\theta} p(D|\theta)p(\theta)$$

```
> ( pData <- sum( L * prior ) )
```

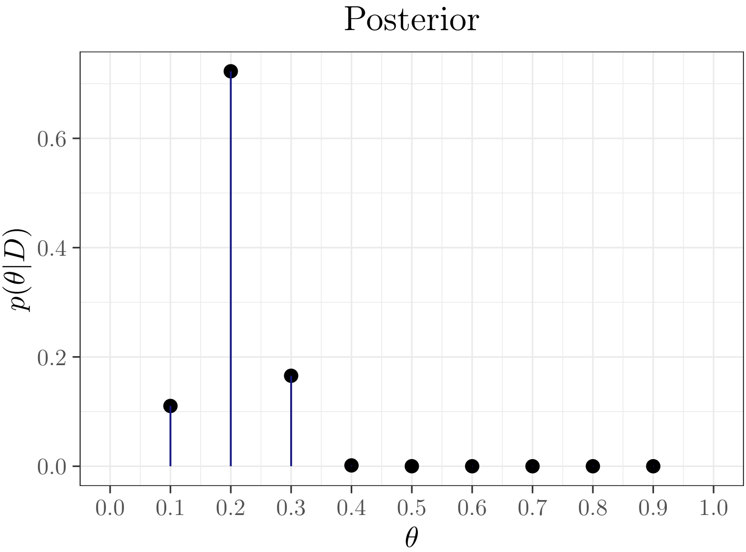
```
[1] 5.81e-11
```

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

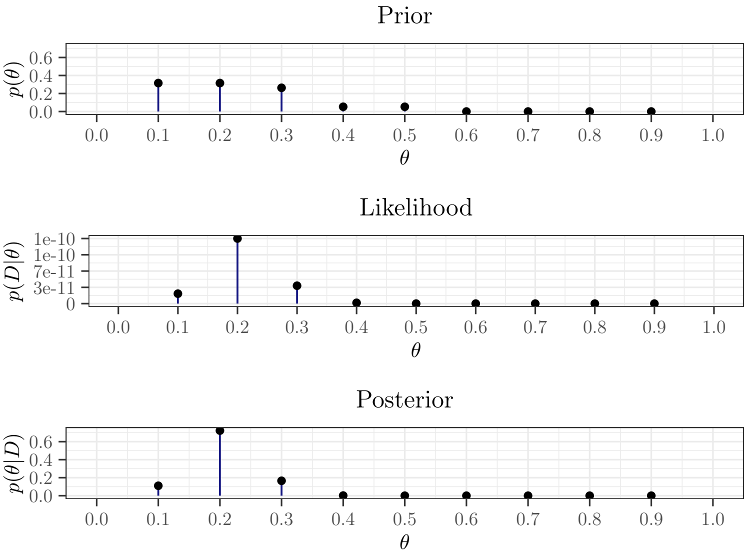
```
> ( post <- L * prior / pData )
```

```
[1] 0.11 0.72 0.17 0.0015 1.3e-05    0    0    0    0
```

```
> plot( theta, post, type = "h", lwd = 3,
+       xlim = c( 0, 1 ), ylim = c( 0, 1 ) )
```


Posterior
$p(\theta|D)$
0.0
0.2
0.4
0.6
0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
θ

Riassumendo



Approcci a confronto

Approccio bayesiano vs NHST

- Una differenza fondamentale tra NHST ed i metodi bayesiani è il modo di trattare i parametri (incogniti). Nel primo caso infatti essi vengono considerati come quantità fisse mentre nel secondo sono trattati come variabili casuali con una distribuzione di probabilità associata.

Approccio bayesiano vs NHST

- Una differenza fondamentale tra NHST ed i metodi bayesiani è il modo di trattare i parametri (incogniti). Nel primo caso infatti essi vengono considerati come quantità fisse mentre nel secondo sono trattati come variabili casuali con una distribuzione di probabilità associata.

Problema 3

Consideriamo un campione di 16 soggetti, di cui 8 maschi, di età intorno ai 17 anni intervistati rispetto alla loro propensione al gioco d'azzardo – calcolata con la frequenza di partecipazione a diversi tipi di gioco (es. scommesse sportive o slot machines) in cui punteggi più alti indicano una maggiore tendenza al gioco d'azzardo. Ci chiediamo se ci sia una differenza di genere su tale propensione.

Dati campionari

I punteggi osservati sono i seguenti:

```
> Ym <- c( 8.8, 19.3, 15.7, 14.6, 8.2, 7.6, 8.7, 8.4 )
```

per i maschi e

```
> Yf <- c( 4.8, 10.6, 11.7, 10.1, 4.8, 8.2, 8.5, 6.8 )
```

per le femmine.

Statistiche del campione

- Soggetti: $n = 16$
- Medie campionarie:

Statistiche del campione

- Soggetti: $n = 16$
- Medie campionarie:

```
> c( mean(Ym), mean(Yf) )
[1] 11.41  8.19
```

- Deviazioni standard:

Statistiche del campione

- Soggetti: $n = 16$
- Medie campionarie:

```
> c( mean(Ym), mean(Yf) )
[1] 11.41  8.19
```

- Deviazioni standard:

```
> c( sd(Ym), sd(Yf) )
[1] 4.45 2.59
```

Problema 3

Sulla base dei dati osservati, possiamo considerare questi soggetti come provenienti da due popolazioni con la stessa media $\mu_m = \mu_f = \mu$?

Problema 3

Sulla base dei dati osservati, possiamo considerare questi soggetti come provenienti da due popolazioni con la stessa media $\mu_m = \mu_f = \mu$?

- In ottica NHST quello che possiamo fare è definire il seguente sistema di ipotesi:

$$H_0 :$$

Problema 3

Sulla base dei dati osservati, possiamo considerare questi soggetti come provenienti da due popolazioni con la stessa media $\mu_m = \mu_f = \mu$?

- In ottica NHST quello che possiamo fare è definire il seguente sistema di ipotesi:

$$\begin{aligned} &H_0 : \mu_m = \mu_f \\ &H_1 : \end{aligned}$$

Problema 3

Sulla base dei dati osservati, possiamo considerare questi soggetti come provenienti da due popolazioni con la stessa media $\mu_m = \mu_f = \mu$?

- In ottica NHST quello che possiamo fare è definire il seguente sistema di ipotesi:

$$\begin{aligned} H_0 &: \mu_m = \mu_f \\ H_1 &: \mu_m > \mu_f \end{aligned}$$

- Sappiamo che la distribuzione campionaria della differenza tra le medie standardizzate sotto H_0

$$\frac{\overline{x}_m - \overline{x}_f}{\sigma_{\overline{x}_m - \overline{x}_f}} \sim$$

- Sappiamo che la distribuzione campionaria della differenza tra le medie standardizzate sotto H_0

$$\frac{\overline{x}_m - \overline{x}_f}{\sigma_{\overline{x}_m - \overline{x}_f}} \sim \text{Student } t_{(n-2)}$$

- Sappiamo che la distribuzione campionaria della differenza tra le medie standardizzate sotto H_0

$$\frac{\overline{x}_m - \overline{x}_f}{\sigma_{\overline{x}_m - \overline{x}_f}} \sim \text{Student } t_{(n-2)}$$

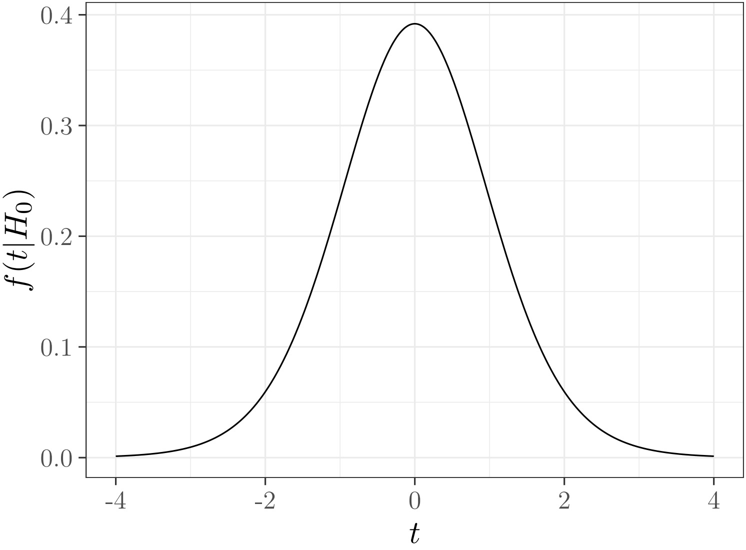
- Nel caso in esame:

$$\frac{11.41 - 8.19}{1.82} = 1.77$$

L'errore standard della differenza tra le medie si calcola con $\sqrt{\frac{\sigma_m^2}{n_m} + \frac{\sigma_f^2}{n_f}}$.

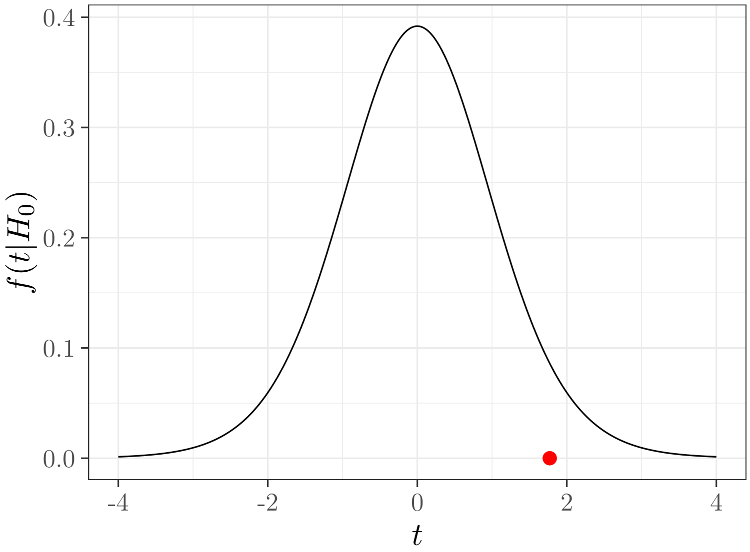
```
> curve( dt( x, 14 ), -4, 4 )
```

```
> curve( dt( x, 14 ), -4, 4 )
```



```
> points( 1.77, 0, pch = 19, col = 'red' )
```

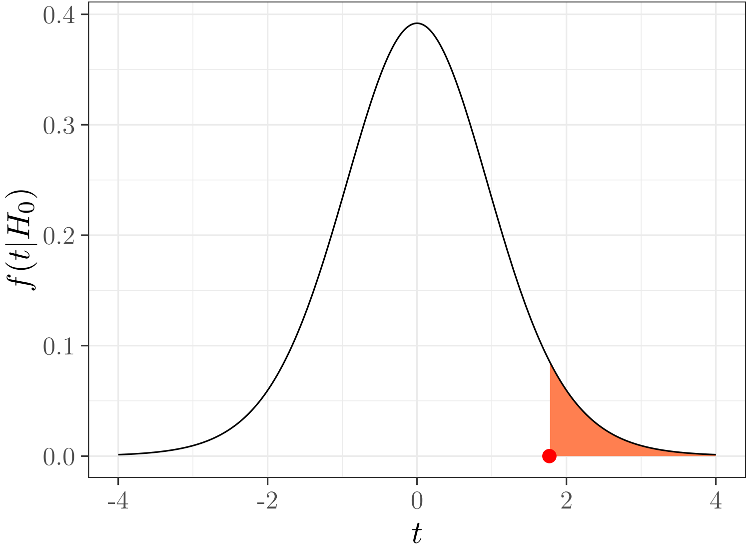
```
> points( 1.77, 0, pch = 19, col = 'red' )
```



```
> pt( 1.77, 14, lower.tail = FALSE )
```

```
> pt( 1.77, 14, lower.tail = FALSE )
[1] 0.0491428
```

```
> pt( 1.77, 14, lower.tail = FALSE )
[1] 0.0491428
```



Questo test statistico è noto come

Questo test statistico è noto come test t di Student.

Questo test statistico è noto come test t di Student.

```
> t.test( Ym, Yf, alternative = "greater",
+                 var.equal = TRUE )
```

Questo test statistico è noto come test t di Student.

```
> t.test( Ym, Yf, alternative = "greater",
+                 var.equal = TRUE )
```

```
     Two Sample t-test

data:  Ym and Yf
t = 1.7712, df = 14, p-value = 0.04914
alternative hypothesis:
     true difference in means is greater than 0
95 percent confidence interval:
 0.01800657        Inf
sample estimates:
mean of x mean of y
  11.4125    8.1875
```

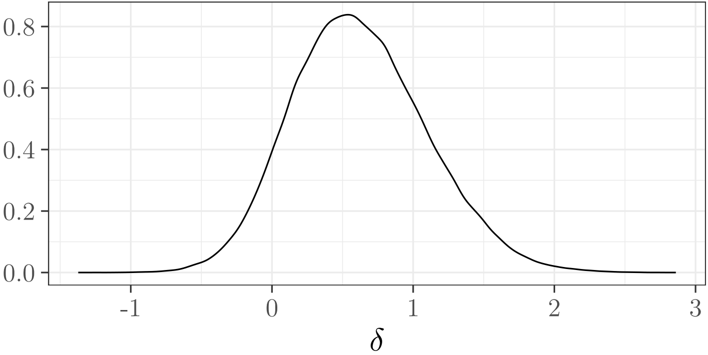
- In ottica bayesiana la differenza $\mu_m - \mu_f = \delta$ è considerata come variabile casuale con una sua distribuzione e, di conseguenza, è possibile stimare una probabilità sia per H_0, che per H_1 e stabilire quale delle due ipotesi risulti più verosimile.
- In altre parole, con un metodo bayesiano possiamo calcolare $Pr(\delta > 0)$, mentre con NHST tale probabilità può essere solo 0 (H_1 è falsa) oppure 1 (H_1 è vera).
- In particolare, utilizzando la distribuzione a posteriori di δ possiamo calcolare la probabilità associata a qualunque ipotesi ci interessi.

Posterior distribution

```
> library( BayesFactor )
> samples <- ttestBF( Ym, Yf, posterior = TRUE,
+                        iterations = 1e5 )
> plot( density( samples[, "delta"] ) )
```

Posterior distribution

```
> library( BayesFactor )
> samples <- ttestBF( Ym, Yf, posterior = TRUE,
+                            iterations = 1e5 )
> plot( density( samples[, "delta"] ) )
```



Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

```
[1] 0.91477
```

Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

```
[1] 0.91477
```

- $Pr(\delta > 2)$

```
> sum( samples[,'delta'] > 2 ) / 100000
```

Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

```
[1] 0.91477
```

- $Pr(\delta > 2)$

```
> sum( samples[,'delta'] > 2 ) / 100000
```

```
[1] 0.00416
```

Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

```
[1] 0.91477
```

- $Pr(\delta > 2)$

```
> sum( samples[,'delta'] > 2 ) / 100000
```

```
[1] 0.00416
```

- $Pr(1 > \delta > 2)$

```
> sum( samples[,'delta'] > 1 &
+   samples[,'delta'] < 2 ) / 100000
```

Sfruttando la distribuzione a posteriori possiamo calcolare le probabilità associate a qualunque ipotesi: ad esempio

- $Pr(\delta > 0)$

```
> sum( samples[,'delta'] > 0 ) / 100000
```

```
[1] 0.91477
```

- $Pr(\delta > 2)$

```
> sum( samples[,'delta'] > 2 ) / 100000
```

```
[1] 0.00416
```

- $Pr(1 > \delta > 2)$

```
> sum( samples[,'delta'] > 1 &
+   samples[,'delta'] < 2 ) / 100000
```

```
[1] 0.20523
```

Used R packages

- `BayesFactor`. Morey R, Rouder J (2024).
- `coda`. Plummer M, Best N, Cowles K, Vines K (2006).
- `geodata`. Hijmans RJ, Barbosa M, Ghosh A, Mandel A (2024).
- `ggplot2`. Wickham H (2016).
- `knitr`. Xie Y (2025).
- `mapIT`. Sommacal NS, Massidda D (2015).
- `Matrix`. Bates D, Maechler M, Jagan M (2025).
- `R`. R Core Team (2025).
- `report`. Makowski D, Lüdecke D, Patil I, Thériault R, Ben-Shachar M, Wiernik B (2023).
- `riskyr`. Neth H, Gaisbauer F, Gradwohl N, Gaissmaier W (2022).
- `sf`. Pebesma E, Bivand R (2023).
- `terra`. Hijmans R (2025).

massimiliano.pastore@unipd.it
https://psicostat.dpss.psy.unipd.it/