# Media

### *Distribuzione di frequenza unitaria*

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

### *Distribuzione di frequenza*

$\bar{x} = \frac{1}{n}\sum_{j=1}^{k} x_j n_j$ dove $x_1, x_2, \ldots, x_k$ sono le $k$ modalità distinte e $n = n_1 + n_2 + \ldots + n_k$ .

# Varianza

### *Distribuzione di frequenza unitaria*

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

### *Distribuzione di frequenza*

$\sigma^2 = \frac{1}{n}\sum_{j=1}^{k}(x_j - \bar{x})^2 n_j$ o $\sigma^2 = \frac{1}{n}\sum_{j=1}^{k} x_j^2 n_j - \bar{x}^2$

dove $x_1, x_2, \ldots, x_k$ sono le $k$ modalità distinte e $n = n_1 + n_2 + \ldots + n_k$ .

# Quartili

$$Q_1 \approx I_{Q1} + \left(\frac{0.25 - F_{Q_1 - 1}}{F_{Q_1} - F_{Q_1 - 1}}\right)\Delta_{Q_1} \qquad Me = Q_2 \approx I_{Q_2} + \left(\frac{0.5 - F_{Q_2 - 1}}{F_{Q_2} - F_{Q_2 - 1}}\right)\Delta_{Q_2}$$

$$Q_3 \approx I_{Q_3} + \left(\frac{0.75 - F_{Q_3 - 1}}{F_{Q_3} - F_{Q_3 - 1}}\right)\Delta_{Q_3}$$

$I_{Q_i}$ è l'estremo inferiore della classe dove cade l'i-esimo quartile

$F_{Q_i}$ è la frequenza relativa cumulata della classe che contiene l'i-esimo quartile

$F_{Q_i - 1}$ è la frequenza relativa cumulata fino alla classe precedente a quella in cui cade l'i-esimo quartile

$\Delta_{Q_i}$ è l'ampiezza della classe dove cade l'i-esimo quartile

# Concentrazione

R=$\dfrac{\sum_{i=1}^{n-1}(F_i - Q_i)}{\sum_{i=1}^{n-1} F_i}$

dove $F_i = \dfrac{i}{n}$, $Q_i = \dfrac{A_i}{A_n}$, $A_i = x_1 + x_2 + \ldots + x_i$ e $x_1 \le x_2 \le \ldots \le x_n$

# Associazione

$n'_{ij} = \dfrac{n_{i0} n_{0j}}{n}$ rappresentano le frequenze teoriche nel caso di indipendenza.

Cov(X,Y)=$\sigma_{xy} = \dfrac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$

$\sigma_{xy} = \dfrac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy}$

$\rho_{xy} = \dfrac{\sigma_{xy}}{\sigma_x \sigma_y}$

# Probabilità

A e B siano due eventi:

$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

A e B sono eventi incompatibili se : $A \cap B = \varnothing$

A e B sono eventi indipendenti se : $P(A \cap B) = P(A) \cdot P(B)$

La probabilità condizionata di A dato B è : $P(A|B) = \dfrac{P(A \cap B)}{P(B)}$

$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$

# V.c. Binomiale

$p(x) = \binom{n}{x} p^x q^{n-x} = \dfrac{n!}{x!(n-x)!} p^x q^{n-x} \qquad x = 0, 1, \ldots, n$

# V.c. Normale

$X \sim N(\mu, \sigma^2)$ standardizzazione $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

$\Phi(z) = P(Z \le z)$

$\Phi(1.282) = 0.90$ $\quad\Phi(1.645) = 0.95$ $\quad\Phi(1.96) = 0.975$

$\Phi(2.326) = 0.99$ $\quad\Phi(2.575) = 0.995$

# Stimatori

Lo stimatore **T** è uno **stimatore corretto** di $\theta$ se $E(T) = \theta$

**Distorsione di uno stimatore**: $B(T) = E(T) - \theta$

**Errore quadratico medio di uno stimatore:** $MSE(T) = E\left[(T-\theta)^2\right] = Var(T) + B(T)^2$

$T_1$ **è più efficiente** di $T_2$ se e solo se $MSE(T_1) < MSE(T_2)$

Lo stimatore $T_n$ è **consistente in media quadratica** se $\lim\limits_{n\to+\infty} MSE(T_n) = \lim\limits_{n\to+\infty} E\left[(T_n-\theta)^2\right] = 0$.

# Intervallo di confidenza

*Intervallo di confidenza della media $\mu$ della popolazione al livello* $1-\alpha$

*Varianza incognita (n >120)*

$$(1-\alpha)IC = \left(\bar{x}_n - z_{\alpha/2}\frac{S}{\sqrt{n}} ; \bar{x}_n + z_{\alpha/2}\frac{S}{\sqrt{n}}\right)$$

*Varianza nota*

$$(1-\alpha)IC = \left(\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} ; \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

dove $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ e $\bar{x}_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

*Intervallo di confidenza per la proporzione $\pi$ di una popolazione bernoulliana al livello* $1-\alpha$ *per n grande (n>30)*

$$(1-\alpha)IC = \left(\bar{x}_n - z_{\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}} ; \bar{x}_n + z_{\alpha/2}\sqrt{\frac{\bar{x}_n(1-\bar{x}_n)}{n}}\right)$$

dove $\bar{x}_n$ o $\hat{\pi} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ .

# Verifica di ipotesi

*Regione di rifiuto al livello di significatività $\alpha$ per la verifica d'ipotesi sulla media $\mu$ della popolazione*

*Varianza incognita(n >120)*

Statistica test $Z = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$

dove $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ .

*Varianza nota*

Statistica test $Z = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu > \mu_0 \end{aligned} \qquad R = \{Z \geq z_\alpha\}$$

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu < \mu_0 \end{aligned} \qquad R = \{Z \leq -z_\alpha\}$$

$$\begin{aligned} H_0 &: \quad \mu = \mu_0 \\ H_1 &: \quad \mu \neq \mu_0 \end{aligned} \qquad R = \{\, |Z| \geq z_{\alpha/2}\}$$

*Regione di rifiuto ad un livello di significatività $\alpha$ per la verifica d'ipotesi sulla proporzione $\pi$ di una popolazione bernoulliana per n grande (n>30):*

Statistica test $Z = \dfrac{\bar{X} - \pi_0}{\sqrt{\pi_0(1-\pi_0)}/\sqrt{n}}$

$$\begin{aligned} H_0 &: \quad \pi = \pi_0 \\ H_1 &: \quad \pi > \pi_0 \end{aligned} \qquad R = \{Z \geq z_\alpha\}$$

$$\begin{aligned} H_0 &: \quad \pi = \pi_0 \\ H_1 &: \quad \pi < \pi_0 \end{aligned} \qquad R = \{Z \leq -z_\alpha\}$$

$$\begin{aligned} H_0 &: \quad \pi = \pi_0 \\ H_1 &: \quad \pi \neq \pi_0 \end{aligned} \qquad R = \{|Z| \geq z_{\alpha/2}\}$$

dove $\bar{X}$ o $\hat{\pi} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ .