Laboratorio di Probabilità e Statistica

lezione 5

Massimo Guerriero – Ettore Benedetti

Indice Lezione

- Prerequisiti dalla lezione scorsa
- Media e varianza campionaria
  - Legge dei grandi numeri
  - Teorema del limite centrale
- Intervallo di confidenza per la media
- Verifica d’ipotesi sulla media

Prerequisiti dalla lezione scorsa

- Variabili casuali discrete e continue
- Scelta corretta della variabile secondo il fenomeno statistico proposto
- Plot della funzione di «densità di probabilità»
- Calcolo della probabilità
  - $P(X \leq x)$
  - $P(X \geq x)$
  - $P(a \leq X \leq b)$

Media e varianza campionaria 1/2

L'**inferenza statistica** è il processo attraverso il quale si traggono conclusioni su un'intera popolazione in base ad un campione.

In molti casi la media μ della popolazione è ignota e occorre stimarla, supponendo sia normalmente distribuita.
Lo stimatore della media, $\bar{X}$, è una variabile casuale con risultati $\overline{x_1}, \overline{x_2}, \ldots, \overline{x_n}$ derivanti da diversi campioni della popolazione.

La media di questi risultati si chiama **media campionaria**.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} \overline{x_i} = \mu$$

Media e varianza campionaria 1/2

La comprensione delle proprietà della distribuzione di $\bar{X}$, ci consente di fare inferenze in base a un singolo campione di dimensione n,
quindi la media campionaria è la media di quel campione.

Se la varianza di una popolazione non è nota, si può stimare con la varianza campionaria che è data dalla formula:

$$\overline{S_n^2} = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{X})^2$$

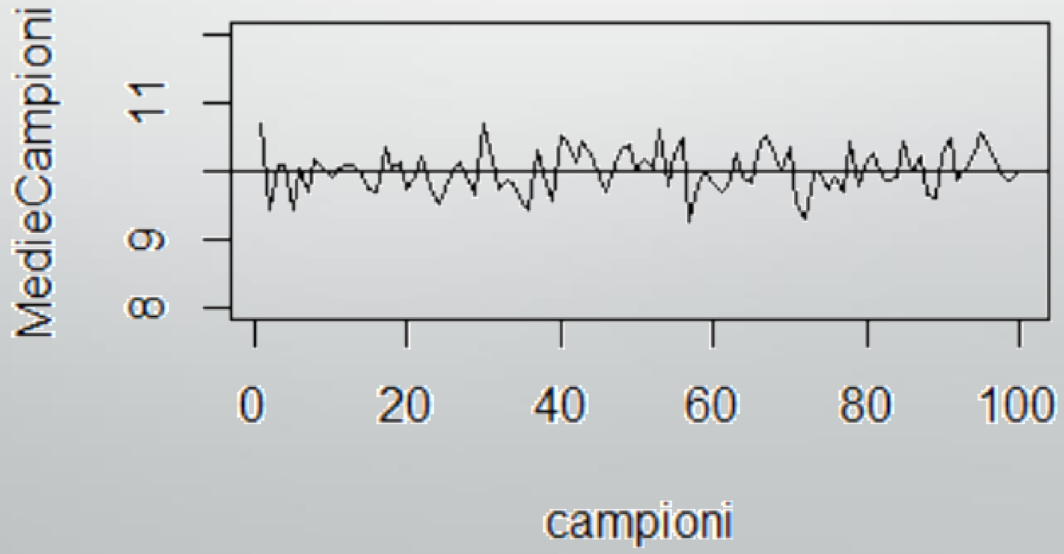
Legge dei grandi numeri

La legge dei grandi numeri afferma che la media che calcoliamo a partire da un numero sufficiente di campioni (almeno 30), è sufficientemente vicina alla media reale della distribuzione da cui prendiamo i campioni.

Es.

100 campioni formati da 40 elementi presi da una normale con media 10 e deviazione standard 2.

(Codice Esempio)



mean(MedieCampioni);
[1] 10.00277

Teorema del limite centrale

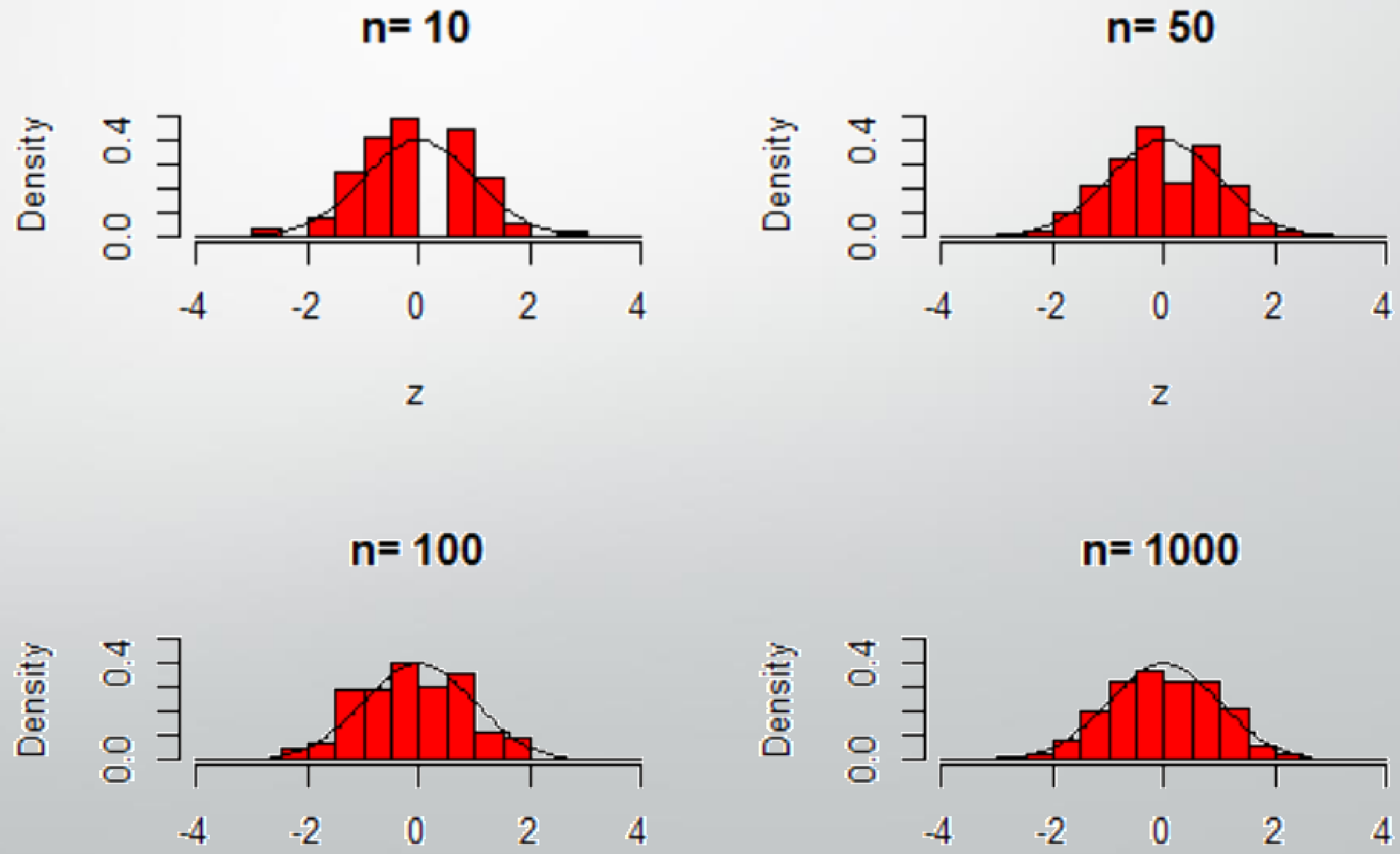
Secondo il teorema del limite centrale, in presenza di grandi campioni, qualsiasi distribuzione probabilistica (continua o discreta) si distribuisce come la normale standard.

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$$

Es.

Esperimento su 500 campioni con diversa numerosità presi dalla distribuzione di bernoulli con p=0.5

[Codice Esempio]



Consegna

1. Estrarre un campione di 100 elementi da una distribuzione normale con media 15 e varianza 3.14, calcolare media e varianza campionaria.
2. Verificare il teorema del limite centrale per un’altra distribuzione di probabilità (discreta o continua).
3. Verificare la legge dei grandi numeri per un’altra distribuzione di probabilità (discreta o continua).

Indice Lezione

- Prerequisiti dalla lezione scorsa
- Media e varianza campionaria
  - Legge dei grandi numeri
  - Teorema del limite centrale
- Intervallo di confidenza per la media
- Verifica d'ipotesi sulla media

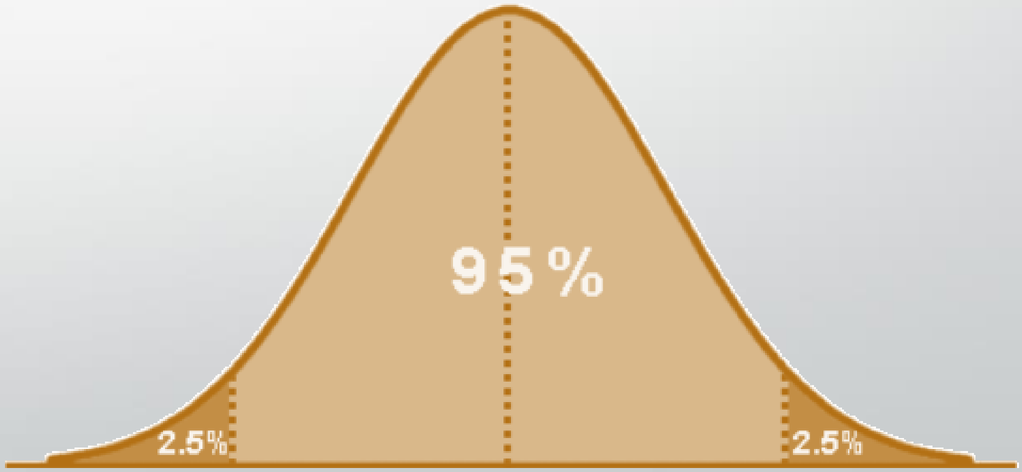
Intervallo di confidenza per la media 1/2

Gli intervalli di confidenza per la media forniscono un campo di variazione centrato sulla media campionaria all'interno del quale ci si aspetta di trovare il parametro incognito μ.

Per la variabile casuale normale standard Z= $\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}}$ con probabilità 1-α:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(\overline{X_n} - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < \mu < \overline{X_n} + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right)$$



Intervallo di confidenza per la media 2/2

Nella maggior parte dei casi la varianza iniziale non è conosciuta, si deve stimare. La distribuzione probabilistica utilizzata è la t di Student

$$\mu \in \left(\overline{X_n} \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \sqrt{\frac{S_n^2}{n}} \right)$$

In R si possono calcolare gli intervalli di confidenza manualmente (creandosi una funzione), oppure utilizzando il comando

t.test*(campione, conf.lev* = *α)*

```
> x<-(c(0.39,0.68,0.82,1.35,1.38,1.62,
1.70,1.71,1.85,2.14,2.89,3.69))
> t.test(x, conf.lev=0.99)

 One Sample t-test

data:  x
t = 6.3305, df = 11, p-value = 5.595e-05
alternative hypothesis: true mean is not equal to 0
99 percent confidence interval:
 0.8583201 2.5116799
sample estimates:
mean of x
    1.685
```

Verifica d'ipotesi sulla media 1/6

Vogliamo rispondere a questa domanda:

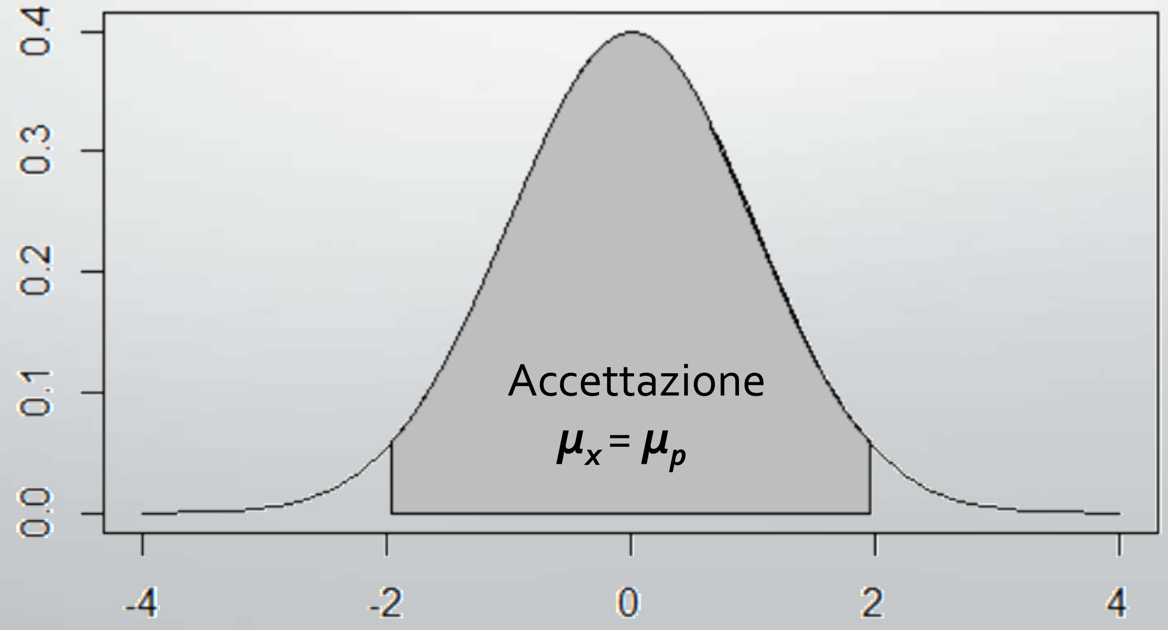
- *disponendo di un campione di numerosità limitata, si può affermare che la media μ_x della popolazione da cui esso è stato estratto è diversa dal valore prestabilito μ_p?*

Per condurre il test si devono effettuare i seguenti tre passi:

1. Si fissa il "**tasso accettabile di rischio**" α. *Es.* $\alpha = 0{,}05$.
2. Si estrae il campione dalla popolazione e si determina la sua **media campionaria**
3. Si individua l'**intervallo di confidenza** ad 1- α mediante la variabile z (normale standard) se la varianza della popolazione è nota oppure mediante la t di Student se la varianza della popolazione è incognita.

Verifica d'ipotesi sulla media 2/6

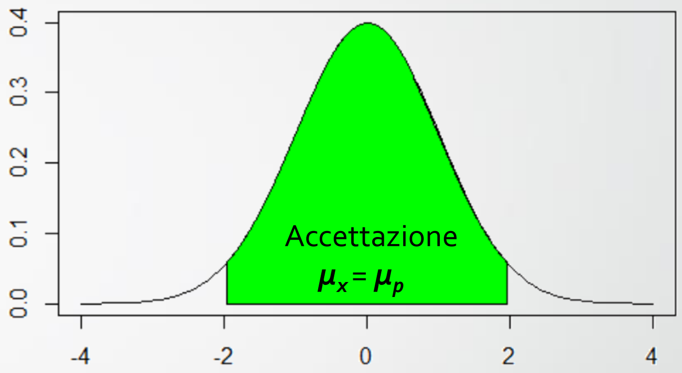
- Se ***z*** (oppure ***t***) è compreso nell'intervallo di confidenza trovato NON si può affermare che $\boldsymbol{\mu_x}$ è diverso dal valore prestabilito $\boldsymbol{\mu_p}$.
- Se invece ***z*** (oppure ***t***) NON è compreso nell'intervallo di confidenza trovato allora SI PUO' affermare, con una probabilità di errare non superiore ad ***α***, che $\boldsymbol{\mu_x}$ è diverso dal valore prestabilito $\boldsymbol{\mu_p}$.



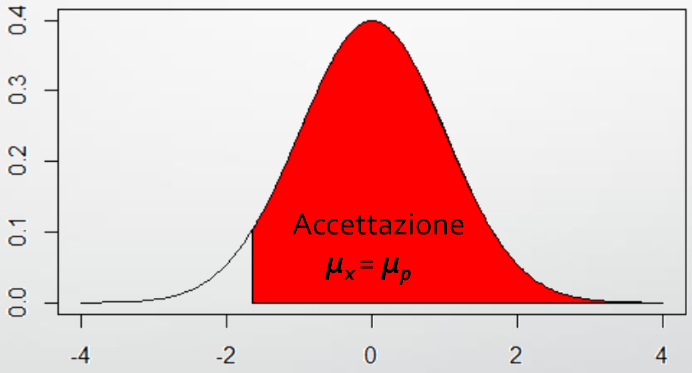
Codice Grafico

Verifica d'ipotesi sulla media 3/6

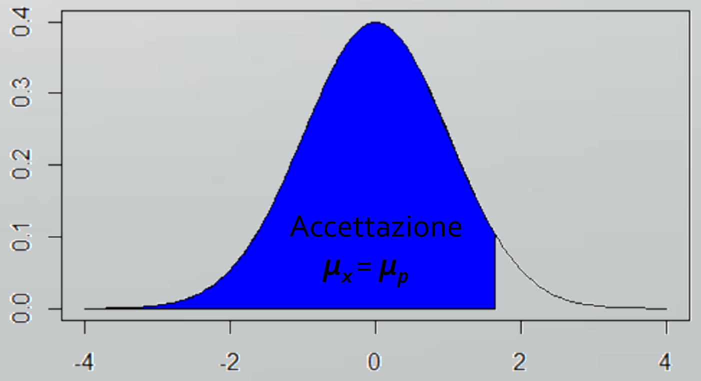
- *disponendo di un campione di numerosità limitata, si può affermare che la media μ_x della popolazione da cui esso è stato estratto è diversa dal valore prestabilito μ_p?*



- *è inferiore al valore prestabilito μ_p?*



- *è superiore al valore prestabilito μ_p?*



Verifica d’ipotesi sulla media 4/6

Es. *Supponiamo di aver rilevato il tempo medio di vita di un campione di 15 lampadine:*

```
x <- c(2928, 2997, 2689, 3081, 3011, 2996, 2962, 3007, 3000, 2953, 2792, 2947, 3094, 2913, 3017)
```

Per poter vendere queste lampadine occorre indicare sulla scatola il tempo medio di vita con un errore dell’1% se le vendo in Italia, 5% se le vendo all’estero.
La ditta di lampadine vende sia in Italia che all’estero e produce un solo tipo di scatola in cui è indicato 3010h come tempo medio di vita.
Verifichiamo se siamo confidenti nell’affermare che $\mu x = \mu p = 3010$*, ovvero che la ditta è in regola secondo le norme internazionali e locali.*

```
mean(x)
[1]  2959.133
```

Come si può notare la media campionaria è inferiore a quella dichiarata ($\mu x < \mu p$)*.*
Vediamo se questa anomalia è dovuta all’inferenza o se l’azienda non rispetta le norme.

Verifica d'ipotesi sulla media 5/6

Per fare questo dobbiamo effettuare un test t-di student ad una coda con la regione di rifiuto a sinistra (μx < μp):
Utilizziamo il comando: ***t.test(****x, mu=3010,* ***alternative="less")***

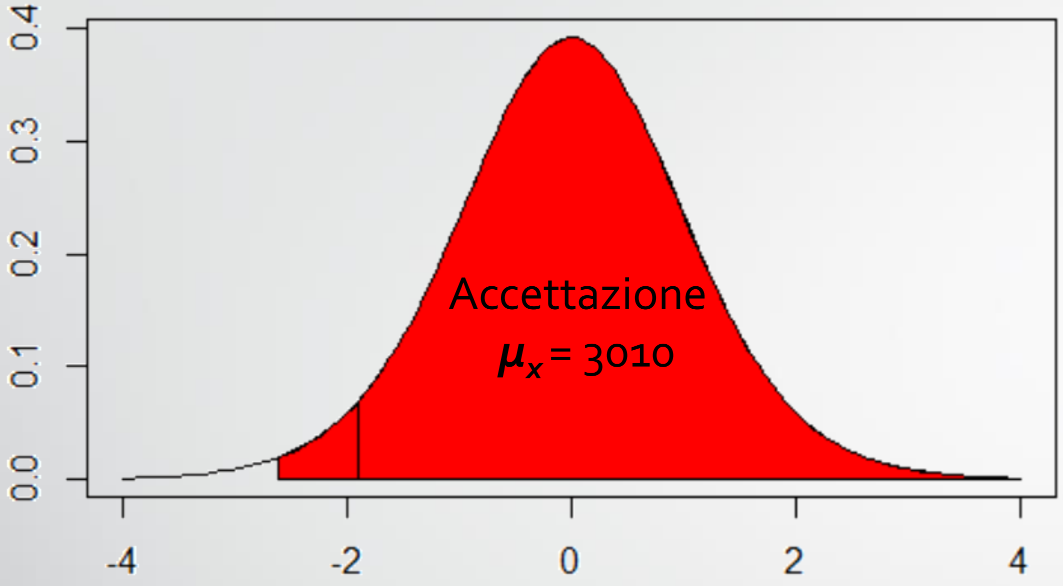
```
> t.test(x,mu=3010, alternative="less")

 One Sample t-test

data:  x
t = -1.9031, df = 14, p-value = 0.0389
alternative hypothesis: true mean is less than 3010
95 percent confidence interval:
     -Inf 3006.211
sample estimates:
mean of x
 2959.133
```

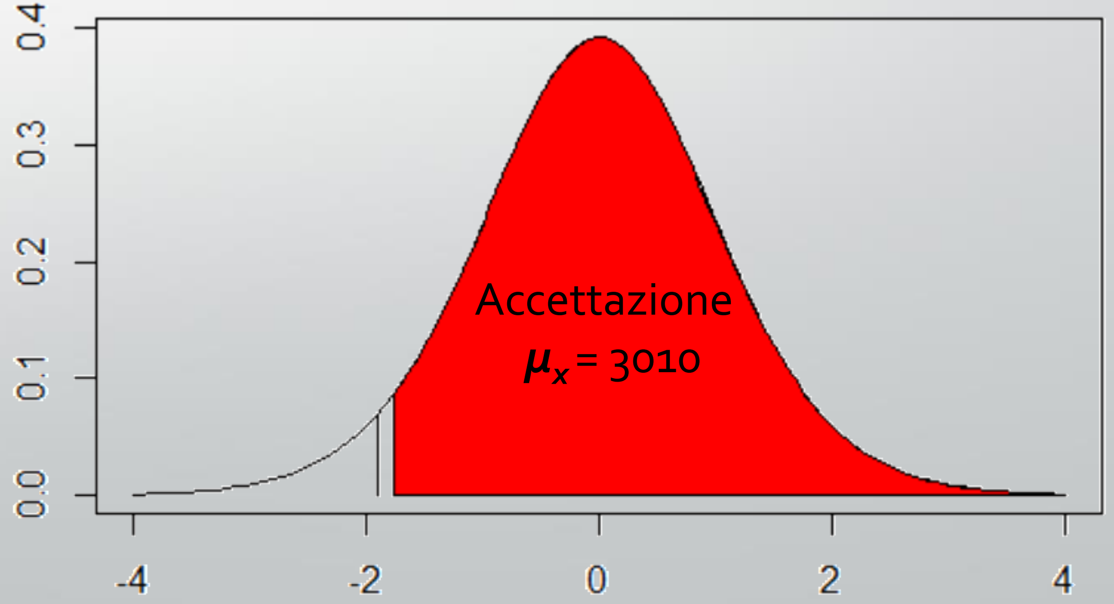
Fissato α=0.01 notiamo che α<p-value per cui l'azienda può vendere queste lampadine in Italia scrivendo 3010 come tempo medio di vita. Ma se fissiamo α=0.05 notiamo che α>p-value per cui l'Azienda non è conforme alle norme estere.

Verifica d'ipotesi sulla media 6/6



Fissato α=0.01 notiamo che α<p-value per cui l'azienda può vendere queste lampadine in Italia

Ma se fissiamo α=0.05 notiamo che α>p-value per cui l'Azienda non è conforme alle norme estere.



[Codice Grafici]

Consegna

1. Creare una funzione per calcolare gli intervalli di confidenza per la normale standard, accettando come parametri alfa e un numero intero da 0 a 2 a seconda che il test sia a due code o a una coda (destra e sinistra).

2. Una fabbrica di funi per arrampicata sportiva ha ottenuto i seguenti risultati espressi in Newton in 25 prove di rottura, per un nuovo tipo di funi:

   1975, 1869, 1879, 1790, 1860, 1895, 1810, 1831, 1759, 1585, 1553, 1774, 1640
   1761, 1946, 1915, 1894, 1971, 1876, 1716, 1652, 1591, 1700, 1842, 1781

   Sapendo che le funi tradizionali hanno una resistenza di rottura pari a 1730N, ci si chiede se il nuovo tipo abbia significativamente migliorato la qualità delle funi con una fiducia del 95%.