



Thống kê mô tả

Bởi:

Phạm Trí Cao

Thống kê mô tả

Mô tả dữ liệu thống kê(Descriptive Statistic)

Có bốn tính chất mô tả phân phối xác suất của một biến ngẫu nhiên như sau:

Xu hướng trung tâm hay “điểm giữa” của phân phối.

Mức độ phân tán của dữ liệu quanh vị trí “điểm giữa”.

Độ trôi(skewness) của phân phối.

Độ nhọn(kurtosis) của phân phối.

Mối quan hệ thống kê giữa hai biến số được mô tả bằng hệ số tương quan.

Xu hướng trung tâm của dữ liệu

Trung bình tổng thể (giá trị kỳ vọng) $\mu_X = E[X]$

Trung bình mẫu

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

Trung vị của tổng thể : X là một biến ngẫu nhiên liên tục, Md là trung vị của tổng thể khi $P(X < Md) = 0,5$.

Trung vị mẫu : Nếu số phân tử của mẫu là lẻ thì trung vị là số “ở giữa” của mẫu sắp theo thứ tự tăng dần hoặc giảm dần.

Nếu số phân tử của mẫu chẵn thì trung vị là trung bình cộng của hai số “ở giữa”.

Thống kê mô tả

Trong kinh tế lượng hầu như chúng ta chỉ quan tâm đến trung bình mà không tính toán trên trung vị.

Độ phân tán của dữ liệu

Phương sai

Phương sai của tổng thể :

$$\sigma_x^2 = E[(X - \mu_x)^2]$$

Phương sai mẫu:

$$S_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

hoặc

$$\hat{\sigma}_x^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

Độ lệch chuẩn

Độ lệch chuẩn tổng thể :

$$\sigma_x = \sqrt{\sigma_x^2}$$

Độ lệch chuẩn mẫu :

$$S_x = \sqrt{S_x^2}$$

hoặc :

$$\hat{\sigma}_x = \sqrt{\hat{\sigma}_x^2}$$

Độ trôi S

Độ trôi tổng thể :

Thống kê mô tả

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]$$

Độ trôi mẫu :

$$S = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma}\right)^3$$

Đối với phân phối chuẩn độ trôi bằng 0.

Độ nhọn K

Độ nhọn của tổng thể

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right]$$

Độ nhọn mẫu

$$K = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - \bar{X}}{\sigma}\right)^4$$

Đối với phân phối chuẩn độ nhọn bằng 3. Một phân phối có K lớn hơn 3 là nhọn, nhỏ hơn 3 là phẳng.

Quan hệ giữa hai biến-Hệ số tương quan

Hệ số tương quan tổng thể

$$\rho_{\mathbf{H}} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y}$$

Hệ số tương quan mẫu

$$r_{\mathbf{H}} = \frac{S_{\mathbf{H}}}{S_X S_Y}$$

với

$$s_{xy} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})$$