Corso di Statistica

francesco mola

Lezione n° 14 (2): Campionamento e distribuzioni campionarie

sommario

- Teorema del Limite Centrale
- Principali distribuzioni campionarie

Teorema Limite Centrale

Sia $\{X_n\}$ una successione di v.c. indipendenti con egual media μ ed egual varianza $\sigma^2 (0 < \sigma^2 < \infty)$, allora data la v.c.

$$S_n = X_1 + X_2 + \cdots X_n$$

con $E(S_n) = n\mu$

$Var(S_n) = n\sigma^2$

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} Z \sim N(0,1)$$

Teorema Limite Centrale (cont.)

E' un teorema importantissimo!

Se consideriamo la successione $\frac{S_n}{n} = \frac{1}{n}\sum X_i = \overline{X}_n$

abbiamo:

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

Media Campionaria

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

Vale sempre, anche se non si conosce la distribuzione

Consideriamo un campione casuale

$$(X_1, X_2, \ldots X_n)$$

$$\bar{X} = \frac{1}{n}\sum_i X_i$$

Media Campionaria

$$E(\bar{X}) = E\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{1}{n}E(X_1 + X_2 + \cdots + X_n) =$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \cdots + E(X_n)\right] = \frac{1}{n}[\mu + \mu + \cdots \mu] = \frac{1}{n}n\mu = \mu$$

$$Var(\bar{X}) = Var\left[\frac{1}{n}(X_1 + X_2 + \cdots + X_n)\right] = \frac{1}{n^2}\mathrm{var}(X_1 + X_2 + \cdots + X_n) =$$

$$= \frac{1}{n^2}\left[\mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n)\right] = \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \cdots \sigma^2\right] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

Media Campionaria

$$\bar{X} = \frac{1}{n}\sum_i X_i \quad\quad \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

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

Normale Standardizzata

Esempio:

$$\Pr\left(\bar{x}_1 \le \bar{X} \le \bar{x}_2\right) = \Pr\left(\frac{\bar{x}_1 - \mu}{\sigma/\sqrt{n}} \le Z \le \frac{\bar{x}_2 - \mu}{\sigma/\sqrt{n}}\right)$$

proporzione campionaria

$$\frac{S_n}{n} \sim B(n,p) \quad S_n = X_1 + X_2 + \ldots + X_n$$

$$E\left(\frac{S}{n}\right) = p \quad VAR\left(\frac{S}{n}\right) = \frac{p(1-p)}{n}$$

$\frac{S_n}{n} =$ **Proporzione campionaria**

Per n grande: $\frac{S_n}{n} \sim N\left(p, \frac{p(1-p)}{n}\right)$

Varianza campionaria

$X \sim N(\mu, \sigma^2)$ $\quad (X_1, X_2, \ldots X_n)$

$$S^2 = \frac{1}{n-1}\sum\left(X_i - \bar{X}\right)^2$$

Si dimostra che: $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

Quindi

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \quad VAR\left(\frac{(n-1)S^2}{\sigma^2}\right) = 2(n-1)$$

$$E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n-1 \Rightarrow E\left(S^2\right) = \sigma^2$$

La varianza campionaria è uno stimatore non distorto

Distribuzione Media Campionaria σ^2 incognita

Se $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}}$ **ma** σ^2 **è incognita allora possiamo sostituire a** σ^2 **la v.c.** S^2

$$T = \frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{(n-1)}$$

DIFFERENZA TRA MEDIE

Consideriamo un carattere X osservato su due popolazioni:

$$X \sim N\left(\mu_1, \sigma_1^2\right) \qquad X \sim N\left(\mu_2, \sigma_2^2\right)$$

Consideriamo dei campioni di ampiezza n_1 e n_2 estratti dalle due popolazioni.

$\overline{X}_1 \longleftarrow$ *Media campionaria del primo campione*

$\overline{X}_2 \longleftarrow$ *Media campionaria del secondo campione*

$$\overline{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right) \qquad \overline{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$$\left(\overline{X}_1 - \overline{X}_2\right) \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

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

DIFFERENZA TRA MEDIE
σ_1^2 σ_2^2 incognite ma supposte uguali

Se σ_1^2 σ_2^2 sono incognite ma supposte uguali allora abbiamo:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = \text{Stima congiunta della varianza}$$

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