



Facoltà di Ingegneria  
Università degli Studi di Firenze  
Dipartimento di Elettronica e Telecomunicazioni

# **Misure Elettriche**

## **- Inferenza statistica -**

### **3. Stima dei Parametri**

**Ing. Andrea Zanobini**

Dipartimento di Elettronica e Telecomunicazioni

# Stime puntuali e stime per intervallo

## Definizioni 1

Se la stima di un parametro della popolazione è data da un singolo numero, tale valore è detto **stima puntuale** del parametro.

Se invece la stima di un parametro della popolazione fornisce gli estremi di un intervallo fra i quali si può supporre, con un certo grado di fiducia, che il parametro sia compreso, tale stima è detta **stima per intervallo** del parametro.

I parametri che più frequentemente accade di dover stimare sono:

- 1 – la media  $\mu$  di una popolazione;
- 2 – la varianza  $\sigma^2$  di una popolazione;
- 3 – la proporzione  $p$  di individui di una popolazione che appartengono a una certa classe di interesse;

Ragionevoli stime puntuali di questi parametri sono:

- 1 – per  $\mu$ , la media campionaria  $\bar{x}$ ;
- 2 – per  $\sigma^2$ , la varianza campionaria  $s^2$ ;
- 3 – per  $p$ , la proporzione campionaria  $\hat{p} = \frac{x}{n}$ , dove  $x$  è il numero di individui in un campione di ampiezza  $n$  appartenenti alla classe di interesse;

# Proprietà degli Stimatori

proprietà della **correttezza o non distorsione**.

## **Definizione 2**

Se la media di una distribuzione campionaria di una statistica è uguale al corrispondente parametro della popolazione, la statistica è detta **stimatore corretto o non distorto** del parametro.

un'ulteriore proprietà, detta **efficienza**,

## **Definizione 3**

Se due statistiche sono entrambe stimatori corretti di un parametro, la statistica per cui la varianza della sua distribuzione campionaria è minore è detta **stimatore più efficiente**.

Si può dimostrare che, fra tutte le statistiche che stimano la media di una popolazione, la media campionaria è la più efficiente. Anche la varianza campionaria è una stima corretta ed efficiente della varianza di una popolazione.

### *Esempio 1*

Dato un campione di 5 misurazioni del diametro di una sferetta in cm

6.33    6.37    6.36    6.32    6.37

trovare stime corrette ed efficienti per la media e la varianza della popolazione.

La stima corretta ed efficiente per la media della popolazione è la media campionaria

$$\bar{x} = \frac{6.33 + 6.37 + 6.36 + 6.32 + 6.37}{5} = 6.35 \text{ cm}$$

Anche per la varianza la stima corretta ed efficiente è la varianza campionaria

$$s^2 = \frac{1}{4} \sum_{i=1}^5 (x_i - \bar{x})^2 = \frac{(0.02)^2 + (0.02)^2 + (0.01)^2 + (0.03)^2 + (0.02)^2}{4} = 0.00055 \text{ cm}^2$$

Poiché non ci si può aspettare che una stima puntuale coincida esattamente con la quantità che essa deve stimare, è spesso preferibile usare una stima per intervallo, ossia un intervallo per il quale si può affermare con un certo grado di fiducia che conterrà il parametro della popolazione che si vuole stimare.

## Intervalli di confidenza per la media (varianza nota)

Poiché nella pratica si estrae un solo campione, e ovviamente non conosciamo la media della popolazione, non possiamo essere certi che le conclusioni a cui si perviene siano corrette. Per risolvere questo problema, ogni stima per intervallo viene calcolata valutando anche la percentuale dei campioni che dà luogo a conclusioni corrette, ossia il **grado di fiducia**.

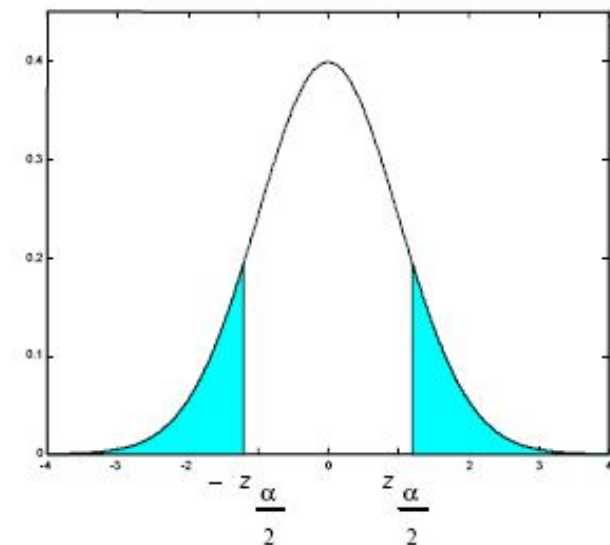
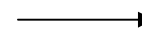
Si consideri una popolazione avente una distribuzione con varianza  $\sigma^2$  nota e media incognita  $\mu$ , e si estragga da questa popolazione un campione di ampiezza  $n$ .

In base al teorema del limite centrale possiamo affermare che, per grandi valori di  $n$ , la statistica

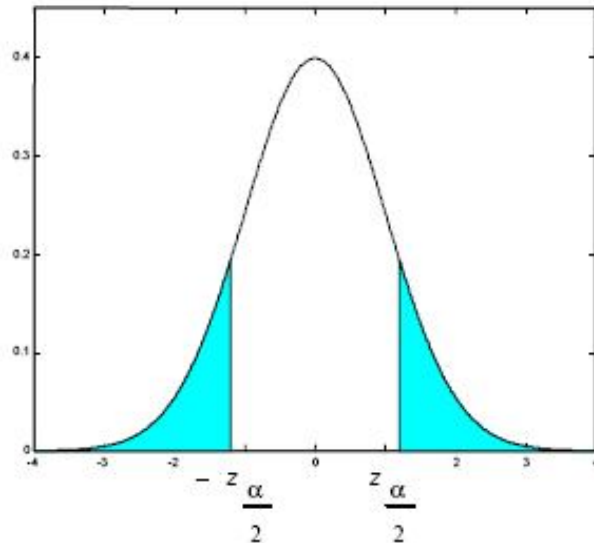
$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \quad (7.1)$$

ha approssimativamente la distribuzione normale standardizzata.

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$



Di conseguenza, si può asserire, con probabilità uguale a  $1 - \alpha$ , che è soddisfatta la disuguaglianza



$$-z_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}$$

$$\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Pertanto, una volta estratto il campione di ampiezza  $n$ , con  $n$  sufficientemente grande ( $n \geq 30$ , **grande campione**) e calcolato il valore  $\bar{x}$  della media del campione, si ottiene la seguente stima per intervallo per la media  $\mu$ , soddisfatta con probabilità  $1 - \alpha$ .

$$\bar{x} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad (7.3)$$

L'intervallo (7.3) è detto anche **intervallo di confidenza per la media  $\mu$ , per grandi campioni**, con **grado di fiducia**  $(1 - \alpha) \cdot 100\%$ .

## Valori critici e grado di fiducia

Il valore di  $z_{\frac{\alpha}{2}}$  che compare nella (7.3) è detto **valore critico** della distribuzione; a ciascun grado di fiducia corrisponde un diverso valore critico.

I valori più comunemente usati per  $1 - \alpha$  sono 0.90, 0.95 e 0.99 ; di solito si usa il termine **grado di fiducia** del 90%, del 95% o del 99% , anziché il termine probabilità uguale a 0.90, a 0.95, oppure a 0.99; i corrispondenti valori di  $z_{\frac{\alpha}{2}}$  sono i seguenti

|                          |                                            |
|--------------------------|--------------------------------------------|
| grado di fiducia del 90% | $z_{\frac{\alpha}{2}} = z_{0.05} = 1.645$  |
| grado di fiducia del 95% | $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$  |
| grado di fiducia del 99% | $z_{\frac{\alpha}{2}} = z_{0.005} = 2.576$ |

## Compromesso tra grado di fiducia e dimensione del campione

La lunghezza di un intervallo di confidenza con grado di fiducia  $(1 - \alpha) \cdot 100\%$  è

$$2z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

e dipende quindi da tre fattori

- $n$ : al crescere dell'ampiezza del campione, la lunghezza dell'intervallo diminuisce, quindi la stima è più precisa;
- $\alpha$ : al crescere del grado di fiducia richiesto, la lunghezza dell'intervallo aumenta, quindi la stima è meno precisa;
- $\sigma$ : al crescere della deviazione standard, che riflette la variabilità del campione, la lunghezza dell'intervallo aumenta.

Normalmente solo  $n$  e  $\alpha$  possono essere controllati, mentre  $\sigma$  dipende dal tipo di dati studiati.

Nelle applicazioni pratiche può non essere facile trovare un buon compromesso tra grado di fiducia e ampiezza del campione: un maggior grado di fiducia comporta una perdita di precisione nella stima; un aumento delle dimensioni del campione può comportare problemi pratici o essere antieconomico. Solo l'esperienza e la conoscenza del problema trattato possono indicare la scelta più opportuna.



**Esempio 2**

Sia dato un campione di ampiezza  $n = 100$  estratto da una popolazione avente scarto quadratico medio  $\sigma = 5.1$ ; la media campionaria sia  $\bar{x} = 21.6$ . Costruire l'intervallo di confidenza al 95% per la media  $\mu$  della popolazione.

Per il grado di fiducia del 95% il valore critico è  $z_{\frac{\alpha}{2}} = z_{0.025} = 1.96$

Applicando la formula (7.3) si ottiene l'intervallo di confidenza

$$21.6 - 1.96 \cdot \frac{5.1}{\sqrt{100}} < \mu < 21.6 + 1.96 \cdot \frac{5.1}{\sqrt{100}}$$
$$20.6 < \mu < 22.6$$

Questo intervallo può anche non contenere  $\mu$ , ma abbiamo un grado di fiducia del 95% che lo contenga. In altre parole, se applichiamo ripetutamente su tutti i campioni di ampiezza  $n = 100$  estraibili dalla popolazione la formula (7.3) per calcolare l'intervallo di confidenza, il 95% degli intervalli di confidenza conterrà la media  $\mu$  della popolazione.

**Fare da soli l' esempio 3, 5 e 6 !!**

#### **Esempio 4**

Le misure dei diametri di un campione casuale di 200 sferette da cuscinetto prodotte da una macchina in una settimana hanno una media campionaria  $\bar{x} = 0.824$  cm e una deviazione standard campionaria  $s = 0.042$  cm.

Determinare gli intervalli di confidenza per la media della popolazione con grado di fiducia del 95% e del 99%.

a – Per il grado di fiducia del 95% il valore critico è  $z_{\frac{\alpha}{2}} = 1.96$

Con la formula (7.3) si ottiene l'intervallo di confidenza

$$0.824 - 1.96 \cdot \frac{0.042}{\sqrt{200}} < \mu < 0.824 + 1.96 \cdot \frac{0.042}{\sqrt{200}}$$
$$0.818 < \mu < 0.830$$

b – Per il grado di fiducia del 99% il valore critico è  $z_{\frac{\alpha}{2}} = 2.576$

Con la formula (7.3) si ottiene l'intervallo di confidenza

$$0.824 - 2.576 \cdot \frac{0.042}{\sqrt{200}} < \mu < 0.824 + 2.576 \cdot \frac{0.042}{\sqrt{200}}$$
$$0.816 < \mu < 0.832$$

Si osservi che aumentando il grado di fiducia l'ampiezza dell'intervallo aumenta, ossia a parità di numero di elementi del campione la stima è meno precisa.

## Ampiezza del campione

La disuguaglianza (7.2), valida con probabilità  $1 - \alpha$ , può anche essere usata per ricavare una formula che consente di determinare l'**ampiezza  $n$  del campione** necessaria per ottenere un errore prefissato.

$$\frac{|\bar{X} - \mu|}{\frac{\sigma}{\sqrt{n}}} < z_{\frac{\alpha}{2}}$$

→

$$|\bar{X} - \mu| < z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$

Indicando il massimo dell'errore con  
 $E = \max |\bar{X} - \mu|$

la stima di  $E$  con probabilità  $1 - \alpha$  è

$$E = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \quad (7.4)$$

Dalla formula (7.4), risolvendo rispetto a  $n$ , si ricava l'ampiezza del campione necessaria per stimare la media con un errore prefissato  $E$  e con un dato grado di fiducia (si ricordi che  $n$  deve essere un intero)

$$n \geq \left( \frac{z_{\frac{\alpha}{2}} \sigma}{E} \right)^2 \quad (7.5)$$

**Esempio 7**

Determinare l'ampiezza campionaria che consente di ottenere un intervallo di confidenza per la media  $\mu$  di una popolazione con grado di fiducia del 95%, con un errore in valore assoluto non superiore a 5, supponendo che lo scarto quadratico medio sia  $\sigma = 15$ .

Per il grado di fiducia del 95% il valore critico è  $z_{\frac{\alpha}{2}} = 1.96$ . Con la formula (7.5) si ottiene

$$n \geq \left( \frac{1.96 \cdot 15}{5} \right)^2 = 34.6$$

Per ottenere la stima con la precisione fissata e con grado di fiducia del 95%, occorre scegliere un campione di ampiezza  $n = 35$ .

Per il grado di fiducia del 99% il valore critico è invece  $z_{\frac{\alpha}{2}} = 2.576$ . Con la formula (7.5) si ottiene

$$n \geq \left( \frac{2.576 \cdot 15}{5} \right)^2 = 59.7$$

Per ottenere la stima con la precisione fissata e con grado di fiducia del 95%, occorre scegliere un campione di ampiezza  $n = 60$ .

**Fare da soli l' esempio 8 !!**

## Intervalli di confidenza per la media (varianza incognita)

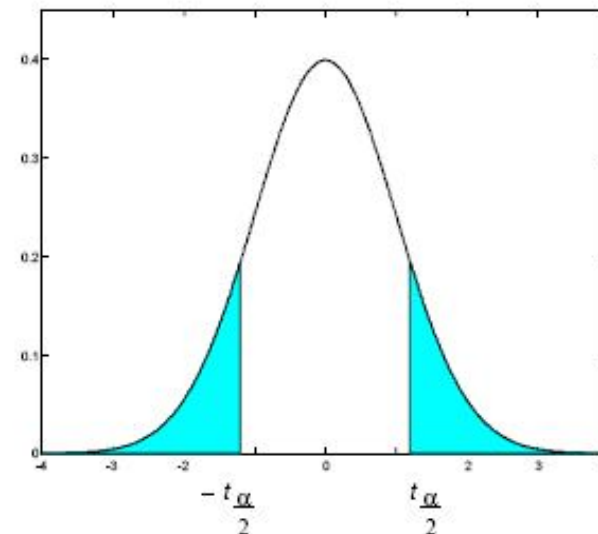
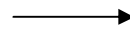
L'applicazione della (7.3) richiede la conoscenza di  $\sigma$ ; se  $\sigma$  non è noto, si è già osservato che per grandi campioni può essere sostituito con lo scarto quadratico medio campionario  $s$ .

Per **piccoli campioni** ( $n < 30$ ), nell'ipotesi che la popolazione da cui si estrae il campione abbia distribuzione normale, ci si può servire del teorema 3, Cap. 6, in base al quale la statistica

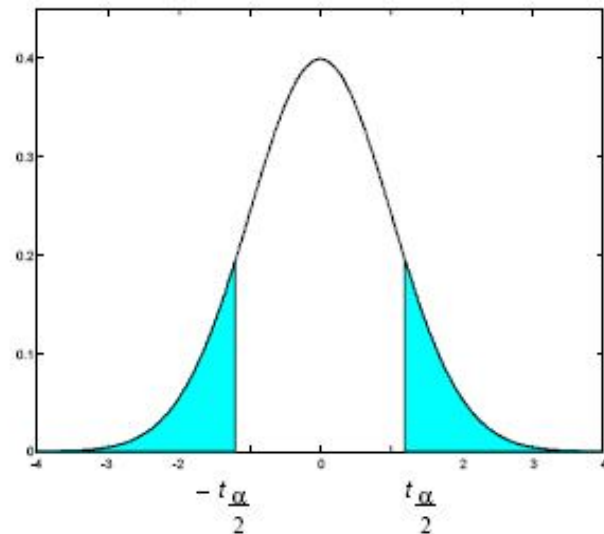
$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} \quad (7.6)$$

è una variabile aleatoria che ha la distribuzione  $t$  di Student con grado di libertà  $\nu = n - 1$ .

$$P\left(-t_{\frac{\alpha}{2}} < Z < t_{\frac{\alpha}{2}}\right) = 1 - \alpha$$



In altre parole si può asserire, con probabilità uguale a  $1 - \alpha$ , che è soddisfatta la disuguaglianza



$$-t_{\frac{\alpha}{2}} < \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}} < t_{\frac{\alpha}{2}}$$

Pertanto, una volta estratto il campione di ampiezza  $n$ , con  $n < 30$ , e calcolati i valori della media  $\bar{x}$  e dello scarto quadratico medio  $s$  del campione, si ottiene la stima per intervallo per la media  $\mu$ , con probabilità  $1 - \alpha$ , o con grado di fiducia  $(1 - \alpha) \cdot 100\%$

$$\bar{x} - t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \quad (7.8)$$

L'intervallo (7.8) è detto **intervallo di confidenza per la media  $\mu$ , per piccoli campioni**, con grado di fiducia  $(1 - \alpha) \cdot 100\%$ .

## Valori critici e grado di fiducia

Il valore di  $t_{\frac{\alpha}{2}}$  che compare nella (7.8) è detto **valore critico** della distribuzione; a ciascun grado di fiducia corrisponde un diverso valore critico, e diversamente dal caso dei grandi campioni, tale valore dipende anche dal grado di libertà della distribuzione  $t$ .

I valori più comunemente usati per  $1 - \alpha$  sono 0.90, 0.95 e 0.99 ; i relativi gradi di fiducia sono il 90%, il 95% e il 99%; i corrispondenti valori di  $t_{\frac{\alpha}{2}}$  sono

|                          |                                    |
|--------------------------|------------------------------------|
| grado di fiducia del 90% | $t_{\frac{\alpha}{2}} = t_{0.05}$  |
| grado di fiducia del 95% | $t_{\frac{\alpha}{2}} = t_{0.025}$ |
| grado di fiducia del 99% | $t_{\frac{\alpha}{2}} = t_{0.005}$ |

Questi valori possono essere letti sulla tabella della distribuzione  $t$  in corrispondenza al grado di libertà  $v = n - 1$ .

### **Esempio 9**

Sia dato un campione di 16 oggetti di cui si misura il peso, trovando un peso medio  $\bar{x} = 3.42$  g e uno scarto quadratico medio  $s = 0.68$  g.

Determinare un intervallo di confidenza con grado di fiducia del 99% per il peso medio della popolazione.

Poiché si tratta di misure, si può ragionevolmente ipotizzare che la popolazione da cui proviene il campione abbia distribuzione normale.

Il campione ha ampiezza  $n = 16$ , perciò il grado di libertà è

$$v = n - 1 = 15.$$

Dalle tavole della distribuzione  $t$  si ottiene

$$t_{0.005} = 2.947.$$

Con la formula (7.8) si ottiene l'intervallo di confidenza

$$3.42 - 2.947 \cdot \frac{0.68}{\sqrt{16}} < \mu < 3.42 + 2.947 \cdot \frac{0.68}{\sqrt{16}}$$
$$2.91 < \mu < 3.93$$

**Fare da soli l' esempio 10 !!**



**Esempio 11**

Le misure in kg del peso di un campione di 10 studenti maschi del primo anno di un'università sono

|    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|
| 60 | 63 | 60 | 68 | 70 | 72 | 65 | 61 | 69 | 67 |
|----|----|----|----|----|----|----|----|----|----|

Trovare un intervallo di confidenza con grado di fiducia del 99% per il peso medio della popolazione universitaria maschile del primo anno di quella università.

Calcoliamo la media e la varianza campionaria

$$\bar{x} = \frac{60 + 63 + 60 + 68 + 70 + 72 + 65 + 61 + 69 + 67}{10} = 65.5$$

$$s^2 = \frac{1}{9} \cdot [60^2 + 63^2 + 60^2 + 68^2 + 70^2 + 72^2 + 65^2 + 61^2 + 69^2 + 67^2 - 10 \cdot 65.5^2] = 18.94$$

Il campione ha ampiezza  $n = 10$ , perciò il grado di libertà è  $v = n - 1 = 9$ .

Per il grado di fiducia del 99% e il grado di libertà  $v = 9$ , si ha  $t_{\frac{\alpha}{2}} = t_{0.005} = 3.250$

$$65.5 - 3.250 \cdot \frac{\sqrt{18.94}}{\sqrt{10}} < \mu < 65.5 + 3.250 \cdot \frac{\sqrt{18.94}}{\sqrt{10}}$$
$$61.02 < \mu < 69.98$$

## ...un'osservazione abbastanza importante

Come già osservato (Cap. 1, pag. 23), la media è sensibile ai valori estremi, ossia quelli che si discostano in modo quantitativamente apprezzabile dalla maggior parte dei dati dell'insieme. Questi valori sono a volte chiamati **outliers**.

Gli stimatori che non sono sensibili agli outliers sono chiamati **stimatori robusti**. Un altro stimatore robusto per la tendenza centrale è la **media trimmed**.

Dato un campione di  $n$  dati, la media trimmed al  $q\%$  si calcola come segue:

1 – si ordinano i dati;

2 – si eliminano i  $q\%$  dati più piccoli e i  $q\%$  dati più grandi. I valori di solito usati sono  $q\% = 10\%$  o  $q\% = 20\%$ ;

3 – si calcola la media dei rimanenti dati.

Dati ordinati

8.5   8.7   9.4   9.8   10.3   11.6   12.1   12.8   13.1   14.1

Mediana

$$M = \frac{10.3 + 11.6}{2} = 10.95$$

$$\text{Media} = 11.04$$

$$\bar{x}_{tr(10)} = \frac{8.7 + 9.4 + 9.8 + 10.3 + 11.6 + 12.1 + 12.8 + 13.1}{8} = 10.98$$

## Intervalli di confidenza per la proporzione

Un caso particolarmente importante di stima della media per una popolazione non normale e per grandi campioni è quello di una popolazione bernoulliana. Si vuole stimare il valore del parametro  $p$  (probabilità di successo), che rappresenta la frequenza relativa o proporzione con cui una certa caratteristica si presenta negli individui di una data popolazione.

Esempi tipici di questa situazione sono i seguenti.

- 1 – Il sondaggio di opinione: si vuole stimare la proporzione  $p$  della popolazione complessiva che è d'accordo con una certa opinione, osservando il valore che questa proporzione ha su un campione di  $n$  individui.
- 2 – La produzione di un dato tipo di oggetto: il produttore vuole poter garantire che la proporzione di pezzi difettosi in una data produzione non superi un certo valore prefissato; occorre quindi determinare, esaminando un campione, un intervallo di confidenza per la proporzione  $p$  di pezzi difettosi in una produzione, ed eventualmente intervenire sulla produzione affinché la proporzione di pezzi difettosi non superi una certa soglia fissata.
- 3 – Lo studio della diffusione di una data malattia: si vuole stimare qual è la proporzione di pazienti di una certa popolazione che ha una data malattia, studiando il valore di questa proporzione su un campione di  $n$  persone appartenenti a quella popolazione.

## Definizione di una statistica per la proporzione

Si estraggono campioni di ampiezza  $n$  dalla popolazione e si considera la proporzione campionaria  $\hat{P} = \frac{X}{n}$ , dove  $X$  è il numero di volte in cui la caratteristica osservata si presenta nel campione.

Questa proporzione campionaria è uno stimatore corretto della proporzione  $p$  della popolazione e viene usato come stima puntuale.

$$Z = \frac{X - np}{\sqrt{np(1-p)}} = \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

ha approssimativamente la distribuzione normale standardizzata per grandi valori di  $n$ .

Quindi quando  $n$  è grande si può costruire un intervallo di confidenza per il parametro  $p$ , usando l'approssimazione normale per la distribuzione binomiale.

$$P\left(-z_{\frac{\alpha}{2}} < Z < z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

→

$$-z_{\frac{\alpha}{2}} < \frac{\hat{P} - p}{\sqrt{\frac{p(1-p)}{n}}} < z_{\frac{\alpha}{2}}$$

In questo modo, estraendo un campione di ampiezza  $n$  da una popolazione bernoulliana e indicando con  $\hat{p}$  la proporzione del campione, si ottiene il seguente **intervallo di confidenza per la proporzione  $p$**  della popolazione bernoulliana, con grado di fiducia  $(1-\alpha)\cdot 100\%$ , valido per grandi campioni.

$$\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} < p < \hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (7.11)$$

Osserviamo che per ottenere l'intervallo di confidenza (7.11) sono state fatte tre approssimazioni:

1 – l'approssimazione normale della binomiale;

2 – l'approssimazione di  $p$  con  $\hat{p} = \frac{x}{n}$ , nell'espressione  $\sqrt{np(1-p)}$ ;

3 – non è stata fatta la correzione di continuità per l'approssimazione normale!

Questo implica che l'intervallo di confidenza trovato è un intervallo approssimato.

Per verificare le condizioni di applicabilità dell'approssimazione della binomiale con la normale, ossia  $np \geq 5$  e  $n(1-p) \geq 5$ , possiamo solo verificare che sia  $n\hat{p} \geq 5$  e  $n(1-\hat{p}) \geq 5$ ; questa verifica si può fare solo dopo aver effettuato il campionamento: se le condizioni precedenti non sono soddisfatte, il risultato è privo di valore, e occorre ripetere il campionamento aumentando l'ampiezza  $n$  del campione.

**Esempio 13**

In un campione di 400 persone a cui è stato somministrato un dato vaccino, 136 di esse hanno avuto effetti collaterali di un certo rilievo. Determinare un intervallo di confidenza con grado di fiducia del 95% per la proporzione della popolazione che soffre di tali effetti collaterali.

Nel campione di  $n = 400$  persone la proporzione campionaria è

$$\hat{p} = \frac{136}{400} = 0.34$$

Per il grado di fiducia del 95% il valore critico è  $z_{\frac{\alpha}{2}} = 1.96$  e con la formula (7.11) si trova

l'intervallo di confidenza

$$0.34 - 1.96 \cdot \sqrt{\frac{0.34 \cdot (1 - 0.34)}{400}} < p < 0.34 + 1.96 \cdot \sqrt{\frac{0.34 \cdot (1 - 0.34)}{400}}$$
$$0.29 < p < 0.39$$

Osserviamo che le condizioni per poter usare l'approssimazione della binomiale con la normale sono verificate, essendo

$$n\hat{p} = 400 \cdot 0.34 = 135 \quad \text{e} \quad n(1 - \hat{p}) = 400 \cdot 0.66 = 264.$$

**Fare da soli l' esempio 14 !!**

## Ampiezza del campione

Con lo stesso procedimento già usato nel caso dell'intervallo di confidenza per la media di un grande campione, si può usare la disuguaglianza (7.10), valida con probabilità  $1-\alpha$ , per ricavare una formula che consente di determinare l'**ampiezza  $n$  del campione** necessaria per ottenere un errore prefissato.

$$\frac{|\hat{P} - p|}{\sqrt{\frac{p(1-p)}{n}}} < z_{\frac{\alpha}{2}} \longrightarrow |\hat{P} - p| < z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

Indicando con

$$E = \max |\hat{P} - p|$$

il massimo dell'errore che si commette approssimando la proporzione della popolazione  $p$  con la proporzione campionaria  $\hat{P} = \frac{X}{n}$ , la stima di  $E$  con probabilità  $1 - \alpha$  è data da

$$E = z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{p(1-p)}{n}} \quad (7.12)$$

In altre parole, se si vuole stimare la proporzione  $p$  della popolazione con la proporzione campionaria  $\hat{p} = \frac{x}{n}$  di un campione di ampiezza  $n$  ( $n \geq 30$ ), si può affermare, con probabilità  $1-\alpha$ , che l'errore  $\left| \frac{X}{n} - p \right|$  sarà al più uguale a  $\frac{z_{\alpha}}{2} \cdot \sqrt{\frac{p(1-p)}{n}}$ .

Dalla formula (7.12), risolvendo rispetto a  $n$ , si ricava l'ampiezza del campione necessaria per stimare la proporzione  $p$  con un errore prefissato  $E$  e con un dato grado di fiducia (si ricordi che  $n$  deve essere un intero)

$$n \geq p(1-p) \left( \frac{z_{\alpha}}{2E} \right)^2 \quad (7.13)$$

Questa formula non può essere usata se non si ha qualche informazione sul valore di  $p$ ; se tali informazioni non sono disponibili, si può far uso del fatto che il valore massimo<sup>2</sup> che può assumere la quantità  $p(1-p)$  è  $\frac{1}{4}$ , corrispondente a  $p = \frac{1}{2}$ .

$$n \geq \frac{1}{4} \left( \frac{z_{\alpha}}{2E} \right)^2 \quad (7.14)$$



**Esempio 15**

Problema del sondaggio di opinione. Supponiamo che si voglia stimare la proporzione di elettori che approva l'operato del capo del governo; su un campione di 150 persone intervistate, 90 si sono dichiarate favorevoli.

Determinare un intervallo di confidenza con grado di fiducia del 95% per la proporzione degli elettori favorevoli al capo del governo e valutare la precisione della stima.

La proporzione campionaria dei favorevoli è

$$\hat{p} = \frac{x}{n} = \frac{90}{150} = 0.6$$

L'intervallo di confidenza con grado di fiducia del 95% è il seguente

$$0.6 - 1.96 \cdot \sqrt{\frac{0.6 \cdot (1 - 0.6)}{150}} < p < 0.6 + 1.96 \cdot \sqrt{\frac{0.6 \cdot (1 - 0.6)}{150}}$$
$$0.52 < p < 0.68$$

La percentuale dei favorevoli, con un grado di fiducia del 95%, è compresa fra il 52% e il 68%: la stima è troppo imprecisa, l'ampiezza dell'intervallo è di 16 punti percentuali.

Può quindi essere utile determinare l'ampiezza del campione necessaria per ottenere una stima con precisione fissata. Stabiliamo ad esempio che si vuole una stima con una precisione dell'1% (corrispondente a un'ampiezza dell'intervallo non superiore a 2 punti percentuali), ossia fissiamo

$$E = 0.01.$$

$$n \geq \frac{1}{4} \left( \frac{1.96}{0.01} \right)^2 = 9604$$

### **Esempio 16**

Supponiamo di voler stimare la proporzione di pezzi difettosi in un lotto di oggetti di un dato tipo con un errore  $E = 0.04$  e un grado di fiducia del 95%; calcolare l'ampiezza necessaria per il campione, nel caso che

- a – non si abbia alcuna informazione su quale possa essere la proporzione effettiva della popolazione;
- b – si sappia che la proporzione della popolazione non supera il 12%.

a – Se non si ha alcuna informazione su  $p$ , si usa la formula (7.14), e con grado di fiducia del 95% si ricava

$$n \geq \frac{1}{4} \left( \frac{1.96}{0.04} \right)^2 = 600.3$$

Occorre quindi un campione di ampiezza  $n = 601$ .

b – Se sappiamo che  $p \leq 0.12$ , con la formula (7.13) e con grado di fiducia del 95% si ottiene

$$n \geq 0.12(1 - 0.12) \left( \frac{1.96}{0.04} \right)^2 = 253.5 .$$

Occorre in questo caso un campione di ampiezza  $n = 254$ .

Questo esempio illustra come il fatto di avere qualche informazione sul possibile valore della proporzione può sensibilmente ridurre la dimensione del campione.

## Intervalli di confidenza per la varianza e per lo scarto quadratico medio

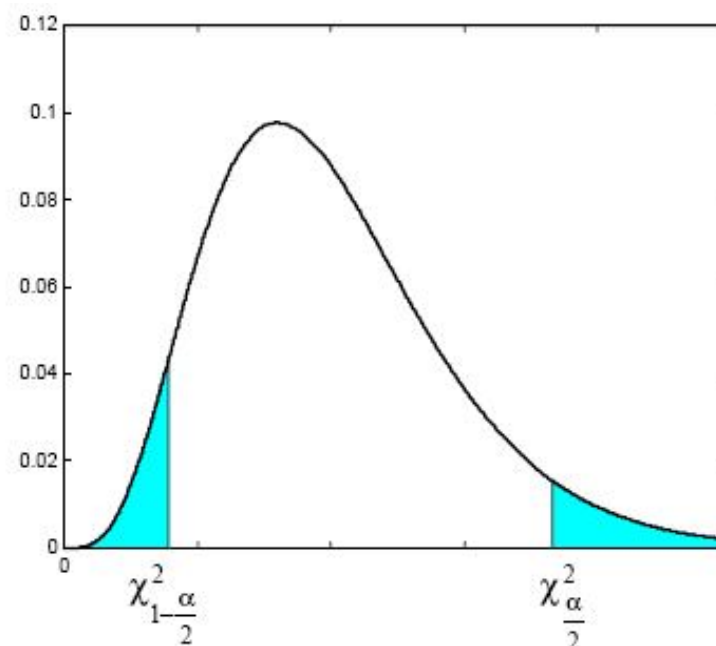
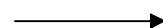
Si consideri una popolazione avente distribuzione normale, e si estraggano da questa popolazione campioni di ampiezza  $n$ .

In base al teorema 4, Cap. 6, si può affermare che la statistica

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \quad (7.23)$$

ha la distribuzione  $\chi^2$  con grado di libertà  $\nu = n - 1$ .

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



In altre parole si può asserire con probabilità  $1 - \alpha$ , ossia con grado di fiducia  $(1 - \alpha) \cdot 100\%$ , che vale la disuguaglianza

$$\chi^2_{1-\frac{\alpha}{2}} < \frac{(n-1) \cdot S^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}$$

Pertanto, indicando con  $s^2$  la varianza di un campione di ampiezza  $n$  estratto da una popolazione normale, e risolvendo questa disuguaglianza rispetto a  $\sigma^2$  si ottiene l'**intervallo di confidenza per la varianza  $\sigma^2$**  con grado di fiducia  $(1 - \alpha) \cdot 100\%$

$$\frac{(n-1) \cdot s^2}{\chi^2_{\frac{\alpha}{2}}} < \sigma^2 < \frac{(n-1) \cdot s^2}{\chi^2_{1-\frac{\alpha}{2}}} \quad (7.24)$$

Estraendo la radice quadrata di ciascun membro della disuguaglianza, si ottiene l'intervallo di confidenza per lo scarto quadratico medio.

## Osservazioni, valori critici e grado di fiducia

L'assunzione che la popolazione abbia distribuzione normale è molto importante: infatti i risultati ottenuti ignorando tale ipotesi possono portare a gravi errori.

Si osservi inoltre che l'intervallo di confidenza non è simmetrico, come invece accade per gli intervalli di confidenza per la media o per la proporzione; ciò è dovuto al fatto che la distribuzione  $\chi^2$  non è simmetrica.

I valori più comunemente usati per  $1 - \alpha$  sono 0.90, 0.95 e 0.99, a cui corrispondono i gradi di fiducia del 90%, del 95% e del 99%; i corrispondenti valori di  $\chi_{\frac{\alpha}{2}}^2$  e di  $\chi_{1-\frac{\alpha}{2}}^2$  sono

|                          |                                              |                                                |
|--------------------------|----------------------------------------------|------------------------------------------------|
| grado di fiducia del 90% | $\chi_{\frac{\alpha}{2}}^2 = \chi_{0.05}^2$  | $\chi_{1-\frac{\alpha}{2}}^2 = \chi_{0.95}^2$  |
| grado di fiducia del 95% | $\chi_{\frac{\alpha}{2}}^2 = \chi_{0.025}^2$ | $\chi_{1-\frac{\alpha}{2}}^2 = \chi_{0.975}^2$ |
| grado di fiducia del 99% | $\chi_{\frac{\alpha}{2}}^2 = \chi_{0.005}^2$ | $\chi_{1-\frac{\alpha}{2}}^2 = \chi_{0.995}^2$ |

### **Esempio 22**

In una scuola è stato scelto a caso un campione di 16 studenti dell'ultimo anno e si è misurata l'altezza di ciascuno di essi. La varianza campionaria della misura delle altezze è  $s^2 = 37.09$  cm. Trovare gli intervalli di confidenza al 95% e al 99% per la varianza della popolazione costituita da tutti gli studenti dell'ultimo anno della scuola.

a – Per il grado di fiducia del 95% e il grado di libertà  $v = n - 1 = 15$ , si ha

$$\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.975} = 6.262 \qquad \chi^2_{\frac{\alpha}{2}} = \chi^2_{0.025} = 27.488$$

Con la formula (7.24) si ottiene l'intervallo di confidenza per la varianza con grado di fiducia del 95%

$$\frac{15 \cdot 37.09}{27.488} < \sigma^2 < \frac{15 \cdot 37.09}{6.262}$$
$$20.23 < \sigma^2 < 88.84$$

b – Per il grado di fiducia del 99% e il grado di libertà  $v = n - 1 = 15$ , si ha

$$\chi^2_{1-\frac{\alpha}{2}} = \chi^2_{0.995} = 4.601 \qquad \chi^2_{\frac{\alpha}{2}} = \chi^2_{0.005} = 32.801$$

Con la formula (7.24) si ottiene l'intervallo di confidenza per la varianza con grado di fiducia del 99%

$$\frac{15 \cdot 37.09}{32.801} < \sigma^2 < \frac{15 \cdot 37.09}{4.601}$$
$$16.96 < \sigma^2 < 120.92$$

**Fare da soli l'  
esempio 23!!**

Il corrispondente intervallo di confidenza per lo scarto quadratico medio è

$$4.11 < \sigma < 11.00$$

**Esempio 24**

Le misure della durata in ore di 10 batterie sono le seguenti

|     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 140 | 136 | 150 | 144 | 148 | 152 | 138 | 141 | 143 | 151 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

Trovare un intervallo di confidenza al 99% per la varianza e per lo scarto quadratico medio della popolazione.

Calcoliamo la media campionaria e la varianza campionaria

$$\bar{x} = \frac{140 + 136 + 150 + 144 + 148 + 152 + 138 + 141 + 143 + 151}{10} = 144.3$$

$$s^2 = \frac{1}{9} [140^2 + 136^2 + 150^2 + 144^2 + 148^2 + 152^2 + 138^2 + 141^2 + 143^2 + 151^2 - 10 \cdot 144.3^2] = 32.233$$

Per il grado di fiducia del 99% e il grado di libertà  $v = n - 1 = 9$ , si ha

$$\chi_{\frac{1-\alpha}{2}}^2 = \chi_{0.995}^2 = 1.735 \quad \chi_{\frac{\alpha}{2}}^2 = \chi_{0.005}^2 = 23.589$$

Con la formula (7.24) si ottiene l'intervallo di confidenza per la varianza con grado di fiducia del 99%

$$\frac{9 \cdot 32.233}{23.589} < \sigma^2 < \frac{9 \cdot 32.233}{1.735}$$

$$12.29 < \sigma^2 < 167.21$$

Per lo scarto quadratico medio si ha

$$3.50 < \sigma < 12.94.$$

**Fare da soli l'  
esempio 25!!**

## Intervallo di confidenza per s.q.m. - Caso di grandi campioni

Per grandi campioni estratti da una popolazione normale, si può dimostrare che la distribuzione campionaria della deviazione standard  $\sigma$  può essere approssimata con una distribuzione normale avente media  $\sigma$  e deviazione standard  $\frac{\sigma}{\sqrt{2n}}$ , ossia la statistica

$$Z = \frac{S - \sigma}{\frac{\sigma}{\sqrt{2n}}}$$

ha approssimativamente la distribuzione normale standardizzata, per  $n$  sufficientemente grande. Si può pertanto asserire che, con probabilità  $1 - \alpha$ , vale la disuguaglianza

Risolvendo la disuguaglianza rispetto a  $\sigma$ , e indicando con  $s$  lo scarto quadratico medio di un campione di ampiezza  $n$ , si trova l'**intervallo di confidenza per lo scarto quadratico medio  $\sigma$** , per **grandi campioni**, con probabilità  $1 - \alpha$ , o con grado di fiducia  $(1 - \alpha) \cdot 100\%$

$$\frac{s}{1 + \frac{z_{\frac{\alpha}{2}}}{\sqrt{2n}}} < \sigma < \frac{s}{1 - \frac{z_{\frac{\alpha}{2}}}{\sqrt{2n}}} \quad (7.26)$$



**Esempio 28**

Un campione di 32 misurazioni del punto di bollitura di una sostanza chimica ha scarto quadratico medio  $s = 0.83^\circ\text{C}$ .

Determinare un intervallo di confidenza al 99% per lo scarto quadratico medio  $\sigma$ .

Poiché l'ampiezza del campione è  $n = 32$ , si può usare la formula (7.26); per il grado di fiducia del 99% si ha  $z_{\frac{\alpha}{2}} = 2.576$ .

L'intervallo di confidenza per lo scarto quadratico medio  $\sigma$  è

$$\frac{0.83}{1 + \frac{2.576}{\sqrt{64}}} < \sigma < \frac{0.83}{1 - \frac{2.576}{\sqrt{64}}}$$
$$0.62 < \sigma < 1.23$$

**Rifare questo  
esempio con la  
chi-quadro!!**

**Fare da soli gli  
esempi 26 e 27!!**

## Intervalli di confidenza per il rapporto di due varianze

Per confrontare fra loro due varianze si costruisce il loro rapporto

$$\frac{\sigma_1^2}{\sigma_2^2}$$

Se le due varianze sono uguali, il loro rapporto sarà uguale a 1; di solito però non si conoscono le varianze delle popolazioni studiate, e il confronto avverrà sulla base delle varianze campionarie, ossia si procede a una stima del rapporto delle varianze delle due popolazioni.

Si considerino due popolazioni avente distribuzione normale, e si estraggano da queste popolazioni campioni indipendenti di ampiezza rispettivamente  $n_1$  e  $n_2$ . Le varianze campionarie siano rispettivamente  $S_1^2$  e  $S_2^2$ .

In base al teorema 5, Cap. 6, si può affermare che la statistica

$$F = \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \quad (7.27)$$

ha la **distribuzione  $F$**  di parametri  $v_1 = n_1 - 1$  e  $v_2 = n_2 - 1$ .

Si può osservare che la distribuzione  $F$  non è simmetrica perciò, con lo stesso tipo di procedimento già utilizzato per ricavare gli intervalli di confidenza per la varianza, usando code di uguale area e indicando con  $\frac{\alpha}{2}$  l'area di ciascuna coda, si ha che

$$P\left(F_{1-\frac{\alpha}{2}} < \frac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} < F_{\frac{\alpha}{2}}\right) = 1 - \alpha \quad \longrightarrow \quad F_{1-\frac{\alpha}{2}} < \frac{S_1^2}{S_2^2} \frac{\sigma_2^2}{\sigma_1^2} < F_{\frac{\alpha}{2}}$$

Risolvendo questa disuguaglianza rispetto a  $\frac{\sigma_2^2}{\sigma_1^2}$  si ha

e prendendo i reciproci dei tre termini si ha

$$\frac{S_2^2}{S_1^2} F_{1-\frac{\alpha}{2}} < \frac{\sigma_2^2}{\sigma_1^2} < \frac{S_2^2}{S_1^2} F_{\frac{\alpha}{2}} \quad \longrightarrow \quad \frac{S_1^2}{S_2^2} \frac{1}{F_{\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{S_1^2}{S_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}}$$

Pertanto, estraendo due campioni indipendenti di ampiezza  $n_1$  e  $n_2$  da due popolazioni normali e indicando con  $s_1^2$  e  $s_2^2$  le varianze dei due campioni, dove  $s_1^2$  è la più grande delle due varianze, si ottiene l'intervallo di confidenza per il rapporto di due varianze  $\frac{\sigma_1^2}{\sigma_2^2}$  con grado di fiducia  $(1 - \alpha) \cdot 100\%$

$$\frac{s_1^2}{s_2^2} \frac{1}{F_{\frac{\alpha}{2}}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} \frac{1}{F_{1-\frac{\alpha}{2}}} \quad (7.28)$$

**Si ricorda la seguente formula  
nell'utilizzo delle tavole**

$$F_{1-\frac{\alpha}{2}}(v_1, v_2) = \frac{1}{F_{\frac{\alpha}{2}}(v_2, v_1)} \quad (7.29)$$

## Osservazioni, valori critici e grado di fiducia

Il metodo descritto per trovare gli intervalli di confidenza per il rapporto di due varianze si applica solo a campioni estratti da popolazioni normali.

Anche in questo caso la verifica dell'ipotesi di normalità delle due popolazioni è di grande importanza.

I valori più comunemente usati per  $1 - \alpha$  sono 0.90, 0.95 e 0.99, a cui corrispondono i gradi di fiducia del 90%, del 95% e del 99%; i corrispondenti valori di  $F_{\frac{\alpha}{2}}$  e di  $F_{1-\frac{\alpha}{2}}$  sono

|                          |                                    |                                      |
|--------------------------|------------------------------------|--------------------------------------|
| grado di fiducia del 90% | $F_{\frac{\alpha}{2}} = F_{0.05}$  | $F_{1-\frac{\alpha}{2}} = F_{0.95}$  |
| grado di fiducia del 95% | $F_{\frac{\alpha}{2}} = F_{0.025}$ | $F_{1-\frac{\alpha}{2}} = F_{0.975}$ |
| grado di fiducia del 99% | $F_{\frac{\alpha}{2}} = F_{0.005}$ | $F_{1-\frac{\alpha}{2}} = F_{0.995}$ |

**Esempio 29**

Si vuole studiare la variabilità dei diametri delle sfere d'acciaio prodotte da due diverse macchine. A tale scopo si estraggono due campioni di sfere prodotte dalle due macchine, di ampiezza rispettivamente  $n_1 = 11$  e  $n_2 = 16$ ; le varianze dei due campioni sono  $s_1^2 = 0.40$  e  $s_2^2 = 0.35$ .

Assumendo che le due popolazioni da cui provengono i campioni abbiano distribuzione normale, trovare gli intervalli di confidenza al 90% e al 95% per il rapporto fra le varianze delle popolazioni.

a – Per il grado di fiducia del 90%, con le tavole e facendo uso della formula (7.29) si ha

$$1 - \alpha = 0.90 \quad \frac{\alpha}{2} = 0.05$$

$$F_{\frac{\alpha}{2}}(10,15) = F_{0.05}(10,15) = 2.54$$

$$F_{1-\frac{\alpha}{2}}(10,15) = F_{0.95}(10,15) = \frac{1}{F_{0.05}(15,10)} = \frac{1}{2.85} = 0.35$$

**Parte a**

Applicando la formula (7.28) si trova l'intervallo di confidenza con grado di fiducia del 90%

$$\frac{0.40}{2.54} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.40}{0.35}$$

$$0.44 < \frac{\sigma_1^2}{\sigma_2^2} < 3.27$$

**Esempio 29**

Si vuole studiare la variabilità dei diametri delle sfere d'acciaio prodotte da due diverse macchine. A tale scopo si estraggono due campioni di sfere prodotte dalle due macchine, di ampiezza rispettivamente  $n_1 = 11$  e  $n_2 = 16$ ; le varianze dei due campioni sono  $s_1^2 = 0.40$  e  $s_2^2 = 0.35$ .

Assumendo che le due popolazioni da cui provengono i campioni abbiano distribuzione normale, trovare gli intervalli di confidenza al 90% e al 95% per il rapporto fra le varianze delle popolazioni.

b – Per il grado di fiducia del 95% si ha invece

$$1 - \alpha = 0.95 \quad \frac{\alpha}{2} = 0.025$$

$$F_{\frac{\alpha}{2}}(10,15) = F_{0.025}(10,15) = 3.06$$

$$F_{1-\frac{\alpha}{2}}(10,15) = F_{0.975}(10,15) = \frac{1}{F_{0.025}(15,10)} = \frac{1}{3.52} = 0.28$$

**Parte b**

Applicando la formula (7.28) si trova l'intervallo di confidenza con grado di fiducia del 95%

$$\frac{0.40}{3.06} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{0.40}{0.28}$$

$$0.37 < \frac{\sigma_1^2}{\sigma_2^2} < 4.09$$

Si osservi che aumentando il grado di fiducia, cresce l'ampiezza dell'intervallo, ossia la stima è meno precisa.