MIE237

Neil Montgomery

2016-03-01

New Topic: Correlation

Preview of "correlation" (from 2016-01-29)

Recall $\mathrm{Cov}(X, Y) = E(X - E(X))(Y - E(Y)) = E(XY) - E(X)E(Y)$.

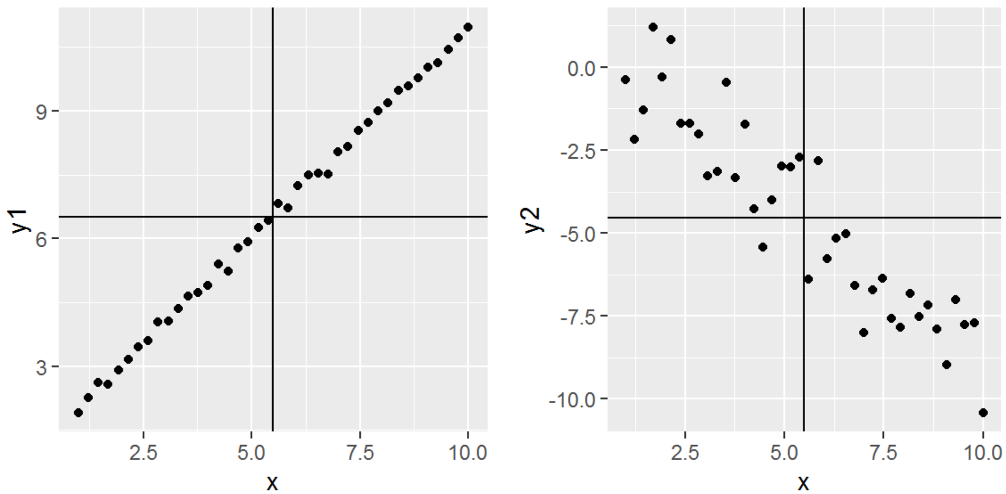
Measures linear relationship between X and Y.

A related measure that divides by the two standard deviations:

$$\rho = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}\,(X)\mathrm{Var}\,(Y)}}$$

"Correlation coefficient" ρ is a measure of the linear relationship between the random variables X and Y. We'd like an analogous measure for .

Correlation



'Twill be based on:

$$S_{xy} = \sum_{i=1}^{n} \left(x_i - \overline{x}\right) \left(y_i - \overline{y}\right)$$

Sample correlation coefficient

Define:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

r is an estimator of ρ, called the "sample correlation coefficient".

In regression analysis the x variable is considered fixed, and y is a function of x.

Correlation is a property of Y and X both random, and correlation is symmetric in X and Y.

Nevertheless we use (misuse?) sample correlation coefficient also in a regression context as a useful numerical summary of a bivariate relationship.

Connection with R^2

Recall that $S_{yy} = \sum(y_i - \bar{y})^2 = SST$ and $R^2 = SSR \big/ SST = 1 - SSE \big/ SST$.

It turns out also that $SSE = S_{yy} - \hat{\beta}_1 S_{xy}$. So:

$$\begin{aligned} R^2 &= 1 - \frac{SSE}{SST} \\ &= 1 - \frac{S_{yy} - \hat{\beta}_1 S_{xy}}{S_{yy}} \\ &= 1 - \left(1 - \frac{\hat{\beta}_1 S_{xy}}{S_{yy}} \right) \\ &= \frac{S_{xy}^2}{S_{xx} S_{yy}} = (r)^2 \end{aligned}$$

Other properties of r

Since $0 \leq R^2 \leq 1$ it follows $-1 \leq r \leq 1$.

Also:

$$r = \hat{\beta}_1 \sqrt{\frac{S_{xy}}{S_{xx}}}$$

It is possible to do things like test $H_0 : \rho = 0$ versus $H_1 : \rho \neq 0$ using r, but we will not cover this.

Examples