

## Gli intervalli di confidenza

- Intervallo di confidenza per la media ( $\sigma^2$  nota) nel caso di popolazione Gaussiana

Sia  $X$  una v.c. Gaussiana di media  $\mu$  e varianza  $\sigma^2$ . Se  $X_1, X_2, \dots, X_n$  è un campione i.i.d. estratto da  $X$  allora l'intervallo di confidenza per  $\mu$  di livello  $1 - \alpha$  si scrive nella seguente forma

$$\mu \in \left( \bar{X}_n \pm z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right)$$

Lo stesso risultato vale se  $X$  è qualunque purché l'ampiezza del campione sia sufficientemente elevata

- Intervallo di confidenza per la media ( $\sigma^2$  incognita) nel caso di popolazione Gaussiana

Se il valore della varianza  $\sigma^2$  non è noto calcoliamo una sua stima attraverso lo stimatore  $\bar{s}_n^2$ .

Se  $X_1, X_2, \dots, X_n$  è un campione i.i.d. estratto da  $X$  allora l'intervallo di confidenza per  $\mu$  di livello  $1 - \alpha$  può essere scritto nella seguente forma

$$\mu \in \left( \bar{X}_n \pm t_{1-\frac{\alpha}{2}}^{(n-1)} \sqrt{\frac{\bar{s}_n^2}{n}} \right)$$

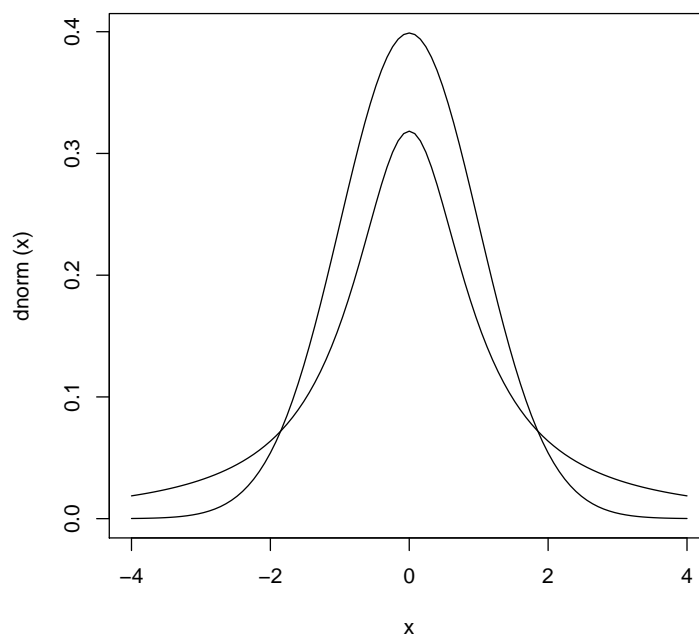
## La variabile $t$ di Student

Il valore  $t_{1-\alpha/2}^{(n-1)}$  è il valore analogo a  $z_{1-\frac{\alpha}{2}}$  ma calcolato per la v.c.  $t$  di Student. Se indichiamo con  $T^{n-1}$  la v.c.  $t$  di Student con  $n - 1$  gradi di libertà abbiamo che

$$P\left(T^{n-1} < t_{1-\alpha/2}^{(n-1)}\right) = 1 - \frac{\alpha}{2}$$

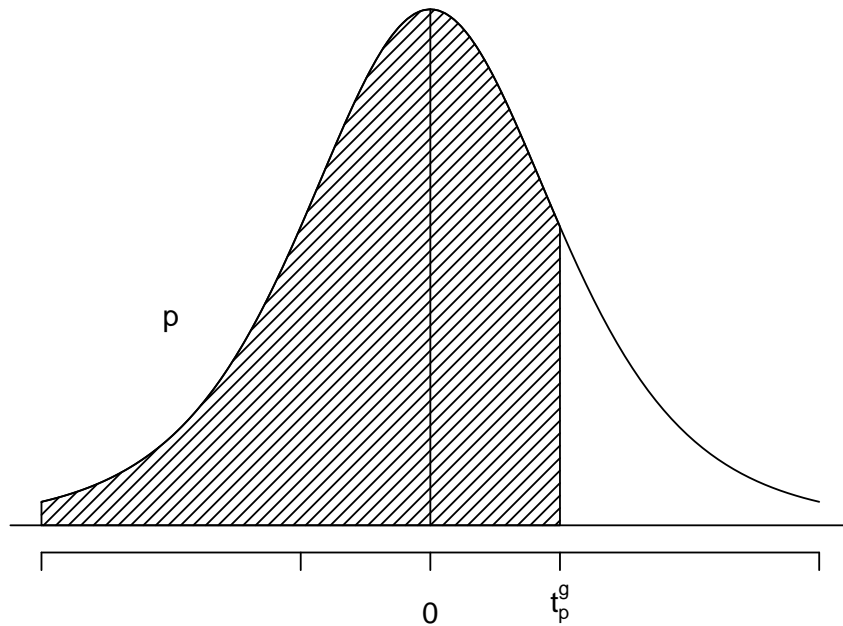
Quel valore di  $t$  si deve cercare nella tavola della tabella che riporta i valori della  $t$  di Student. La distribuzione  $t$  di Student è molto simile ad una Gaussiana per forma ma è leggermente più appiattita.

## La variabile $t$ di Student e la $N(0, 1)$



## La variabile $t$ di Student

$$p = P(t \leq t_p^g) = \int_{-\infty}^{t_p^g} f(x) dx$$



Per trovare il quantile della distribuzione  $t$  di Student si procede come per i quantili della distribuzione gaussiana. L'area tratteggiata corrisponde a

$$p = P\left(T^g < t_p^{(g)}\right)$$

dove  $g$  sono i gradi di libertà e  $p$  la probabilità

**Esempio:** il peso, in grammi, di alcuni granelli di polvere identificati su una piastra di silicio si suppone distribuito come una variabile casuale normale di parametri  $\mu$  e  $\sigma^2$ . I dati sono riportati di seguito:

0.39	0.68	0.82	1.35	1.38	1.62
1.70	1.71	1.85	2.14	2.89	3.69

Dopo aver determinato una stima per  $\mu$  si costruiscono gli intervalli di confidenza per la media al livello 95% e 99% supponendo  $\sigma^2$  ignota

Calcoliamo la media campionaria  $\bar{x}_n$

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i = \frac{0.39 + 0.68 + \dots + 3.69}{12} = 1.685$$

Poiché  $\sigma^2$  è incognita la stimiamo utilizzando la statistica  $\bar{s}_n^2$

$$\bar{s}_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2 = 0.85$$

L'intervallo di confidenza si ottiene attraverso la formula

$$\mu \in \left( \bar{x}_n \pm \sqrt{\frac{\bar{s}_n^2}{n}} t_{1-\frac{\alpha}{2}}^{(n-1)} \right)$$

Nei due casi i valori della  $t$  di Student sono

$$t_{1-\frac{0.05}{2}}^{(11)} = t_{0.975}^{(11)} = 2.201$$

e

$$t_{1-\frac{0.01}{2}}^{(11)} = t_{0.995}^{(11)} = 3.106$$

Sostituendo i valori si ottiene

$$\mu \in (1.10, 2.27) \quad \text{di livello } 95\%$$

$$\mu \in (0.86, 2.51) \quad \text{di livello } 99\%$$

**Osservazione 1:** a parità di  $n$  l'ampiezza dell'intervallo cresce al crescere dell'intervallo di fiducia

**Osservazione 2:** a parità di  $n$  e di livello di fiducia  $1 - \alpha$  si hanno intervalli più ampi nel caso in cui la varianza non è nota

**Osservazione 3:** l'ampiezza di un intervallo di confidenza dipende solo da due quantità: l'ampiezza campionaria  $n$  e il livello di confidenza  $1 - \alpha$

## La giusta scelta dell'ampiezza campionaria

La lunghezza dell'intervallo di confidenza per la media nel caso in cui  $\sigma^2$  è noto

$$\begin{aligned} L(n, \alpha) &= \left( \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} - \left( \bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right) \right) \\ &= 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Mentre la lunghezza dell'intervallo di confidenza per la media nel caso in cui  $\sigma^2$  non è noto

$$\begin{aligned} L(n, \alpha) &= \left( \bar{X}_n + t_{1-\frac{\alpha}{2}}^{n-1} \sqrt{\frac{\bar{s}_n^2}{n}} - \left( \bar{X}_n - t_{1-\frac{\alpha}{2}}^{n-1} \sqrt{\frac{\bar{s}_n^2}{n}} \right) \right) \\ &= 2t_{1-\frac{\alpha}{2}}^{n-1} \sqrt{\frac{\bar{s}_n^2}{n}} \end{aligned}$$

$L(n, \alpha)$  non dipende dal valore assunto da  $\bar{X}_n$ . L'intervallo avrà sempre la stessa ampiezza a parità di ampiezza campionaria  $n$  e livello di confidenza  $1 - \alpha$

Si vuole trovare  $n$  tale per cui  $L(n, \alpha) < C$

$$L(n, \alpha) = 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} < C \quad 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{C} < \sqrt{n}$$

e quindi

$$n > \left( 2z_{1-\frac{\alpha}{2}} \frac{\sigma}{C} \right)^2$$

- Intervallo di confidenza per la proporzione incognita nel caso di prove Bernoulliane

Sia  $X$  una variabile casuale di Bernoulli di media  $p$  incognita. Se  $X_1, X_2, \dots, X_n$  è un campione i.i.d. estratto da  $X$ , allora  $\sum_{i=1}^n X_i \sim \text{Bin}(n, p)$ . Lo stimatore di  $p$ ,  $\hat{p}_n$  è una Binomiale moltiplicata per il fattore  $1/n$ . Il teorema del limite centrale applicato a  $\hat{p}_n$  ci dà

$$Z = \frac{\hat{p}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \sim N(0, 1)$$

Ma i calcoli per ottenere l'intervallo di confidenza per  $p$  a livello  $1 - \alpha$  ci danno

$$p \in \left( \hat{p}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right)$$

che non è possibile calcolare in alcun caso essendo  $p$  incognito

Se sostituiamo  $p$  con la sua stima  $\hat{p}_n$  si può mostrare che vale ancora l'approssimazione alla variabile Gaussiana e l'intervallo di confidenza per  $p$  di livello  $1 - \alpha$  è

$$p \in \left( \hat{p}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}} \right)$$

**Esempio:** su 150 pezzi prodotti da una macchina ve ne sono 18 non conformi. Costruire un intervallo di fiducia a livello  $1 - \alpha = 0.90$  per la percentuale di pezzi prodotti non in conformità

La stima della proporzione di pezzi non conformi è

$$\hat{p} = \frac{18}{150} = 0.12$$

L'intervallo di confidenza ha la forma

$$p \in \left( \hat{p}_n \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{p}_n(1 - \hat{p}_n)}{n}} \right)$$

dove

$$z_{1-\frac{\alpha}{2}} = z_{0.95} = 1.64$$

$$p \in \left( \hat{p}_n \pm 1.64 \sqrt{\frac{0.12(1 - 0.12)}{150}} \right)$$

$$p \in \left( 0.12 \pm 1.64 \sqrt{\frac{0.12(1 - 0.12)}{150}} \right)$$

$$p \in (0.076, 0.164)$$

è l'intervallo di confidenza per  $p$  con un grado di fiducia pari a 0.90



**Esercizio 1** un laboratorio produce molle per orologi con due macchine diverse denotate con  $A$  e  $B$ . La seguente tabella riporta i dati delle resistenze delle molle prodotte dalle due macchine.

A	22	25	24	25	25	27	26	26	22	26	23	25
B	23	24	23	26	24	22	25	24	24			

1. Fissata l'attenzione sul primo gruppo, e supponendo che la varianza sia nota e pari a  $\sigma^2 = 9$ , costruire un intervallo di confidenza per la media incognita con un livello di fiducia pari a  $1 - \alpha = 0.95$ .
2. Di quanto deve essere aumentata la numerosità campionaria affinché, a parità di tutte le condizioni, l'intervallo di confidenza sia ampio 2.5?
3. Supponendo che la varianza sia non nota, calcolare l'intervallo di confidenza con lo stesso livello di fiducia.
4. Risolvere i punti 1. e 3. precedenti per il secondo gruppo. Dire inoltre di quanto deve essere aumentata la numerosità campionaria affinché l'ampiezza dell'intervallo calcolata al punto 1. per il secondo gruppo raddoppi.

**Esercizio 2** La seguente tabella riporta i dati relativi ad una prova di produzione svolta in un'azienda tessile. Un particolare tipo di stoffa è stato sottoposto a due diversi trattamenti di tintura (denotati con T1 e T2) e si vuole vedere se le stoffe trattate con il trattamento T2 hanno sensibilmente ridotto la presenza di un fastidioso difetto.

	T1	T2
senza difetto	30	38
con difetto	10	2

1. Costruire un'intervallo di confidenza per  $p_2$ , cioè la proporzione di stoffe con difetto che hanno subito il trattamento 2, ad un livello di fiducia  $1 - \alpha = 0.80$ .
2. Costruire un'intervallo di confidenza per  $p_1$ , cioè la proporzione di stoffe con difetto che hanno subito il trattamento 1, ad un livello di fiducia  $1 - \alpha = 0.80$ .