

Corso di Statistica Facoltà di Economia

a.a. 2010-2011

francesco mola

Lezione n° 18

Esempio 1: Su un campione casuale di 81 negozi di articoli sportivi la media settimanale di scarpette vendute da ciascun negozio risultata pari a 100. Si sa che nella popolazione la deviazione standard delle vendite settimanali è pari a 8. Si assuma l'ipotesi di normalità del numero di scarpe vendute in ogni negozio.

Si costruisca un intervallo di confidenza al livello del 90% per la vera media settimanale di scarpette vendute da un singolo negozio.

Esempio 2. Due diversi istituti (I1 e I2) sono incaricati di effettuare un'indagine pre-elettorale sul ballottaggio per la elezione del candidato sindaco di una città.

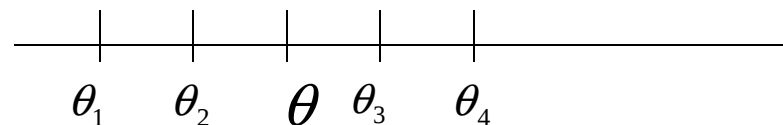
a) l'istituto I1 effettua l'indagine su 150 persone intervistate e trova che 50 di queste dichiarano di votare per il candidato sindaco A mentre il rimanente dichiara di votare per il candidato sindaco B. Considerando un livello di fiducia di 0.9, qual è l'intervallo che trova?

b) l'istituto I2 effettua l'indagine su 450 persone intervistate e trova che 150 di queste dichiarano di votare per il candidato sindaco A mentre il rimanente dichiara di votare per il candidato sindaco B. Considerando un livello di fiducia di 0.9, qual è l'intervallo che trova?

c) Che differenze ci sono nei risultati raggiunti dai due istituti?

INTERVALLI DI CONFIDENZA

- **Stime puntuali**
- **Stime intervallari**



Soluzione esercizio 2:a)

Livello di confidenza $1 - \alpha = 0.9$

Distribuzione di riferimento = proporzione Campionaria

Dimensione del campione $n = 150$

Proporzione votanti
candidato A $\hat{p} = 50 / 150 = 1 / 3 = 0.33$

Proporzione di non
votanti il candidato A $(1 - \hat{p}) = \hat{q} = 100 / 150 = 2 / 3 = 0.66$

Verifica
approssimazione ad
una normale $n\hat{p}\hat{q} = 150 * 0.33 * 0.66 > 10$

Proporzione nella
popolazione π

$$\begin{aligned} \Pr \left\{ \hat{p} - z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq \pi \leq \hat{p} + z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \right\} &= 1 - \alpha \\ &= \Pr \left\{ 0.33 - 1.645 * \sqrt{\frac{0.33 * 0.66}{150}} \leq \pi \leq 0.33 + 1.645 * \sqrt{\frac{0.33 * 0.66}{150}} \right\} = 0.9 \\ &= \Pr \{ 0.27 \leq \pi \leq 0.39 \} = 0.9 \\ &= [0.27; 0.39] = I.C \end{aligned}$$

Soluzione esercizio 2:b)

Livello di confidenza $1 - \alpha = 0.9$

Distribuzione di riferimento = proporzione Campionaria

Dimensione del campione $n = 450$

Proporzione votanti
candidato A $\hat{p} = 150 / 450 = 1 / 3 = 0.33$

Proporzione di non
votanti il candidato A $(1 - \hat{p}) = \hat{q} = 300 / 450 = 2 / 3 = 0.66$

Verifica
approssimazione ad
una normale $n\hat{p}\hat{q} = 450 * 0.33 * 0.66 > 10$

Proporzione nella
popolazione π

$$\begin{aligned} \Pr \left\{ \hat{p} - z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \leq \pi \leq \hat{p} + z_{\alpha/2} * \sqrt{\frac{\hat{p}\hat{q}}{n}} \right\} &= 1 - \alpha \\ &= \Pr \left\{ 0.33 - 1.645 * \sqrt{\frac{0.33 * 0.66}{450}} \leq \pi \leq 0.33 + 1.645 * \sqrt{\frac{0.33 * 0.66}{450}} \right\} = 0.9 \\ &= \Pr \{ 0.29 \leq \pi \leq 0.37 \} = 0.9 \\ &= [0.29; 0.37] = I.C \end{aligned}$$

c) Aumentando la dimensione del campione l'intervallo si riduce di ampiezza

Soluzione esercizio 1)

Livello di confidenza $1 - \alpha = 0.9$

Distribuzione di riferimento = Media Campionaria

Dimensione del campione $n = 81$

Media nel campione $\bar{x} = 100$

Deviazione standard
nella popolazione $\sigma = 8$

Varianza nella
popolazione $\sigma^2 = 64$

$$\Pr\left\{\bar{X} - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}\right\} = 1 - \alpha$$

$$\Pr\left\{100 - 1.645 * \frac{8}{\sqrt{81}} \leq \mu \leq 100 + 1.645 * \frac{8}{\sqrt{81}}\right\} = 0.9$$

$$= \Pr\{98.54 \leq \mu \leq 101.46\} = 0.9$$

$$= [98.54; 101.56] = I.C$$

c) Aumentando la dimensione del campione l'intervallo si riduce di ampiezza

Intervalli di confidenza

$1 - \alpha$ Livello di confidenza o
Livello fiduciario

consideriamo $X \sim N(\mu, \sigma^2)$

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

**Se consideriamo le tavole della Z, possiamo
Individuare 2 valori c_1 e c_2 :**

$$\Pr(c_1 \leq Z \leq c_2) = 1 - \alpha$$

con $0 \leq \alpha \leq 1$

Quante sono le coppie di c_1 e c_2 che soddisfano la relazione?

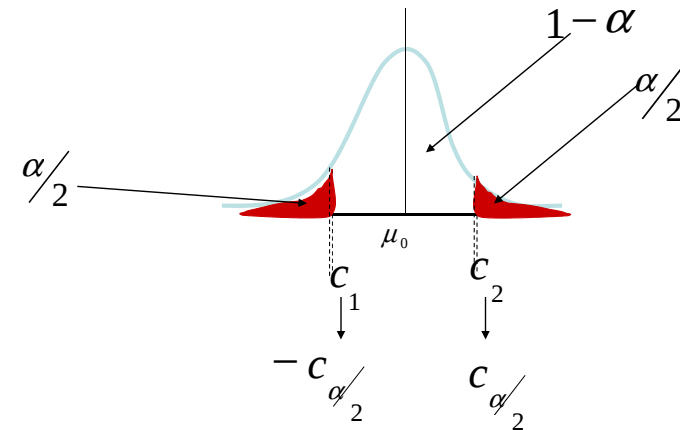
∞

Se imponiamo il vincolo che $[C_1, C_2]$ sia centrale, cioè

$$P\{Z \leq c_1\} = \frac{\alpha}{2} \quad \text{e} \quad P\{Z \geq c_2\} = \frac{\alpha}{2}$$

allora si ha solo c_1 e c_2

Caso della distribuzione Normale



$$P\left(-c_{\alpha/2} \leq Z \leq c_{\alpha/2}\right) = 1 - \alpha$$



Caso della distribuzione Normale (cont..)

$$P\left(-c_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq c_{\alpha/2}\right) =$$

$$= P\left(\mu - c_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + c_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) =$$

$$= P\left(\bar{X} - c_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + c_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

L'intervallo casuale $\left[\bar{X} - c_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + c_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$ comprende il valore incognito μ con probabilità $1 - \alpha$, indipendentemente dal valore di μ .

Se sostituiamo ad $\bar{X}$ v.c. il valore $\bar{x}$ calcolato sul campione abbiamo :

$$\left[\bar{x} - c_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + c_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

intervallo di confidenza per μ al livello di confidenza di $1 - \alpha$

Es :

$$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow c_{\alpha/2} = 1.96 \quad 1 - \alpha = 0.95$$

$$P\left(\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

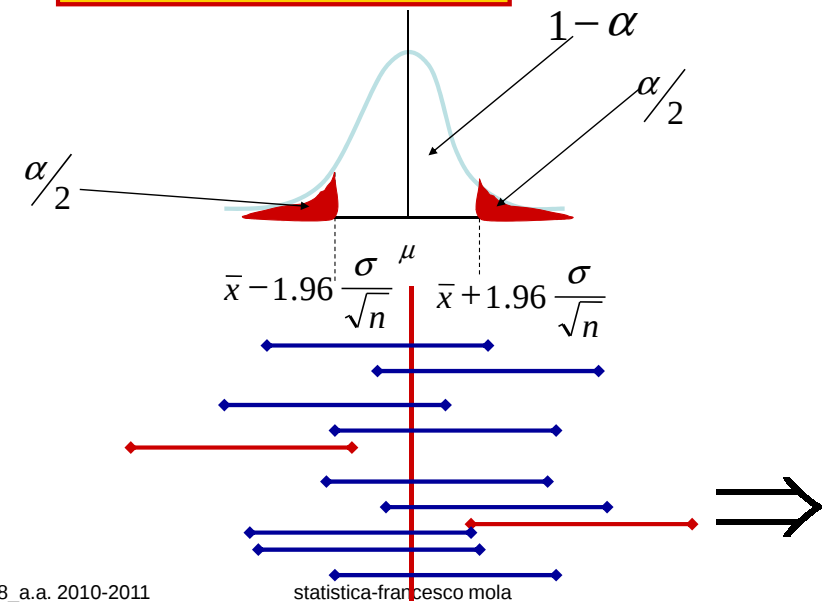
$$\left[\bar{X} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

Intervallo casuale

$$\left[\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}\right]$$

Intervallo di confidenza

Cosa significa $1 - \alpha$?



Esempio 1.

Intervallo di confidenza per la media con σ^2 nota

$$X \sim N(\mu, \sigma^2)$$

$$\left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right]$$

Esempio 2.

Intervallo di confidenza per la media con σ^2 NON nota

Si considera la v.c. T di Student

$$T = \frac{\bar{X} - \mu}{\tilde{S}/\sqrt{n}} \quad \tilde{S}^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

Abbiamo:

$$P\left\{-t_{\alpha/2} \leq T \leq +t_{\alpha/2}\right\} = P\left\{-t_{\alpha/2} \leq \frac{\bar{X} - \mu}{\tilde{S}/\sqrt{n}} \leq +t_{\alpha/2}\right\} =$$

$$= P\left\{\bar{X} - t_{\alpha/2} \tilde{S}/\sqrt{n} \leq \mu \leq \bar{X} + t_{\alpha/2} \tilde{S}/\sqrt{n}\right\} = 1 - \alpha$$

Cioè:

$$\left[\bar{x} - t_{\alpha/2} \tilde{S} / \sqrt{n} \leq \mu \leq \bar{x} + t_{\alpha/2} \tilde{S} / \sqrt{n} \right]$$

Esempio 3.

Intervallo di confidenza per la varianza con μ non nota

$$X \sim N(\mu, \sigma^2)$$

Si ricordi che

$$\frac{nS^2}{\sigma^2} \sim \chi_{n-1}^2 \quad S^2 = \frac{1}{n} \sum (X_i - \bar{X})^2$$

Considerando le tavole della distribuzione χ^2 possiamo individuare $y_{1,\alpha/2}$ e $y_{2,\alpha/2}$ tali che:

$$P(Y \leq y_{1,\alpha/2}) = \alpha/2 = P(Y \geq y_{2,\alpha/2})$$

$$P\left(\frac{nS^2}{y_{2,\alpha/2}} \leq \sigma^2 \leq \frac{nS^2}{y_{1,\alpha/2}} \right) = 1 - \alpha \quad \left[\frac{nS^2}{y_{2,\alpha/2}}, \frac{nS^2}{y_{1,\alpha/2}} \right]$$

Esempio 4.

Intervallo di confidenza per la proporzione campionaria

$$\frac{S_n}{n} \sim B(n, p)$$

$$E\left(\frac{S}{n}\right) = p \quad \text{VAR}\left(\frac{S}{n}\right) = \frac{pq}{n}$$

$$\frac{S_n}{n} \sim N\left(p, \frac{pq}{n}\right)$$

Quindi si può affermare che:

$$P\left(-c_{\alpha/2} \leq \frac{\frac{S_n}{n} - p}{\sqrt{\frac{pq}{n}}} \leq c_{\alpha/2} \right) \cong 1 - \alpha$$

Risolvendo la disuguaglianza si ha:

$$P\left\{\frac{n}{n+c_{\alpha/2}^2}\left[\frac{S_n}{n}+\frac{c_{\alpha/2}^2}{2n}-c_{\alpha/2}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}+\frac{c_{\alpha/2}^2}{4n^2}\right]\leq p\leq\right. \\ \left.\frac{n}{n+c_{\alpha/2}^2}\left[\frac{S_n}{n}+\frac{c_{\alpha/2}^2}{2n}+c_{\alpha/2}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}+\frac{c_{\alpha/2}^2}{4n^2}\right]\right\}\cong 1-\alpha$$

Quando n è abbastanza grande si ha:

$\frac{n}{n+c_{\alpha/2}^2} \cong 1$ $\frac{c_{\alpha/2}^2}{2n}$ e $\frac{c_{\alpha/2}^2}{4n^2}$ possono essere trascurati, quindi :

$$P\left\{\left[\frac{S_n}{n}-c_{\alpha/2}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}\leq p\leq\frac{S_n}{n}+c_{\alpha/2}\sqrt{\frac{S_n}{n}\left(1-\frac{S_n}{n}\right)}\right]\right\}\cong 1-\alpha$$

$$C_{\alpha/2} = Z_{\alpha/2}$$

Esempio 5.

Intervallo di confidenza per differenze tra medie

$$X \sim N(\mu_1, \sigma_1^2)$$

$$X \sim N(\mu_2, \sigma_2^2)$$

Caso 1: σ_1^2 e $\sigma_2^2 = \text{note}$

È noto che:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

Quindi fissato $1-\alpha$ si ha:

$$P\left\{(\mu_1 - \mu_2) - z_{\alpha/2}\sigma_{\bar{X}_1 - \bar{X}_2} \leq (\bar{X}_1 - \bar{X}_2) \leq (\mu_1 - \mu_2) + z_{\alpha/2}\sigma_{\bar{X}_1 - \bar{X}_2}\right\} = 1-\alpha$$

Avendo posto

$$\sigma_{\bar{X}_1 - \bar{X}_2}^2 = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

L'espressione precedente può essere riscritta così:

$$P\left\{(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2}\right\} = 1 - \alpha$$

$$\left[(\bar{X}_1 - \bar{X}_2) - z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2}, (\bar{X}_1 - \bar{X}_2) + z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2} \right] =$$

Intervallo casuale

$$\left[(\bar{x}_1 - \bar{x}_2) - z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2}, (\bar{x}_1 - \bar{x}_2) + z_{\alpha/2} \sigma_{\bar{X}_1 - \bar{X}_2} \right] =$$

Intervallo di confidenza

Esempio 5.

Intervallo di confidenza per differenze tra medie

Caso 2: σ_1^2 e $\sigma_2^2 = \text{incognite}$

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \sim t_{(n_1 + n_2 - 2)}$$

Se indichiamo con

$$S_p^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

$$P\left\{(\mu_1 - \mu_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\bar{X}_1 - \bar{X}_2) \leq (\mu_1 - \mu_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\} =$$

$$= P\left\{(\bar{X}_1 - \bar{X}_2) - t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq (\mu_1 - \mu_2) \leq (\bar{X}_1 - \bar{X}_2) + t_{\alpha/2} S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right\} = 1 - \alpha$$

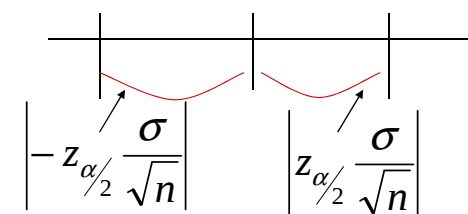
$$g.l. = n_1 + n_2 - 2$$

Per la differenza tra proporzioni si opera analogamente

Riconsideriamo il caso: $X \sim N(\mu, \sigma^2)$
con σ^2 nota

A livello $1 - \alpha$ si può vedere che l'ampiezza dell'intervallo è:

$$A = 2 z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$



Se si risolve l'espressione precedente rispetto ad n:

$$n = \left[\frac{2z_{\alpha/2}\sigma}{A} \right]^2$$

In tal modo si determina la **numerosità minima del campione**

Esempio:

$$\sigma^2 = 4$$

$$1 - \alpha = 0.95$$

$$A = 2$$

$$n = \left[\frac{2 \bullet 1.96 \bullet 2}{2} \right]^2 = 15.36 \cong 16$$