11-Correlation

ENSY SILVER[1]

Thursday 10th September, 2020

[1]Thanks to my family, my friend and freedom.

1 The correlation coefficient

Theorem 1.1. *Let X and Y be any two random variables. The correlation coefficient of X and Y, denoted $\rho(X,Y)$, is given by*

$$\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma_X \sigma_Y}$$

Corollary 1.1.1. *For any two variables X and Y,*

1. $|\rho(X,Y)| \leq 1$.
2. $|\rho(X,Y)| = 1$ *if and only if* $Y = aX + b$ *for some constants a and b, where* $a \neq 0$.

Since the correlation coefficient is defined by

$$\rho(X,Y) = \frac{E(XY) - E(X)E(Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

and we define the **sample correlation coefficient** by

$$R = \frac{\frac{1}{n}\sum_{i=1}^{n} X_i Y_i - \bar{X}\bar{Y}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

Corollary 1.1.2. *Let $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$ be a set of measurements whose sample correlation coefficient is r. Then*

$$r = \hat{\beta}_1 \sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}}$$

where $\hat{\beta}_1$ is the maximum likelihood estimate for the slope.

2 Bivariate normal distribution

We omit the deduction of bivariate normal distribution and give the definition now.

Theorem 2.1. *Let X and Y be random variables with joint pdf*

$$\begin{aligned} f_{X,Y}(x,y) = &\frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \cdot \\ &\exp\left\{-\frac{1}{2}\left(\frac{1}{1-\rho^2}\right)\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{x-\mu_X}{\sigma_X}\cdot\frac{y-\mu_Y}{\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\} \end{aligned}$$

Bivariate normal distribution has several properties.

Theorem 2.2. *Suppose that X and Y are random variables having the bivariate normal distribution. Then*

1. *$f_X(x)$ is a normal pdf with means μ_X and variance σ_X^2, $f_Y(y)$ is a normal pdf with mean μ_Y and variance σ_Y^2.*

2. *ρ is the correlation coefficient between X and Y.*

3. *$E(Y|x) = \mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X)$.*

4. *$\mathrm{Var}(Y|x) = (1 - \rho^2)\sigma_Y^2$.*

3 Estimating parameters in the bivariate normal pdf

Theorem 3.1. *Given that $f_{X,Y}(x, y)$ is a bivariate normal pdf, the maximum likelihood estimators for $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2, \rho$, are $\bar{X}, \bar{Y}, \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2, \frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})^2$ and R, respectively.*

Theorem 3.2. *Let $(X_1, Y_1), (X_2, Y_2), \cdots, (X_n, Y_n)$ be a random sample of size n drawn from a bivariate normal distribution, and let R be the sample correlation coefficient. Under the null hypothesis that $\rho = 0$, the statistic*

$$T_{n-2} = \frac{\sqrt{n-2}R}{\sqrt{1-R^2}}$$

has a Student t distribution with $n - 2$ degrees of freedom.