Statistica Inferenziale

X v.a. = POPOLAZIONE

ESEMPIO: X = MASCHI DI 21 ITALIANI

Voglio conoscere valore di una caratteristica di X non è nota

ESEMPIO: Media dell' altezza

DEF: Un campione casuale di taglia n estratto da una popolaz. X è una n-upla di v.a. $X_1 \dots X_n$ iid $X_i \sim X$.

DEF: Dopo l'osservazione avrò $X_1 = x_1 \dots X_n = x_n$ valori numerici, $(x_1 \dots x_n)$ si chiama realizzazione del campione.

PRIMA DELL'OSSERVAZIONE	DOPO L'OSSERVAZIONE
$X_1 \dots X_n$	$x_1 \dots x_n$ NUMERI
Campione casuale	Realizzazione del campione

popolazione = X v.a. $\sim F$ ← f. ripartizione

$X_1 \dots X_n$ campione casuale estratto da $X \sim F$ è una n-upla di v.a. iid $X_i \sim F$ $\forall i: 1 \dots n$

Se osservo $X_1 = x_1$, $X_2 = x_2 \dots$, $X_n = x_n$ $(x_1 \dots x_n)$ è detta realizzazione del campione casuale/osservazione del campione

Taglia del c.c. $(X_1, \dots, X_n) = n$

Sinonimi di taglia = ampiezza, dimensione

Inferenza
- Ⓐ PARAMETRICA: se conosco F a meno di parametri
- Ⓑ NON PARAMETRICA: se non conosco F

ESEMPIO

X = altezza dei 21enni

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

CASO Ⓐ

(2) pezzi difettosi prodotti da uno stabilimento

mi interessa solo sapere se il pezzo è difettoso o meno

X
- 1 pezzo è difettoso
- 0 altrimenti

→ $X \sim Be(p)$

Caso A ← l'incognito è il parametro p

estraggo n << #pezzi prodotti

$X_1 \dots X_n$ c.c. estratto da $X \sim Be(p)$

Possibile realizzazione $(n=5)$: 0,1,0,0,1 ; il $\frac{\#\text{pezzi difettosi}}{\#\text{pezzi totali}}$ è $\frac{X_1 + \dots + X_n}{n}$

$X_1 + \dots + X_n \sim Binom(n, p)$

esempio: X = altezza di un 21 enne a caso

$X \sim N(\mu, \sigma^2)$ voglio dire qualcosa su μ

$X_1 \dots X_n$ c.c. $X_K \sim N(\mu, \sigma^2)$ $\qquad \dfrac{X_1 + \dots + X_n}{n} \sim N(\mu, \sigma^2/n)$ gaussiana (somma di n Gauss. indip.)

Scarto quadratico medio $\frac{1}{n}\sum_{K=1}^{n}(x_K - \bar{x}_n)^2$ più è grande più i dati si disperdono

$$\frac{1}{n}\sum_{K=1}^{n}(X_K - \bar{X}_n)^2$$

DATO UN CAMPIONE C. $X_1 \dots X_n$ estratto da X (popolazione), si chiama:

MEDIA CAMPIONARIA $\bar{X}_n = \dfrac{X_1 + \dots + X_n}{n}$

VARIANZA CAMPIONARIA $S_n^2 = \dfrac{1}{n-1}\sum_{K=1}^{n}(x_K - \bar{x}_n)^2$

DEF: Dato un c.c. $X_1 \dots X_n$ estratto da $X \sim F$, una statistica è una v.a. che è funzione del c.c. $X_1 \dots X_n$,

$$D_n = d_n(X_1 \dots X_n) \qquad d_n : \mathbb{R}^n \to \mathbb{R}^K$$

ESEMPI 1) $\bar{X}_n = d_n(X_1 \dots X_n)$ dove

$d_n(x_1 \dots x_n) = \dfrac{x_1 + \dots + x_n}{n}$

$\to \bar{X}_n$ è una statistica del campione

2) $S_n^2 : \dfrac{1}{n-1}\sum_{K=1}^{n}(X_K - \bar{X}_n)^2 = g_n(X_1 \dots X_n) \qquad g_n(x_1 \dots x_n) = \dfrac{1}{n-1}\sum_{K=1}^{n}(x_K - \bar{x}_n)^2$

OSS Dopo l'osservazione del campione c conosco il valore di qualsiasi statistica.

Altre statistiche importanti

Momento campionario di ordine $h \in \mathbb{N}$

ORDINE DEL MOMENTO

$$M_n^{(h)} := \frac{1}{n}\sum_{K=1}^{n} x_K^h = \frac{X_1^h + X_2^h + \dots + X_n^h}{n}$$

Taglia del campione

$h = 1$

$M_n^{(1)} = \bar{X}_n$

PROP: Sia $(X_1 \dots X_n)$ un c.c.

$$S_n^2 = \frac{n}{n-1}\left[M_n^{(2)} - (M_n^{(1)})^2\right]$$

$$= \frac{n}{n-1}\,Var(X) = \frac{n}{n-1}\left[E(X^2) - (E(X))^2\right] = \frac{n}{n-1}\left[M_n^{(2)} - (M_n^{(1)})^2\right]$$

DIM $S_n^2 = \dfrac{1}{n-1}\left[\sum_{K=1}^{n}(X_K - \bar{X}_n)^2\right] = \dfrac{1}{n-1}\left[\sum_{K=1}^{n} X_K^2 - 2\bar{X}_n\sum_{K=1}^{n} X_K + \sum_{K=1}^{n}(\bar{X}_n)^2\right]$

$= \dfrac{1}{n-1}\left[n M_n^{(2)} - 2n(M^{(1)})^2 + n(M_n^{(1)})^2\right]$

$= \dfrac{1}{n-1}\left[n M_n^{(2)} - n(M_n^{(1)})^2\right]$

$\sum_{K=1}^{n} X_K^2 = n M_n^{(2)}$

$\sum_{K=1}^{n} X_K = n M_n^{(1)}$

$\bar{X}_n = M_n^{(1)}$

PROP: Sia $X_1 \dots X_m$ un c.c. estratto da una popolazione X con $E(X) = \mu$ e $Var(X) = \sigma^2$

Allora: $E(S_m^2) = \sigma^2$ $\qquad Var(S_m^2) = \frac{1}{m}\left[\mu_4 - \frac{(m-3)}{m-1}\sigma^4\right]$

$\mu_4 = E\left[(X-\mu)^4\right] \qquad m > 1$

DIM:

$$S_m^2 = \frac{m}{m-1}\left[M_m^{(2)} - (M_m^{(1)})^2\right] \qquad \sigma^4 = (\sigma^2)^2$$

$$E(S_m^2) = \frac{m}{m-1}\left\{E(M_m^{(2)}) - E\left[(M_m^{(1)})^2\right]\right\}$$

$$E(M_m^{(2)}) = E\left(\frac{1}{m}\sum_{k=1}^{m} X_k^2\right) = \frac{1}{m}\sum_{k=1}^{m} E(X_k^2)$$

$$E(X_k^2) = E(X^2) = Var(X) + (E(X))^2 = \sigma^2 + \mu^2 \quad \forall k: 1 \dots m$$

$$\to E(M_m^{(2)}) = \frac{1}{m}\sum_{k=1}^{m}(\sigma^2 + \mu^2) = \sigma^2 + \mu^2$$

$$E\left[(M_m^{(1)})^2\right] = E\left[\bar{X}_m^2\right] = Var(\bar{X}_m) + (E(\bar{X}_m))^2 = \frac{\sigma^2}{m} + \mu^2$$

$$\to E(S_m^2) = \frac{m}{m-1}\left[\sigma^2 + \cancel{\mu^2} - \frac{\sigma^2}{m} - \cancel{\mu^2}\right] = \frac{m}{m-1}\,\frac{(m-1)}{m}\sigma^2 = \sigma^2$$

OSS. $Var(S_m^2) = \frac{1}{m}\left[\tilde{\mu}_4 - \frac{m-3}{m-1}\sigma^4\right] \xrightarrow[m\to\infty]{} 0 \qquad \tilde{\mu}_4 : E((X-\mu)^4)$

$\exists\ E((X-\mu)^4)$, $Var(X)$

OSS. $E(M_m^{(2)}) = \mu_2 \qquad E(M_m^{(k)}) = \mu_k$

OSS. Abbiamo dimostrato che

$E(\bar{X}_m) = \mu = E(X)$

$E(S_m^2) = \sigma^2 = Var(X)$

Per un c.c. $X_1, \dots X_m$ estratto da una X qualsiasi, purchè esistano μ e σ^2

PROP: $X_1 \dots X_m$ c.c. estratto da $X \sim N(\mu, \sigma^2)$

Allora:

(1) $\bar{X}_m \sim N(\mu, \sigma^2/m) \quad \forall m \in \mathbb{N}$

(2) $\bar{X}_m \perp\!\!\!\perp S_m^2$

(3) $(m-1)\frac{S_m^2}{\sigma^2} \sim \chi^2(m-1)$

(4) $\frac{\bar{X}_m - \mu}{\sqrt{\frac{S_m^2}{m}}} \sim t(m-1)$

(5) $X_1 \dots X_m$ estratto da $X \sim N(\mu_X, \sigma_X^2)$

$Y_1 \dots Y_m$ " " $Y \sim N(\mu_Y, \sigma_Y^2)$

con $(X_1 \dots X_m) \perp\!\!\!\perp (Y_1 \dots Y_m)$

Allora $S_m^{2,X} = \frac{1}{m-1}\sum_{k=1}^{m}(X_k - \bar{X}_m)^2 \perp\!\!\!\perp S_m^{2,Y} = \frac{1}{m-1}\sum_{h=1}^{m}(Y_h - \bar{Y}_m)^2$

$(m-1)\frac{S_m^{2,X}}{\sigma^2} \sim \chi^2(m-1)$ e $(m-1)\frac{S_m^{2,Y}}{\sigma_Y^2} \sim \chi^2(m-1)$

se $\sigma_X^2 = \sigma_Y^2 \longrightarrow \frac{S_m^{2,X}}{S_m^{2,Y}} \sim F_{m-1,m-1} = \mathbb{F}(m-1, m-1)$