STATISTICA E ANALISI DEI DATI

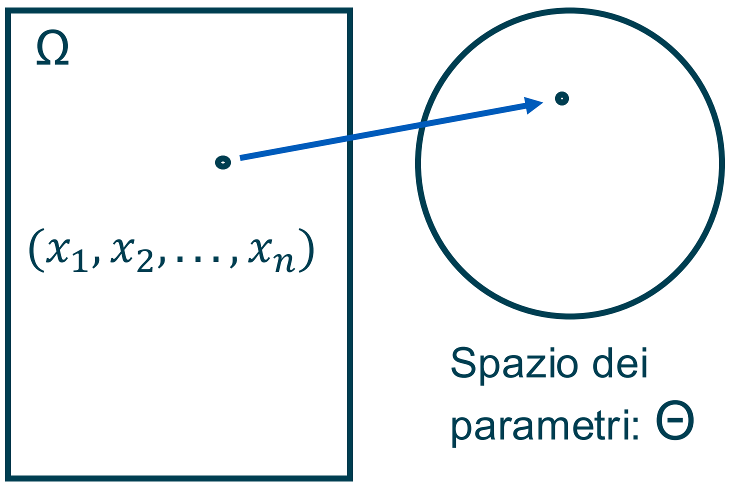
Capitolo 11 – Intervalli di Confidenza

Dott. Stefano Cirillo
Dott. Luigi Di Biasi

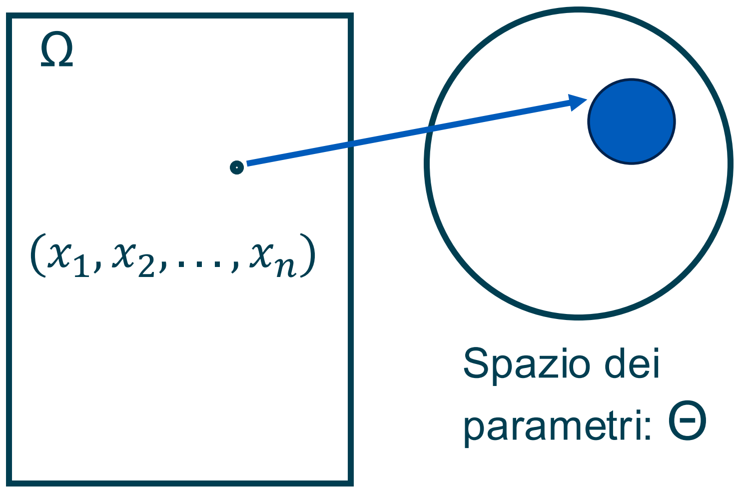
a.a. 2025-2026

Campione casuale della Popolazione

- Il **campione** selezionato deve essere rappresentativo della popolazione
- L'inferenza statistica si basa su due metodi fondamentali di indagine:
  - **Stima dei parametri**: sulla base del campione si assegna al parametro di interesse (ϑ) un valore (**stima puntuale**) o un insieme di valori (**stima per intervallo**)
- Ovviamente quanto più sarà grande il campione, più esso sarà rappresentativo della popolazione
- Ciò significa che le stime fatte saranno sempre più rappresentative della totalità della popolazione



Stima puntuale



Stima per intervallo

Stima Statistica

- Alla stima puntuale di un parametro non noto di una popolazione (costituita da un singolo valore reale) spesso si preferisce sostituire un intervallo di valori, detto **intervallo di confidenza** o intervallo di fiducia
  - si cerca di determinare in base ai dati del campione
    - Un **limite inferiore ed uno superiore** entro i quali sia compreso il parametro non noto con un certo **coefficiente di confidenza** (detto anche grado di fiducia)

Stima Statistica

- Alla stima puntuale di un parametro non noto di una popolazione (costituita da un singolo valore reale) spesso si preferisce sostituire un intervallo di valori, detto **intervallo di confidenza** o intervallo di fiducia
  - si cerca di determinare in base ai dati del campione
    - Un **limite inferiore ed uno superiore** entro i quali sia compreso il parametro non noto con un certo **coefficiente di confidenza** (detto anche grado di fiducia)
- Formalmente:
  - Sia $X_1, X_2, \dots, X_n$ un **campione casuale** di ampiezza n estratto da una popolazione con funzione di probabilità (nel caso discreto) oppure densità di probabilità (nel caso continuo) $f(x, \vartheta)$ dove ϑ denota il parametro non noto della popolazione
  - Denotiamo con $C_n = g_1(X_1, X_2, \dots, X_n)$ e con $\overline{C_n} = g_2(X_1, X_2, \dots, X_n)$ due statistiche che soddisfino la condizione $C_n < \overline{C_n}$ cioè che godono della proprietà che per ogni possibile campione osservato $x = (x_1, x_2, \dots, x_n)$ risulti che $g_1(x) < g_2(x)$

Intervallo di Confidenza

- Un **intervallo di confidenza** è un intervallo di valori calcolato a partire da un campione statistico che serve a stimare un parametro **non noto** della popolazione (ad esempio, la media o λ).
- L'intervallo è costruito in modo che, con una certa probabilità (detta **livello di confidenza**, tipicamente 90%, 95% o 99%), il parametro vero della popolazione si trovi all'interno di esso
- Fissato un coefficiente di confidenza $1-\alpha$ $(0<\alpha<1)$ se è possibile scegliere le statistiche C_n e $\overline{C_n}$ in modo tale che:

$$P\left(C_n < \vartheta < \overline{C_n}\right) = \boxed{1-\alpha} \longleftarrow$$

La probabilità che ϑ sia compresa nel range C_n e $\overline{C_n}$

Allora si dice che $(C_n, \overline{C_n})$ è un intervallo di confidenza o fiducia di grado $1-\alpha$ per ϑ

- Le statistiche:
  - C_n **limite inferiore**
  - $\overline{C_n}$ **limite superiore**

dell'intervallo di confidenza

Metodo Pivotale

- Un metodo per la costruzione degli intervalli di confidenza è il **metodo pivotale**
- Il metodo pivotale afferma che:
  - Quando si desidera effettuare inferenze su un parametro di interesse in una popolazione, è possibile utilizzare una **variabile pivot** o **quantità pivotale** che segue una distribuzione di probabilità **nota**, indipendentemente dai parametri della popolazione
  - Questo consente di costruire intervalli di confidenza senza dover conoscere la distribuzione esatta dei dati campionari
  - In termini più specifici, **il metodo pivotale consente di formulare un'equazione in cui una quantità basata sui dati campionari e il parametro di interesse è espressa come funzione di una variabile pivot**

Metodo Pivotale

- Il metodo pivotale consiste nel determinare una variabile aleatoria di pivot

$$\gamma(X_1, X_2, \dots, X_n; \vartheta)$$

che:

- Dipende dal campione casuale $X_1, X_2, \dots, X_n$
  - Questo significa che γ è una funzione delle osservazioni raccolte, cioè i **dati**
- Dipende dal parametro non noto ϑ
  - La variabile pivotale coinvolge il parametro che stiamo cercando di stimare (ad esempio, media, varianza, ecc.)
- La sua funzione di distribuzione non contiene il parametro ϑ da stimare
  - La funzione di distribuzione di γ, una volta calcolata, è indipendente dal parametro ϑ

- **Nota**: La variabile aleatoria di pivot non è una statistica poiché dipende dal parametro non noto ϑ e quindi non è osservabile

Metodo Pivotale

- Il metodo pivotale afferma che:
  - Per ogni coefficiente fissato α $(0 < \alpha < 1)$ con un intervallo di confidenza al $100 \cdot (1-\alpha)\%$, siano α_1 e α_2 e $(\alpha_1 < \alpha_2)$ i **limiti nella distribuzione pivotale**, cioè due valori dipendenti dal coefficiente fissato α tali che per ogni $\vartheta \in \Theta$ si ha:

$$P(\alpha_1 < \gamma(X_1, X_2, \dots, X_n; \vartheta) < \alpha_2) = 1 - \alpha$$

Metodo Pivotale

- Il metodo pivotale afferma che:
  - Per ogni coefficiente fissato α $(0 < \alpha < 1)$ con un intervallo di confidenza al $100 \cdot (1 - \alpha)\%$, siano α_1 e α_2 e $(\alpha_1 < \alpha_2)$ i **limiti nella distribuzione pivotale**, cioè due valori dipendenti dal coefficiente fissato α tali che per ogni $\vartheta \in \Theta$ si ha:
  $$P(\alpha_1 < \gamma(X_1, X_2, \dots, X_n; \vartheta) < \alpha_2) = 1 - \alpha$$
  - Se per ogni possibile campione osservato $x = (X_1, X_2, \dots, X_n)$ e per ogni $\vartheta \in \Theta$ si dimostra che:
  $$\alpha_1 < \gamma(x; \vartheta) < \alpha_2 \quad se\ e\ solo\ se \quad g_1(x) < \vartheta < g_2(x)$$
  con $g_1(x)$ e $g_2(x)$ dipendenti solo dal campione osservato, allora si ha che:
  $$P(\alpha_1 < \gamma(X_1, X_2, \dots, X_n; \vartheta) < \alpha_2)$$
  È equivalente a:
  $$P\big(g_1(X_1, X_2, \dots, X_n) < \vartheta < g_2(X_1, X_2, \dots, X_n)\big) = 1 - \alpha$$
- Se consideriamo $C_n = g_1(X_1, X_2, \dots, X_n)$ e $\overline{C_n} = g_2(X_1, X_2, \dots, X_n)$ segue che $(C_n, \overline{C_n})$ è un intervallo di confidenza di grado $1 - \alpha$ per il parametro non noto ϑ della popolazione

STATISTICA E ANALISI DEI DATI

Popolazione Normale

Stimatore e Stima dei parametri

- Considerato $X_1, X_2, \dots, X_n$ un campione casuale di ampiezza n estratto da una popolazione normale descritta da una variabile aleatoria $X = N(\mu, \sigma^2)$ con valore medio μ e varianza σ^2
- Vogliamo determinare un intervallo di confidenza per la media μ di grado $1 - \alpha$
  - La media del campione casuale estratto (**media campionaria**) $\overline{X}$ è uno stimatore **non distorto** di μ (media di tutta la popolazione)
    - Significa che la sua media corrisponde alla media della popolazione: $E(\overline{X}) = \mu$
  - La varianza del campione casuale estratto (**varianza campionaria**) $Var(X)$ è uno stimatore **non distorto** di σ (varianza di tutta la popolazione): $Var(X) = S^2 = \sigma^2$

Stimatore e Stima dei parametri

- Considerato $X_1, X_2, \dots, X_n$ un campione casuale di ampiezza n estratto da una popolazione normale descritta da una variabile aleatoria $X = N(\mu, \sigma^2)$ con valore medio μ e varianza σ^2
- Vogliamo determinare un intervallo di confidenza per la media μ di grado $1 - \alpha$
  - La media del campione casuale estratto (**media campionaria**) $\overline{X}$ è uno stimatore **non distorto** di μ (media di tutta la popolazione)
    - Significa che la sua media corrisponde alla media della popolazione: $E(\overline{X}) = \mu$
  - La varianza del campione casuale estratto (**varianza campionaria**) $Var(X)$ è uno stimatore **non distorto** di σ (varianza di tutta la popolazione): $Var(X) = S^2 = \sigma^2$
  - Vogliamo determinare un **intervallo** di confidenza di grado $1 - \alpha$:
    1. Per la media μ nel caso la varianza σ^2 è nota
    2. Per la media μ nel caso la varianza σ^2 NON è nota
    3. Per la varianza σ^2 nel caso la media μ è nota
    4. Per la varianza σ^2 nel caso la media μ NON è nota

STATISTICA E ANALISI DEI DATI

Popolazione Normale - I.C. per il valore medio μ con varianza σ^2 nota

I.C. per il valore medio μ con varianza σ^2 nota

- Sappiamo che il campione di ampiezza n è estratto da una popolazione distribuita come una normale $X = N(\mu, \sigma^2)$ di media μ e varianza σ^2
- La media del campione casuale estratto (media campionaria) $\overline{X}$ è uno stimatore **non distorto** di μ
  - Significa che la sua media corrisponde alla media della popolazione:

$$E(\overline{X}) = \mu$$

  - Inoltre sappiamo che la media campionaria $\overline{X}$ si distribuisce come una variabile aleatoria normale:

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

- La **varianza della media campionaria** $\overline{X}$ viene calcolata come $\frac{\sigma^2}{n}$ perché rappresenta la variabilità della media campionaria in base alla dimensione del campione n

I.C. per il valore medio μ con varianza σ^2 nota

- A partire da questo, possiamo costruire una nuova variabile aleatoria:

$$Z_n = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

$\overline{X}$ = media campionaria, μ = media della popolazione,

σ = deviazione standard della popolazione, *n* = dimensione del campione

- La cui distribuzione $N(0,1)$ non dipende più da alcun parametro
- Z_n è una quantità pivotale, cioè una variabile aleatoria di cui la distribuzione è nota (Normale Standard) ma non dipende dal parametro μ che si vuole stimare
  - La variabile Z_n ci servirà per stimare μ

I.C. per il valore medio μ con varianza σ^2 nota

- A partire da questo, <u>possiamo costruire una nuova variabile aleatoria</u>:

$$\boxed{Z_n = \frac{\overline{X} - \boxed{\mu}}{\frac{\sigma}{\sqrt{n}}}} \sim N(0,1)$$

Quantità Pivotale

Parametro da stimare

$\overline{X}$ = media campionaria, μ = media della popolazione,

σ = deviazione standard della popolazione, n = dimensione del campione

- La cui distribuzione $N(0,1)$ non dipende più da alcun parametro
- Z_n è una quantità pivotale, cioè una variabile aleatoria di cui la distribuzione è nota (Normale Standard) ma non dipende dal parametro μ che si vuole stimare
  - <u>La variabile Z_n ci servirà per stimare μ</u>

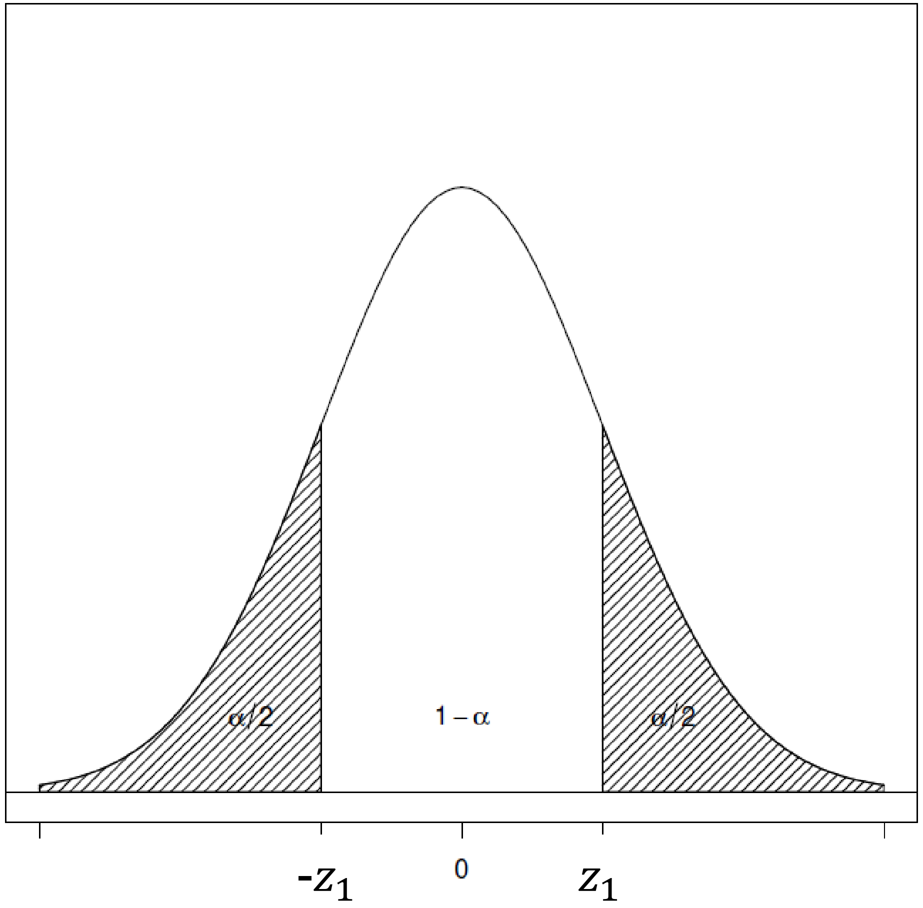
I.C. per il valore medio μ con varianza σ^2 nota

- Vogliamo calcolare la probabilità che μ sia in un range a e b con una certa probabilità $1-\alpha$

$$P(a < \mu < b) = 1 - \alpha$$

- Pertanto usiamo la quantità Pivotale:
  - Calcoliamo la probabilità che la quantità pivotale sia compresa tra due **quantili** $-z_1$ e z_1:

$$P\left(-z_1 < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 1 - \alpha$$



I.C. per il valore medio μ con varianza σ^2 nota

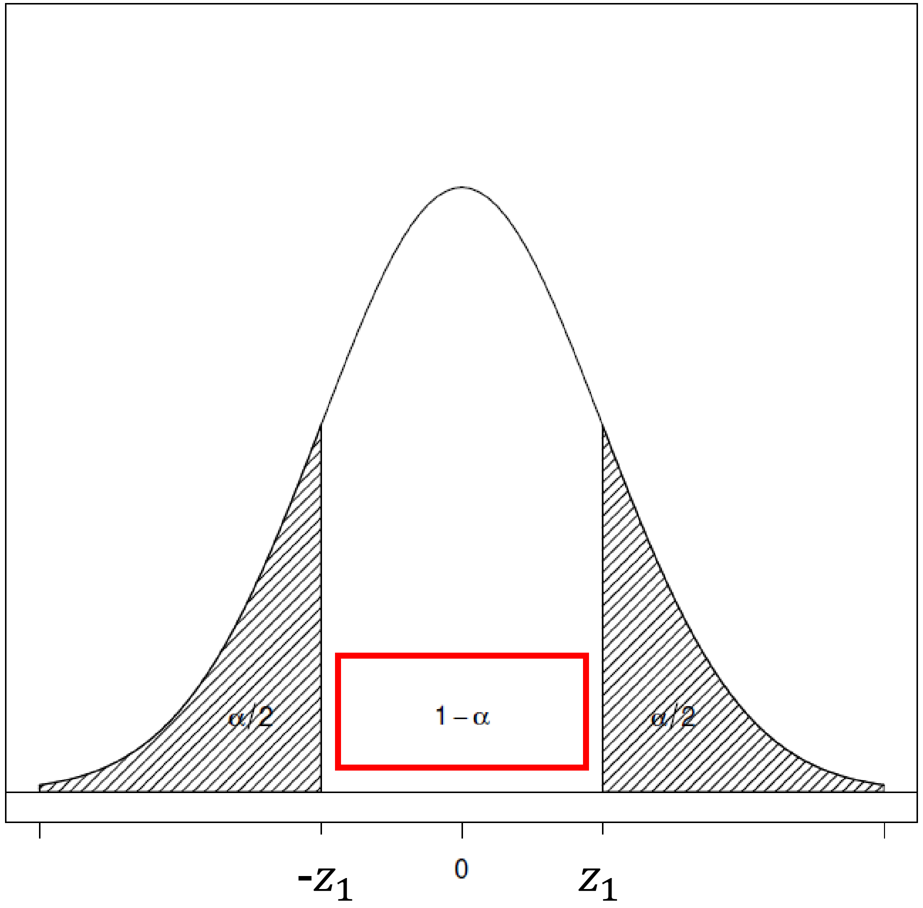
- Vogliamo calcolare la probabilità che μ sia in un range a e b con una certa probabilità $1 - \alpha$

$$P(a < \mu < b) = 1 - \alpha$$

- Pertanto usiamo la quantità Pivotale:
  - Calcoliamo la probabilità che la quantità pivotale sia compresa tra due **quantili** $-z_1$ e z_1:

$$P\left(-z_1 < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 1 - \alpha$$

- Poiché $Z_n = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ è una normale standard, si ha una distribuzione simmetrica
  - Dove la probabilità che Z_n sia compresa tra i quantili $-z_1$ e z_1 è $1 - \alpha$
    - Dove $1 - \alpha$ è l'area sottesa alla curva tra $-z_1$ e z_1
    - Poiché l'area sottesa in tutta la curva è 1, l'area sottesa sotto le due code sarà $\frac{\alpha}{2}$

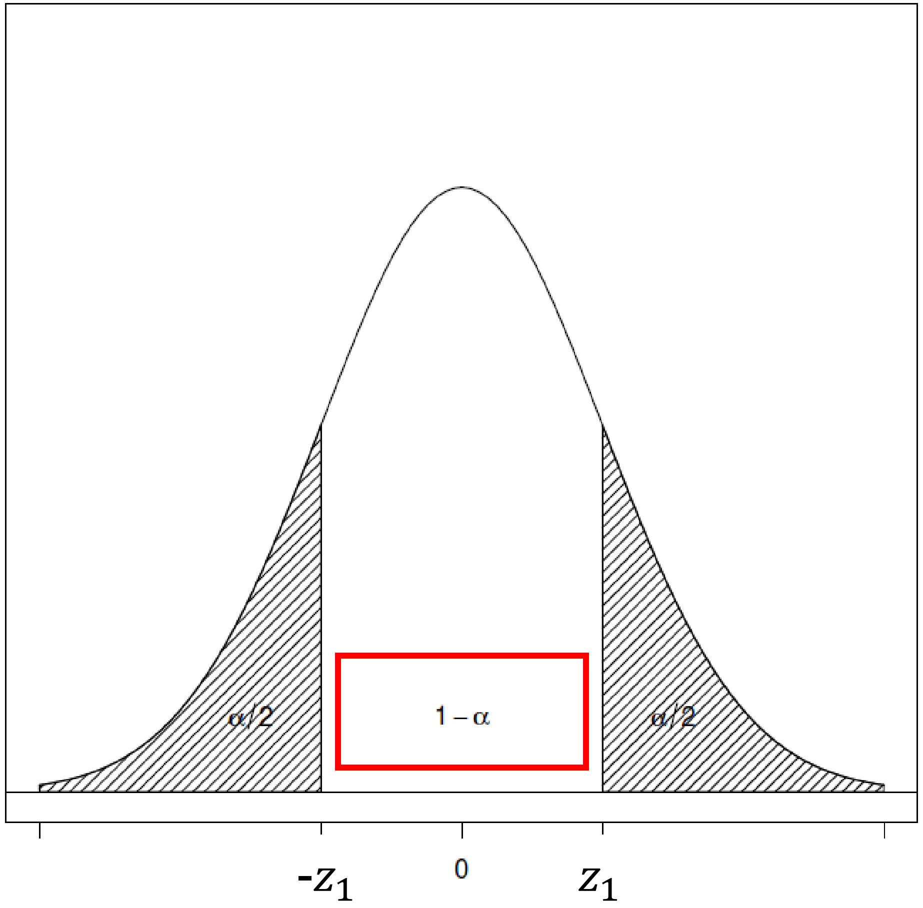


I.C. per il valore medio μ con varianza σ^2 nota

- A partire da:

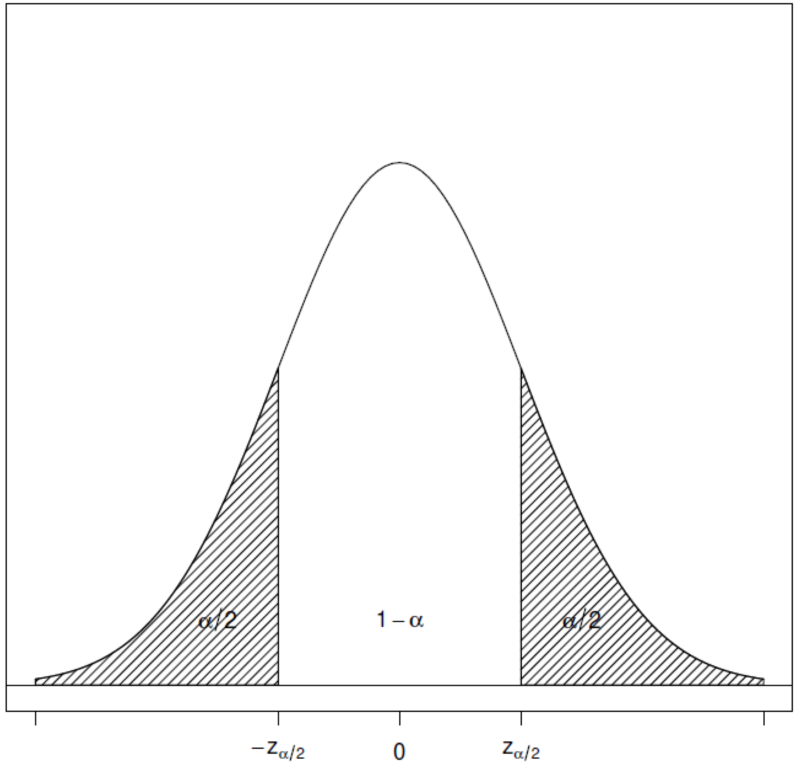
$$P\left(-z_1 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 1-\alpha$$

- Possiamo calcolare i quantili $-z_1$ e z_1
- Dopodichè isoliamo il parametro incognito μ ed otteniamo l'intervallo di confidenza di grado $1-\alpha$



Esempio

- **Esempio**:
  - È noto che la deviazione standard della vita media di una marca di lampadine equivale a 100 ore
  - Un campione casuale di 169 lampadine ha una vita media di 1350 ore.
  - Determina un intervallo di confidenza della vita media di queste lampadine al 90%

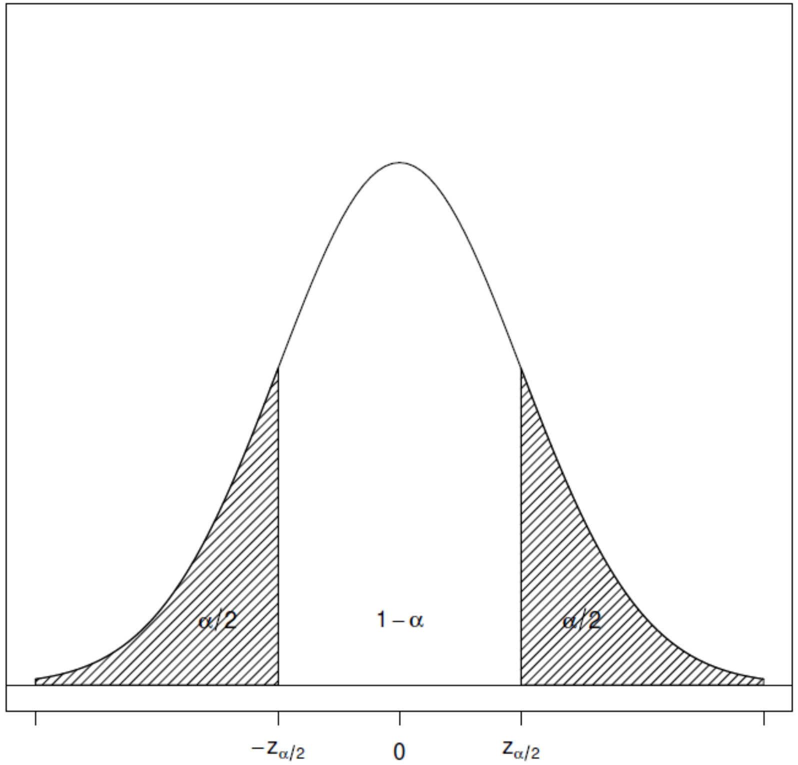


Esempio

- **Esempio**:
  - È noto che la deviazione standard della vita media di una marca di lampadine equivale a 100 ore
  - Un campione casuale di 169 lampadine ha una vita media di 1350 ore.
  - Determina un intervallo di confidenza della vita media di queste lampadine al 90%

Conosciamo che:

- Deviazione Standard: $\sigma = \sqrt{\sigma^2} = 100$ ore
- $n = 169$
- Vita media delle 169 lampadine $\overline{X} = 1350$



Esempio

- **Esempio**:
  - È noto che la deviazione standard della vita media di una marca di lampadine equivale a 100 ore
  - Un campione casuale di 169 lampadine ha una vita media di 1350 ore.
  - Determina un intervallo di confidenza della vita media di queste lampadine al 90%

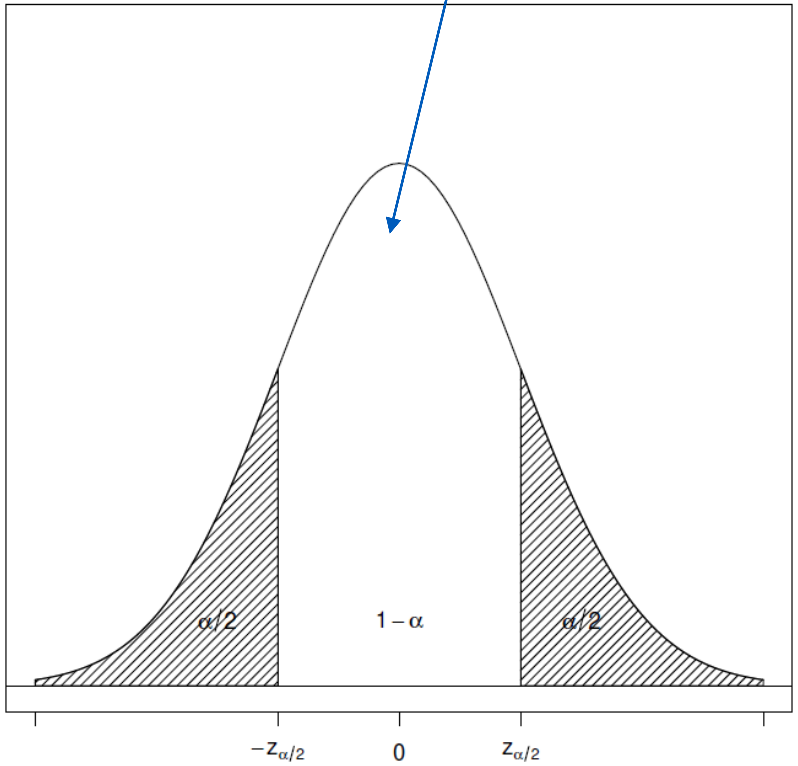
Conosciamo che:

  - Deviazione Standard: $\sigma = \sqrt{\sigma^2} = 100$ ore
  - $n = 169$
  - Vita media delle 169 lampadine $\overline{X} = 1350$

- Vogliamo stimare μ (intervallo di conf. Della vita media) al $90\% = 1 - \alpha$
  - La quantità pivotale è: $Q = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ ed è una variabile aleatoria **normale standard**
  - Vogliamo la probabilità che $P\left(-z_1 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 0.9$

- Trovare i percentili $-z_1$ e $z_1$tale che $1 - \alpha = 0.9$

L'area deve essere uguale a 0.9



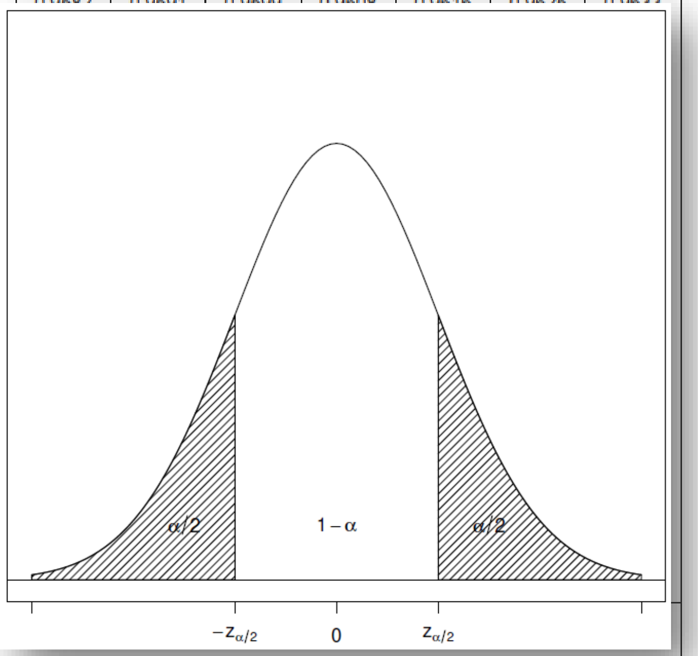
Esempio

Le tavole ci danno il valore del percentile sulla base dell'area sottesa fino alla percentuale richiesta

- **Esempio**:
- Vogliamo stimare μ al $90\% = 1 - \alpha$
  - La quantità pivotale è: $Q = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$
  - Vogliamo la probabilità che $P\left(-z_1 < \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 0.9$
- Trovare i percentili $-z_1$ e $z_1$tale che l'area di $1 - \alpha$ sia 0.9
  - Poiché $1 - \alpha = 0.9 \Rightarrow \alpha = 1 - 0.9 \Rightarrow \alpha = 0.1$
    - Si ha che $\frac{\alpha}{2} = 0.05$
  - Con la tavola della normale standard si ha che:

Il quantile $z_1 = \varphi\left(1 - \alpha + \frac{\alpha}{2}\right) = \varphi(0.9 + 0.05)$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.8	0.9641	0.9649	0.9656							
1.9	0.9713	0.9719	0.9726							
2.0	0.9772	0.9778	0.9783							
2.1	0.9821	0.9826	0.9830							
2.2	0.9861	0.9864	0.9868							
2.3	0.9893	0.9896	0.9898							
2.4	0.9918	0.9920	0.9922							
2.5	0.9938	0.9940	0.9941							
2.6	0.9953	0.9955	0.9956							
2.7	0.9965	0.9966	0.9967							
2.8	0.9974	0.9975	0.9976							
2.9	0.9981	0.9982	0.9982							
3.0	0.9987	0.9987	0.9987							
3.1	0.9990	0.9991	0.9991							
3.2	0.9993	0.9993	0.9994							
3.3	0.9995	0.9995	0.9995							
3.4	0.9997	0.9997	0.9997							
3.5	0.9998	0.9998	0.9998							
3.6	0.9998	0.9998	0.9999							
3.7	0.9999	0.9999	0.9999							
3.8	0.9999	0.9999	0.9999							
3.9	0.99995	1.0000	1.0000							
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									



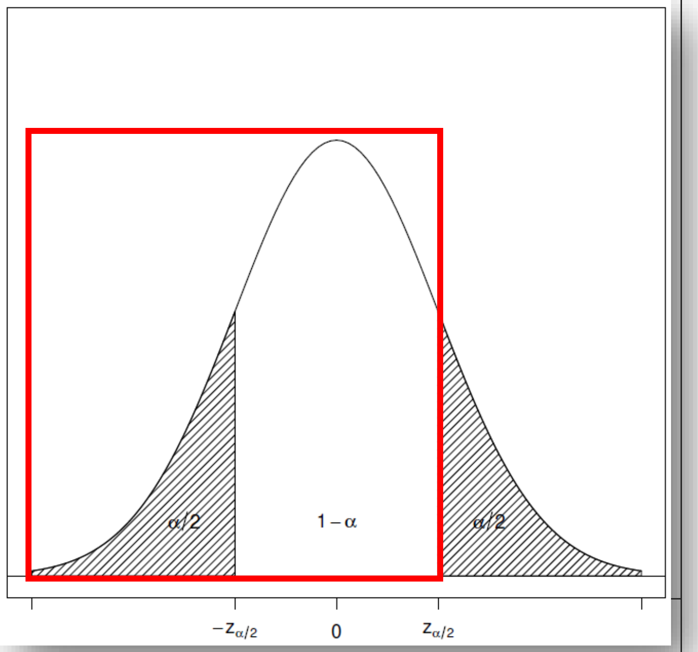
Esempio

Le tavole ci danno il valore del percentile sulla base dell'area sottesa fino alla percentuale richiesta

- **Esempio**:
- Vogliamo stimare μ al $90\% = 1-\alpha$
  - La quantità pivotale è: $Q = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$
  - Vogliamo la probabilità che $P\left(-z_1 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 0.9$
- Trovare i percentili $-z_1$ e $z_1$tale che l'area di $1-\alpha$ sia 0.9
  - Poiché $1-\alpha = 0.9 \Rightarrow \alpha = 1-0.9 \Rightarrow \alpha = 0.1$
    - Si ha che $\frac{\alpha}{2} = 0.05$
  - Con la tavola della normale standard si ha che:

Il quantile $z_1 = \varphi\left(1-\alpha+\frac{\alpha}{2}\right) = \varphi(0.9+0.05) = \varphi(0.95) \approx 1.6 + 0.05$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	$\boxed{0.9505}$	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656							
1.9	0.9713	0.9719	0.9726							
2.0	0.9772	0.9778	0.9783							
2.1	0.9821	0.9826	0.9830							
2.2	0.9861	0.9864	0.9868							
2.3	0.9893	0.9896	0.9898							
2.4	0.9918	0.9920	0.9922							
2.5	0.9938	0.9940	0.9941							
2.6	0.9953	0.9955	0.9956							
2.7	0.9965	0.9966	0.9967							
2.8	0.9974	0.9975	0.9976							
2.9	0.9981	0.9982	0.9982							
3.0	0.9987	0.9987	0.9987							
3.1	0.9990	0.9991	0.9991							
3.2	0.9993	0.9993	0.9994							
3.3	0.9995	0.9995	0.9995							
3.4	0.9997	0.9997	0.9997							
3.5	0.9998	0.9998	0.9998							
3.6	0.9998	0.9998	0.9999							
3.7	0.9999	0.9999	0.9999							
3.8	0.9999	0.9999	0.9999							
3.9	0.99995	1.0000	1.0000							
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									



Esempio

Le tavole ci danno il valore del percentile sulla base dell'area sottesa fino alla percentuale richiesta

Il quantile $z_1 = \varphi\left(1-\alpha+\frac{\alpha}{2}\right) = \varphi(0.9+0.05) = \varphi(0.95) \approx 1.6 + 0.05$

- Riprendiamo:

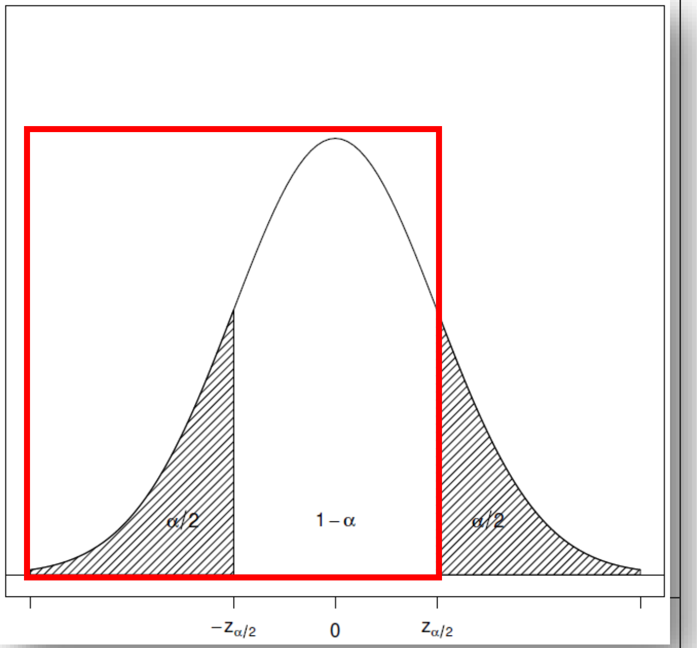
$$P\left(-z_1 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 0.9 \Rightarrow P\left(-1.65 * \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.65 * \frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$$\Rightarrow P\left(\overline{X} - 1.65 * \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.65 * \frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$$\Rightarrow P\left(1350 - 1.65 * \frac{100}{\sqrt{169}} < \mu < 1350 + 1.65 * \frac{100}{\sqrt{169}}\right) = 0.9$$

$$\Rightarrow P\left(1350 - 1.65 * \frac{100}{13} < \mu < 1350 + \frac{1.65 * 100}{13}\right) = 0.9$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	$\boxed{0.9505}$	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.8	0.9641	0.9649	0.9656							
1.9	0.9713	0.9719	0.9726							
2.0	0.9772	0.9778	0.9783							
2.1	0.9821	0.9826	0.9830							
2.2	0.9861	0.9864	0.9868							
2.3	0.9893	0.9896	0.9898							
2.4	0.9918	0.9920	0.9922							
2.5	0.9938	0.9940	0.9941							
2.6	0.9953	0.9955	0.9956							
2.7	0.9965	0.9966	0.9967							
2.8	0.9974	0.9975	0.9976							
2.9	0.9981	0.9982	0.9982							
3.0	0.9987	0.9987	0.9987							
3.1	0.9990	0.9991	0.9991							
3.2	0.9993	0.9993	0.9994							
3.3	0.9995	0.9995	0.9995							
3.4	0.9997	0.9997	0.9997							
3.5	0.9998	0.9998	0.9998							
3.6	0.9998	0.9998	0.9999							
3.7	0.9999	0.9999	0.9999							
3.8	0.9999	0.9999	0.9999							
3.9	0.99995	1.0000	1.0000							
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									



Esempio

Le tavole ci danno il valore del percentile sulla base dell'area sottesa fino alla percentuale richiesta

Il quantile $z_1 = \varphi\left(1 - \alpha + \frac{\alpha}{2}\right) = \varphi(0.9 + 0.05) = \varphi(0.95) \approx 1.6 + 0.05$

- Riprendiamo:

$$P\left(-z_1 < \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} < z_1\right) = 0.9 \Rightarrow P\left(-1.65 * \frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.65 * \frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$$\Rightarrow P\left(\overline{X} - 1.65 * \frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.65 * \frac{\sigma}{\sqrt{n}}\right) = 0.9$$

$$\Rightarrow P\left(1350 - 1.65 * \frac{100}{\sqrt{169}} < \mu < 1350 + 1.65 * \frac{100}{\sqrt{169}}\right) = 0.9$$

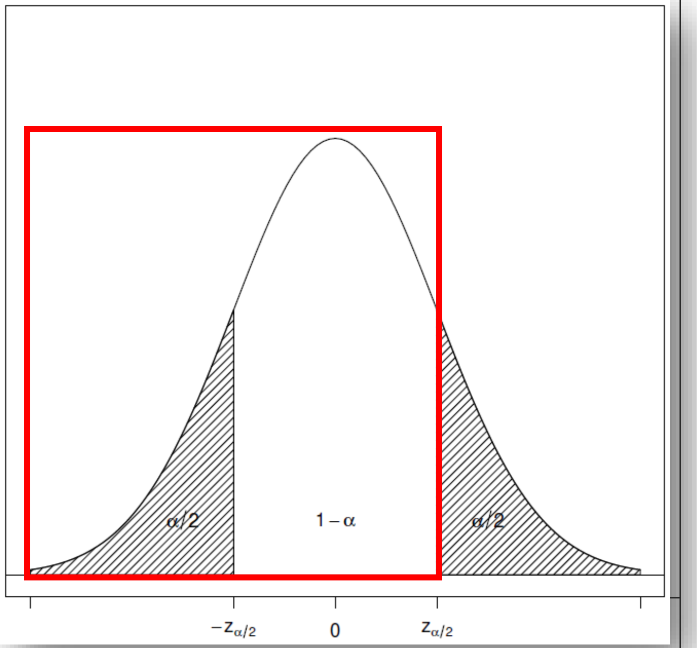
$$\Rightarrow P\left(1350 - 1.65 * \frac{100}{13} < \mu < 1350 + \frac{1.65 * 100}{13}\right) = 0.9$$

- L'intervallo di confidenza al 90% è quindi:

$$(1350 - 12.65; 1350 + 12.65) = (1337.35; 1362.65)$$

Significa che la **media sull'intera popolazione sarà compresa in questo intervallo** al 90%

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	$\boxed{0.9505}$	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1.8	0.9641	0.9649	0.9656							
1.9	0.9713	0.9719	0.9726							
2.0	0.9772	0.9778	0.9783							
2.1	0.9821	0.9826	0.9830							
2.2	0.9861	0.9864	0.9868							
2.3	0.9893	0.9896	0.9898							
2.4	0.9918	0.9920	0.9922							
2.5	0.9938	0.9940	0.9941							
2.6	0.9953	0.9955	0.9956							
2.7	0.9965	0.9966	0.9967							
2.8	0.9974	0.9975	0.9976							
2.9	0.9981	0.9982	0.9982							
3.0	0.9987	0.9987	0.9987							
3.1	0.9990	0.9991	0.9991							
3.2	0.9993	0.9993	0.9994							
3.3	0.9995	0.9995	0.9995							
3.4	0.9997	0.9997	0.9997							
3.5	0.9998	0.9998	0.9998							
3.6	0.9998	0.9998	0.9999							
3.7	0.9999	0.9999	0.9999							
3.8	0.9999	0.9999	0.9999							
3.9	0.99995	1.0000	1.0000							
4.0	0.99997									
5.0	0.9999997									
6.0	0.9999999990									



(i) Esempio (R)

- Un urbanista è interessato alla superficie media μ delle abitazioni di una certa città
- A questo scopo osserva un campione di 50 appartamenti

```
> campnorm<-c(112.6, 118.2, 124.8, 122.1, 137.5, 106.7, 123.7,
+ 127.3, 123.2, 125.1, 120.8, 112.9, 117.0, 128.1, 102.9, 119.1,
+ 127.2, 124.8, 118.0, 131.4, 117.0, 118.2, 125.8, 116.2, 118.5,
+ 120.8, 127.1, 125.0, 131.2, 120.2, 126.0, 119.2, 112.4, 124.6,
+ 117.7, 116.1, 125.3, 115.5, 129.6, 119.1, 130.6, 125.3, 128.7,
+ 134.6, 124.5, 117.2, 126.1, 116.1, 116.0, 125.6)
>
> mean(campnorm)
[1] 121.872
```

- Supponendo che la popolazione da cui proviene il campione sia normale con deviazione standard nota:

$$\sigma = 8\, m^2$$

determinare una stima dell'intervallo di confidenza di grado

$$1 - \alpha = 0.95$$

per la superficie media μ delle abitazioni

(i) Esempio (R)

- Un urbanista è interessato alla superficie media μ delle abitazioni di una certa città
- A questo scopo osserva un campione di 50 appartamenti

```
> campnorm<-c(112.6, 118.2, 124.8, 122.1, 137.5, 106.7, 123.7,
+ 127.3, 123.2, 125.1, 120.8, 112.9, 117.0, 128.1, 102.9, 119.1,
+ 127.2, 124.8, 118.0, 131.4, 117.0, 118.2, 125.8, 116.2, 118.5,
+ 120.8, 127.1, 125.0, 131.2, 120.2, 126.0, 119.2, 112.4, 124.6,
+ 117.7, 116.1, 125.3, 115.5, 129.6, 119.1, 130.6, 125.3, 128.7,
+ 134.6, 124.5, 117.2, 126.1, 116.1, 116.0, 125.6)
>
> mean(campnorm)
[1] 121.872
```

Media campionaria su 50 appartamenti: $\bar{x}_{50} = 121.872$

- Supponendo che la popolazione da cui proviene il campione sia normale con deviazione standard nota:

$$\sigma = 8\, m^2$$

determinare una stima dell'intervallo di confidenza di grado

$$1 - \alpha = 0.95$$

per la superficie media μ delle abitazioni

(i) Esempio (R)

- Poiché $\alpha = 1 - 0.95 = 0.05$ e $\frac{\alpha}{2} = 0.025$
- Dobbiamo determinare il valore $z_{\frac{\alpha}{2}} = z_{0.025} = 1.95$

```
> alpha<-1-0.95
> qnorm(1-alpha/2,mean=0,sd=1)
[1] 1.959964
```

- Considerando che il valore di μ è:

$$\overline{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Calcoliamo l'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la superficie media μ delle abitazioni:

```
> n<-length(campnorm)
> mean(campnorm)-qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 119.6546
> mean(campnorm)+qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 124.0894
```

(i) Esempio (R)

- Poiché $\alpha = 1 - 0.95 = 0.05$ e $\frac{\alpha}{2} = 0.025$
- Dobbiamo determinare il valore $z_{\frac{\alpha}{2}} = z_{0.025} = 1.95$

```
> alpha<-1-0.95
> qnorm(1-alpha/2,mean=0,sd=1)
[1] 1.959964
```

- Considerando che il valore di μ è:

$$\boxed{\overline{x}_n} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \boxed{\overline{x}_n} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Calcoliamo l'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la superficie media μ delle abitazioni:

```
> n<-length(campnorm)
> mean(campnorm)-qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 119.6546
> mean(campnorm)+qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 124.0894
```

(i) Esempio (R)

- Poiché $\alpha = 1 - 0.95 = 0.05$ e $\frac{\alpha}{2} = 0.025$
- Dobbiamo determinare il valore $z_{\frac{\alpha}{2}} = z_{0.025} = 1.95$

```
> alpha<-1-0.95
> qnorm(1-alpha/2,mean=0,sd=1)
[1] 1.959964
```

- Considerando che il valore di μ è:

$$\boxed{\overline{x}_n} - \boxed{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}} < \mu < \boxed{\overline{x}_n} + \boxed{z_{\alpha/2}\frac{\sigma}{\sqrt{n}}}$$

- Calcoliamo l'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la superficie media μ delle abitazioni:

```
> n<-length(campnorm)
> mean(campnorm)-qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 119.6546
> mean(campnorm)+qnorm(1-alpha/2,mean=0,sd=1)*8/sqrt(n)
[1] 124.0894
```

- L'intervallo di confidenza è quindi (119.7, 124.1)

STATISTICA E ANALISI DEI DATI

Popolazione Normale - I.C. per il valore medio μ con varianza σ^2 incognita

I.C. per il valore medio μ con varianza σ^2 incognita

- Sappiamo che la popolazione che stiamo considerando è distribuita come una normale $X = N(\mu, \sigma^2)$ di media μ e varianza σ^2
- Consideriamo un campione casuale di ampiezza n
- La **varianza campionaria** S^2 è uno stimatore uno stimatore **non distorto** della varianza effettiva σ^2 della popolazione, dove:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

dove:

- n lunghezza del campione
- X_i valori nel campione estratto
- $\overline{X}$ media campionaria

I.C. per il valore medio μ con varianza σ^2 incognita

- A partire da questo, possiamo costruire una nuova variabile aleatoria di **Student** con $n-1$ gradi di libertà:

$$\frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} \sim T_{(n-1)}$$

$\bar{X}$ = media campionaria, μ = media della popolazione, S = radice quadrata della varianza campionaria, n = dimensione del campione

- La variabile aleatoria $T_{(n-1)}$ dipende dal campione casuale e dal parametro non noto S e la sua legge di probabilità **non dipende** dal parametro μ che si vuole stimare
- Quindi, $T_{(n-1)}$ può essere interpretata come una **quantità pivotale**
- **Nota:**
  1. Quando la **varianza** σ^2 **della popolazione non è nota**, dobbiamo stimarla dai dati usando la **varianza campionaria** S^2 che è presente nella distribuzione di Student
  2. I gradi di libertà rappresentano il numero di valori indipendenti che possiamo variare nel calcolo di una statistica e in questo caso sono $n-1$ perché quando calcoliamo la varianza campionaria S^2, la media campionaria $\bar{X}$ è già stata calcolata, e questo vincola uno dei valori del campione

I.C. per il valore medio μ con varianza σ^2 incognita

- Vogliamo calcolare la probabilità che μ sia in un range a e b con una certa probabilità $1-\alpha$

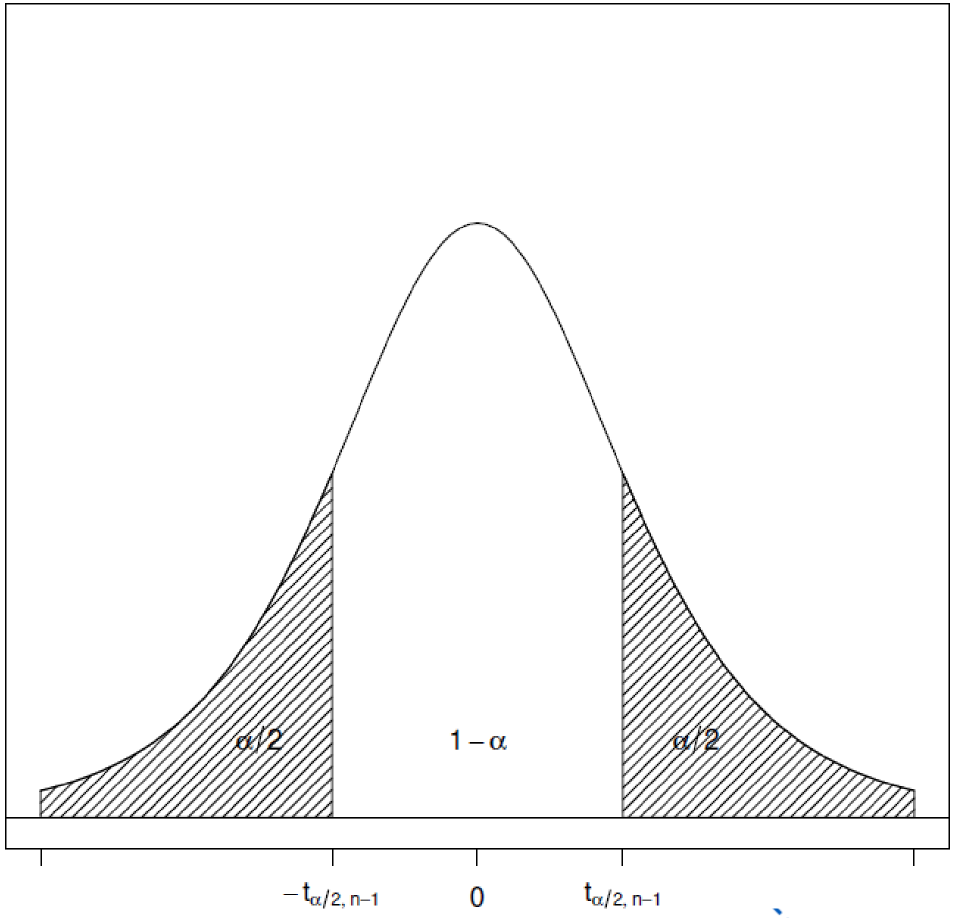
$$P(a < \mu < b) = 1-\alpha$$

- Pertanto usiamo la quantità Pivotale con la variabile di Student:
  - Calcoliamo la probabilità che la quantità pivotale sia compresa tra due quantili (Percentili) $-t_1$ e t_1:

$$P\left(-t_1 < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_1\right) = 1-\alpha$$

- Poiché $T_n = \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}}$ è una normale, si ha una campana simmetrica
  - Dove la probabilità che T_n sia compresa tra $-t_1$ e t_1 è $1-\alpha$
    - Dove $1-\alpha$ è l'area sottesa alla curva tra $-t_1$ e t_1
    - Poiché l'area sottesa in tutta la curva è 1, l'area sottesa sotto le due code sarà $\frac{\alpha}{2}$ (poiché $\frac{\alpha}{2}+\frac{\alpha}{2} = \alpha \Rightarrow 1-\alpha+\alpha = 1$)

Densità di Student con n−1 gradi di libertà



Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

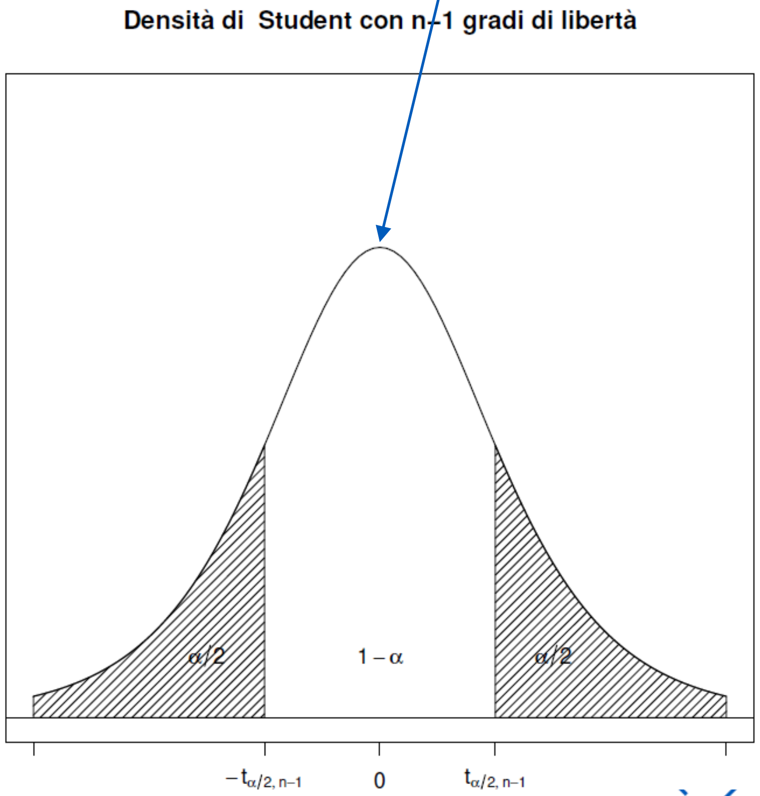
285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$

L'area deve essere uguale a 0.9



Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\boxed{X_i - \overline{X}})^2$$

X_i	$\boxed{X_i - \overline{X}}$	$(X_i - \overline{X})^2$
285		
279		
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i)$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285		
279		
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8}$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285		
279		
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285		
279		
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279		
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280		
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288		
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279		
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279	279 - 282.5 = -3.5	12.25
286		
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} (X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279	279 - 282.5 = -3.5	12.25
286	286 - 282.5 = 3.5	12.25
284		
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

La media campionaria è:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}(X_i) = \frac{285+279+280+288+279+286+284+279}{8} = \frac{2260}{8} = 282.5$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279	279 - 282.5 = -3.5	12.25
286	286 - 282.5 = 3.5	12.25
284	284 - 282.5 = 1.5	2.25
279		

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279	279 - 282.5 = -3.5	12.25
286	286 - 282.5 = 3.5	12.25
284	284 - 282.5 = 1.5	2.25
279	279 - 282.5 = -3.5	1.25

Esempio

- **Esempio**:
  - Sono stati riportati i risultati della lunghezza di un campione casuale di cetacei scelti da una popolazione con distribuzione normale
  - Determina un intervallo di confidenza per la media μ della lunghezza del campione al 90%
- La popolazione è:

285, 279, 280, 288, 279, 286, 284, 279

Conosciamo che:

- $n = 8$

- Vogliamo stimare $\overline{X}$ e la varianza campionaria S^2 al $90\% = 1 - \alpha$ sapendo:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 =$$

$$\frac{1}{8-1}\boxed{(6.25+12.25+6.25+30.25+12.25+12.25+2.25+1.25)} = \frac{94}{7} = 13{,}428$$

- Da cui: $S = \sqrt{13{,}428} \approx 3{,}664$

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
285	285 - 282.5 = 2.5	6.25
279	279 - 282.5 = -3.5	12.25
280	280 - 282.5 = -2.5	6.25
288	288 - 282.5 = 5.5	30.25
279	279 - 282.5 = -3.5	12.25
286	286 - 282.5 = 3.5	12.25
284	284 - 282.5 = 1.5	2.25
279	279 - 282.5 = -3.5	1.25

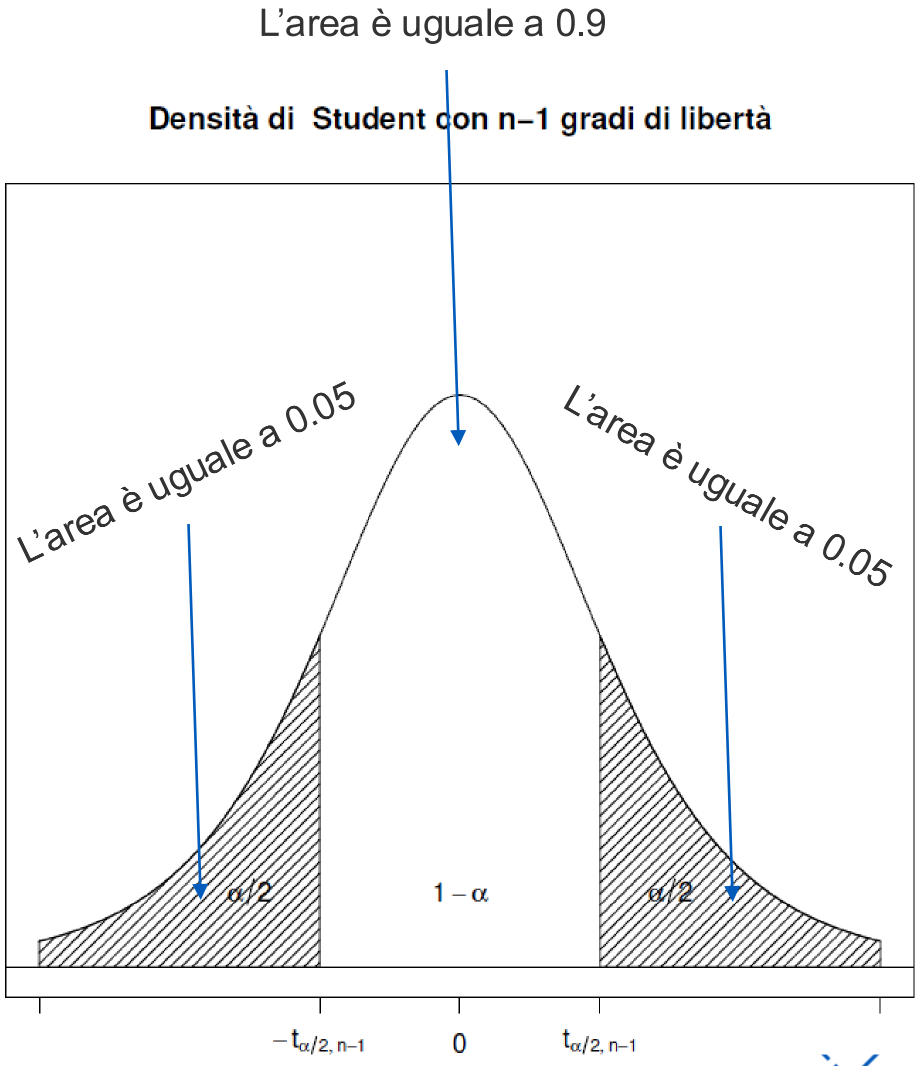
Esempio

- A questo punto conosciamo:
  - $n = 8$
  - $S^2 = 13{,}428$
  - $S = 3{,}664$
  - $\overline{X} = 282{,}5$
- Riprendiamo la definizione della variabile di Student:

$$Q = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim T_{(n-1)}$$

E la probabilità:

$$P\left(-t_1 < \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} < t_1\right) = 0.90$$



Esempio

- A questo punto conosciamo:
  - $n = 8$
  - $S^2 = 13{,}428$
  - $S = 3{,}664$
  - $\overline{X} = 282{,}5$
- Riprendiamo la definizione della variabile di Student:

$$Q = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim T_{(n-1)}$$

E la probabilità:

$$P\left(-t_1 < \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} < t_1\right) = 0.90$$

Gradi di libertà

Area

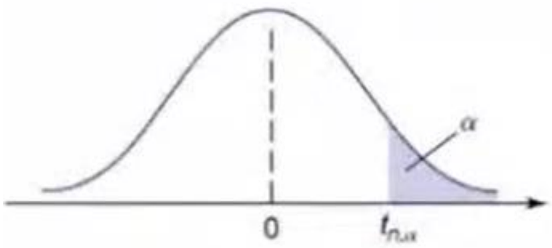


Table D.2 Percentiles $t_{n,\alpha}$ of t Distributions

	α									
n	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015

Esempio

- A questo punto conosciamo:
  - $n = 8$
  - $S^2 = 13{,}428$
  - $S = 3{,}664$
  - $\overline{X} = 282{,}5$
- Riprendiamo la definizione della variabile di Student:

$$Q = \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \sim T_{(n-1)}$$

E la probabilità:

$$P\left(-t_1 < \frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} < t_1\right) = 0.90$$

Si ha quindi: $t_1 = 1.895$

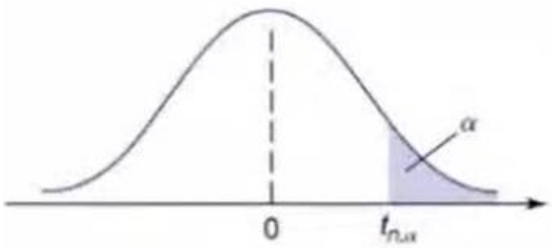


Table D.2 Percentiles $t_{n,\alpha}$ of t Distributions

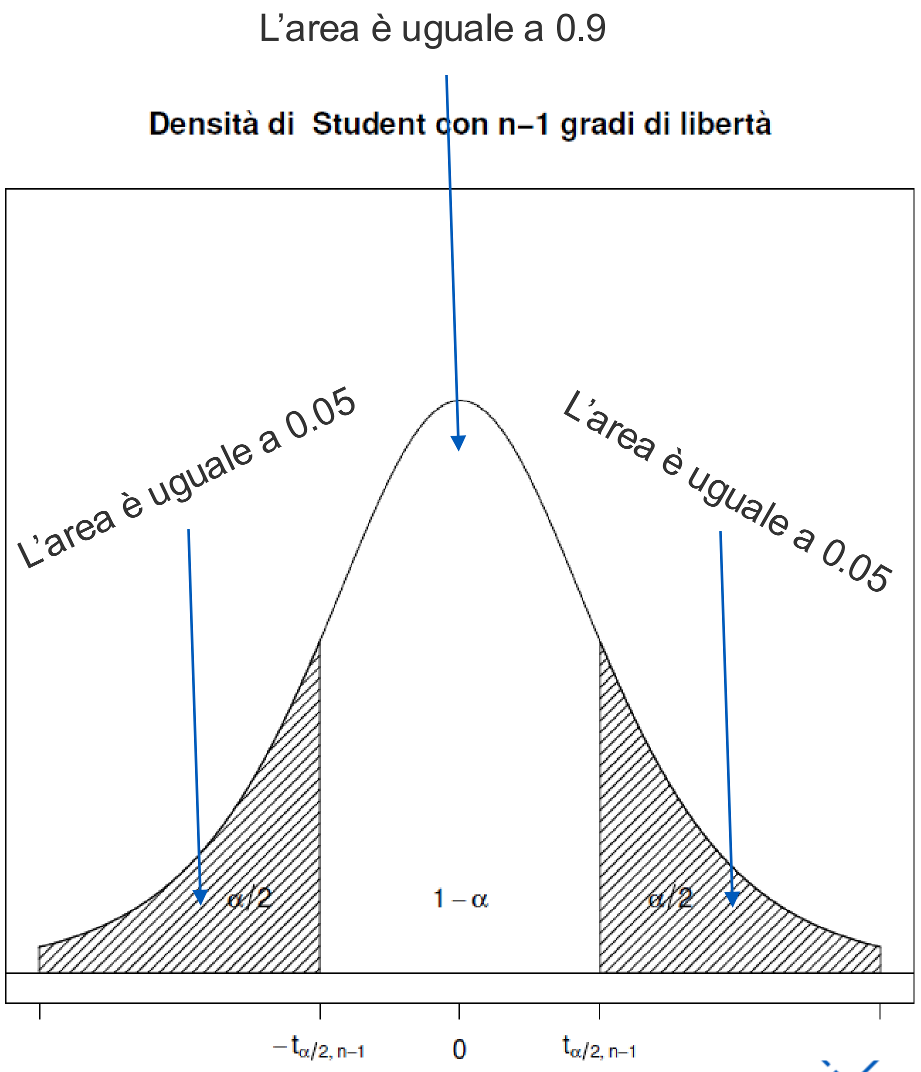
n	α = 0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	$\boxed{1.895}$	2.365	2.998	3.499	4.029	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015

Esempio

- Riprendiamo la probabilità:

$$P\left(-t_1 < \frac{\bar{X}-\mu}{\frac{S}{\sqrt{n}}} < t_1\right) = P\left(-t_1\frac{S}{\sqrt{n}} < \bar{X}-\mu < t_1\frac{S}{\sqrt{n}}\right) =$$

$$P\left(\bar{X} - t_1\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_1\frac{S}{\sqrt{n}}\right)$$



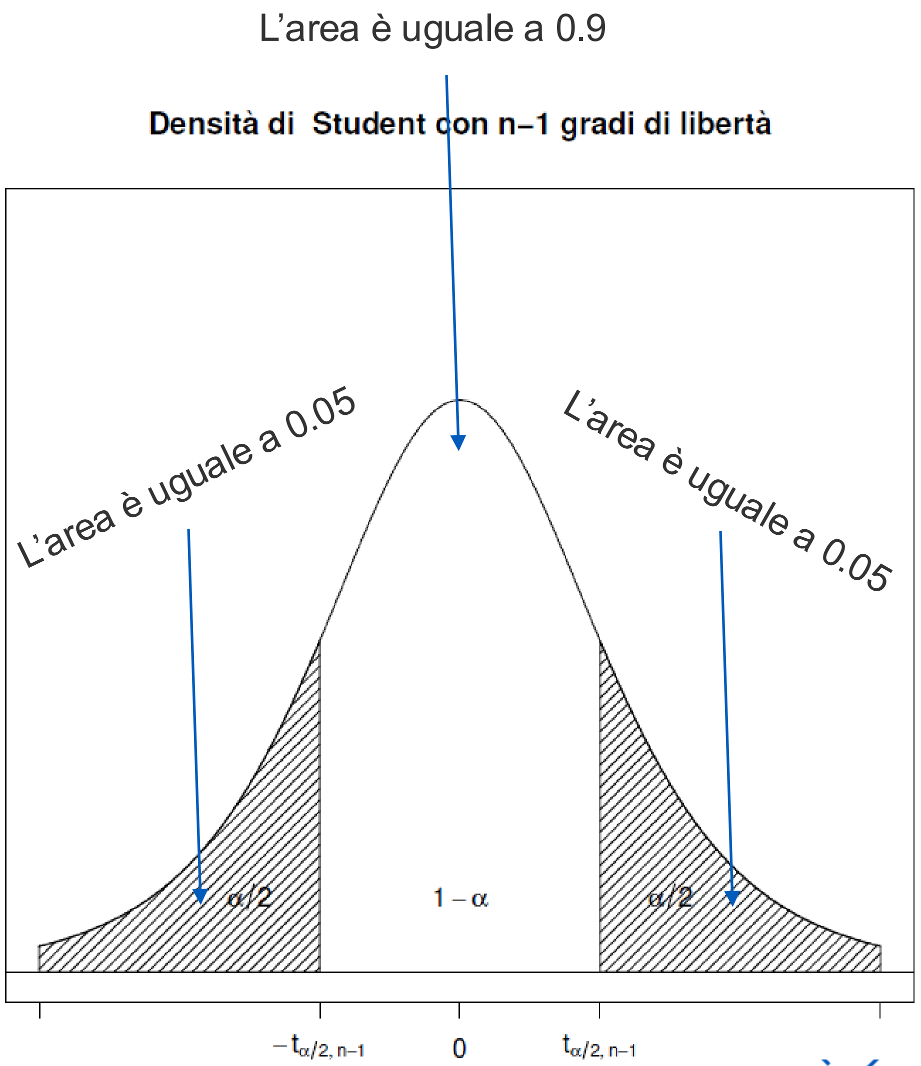
Esempio

- Riprendiamo la probabilità:

$$P\left(-t_1 < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_1\right) = P\left(-t_1\frac{S}{\sqrt{n}} < \overline{X}-\mu < t_1\frac{S}{\sqrt{n}}\right) =$$

$$P\left(\overline{X} - t_1\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_1\frac{S}{\sqrt{n}}\right)$$

$$\Rightarrow P\left(282{,}5 - 1{,}895\frac{3{,}664}{\sqrt{8}} < \mu < 282{,}5 + 1{,}895\frac{3{,}664}{\sqrt{8}}\right) = 0.9$$



Esempio

- Riprendiamo la probabilità:

$$P\left(-t_1 < \frac{\overline{X}-\mu}{\frac{S}{\sqrt{n}}} < t_1\right) = P\left(-t_1\frac{S}{\sqrt{n}} < \overline{X}-\mu < t_1\frac{S}{\sqrt{n}}\right) =$$

$$P\left(\overline{X} - t_1\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_1\frac{S}{\sqrt{n}}\right)$$

$$\Rightarrow P\left(282{,}5 - 1{,}895\frac{3{,}664}{\sqrt{8}} < \mu < 282{,}5 + 1{,}895\frac{3{,}664}{\sqrt{8}}\right) = 0.9$$

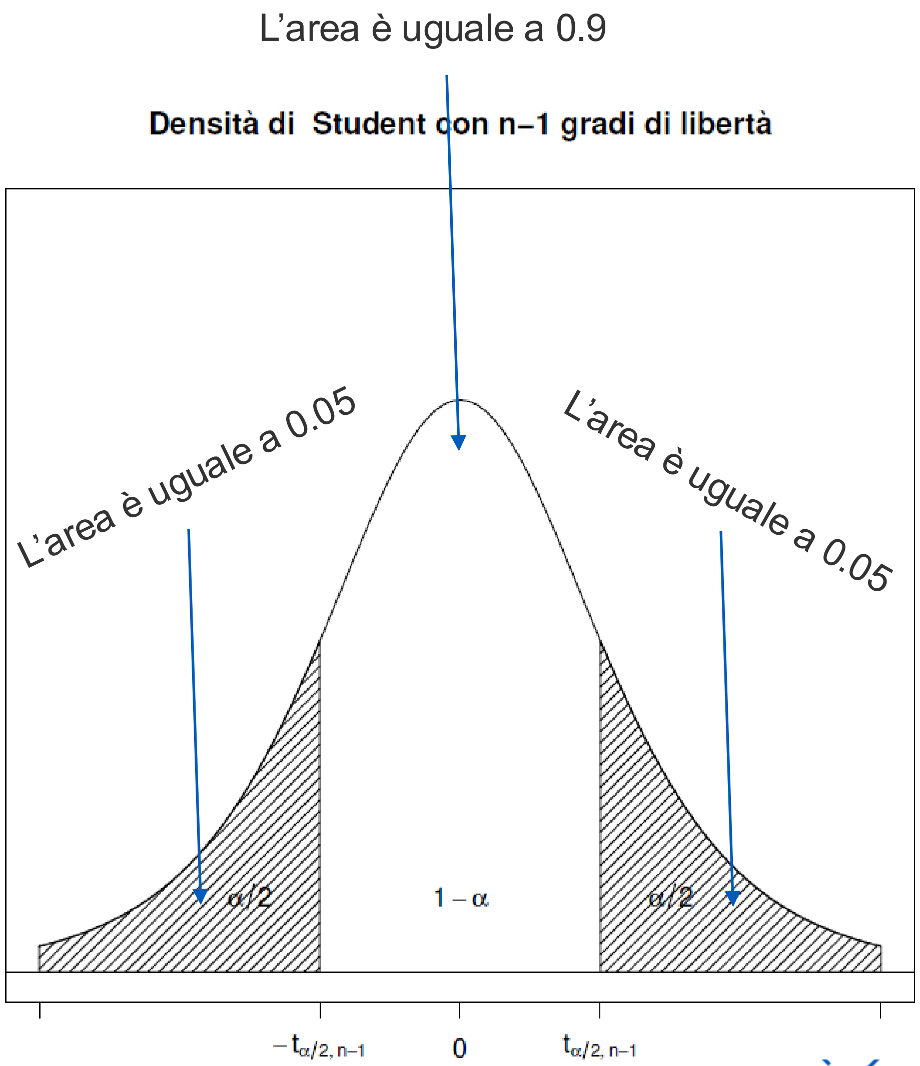
Cioè:

$$P(280{,}045 < \mu < 284{,}955) = 0.9$$

Da cui l'intervallo di confidenza per μ al 90% è:

$$(280{,}045; 284{,}955)$$

- In altri termini la media di tutta la popolazione μ sarà al 90% in questo intervallo



(i) Esempio (R)

- Un produttore di una certa marca di sigarette desidera controllare il quantitativo medio di nicotina in esse contenuto
- A questo scopo egli osserva un campione di 30 sigarette

```
> campnorm<-c(10.2, 11.4,  9.7, 10.9, 11.0, 11.3,  9.8, 10.1,
+ 10.8, 10.43, 11.4, 10.8, 11.5, 10.9, 10.0, 11.2, 11.8, 11.8,
+ 10.9, 10.9, 10.9, 11.2, 11.3, 10.6, 10.9, 11.2, 11.5, 11.6,
+ 10.3, 10.8)
>
> mean(campnorm)
[1] 10.90433
> sd(campnorm)
[1] 0.563864
```

- Si ha quindi che $\bar{x}_{30} = 10.90$ e $s_{30} = 0.56$
- Supponendo che la popolazione da cui proviene il campione sia normale, determinare una stima dell'intervallo di confidenza di grado $1 - \alpha = 0.99$ per il quantitativo medio di nicotina contenuto in una sigaretta

(i) Esempio (R)

- Si ha quindi:

  - $\alpha = 1 - 0.99 = 0.01$

  - $\alpha = \frac{\alpha}{2} = 0.005$

- Il valore valore di t_1 può essere determinato in R con:

```
> alpha<-1-0.99
> n<-length(campnorm)
> qt(1-alpha/2,df=n-1)
[1] 2.756386
> mean(campnorm)-qt(1-alpha/2,df=n-1)*sd(campnorm)/sqrt(n)
[1] 10.62057
> mean(campnorm)+qt(1-alpha/2,df=n-1)*sd(campnorm)/sqrt(n)
[1] 11.1881
```

Quantili di Student

- Si ha quindi che $t_1 = 2.756$ e la stima dell'intervallo di confidenza di grado $1 - \alpha = 0.99$ per il quantitativo medio di nicotina contenuto in una sigaretta è:

(10.62; 11.19)

(ii) Esempio (R)

- Ad un campione di 100 studenti che frequentano un certo corso universitario è stato chiesto di assegnare un voto da 1 (pessimo) a 10 (ottimo) come valutazione del corso
  - La media campionaria del punteggio su una popolazione con 100 campioni è $\bar{X}_{100} = 7.2$
  - Mentre la varianza campionaria è $S_{100}^2 = 2.25$
- Si desidera
  1. Costruire un intervallo di confidenza del 95% per il punteggio medio assegnato dai 100 studenti;
  2. Costruire un intervallo di confidenza del 99% per il punteggio medio assegnato dai 100 studenti;
  3. Se invece di 100 studenti si considerano 200 studenti e risulta $\bar{X}_{200} = 7.2$ e $S_{100}^2 = 2.25$, costruire un intervallo di confidenza del 95% per il punteggio medio assegnato dagli studenti

(ii) Esempio (R)

1. Costruire un intervallo di confidenza del 95% per il punteggio medio assegnato dai 100 studenti;
   - Abbiamo che:
     - $\bar{x}_{100} = 7.2$
     - $s^2_{100} = 2.25$
     - $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

- Con R si ha che:

```
> m<-7.2
> s2<-2.25
> alpha<-1-0.95
> n<-100
> qt(1-alpha/2,df=n-1)
[1] 1.984217
>
> m-qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 6.902367
> m+qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 7.497633
```

- L'intervallo di confidenza è: (6.902, 7.498)

(ii) Esempio (R)

2. Costruire un intervallo di confidenza del 99% per il punteggio medio assegnato dai 100 studenti;
    - Abbiamo che:
        - $\bar{x}_{100} = 7.2$
        - $s_{100}^2 = 2.25$
        - $1 - \alpha = 0.99 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.005$

- Con R si ha che:

```
> m<-7.2
> s2<-2.25
> alpha<-1-0.99
> n<-100
> qt(1-alpha/2,df=n-1)
[1] 2.626405
>
> m-qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 6.806039
> m+qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 7.593961
```

- L'intervallo di confidenza è: (6.806, 7.594)

(ii) Esempio (R)

3. se invece di 100 studenti si considerano 200 studenti e risulta $\bar{x}_{200} = 7.2$ e $s^2_{100} = 2.25$, costruire un intervallo di confidenza del 95% per il punteggio medio assegnato dagli studenti

- Abbiamo che:
  - $\bar{x}_{200} = 7.2$
  - $s^2_{200} = 2.25$
  - $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

- Con R si ha che:

```
> m<-7.2
> s2<-2.25
> alpha<-1-0.95
> n<-200
> qt(1-alpha/2,df=n-1)
[1] 1.971957
>
> m-qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 6.990842
> m+qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 7.409158
```

- L'intervallo di confidenza è: (6.991, 7.409)

(ii) Esempio (R)

3. se invece di 100 studenti si considerano 200 studenti e risulta $\bar{x}_{200} = 7.2$ e $s^2_{100} = 2.25$, costruire un intervallo di confidenza del 95% per il punteggio medio assegnato dagli studenti

- Abbiamo che:
  - $\bar{x}_{200} = 7.2$
  - $s^2_{200} = 2.25$
  - $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05 \Rightarrow \frac{\alpha}{2} = 0.025$

- Con R si ha che:

```
> m<-7.2
> s2<-2.25
> alpha<-1-0.95
> n<-200
> qt(1-alpha/2,df=n-1)
[1] 1.971957
>
> m-qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 6.990842
> m+qt(1-alpha/2,df=n-1)*sqrt(s2)/sqrt(n)
[1] 7.409158
```

- L'intervallo di confidenza è: (6.991, 7.409) → All'aumentare del campione, l'intervallo si riduce

STATISTICA E ANALISI DEI DATI

I.C. per la varianza σ^2 con il valore medio μ noto

I.C. per la varianza σ^2 con il valore medio μ noto

- Sappiamo che la popolazione che stiamo considerando è distribuita come una normale $X = N(\mu, \sigma^2)$ di media μ e varianza σ^2 (e varianza campionaria S^2)
- Consideriamo un campione casuale di ampiezza n
  - Allora si ha che la statistica:

$$\frac{nS^2}{\sigma^2} \text{ si distribuisce come } V_n = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2$$

è **distribuita con legge chi-quadrato** χ^2 con n gradi di libertà, essendo costituita dalla somma dei quadrati di n variabili aleatorie normali standard

- S^2 varianza campionaria

- n numero di campioni nella popolazione

- μ valore medio

- σ^2 varianza

I.C. per la varianza σ^2 con il valore medio μ noto

- Vogliamo calcolare la probabilità che V_n sia compresa in un range $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ con una certa probabilità $1-\alpha$

$$P\left(\chi^2_{1-\frac{\alpha}{2}} < V_n < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$$

- Poichè V_n si può scrivere in forma alternativa in termini della media campionaria e della varianza campionaria:

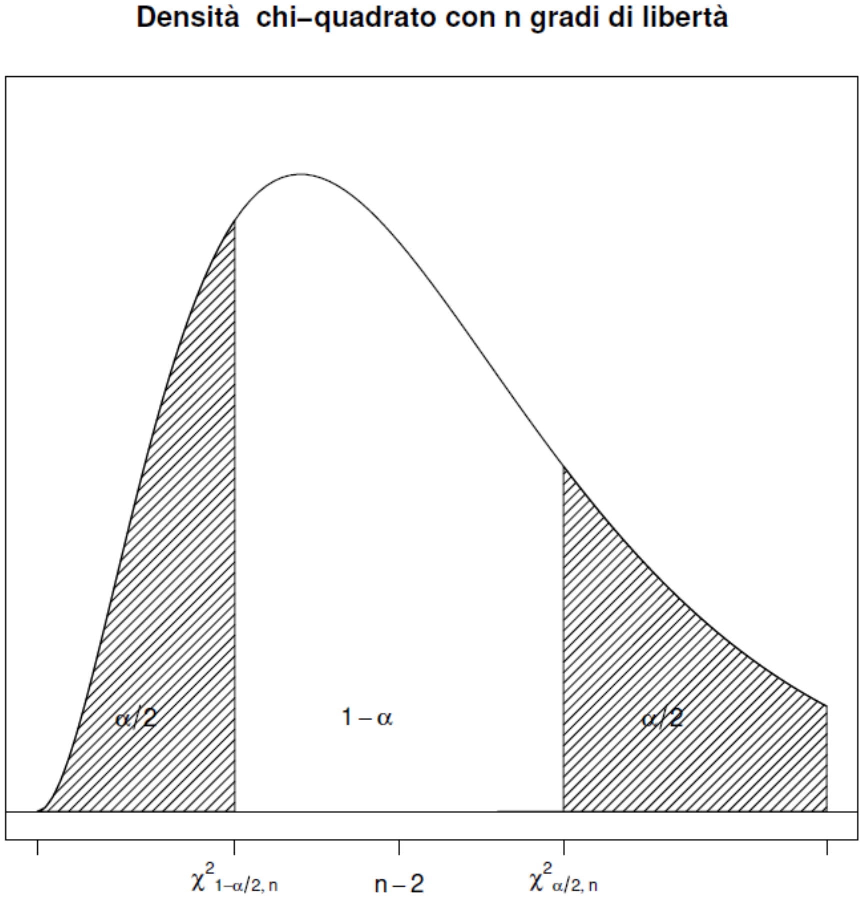
$$V_n = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X} + \overline{X} - \mu}{\sigma}\right)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \overline{X}}{\sigma}\right)^2 + \sum_{i=1}^{n}\left(\frac{\overline{X} - \mu}{\sigma}\right)^2$$

$$= \frac{(n-1)\,S_n^2}{\sigma^2} + \left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2,$$

- Si ha che:

$$P\left(\chi^2_{1-\alpha/2,n} < \frac{(n-1)\,S_n^2}{\sigma^2} + \left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}\right)^2 < \chi^2_{\alpha/2,n}\right) = 1-\alpha$$

- Da cui isoliamo la varianza: $P\left(\frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{\alpha/2,n}} < \sigma^2 < \frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{1-\alpha/2,n}}\right) = 1-\alpha$



I.C. per la varianza σ^2 con il valore medio μ noto

- Da cui isoliamo la varianza:

$$P\left(\frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{\alpha/2,n}} < \sigma^2 < \frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{1-\alpha/2,n}}\right) = 1 - \alpha$$

- Ponendo:

$$\underline{C}_n = \frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{\alpha/2,n}}, \qquad \overline{C}_n = \frac{(n-1)S_n^2 + n\left(\overline{X}_n - \mu\right)^2}{\chi^2_{1-\alpha/2,n}}$$

- Si ottiene che $(\underline{C}_n;\ \overline{C}_n)$ è l'intervallo di confidenza di grado $1-\alpha$ per σ^2 e $\underline{C}_n$ e $\overline{C}_n$ sono il limite inferiore ed il limite superiore di tale intervallo

(i) Esempio (R)

- Osservando un campione contenente il peso in grammi di 12 uova prodotte da un'azienda agricola

```
> campnorm<-c(69.6, 82.2, 64.4, 74.8, 71.2, 70.2, 71.3, 70.6,
+ 72.0, 65.8, 70.3, 63.5)
>
> mean(campnorm)
[1] 70.49167
> var(campnorm)
[1] 24.36447
```

- Si nota quindi che $\bar{x}_{12} = 70.49gr$ e $s_{12}^2 = 24.36gr$
- Supponendo che il peso sia distribuito normalmente con valore medio $\mu = 70gr$ e varianza non nota
  - Determinare una stima dell'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la varianza σ^2
- <u>Soluzione</u>:
  - $1 - \alpha = 0.95 \Rightarrow \alpha = 1 - 0.95 = 0.05$ cioè $\frac{\alpha}{2} = 0.025$ e $1 - \frac{\alpha}{2} = 0.975$
  - Per ottenere i valori $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ usiamo R

(i) Esempio (R)

- Soluzione:
  - $1-\alpha = 0.95 \Rightarrow \alpha = 1-0.95 = 0.05$ cioè $\frac{\alpha}{2} = 0.025$ e $1-\frac{\alpha}{2} = 0.975$
  - Per ottenere i valori $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ usiamo R

```
> n<-length(campnorm)
> mu<-70
>
> alpha<-1-0.95
> qchisq(alpha/2,df=n)
[1] 4.403789
> qchisq(1-alpha/2,df=n)
[1] 23.33666
>
> ((n-1)*var(campnorm)+n*(mean(campnorm)-mu)**2)/qchisq(1-alpha/2,
   df=n)
[1] 11.60877
> ((n-1)*var(campnorm)+n*(mean(campnorm)-mu)**2)/qchisq(alpha/2,df=
   n)
[1] 61.51749
```

  - La stima dell'intervallo di confidenza di grado $1-\alpha = 0.95$ per la varianza della popolazione normale è quindi $(11.61, 61.51)$

STATISTICA E ANALISI DEI DATI

I.C. per la varianza σ^2 con il valore medio μ incognito

I.C. per la varianza σ^2 con il valore medio μ noto

- Sappiamo che la popolazione che stiamo considerando è distribuita come una normale $X = N(\mu, \sigma^2)$ di media μ e varianza σ^2 (e varianza campionaria S^2)
- Consideriamo un campione casuale di ampiezza n
- Allora si ha che la statistica o anche variabile Pivot è:

$$Q_n = \frac{(n-1)S_n^2}{\sigma^2} = \frac{(n-1)}{\sigma^2} * \frac{1}{(n-1)} \sum_{i=1}^{n} (X_i - \bar{X}_i)^2 = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \bar{X}_i)^2$$

è **distribuita con legge chi–quadrato** χ^2 con $n-1$ gradi di libertà, essendo costituita dalla somma dei quadrati di $n-1$ variabili aleatorie normali standard

- S^2 varianza campionaria

- n numero di campioni nella popolazione

- μ valore medio

- σ^2 varianza

I.C. per la varianza σ^2 con il valore medio μ noto

- Vogliamo calcolare la probabilità che Q_n sia compresa in un range $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ con una certa probabilità $1-\alpha$

$$P\left(\chi^2_{1-\frac{\alpha}{2}} < Q_n < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$$

- Da cui:

$$P\left(\chi^2_{1-\frac{\alpha}{2}} < \frac{(n-1)S_n^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}\right) = 1-\alpha$$

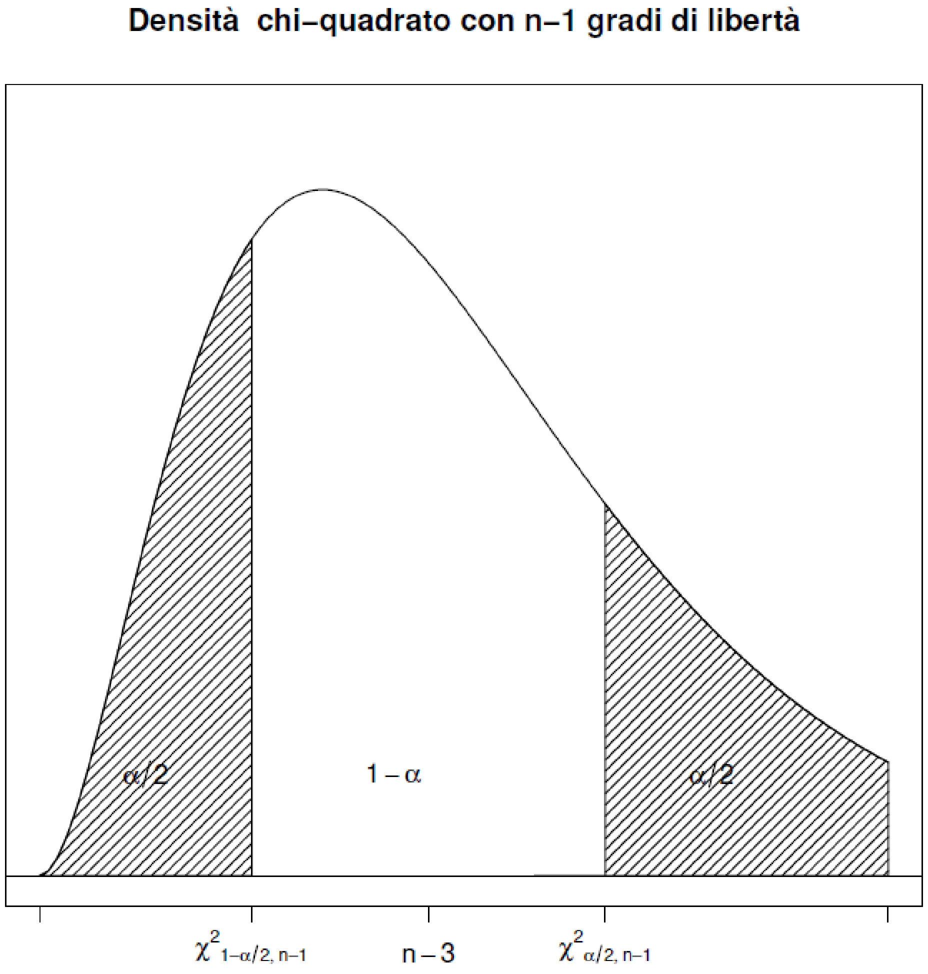
- Da cui isoliamo la varianza:

$$P\left(\frac{(n-1)\,S_n^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)\,S_n^2}{\chi^2_{1-\alpha/2,n-1}}\right) = 1-\alpha$$

- Ponendo:

$$\underline{C}_n = \frac{(n-1)\,S_n^2}{\chi^2_{\alpha/2,n-1}}, \qquad \overline{C}_n = \frac{(n-1)\,S_n^2}{\chi^2_{1-\alpha/2,n-1}}$$

- Si ottiene che $(\underline{C}_n;\ \overline{C}_n)$ è l'intervallo di confidenza di grado $1-\alpha$ per σ^2 e $\underline{C}_n$ e $\overline{C}_n$ sono il limite inferiore ed il limite superiore di tale intervallo



(i) Esempio (R)

- Si supponga che l'errore mensile misurato in secondi commesso da un certo tipo di orologi sia distribuito normalmente con valore medio e varianza non noti
- Osservando un campione di 20 orologi:

```
> campnorm<-c(-0.47, -0.33,  0.53, -0.32,  0.47,  0.52,  0.21,
+ 0.72,  0.54, -0.06,  0.33, -0.09,  0.37,  0.27, -0.07, -0.51,
+ 0.27, -0.13,-0.04, -0.13)
> mean(campnorm)
[1] 0.104
> var(campnorm)
[1] 0.1339937
```

- Si nota quindi che $s_{20}^2 = 0.13gr$
  - Determinare una stima dell'intervallo di confidenza di grado $1 - \alpha = 0.95$ per la varianza σ^2
- <u>Soluzione</u>:
  - $1 - \alpha = 0.95 \Rightarrow \alpha = 1 - 0.95 = 0.05$ cioè $\frac{\alpha}{2} = 0.025$ e $1 - \frac{\alpha}{2} = 0.975$
  - Per ottenere i valori $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ usiamo R

(i) Esempio (R)

- Soluzione:
  - $1-\alpha = 0.95 \Rightarrow \alpha = 1-0.95 = 0.05$ cioè $\frac{\alpha}{2} = 0.025$ e $1-\frac{\alpha}{2} = 0.975$
  - Per ottenere i valori $\chi^2_{1-\frac{\alpha}{2}}$ e $\chi^2_{\frac{\alpha}{2}}$ usiamo R

```
> n<-length(campnorm)
> alpha<-1-0.95
>
> qchisq(alpha/2,df=n-1)
[1] 8.906516
> qchisq(1-alpha/2,df=n-1)
[1] 32.85233
>
> (n-1)*var(campnorm)/qchisq(1-alpha/2,df=n-1)
[1] 0.07749466
> (n-1)*var(campnorm)/qchisq(alpha/2,df=n-1)
[1] 0.2858446
```

  - La stima dell'intervallo di confidenza di grado $1-\alpha = 0.95$ per la varianza della popolazione normale è quindi $(0.077, 0.286)$

Sommario

Intervallo di confidenza per μ con varianza σ^2 nota	$\overline{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ $z_{\alpha/2}$ si calcola con $\mathsf{qnorm(1 - alpha/2, mean = 0, sd = 1)}$
Intervallo di confidenza per μ con varianza non nota	$\overline{x}_n - t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}} < \mu < \overline{x}_n + t_{\alpha/2,n-1}\frac{s_n}{\sqrt{n}}$ $t_{\alpha/2,n-1}$ si calcola con $\mathsf{qt(1 - alpha/2, df = n - 1))}$
Intervallo di confidenza per σ^2 con valore medio μ noto	$\frac{(n-1)s_n^2 + n\left(\overline{x}_n - \mu\right)^2}{\chi^2_{\alpha/2,n}} < \sigma^2 < \frac{(n-1)s_n^2 + n\left(\overline{x}_n - \mu\right)^2}{\chi^2_{1-\alpha/2,n}}$ $\chi^2_{1-\alpha/2,n}$ si calcola con $\mathsf{qchisq(alpha/2, df = n)}$ $\chi^2_{\alpha/2,n}$ si calcola con $\mathsf{qchisq(1 - alpha/2, df = n)}$
Intervallo di confidenza per σ^2 con valore medio non noto	$\frac{(n-1)\,s_n^2}{\chi^2_{\alpha/2,n-1}} < \sigma^2 < \frac{(n-1)\,s_n^2}{\chi^2_{1-\alpha/2,n-1}}$ $\chi^2_{1-\alpha/2,n-1}$ si calcola con $\mathsf{qchisq(alpha/2, df = n - 1)}$ $\chi^2_{\alpha/2,n-1}$ si calcola con $\mathsf{qchisq(1 - alpha/2, df = n - 1)}$