Econometria 6089 - CLES

FORMULARIO DI STATISTICA

Valore atteso:

$$x \sim f(x) : E[x] = \sum_x x f(x) \text{ oppure } E[x] = \int_x x f(x) dx$$

Inoltre: $E[y|x] = \sum_y y f(y|x)$ oppure $E[y|x] = \int_y y f(y|x) dy$, dove $f(y|x) = \frac{f(x,y)}{f(x)}$

Nota bene: tutte le proprietà che seguono valgono anche nel caso dei valori attesi condizionali, anche se, per semplificare la trattazione, non è esplicitamente indicato.

$$E[x + y] = E[x] + E[y]$$

$$E[cx + a] = cE[x] + a$$

Varianza

$$Var(X) = \sigma^2 = E[X - E[X]]^2 = E[X^2] - E[X]^2$$

$$Var(a + cX) = E[(a + cX) - E[(a + cX)]]^2 = c^2 Var[X]$$

Covarianza

$$cov(x,y) = \sigma_{xy} = E[(X - E[X])(Y - E[Y])] = \sum_x \sum_y [x - E(X)][y - E(Y)] f(x,y)$$

Si può facilmente verificare che:

$$cov(x,y) = \sigma_{xy} = E[XY] - E[X]E[Y]$$

Correlazione

$$\rho = \frac{cov(X,Y)}{\sqrt{var(X)var(Y)}}$$

<u>Nota Bene</u>:

X e Y si diranno non correlati quando: $\rho = 0 \Leftrightarrow cov(x,y) = 0$

Varianza di una combinazione lineare di variabili aleatorie

$$Var[aX + bY] = a^2Var[X] + b^2Var[Y] + +2ab\,cov[X,Y]$$

Se X e Y sono indipendenti, o non correlate, la loro covarianza è uguale a zero, e quindi:

$$Var[aX + bY] = a^2Var[X] + b^2Var[Y]$$

Distribuzione normale

$$X \sim \mathrm{N}(\mu,\sigma^2): f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[-\frac{1}{2}\frac{(x-\mu)}{\sigma^2}\right]$$

Se $X \sim \mathrm{N}(\mu,\sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim \mathrm{N}(0,1)$

Se $X \sim \mathrm{N}(\mu,\sigma^2)$ ed a è una costante, allora $P[X \leq a] = P\left[Z \leq \frac{a-\mu}{\sigma}\right]$

Se $X \sim \mathrm{N}(\mu,\sigma^2)$ ed a e b sono costanti, allora $P[a \leq X \leq b] = P\left[\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right]$

Se $X \sim \mathrm{N}(\mu_x,\sigma_x{}^2)$ e $Y \sim \mathrm{N}(\mu_y,\sigma_y{}^2) \Rightarrow aY + BX \sim \mathrm{N}(a\mu_y + \mathrm{b}\mu_x; \mathrm{a}^2\sigma_y{}^2 + \mathrm{b}^2\sigma_x{}^2 + 2a\mathrm{b}\sigma_{yx})$

Distribuzioni notevoli

Distribuzione Chi-quadrato:

$$y_j \text{ iid } \mathrm{N}(0,1) \Rightarrow X_J^2 : \xi = \sum_{j=1}^{J} y_j^2$$

J indica il numero di gradi di libertà. La distribuzione Chi-quadrato, essendo ottenuta come combinazione lineare di normali al quadrato, non può essere negativa, ed ha una coda spostata verso destra.

Distribuzione t-Student:

$$Y \sim \mathrm{N}(0,1),\ \xi \sim \mathrm{X}_J^2,\ Y \text{ e } \xi \text{ indipendenti} \Rightarrow t-student : t = \frac{Y}{\sqrt{\xi / J}}$$

Anche in questo caso J indica il numero di gradi di libertà. Come la Normale è simmetrica intorno alla media, ma è meno "alta" ed ha code più ampie.

Distribuzione F:

$$\xi_1 \sim \mathrm{X}_{J_1}^2,\ \xi_2 \sim \mathrm{X}_{J_2}^2,\ \xi_1 \text{ e } \xi_2 \text{ indipendenti} \Rightarrow F : f = \frac{\xi_1 / J_1}{\xi_2 / J_2}$$

Essendo il rapporto tra due Chi-quadrato, non può essere negativa. Non è simmetrica e la coda è spostata verso destra. J_1 e J_2 sono anche in questo caso i gradi di libertà.