CAMPIONE STATISTICO, MEDIA E VARIANZA CAMPIONARIA

Popolazione di n individui su cui rilevo un certo carattere numerico $\Rightarrow$ ho un vettore delle osservazioni
$$x = (x_1 \_\_\_ x_n)$$
Suppongo che $\exists$ uno spazio di probabilità $(\Omega, \mathcal{E}, \mathbb{P})$ e una famiglia $X_1 \_\_\_ X_n$ di v.a. su $(\Omega, \mathcal{E}, \mathbb{P})$ con le seguenti proprietà

1) $\exists\, \omega \in \Omega \quad x_1 = X_1(\omega) \; \dots \; x_n = X_n(\omega)$
2) le $X_1 \_\_\_ X_n$ sono i.i.d. (con valore atteso e varianza finiti)

Le $X_1 \_\_ X_n$ si dicono un CAMPIONE STATISTICO

N.B. Se $X_1 \_\_ X_n$ ha distribuzione A.C. con densità $f(x)$: $\mathbb{P}_{X_i} = f(x)\,dx \quad \forall i = 1 \_\_ n$, allora per l'indipendenza $(X_1, \_\_\_ X_n)$ è una v.a. con distribuzione A.C. e
$$\mathbb{P}_{X_1 \_\_ X_n} = f(x_1) f(x_2) \_\_ f(x_n)\, dx_1 \_\_ dx_n$$

Analogamente se $X_1 \_\_ X_n$ ha distribuzione discreta concentrata su $\{t_j\}_{j \in \mathbb{N}\,(j=1\_\_N)}$ con densità $p_j = p(t_j)$ allora $(X_1 \_\_ X_n)$ è distribuito su $\{t_j\}_{j\in\mathbb{N}(j=1\_\_N)} \times \_\_\_ \times \{t_j\}_{j\in\mathbb{N}(j=1\_\_N)}$

e densità $p(t_j^{(1)} \dots, t_j^{(n)}) = p(t_j^{(1)}) \dots p(t_j^{(n)})$

In ogni caso, la comune distribuzione di $X_1 \_\_ X_n$ si dice DISTRIBUZIONE CAMPIONARIA.

Sia $X_1 \_\_ X_n$ un campione statistico e sia $h: \mathbb{R}^n \to \mathbb{R}$ una funzione Borel-misurabile.
Allora la v.a. $Y := h \circ (X_1 \_\_ X_n)$ si dice una STATISTICA del campione.

DEF (MEDIA CAMPIONARIA E VARIANZA CAMPIONARIA)
Sia $X_1 \_\_ X_n$ un campione statistico.
Chiamiamo MEDIA CAMPIONARIA di $X_1 \_\_ X_n$ la v.a.

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Chiamiamo VARIANZA CAMPIONARIA di $X_1 \_\_ X_n$ la v.a.

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

PROP Sia $X_1 \_\_ X_n$ un campione statistico di cardinalità n con valore atteso μ e varianza σ^2 entrambi finiti.
Siano $\overline{X}$ e S^2 la media e la varianza campionarie di $X_1 \_\_ X_n$.
Allora

$$E[\overline{X}] = \mu \qquad Var[\overline{X}] = \frac{\sigma^2}{n}$$

$$E[S^2] = \sigma^2$$

DIM

$$E[\overline{X}] = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n}\sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu \ \forall i} = \frac{1}{n} n\mu = \mu$$

$$Var[\overline{X}] = Var\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^{n} X_i\right] =$$

$$= \frac{1}{n^2}\Big(\sum_{i=1}^{n} \underbrace{Var[X_i]}_{\sigma^2 \ \forall i} + 2 \sum_{i<j} \underbrace{Cov(X_i, X_j)}_{=0 \ \forall i \neq j} \Big)$$

$$= \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i^2 - 2X_i\overline{X} + \overline{X}^2 \right) =$$

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

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - 2\overline{X}\, n\overline{X} + n\overline{X}^2 \right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right)$$

$$(n-1) S^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2$$

$$(n-1)\mathbb{E}[S^2] = \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 - n\overline{X}^2 \right] =$$

$$= \mathbb{E}\left[\sum_{i=1}^{n} (X_i - \mu + \mu)^2 - n(\overline{X} - \mu + \mu)^2 \right] =$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu + \mu)^2 \right] - n\,\mathbb{E}\left[(\overline{X} - \mu + \mu)^2 \right] =$$

$$= \sum_{i=1}^{n} \mathbb{E}\left[(X_i - \mu)^2 + 2\mu(X_i - \mu) + \mu^2 \right] - n\,\mathbb{E}\left[(\overline{X} - \mu)^2 + \mu^2 + 2\mu(\overline{X} - \mu) \right]$$

$$= \sum_{i=1}^{n} \left(\mathbb{E}[(X_i - \mu)^2] + 2\mu \underbrace{\cancel{\mathbb{E}[X_i - \mu]}}_{=0} + \mu^2 \right)$$

$$- n\left(\mathbb{E}[(\overline{X} - \mu)^2] + \mu^2 + 2\mu \underbrace{\cancel{\mathbb{E}[\overline{X} - \mu]}}_{=0} \right)$$

$$= \sum_{i=1}^{n} \Big(\overbrace{Var[X_i]}^{\sigma^2 \ \forall i} + \mu^2 \Big) - n\Big(\overbrace{Var[\overline{X}]}^{\sigma^2/n} + \mu^2 \Big)$$

$$= n(\sigma^2 + \cancel{\mu^2}) - n\left(\frac{\sigma^2}{n} + \cancel{\mu^2} \right) = (n-1)\sigma^2$$

$$\Rightarrow \mathbb{E}[S^2] = \sigma^2$$

DEF Dato un campione statistico $X_1 \_\_ X_n$ con un parametro caratterizzante θ, dico che una statistica Y del campione è uno stimatore esatto di θ se $E[Y]=\theta$.

OSSERVAZIONE La media campionaria è uno stimatore esatto del valore atteso del campione;

la varianza campionaria è uno stimatore esatto della varianza del campione.