Raccolta Formule e Dimostrazioni

NB. Non può essere usato durante la prova scritta

Media aritmetica

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Per distribuzioni di frequenza si ha

$$\overline{x} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_n n_n}{n_1 + n_2 + \ldots + n_n} = \frac{\sum_{i=1}^{N} x_i n_i}{\sum_{i=1}^{n} n_i = N}$$

Media armonica

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

Per distribuzioni di frequenza si ha:

$$Mar = \frac{N}{\sum_{i=1}^{n} \frac{n_i}{x_i}}$$

Media geometrica

$$M_g = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Per distribuzioni di frequenza: $M_g = \sqrt{x_1^{n1} x_2^{n2} \ldots x_n^{nn}} = \sqrt[N]{\prod_{i=1}^{N} x_i^{n_i}}$.

Media quadratica

$$M_q = \sqrt{\frac{\sum_{i=1}^{N} x_i^2 n_i}{N}}$$

Dimostrazioni valori medi

1) **La somma algebrica degli scarti di ciascuna modalità dalla media è uguale a 0 (Prima proprietà):**

   $\sum_{i=i}^{n}(x_i - \overline{x}) = 0$. Si ha che: $\sum_{i=i}^{n} x_i - n\overline{x} = 0$, ovvero $\sum_{i=i}^{n} x_i = n\overline{x}$.

2) **La somma dei quadrati degli scarti di ciascuna modalità dalla media aritmetica è un minimo (Seconda proprietà):**

   $\sum_{i=1}^{n}(x_i - \overline{x})^2 =$ minimo

   Al fine di dimostrare la proprietà enunciata, consideriamo un generico valore *a* diverso da $\overline{x}$ e poniamo la relazione seguente:

   $a - \overline{x} = d$, da cui:

   $a = \overline{x} + d$

   Pertanto, la somma dei quadrati degli scostamenti dei valori x_i da *a* può essere scritta come segue:

   $$\sum_{i=1}^{n}(x_i - a)^2 = \sum_{i=1}^{n}[x_i - (\overline{x} + d)]^2 = \sum_{i=1}^{n}(x_i - \overline{x} - d)^2 = \sum_{i=1}^{n}[(x_i - \overline{x}) - d]^2.$$

   Svolgendo il quadrato si ottiene: $\sum_{i=1}^{n}[(x_i - \overline{x})^2 + d^2 - 2d(x_i - \overline{x})]$.

Da cui si giunge a dimostrare che: $\sum_{i=1}^{n}(x_i - a)$ = $\sum_{i=1}^{n}(x_i - \overline{x})^2$ + nd^2 - $2d$ $\sum_{i=1}^{n}(x_i - \overline{x})$.

Da tale relazione è agevole rilevare che la somma dei quadrati degli scarti dei valori x_i da un generico valore a è maggiore rispetto alla somma dei quadrati degli scarti delle x_i dalla media, ossia, in simboli

$\sum_{i=1}^{n}(x_i - a)^2 > \sum_{i=1}^{n}(x_i - \overline{x})^2$. Infatti la quantità nd^2 che compare al secondo membro della relazione è sicuramente positiva, mentre il terzo termine che compare al secondo membro, $2d \sum_{i=1}^{n}(x_i - \overline{x})$, è nullo per la proprietà 4) in precedenza dimostrata. Risulta così dimostrato che: $\sum_{i=1}^{n}(x_i - \mu)^2$ = min.

Mediana

– per caratteri discreti:

Me =	Se n è dispari $\rightarrow$ la mediana è l'intensità individuata dal posto centrale $x\left(\frac{n+1}{2}\right)$
	Se n è pari $\rightarrow$ la mediana è data dalla semisomma delle intensità individuate dai due posti centrali $x\left(\frac{n}{2}\right)$ e $x\left(\frac{n}{2}+1\right)$

Nel caso di distribuzione di frequenza con modalità rappresentate da classi si definisce classe mediana quella a cui corrisponde la prima frequenza cumulata non minore di N/2.

Supponendo l'uniforme distribuzione delle osservazioni nelle classi è possibile calcolare un valore approssimato della mediana attraverso la formula.

$$Me = x_i + \frac{x_{i+1} - x_i}{ni}\left(\frac{N}{2} - N_{i-1}\right)$$

Variabilità

Intervalli di Variazione

Campo di variazione

$W = x_n - x_1$

Differenza Interquartilica

$\delta = Q_3 - Q_1$

Scostamenti Medi

Varianza

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

Nel caso si abbia una distribuzione di frequenza si ha: $\sigma^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}$.

Scarto quadratico medio (s.q.m.)

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

Nel caso di distribuzione di frequenza si ha: $\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 n_i}{N}}$.

Devianza

E' il numeratore della varianza. La sua espressione analitica è: $Dev(X) = \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Per distribuzioni di frequenza si ha: $Dev(X) = \sum_{i=1}^{N} (x_i - \bar{x})^2 n_i$.

Differenze Medie

Differenza semplice media senza ripetizione

$$\Delta = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|}{n(n-1)}$$

Differenza semplice media con ripetizione

$$\Delta_R = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \left| x_i - x_j \right|}{n^2}$$

Differenza quadratica media senza ripetizione

$$^2\Delta = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2}{n(n-1)}}$$

Differenza quadratica media con ripetizione

$$^2\Delta_R = \sqrt{\frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (x_i - x_j)^2}{n^2}}$$

Variabilità relativa

Coefficiente di variazione

$$CV = \frac{\sigma}{\mu} \cdot 100$$

Distribuzione Massimante

Se imponiamo che la distribuzione massimante abbia la stessa media aritmetica $\bar{x}$ della distribuzione di partenza si ricava che:

$$\bar{x} = \frac{x_1 \cdot p + x_n \cdot (N - p)}{N},$$

da cui risolvendo rispetto a p, si ottiene:

$$p = \frac{N(x_n - \bar{x})}{x_n - x_1}$$

$$q = \frac{N(\bar{x} - x_1)}{x_n - x_1}$$

Rapporto di concentrazione

$$R == \frac{\sum_{i=1}^{n-1}(p_i - q_i)}{\sum_{i=1}^{n-1} p_i}$$

Coefficiente di regressione

$$b_{y/x} == \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

$$b_{y/x} = \frac{Codev(X,Y)}{Dev(x)} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}$$

$$b_{y/x} = \frac{Co\,\mathrm{var}(X,Y)}{Var(X)} = \frac{\dfrac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{n}}{\dfrac{\sum_{i=1}^{n} (x_i - \overline{x})^2}{n}}$$

$$b_{x/y} = \frac{Codev(X,Y)}{Dev(Y)} = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

Dimostrazioni

$$\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}) = \sum_{i=1}^{n} x_i y_i - n\overline{xy}$$

$$\sum_{i=1}^{n} (x_i y_i - x_i \overline{y} - \overline{x} y_i + \overline{xy}) = \sum_{i=1}^{n} x_i y_i - n\overline{xy} + n\overline{xy} - n\overline{xy}$$

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

Indice di determinazione

$$R^2 = \frac{Dev(R)}{Dev(Y)} = 1 - \frac{Dev(E)}{Dev(Y)}$$

Scomposizione devianza

$$Dev(y) = \sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} \left(y_i - y_i^{*} + y_i^{*} - \overline{y}\right)^2 =$$

$$= \sum_{i=1}^{n} \left(y_i - y_i^{*}\right)^2 + \sum_{i=1}^{n} \left(y_i^{*} - \overline{y}\right)^2 + 2\sum_{i=1}^{n} \left(y_i - y_i^{*}\right)\left(y_i^{*} - \overline{y}\right)$$

Per dimostrare l’uguaglianza suddetta occorre dimostrare che il doppio prodotto è nullo:

$$\sum_{i=1}^{n} \left(y_i - y_i^{*}\right)\left(y_i^{*} - \overline{y}\right) = \sum_{i=1}^{n} \left(y_i - y_i^{*}\right) y_i^{*} - \sum_{i=1}^{n} \left(y_i - y_i^{*}\right)\overline{y}$$

Per la proprietà dei minimi quadrati che assicura l'uguaglianza tra $\sum_{i=1}^{n} y_i$ e $\sum_{i=1}^{n} y_i^*$ (la media dei valori osservati coincide con la media dei valori teorici), l'ultima sommatoria è nulla. Inoltre, essendo $y_i^* = a + b_1 x_i$ si ha:

$$\underbrace{\sum_{i=1}^{n}\left(y_i - y_i^*\right)y_i^*}_{=0} = \sum_{i=1}^{n}\left(y_i - y_i^*\right)\left(a + b_1 x_i\right) = \sum_{i=1}^{n}\left(y_i - y_i^*\right)a + \sum_{i=1}^{n}\left(y_i - y_i^*\right)bx_i =$$

$$= b\sum_{i=1}^{n}\left(y_i x_i - y_i^* x_i\right) = b\left[\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i\left(a + bx_i\right)\right] = b\left[\sum_{i=1}^{n} x_i y_i - a\sum_{i=1}^{n} x_i - b\sum_{i=1}^{n} x_i^2\right].$$

L'espressione contenuta nelle parentesi quadre coincide con la derivata parziale della $\Phi(a,b) = \sum_{i=1}^{n}\left(y_i - a - b_1 x_i\right)^2$ calcolata rispetto al parametro b che è stata posta pari a zero per la stima del parametro stesso con il metodo dei minimi quadrati, ossia:

$$\frac{\partial \Phi\, a,b}{\partial b} = -2\sum_{i=1}^{n}\left(y_i - a - bx_i\right)x_i = 0$$

Si giunge quindi a dimostrare che Dev(Y)=Dev(E)+Dev(R), da cui si ricava l' indice di determinazione.

Coefficiente di correlazione

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

$$r_{xy} = \pm\sqrt{b_{y/x} \cdot b_{x/y}}$$

Probabilità e variabili casuali

- **Postulato** (o legge) delle probabilità totali:
  $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\}\ se\ A \cap B = \varnothing$ (eventi incompatibilii o disgiunti)

  $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{A \cap B\}\ se\ A \cap B \neq \varnothing$ (eventi compatibili)

- **Postulato** (o legge) delle probabilità composte:
  $\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B\}$ (eventi indipendenti)

  $\Pr\{A \cap B\} = \Pr\{A\}\Pr\{B/A\}$ (eventi dipendenti)

Teorema di Bayes

$$\Pr(C_i/E) = \frac{\Pr(C_i \cap E)}{\Pr(E)} = \frac{\Pr(C_i)\Pr(E/C_i)}{\sum \Pr(C_i)\Pr(E/C_i)}$$

Media e varianza v.c. normale standardizzata

$$Z = \frac{X-\mu}{\sigma}$$

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}E(X-\mu) = \frac{1}{\sigma}[E(X)-\mu] = \frac{1}{\sigma}(\mu-\mu) = 0$$

$$Var(Z) = E[Z-E(Z)]^2 = E\left[\frac{X-\mu}{\sigma}\right]^2 = \frac{1}{\sigma^2}E(X-\mu)^2 = \frac{\sigma^2}{\sigma^2} = 1$$

Media e varianza della v.c di Bernoulli

$p(x) = p^x(1-p)^{1-x},\ x = 0,\ 1.$

E(X)=$1 \cdot p + 0 \cdot (1-p) = p$

$Var(X) = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p \cdot (1-p)$

Media e varianza della v.c. Binomiale

$$\Pr(x)=\binom{n}{x}p^x(1-p)^{n-x}$$

La distribuzione binomiale può essere ottenuta considerando la somma di n v.c di Bernoulli, indipendenti ed identicamente distribuite ($X_1, X_2,...X_n$).

$$E(X)=E(X_1+X_2+....+X_n)=p+p+...+p=np$$

$$Var(X)=Var(X_1+X_2+...+X_n)=p(1-p)+p(1-p)+....+p(1-p)=np(1-p)$$

Media e varianza della v.c. media campionaria

$$E(\overline{X})=E\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\frac{1}{n}\sum_{i=1}^{n}\mu=\mu$$

$$Var(\overline{X})=Var\left(\frac{\sum_{i=1}^{n}X_i}{n}\right)=\frac{1}{n^2}\sum_{i=1}^{n}Var(X_i)=\frac{1}{n^2}\sum_{i=1}^{n}\sigma^2=\frac{\sigma^2}{n}$$

Media e varianza della v.c proporzione o percentuale campionaria

$$E(P)=E\left(\frac{X}{n}\right)=\frac{E(X)}{n}=\frac{np}{n}=p$$

$$Var(P)=Var\left(\frac{X}{n}\right)=\frac{Var(X)}{n^2}=\frac{np(1-p)}{n^2}=\frac{p(1-p)}{n}$$

Intervallo di confidenza per la media di una popolazione normale con varianza σ^2 nota.

$$\Pr\left[\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \le \mu \le \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] = 1-\alpha$$

Intervallo di confidenza per la media di una popolazione normale con varianza σ^2 non nota.

$$\Pr\left[\overline{X} - t_{\alpha/2;n-1}\frac{s}{\sqrt{n}} \le \mu \le \overline{X} + t_{\alpha/2;n-1}\frac{s}{\sqrt{n}}\right] = 1-\alpha$$

Intervallo di confidenza per la proporzione (o percentuale) di una popolazione normale.

$$\Pr\left[\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}} \le \pi \le \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}\left(1-\hat{P}\right)}{n}}\right] = 1-\alpha$$

Statistica test verifica di ipotesi sulla media di una popolazione con varianza.nota

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}}$$

Statistica test verifica di ipotesi sulla media di una popolazione con varianza.non nota

$$T = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

Statistica Test Verifica di ipotesi sulla proporzione o percentuale campionaria (grandi campioni)

$$Z = \frac{\hat{P} - \pi_0}{\sqrt{\frac{\pi_0\left(1-\pi_0\right)}{n}}}$$

Statistica test confronto fra medie due campioni indipendenti (Popolazioni Normali - Varianze note)

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

Statistica test confronto fra medie campioni indipendenti (- Varianze non note)

$$T = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S_p^2 (1/n_1 + 1/n_2)}} \sim t\text{-}Student(n_1 + n_2 - 2)$$

$$S_p^2 = \frac{(n_1 - 1)\mathrm{S}_1^2 + (n_2 - 1)\mathrm{S}_2^2}{n_1 + n_2 - 2}$$

Statistica test confronto fra percentuali per due campioni indipendenti

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}_p (1 - \hat{P}_p)(1/n_1 + 1/n_2)}} \sim N(0,1)$$

$$\hat{P}_p = \frac{n_1 \hat{P}_1 + n_2 \hat{P}_2}{n_1 + n_2}$$