Simple linear regression model

written by The Coué method on Functor Network
original link: https://functor.network/user/1751/entry/653

Goal

Write a mathematical model $y = \alpha + \beta x$ that describes the relationship between two variables x and y.

Setup

Given observations (x_i, y_i), $i = 1, \ldots, n$, consider a model of the form

$$y_i = \alpha + \beta x_i + e_i$$

where e_i is the random part of the model. The only assumption is that the mean of e_i's is 0. The aim is to find estimates for α and β.

Comments

- The model assumes that the x_i's are known exactly and that the error terms appear only in the y_i's.
- Note that y_i is the actual value and that $\alpha + \beta x_i$ is the predicted value, so $e_i = y_i - (\alpha + \beta x_i)$ is the ith residual.

The function to minimize

The residual sum of squares function $\mathrm{RSS}_{x_1,x_2,\ldots,x_n,y_1,y_2,\ldots,y_n}(\alpha, \beta)$ denoted RSS is

$$\mathrm{RSS} = \sum_{i=1}^{n} e_i^2$$

or equivalently

$$\mathrm{RSS} = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2$$

Solving for α

Differentiating RSS with respect to α gives

$$\frac{\partial \text{RSS}}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = -2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) \tag{1}$$

and setting $\frac{\partial \text{RSS}}{\partial \alpha}$ to zero yields

$$\frac{\partial \text{RSS}}{\partial \alpha} = 0 \implies \alpha = \frac{1}{n} \sum_{i=1}^{n} (y_i - \beta x_i) = \bar{y} - \beta \bar{x}$$

where $(\bar{x}, \bar{y}) = (\frac{1}{n} \sum_{i=1}^{n} x_i, \frac{1}{n} \sum_{i=1}^{n} y_i)$ is the centroid of the n observations.

Hence the estimate for α is

$$\hat{\alpha} = \bar{y} - \hat{\beta} \bar{x} \tag{2}$$

Comments

- Note that the right hand side of equation (1) is equivalent to $-2 \sum_{i=1}^{n} e_i$, so setting $\frac{\partial \text{RSS}}{\partial \alpha}$ to 0 implies that

$$\sum_{i=1}^{n} e_i = 0 \tag{3}$$

- Note that this is the assumption that the mean of e_i's is 0. So, in some sense, this assumption follows from the given model.
- When you plot the n observations (x_i, y_i), $i = 1, \ldots, n$ in the xy-plane, the line $y = \hat{\alpha} + \hat{\beta} x$ passes through the centroid $(\bar{x}, \bar{y})$ of the n observations.
- It can be useful to think of the centroid $(\bar{x}, \bar{y})$ as a fixed fulcrum and to think of the line $y = \hat{\alpha} + \hat{\beta} x$ as a lever moving on this fulcrum. To completely determine the line, you'd need to find an estimate for β, which is the slope of the line.

Solving for β

Differentiating RSS with respect to β gives

$$\frac{\partial \text{RSS}}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 = \sum_{i=1}^{n} 2(y_i - \alpha - \beta x_i)(-x_i) \tag{4}$$

and setting $\frac{\partial \text{RSS}}{\partial \beta}$ to zero yields

$$\begin{aligned}\frac{\partial \text{RSS}}{\partial \beta} = 0 &\implies \sum_{i=1}^{n} x_i y_i = \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2 \\ &\implies \sum_{i=1}^{n} x_i y_i = \alpha n \bar{x} + \beta \sum_{i=1}^{n} x_i^2\end{aligned}$$

Using (2) yields

$$\sum_{i=1}^{n} x_i y_i = (\bar{y} - \beta \bar{x}) n \bar{x} + \beta \sum_{i=1}^{n} x_i^2$$

or

$$(\sum_{i=1}^{n} x_i y_i) - n \bar{x} \bar{y} = \beta ((\sum_{i=1}^{n} x_i^2) - n \bar{x}^2)$$

or

$$\beta = \frac{(\sum_{i=1}^{n} x_i y_i) - n \bar{x} \bar{y}}{(\sum_{i=1}^{n} x_i^2) - n \bar{x}^2}$$

or, after dividing both numerator and denominator by n,

$$\beta = \frac{\frac{1}{n}(\sum_{i=1}^{n} x_i y_i) - \bar{x} \bar{y}}{\frac{1}{n}(\sum_{i=1}^{n} x_i^2) - \bar{x}^2}$$

Note that the numerator of the fraction in the previous expression is $\text{Cov}(x, y)$ and the denominator is $\text{Cov}(x, x)$, as shown by the following computations.

$$\begin{aligned}
\mathrm{Cov}(x,y) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \bar{y} - \bar{x} \cdot \sum_{i=1}^{n} y_i + n\bar{x}\bar{y}) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i y_i - n\bar{x} \cdot \bar{y} - \bar{x} \cdot n\bar{y} + n\bar{x}\bar{y}) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i y_i) - \bar{x}\bar{y}
\end{aligned}$$

$$\begin{aligned}
\mathrm{Cov}(x,x) &= \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i^2 - 2\bar{x} \cdot \sum_{i=1}^{n} x_i + n\bar{x}^2) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i^2 - 2\bar{x} \cdot n\bar{x} + n\bar{x}^2) \\
&= \frac{1}{n}((\sum_{i=1}^{n} x_i^2) - n\bar{x}^2) \\
&= \frac{1}{n}(\sum_{i=1}^{n} x_i^2) - \bar{x}^2
\end{aligned}$$

Therefore

$$\beta = \frac{\mathrm{Cov}(x,y)}{\mathrm{Cov}(x,x)}$$

Using $\mathrm{Cov}(x,x) = \sigma_x^2$ and $\mathrm{Cov}(x,y) = \sigma_x\sigma_y r_{xy}$, where σ_x is the standard deviation of the x_i's, σ_y is the standard deviation of the y_i's, and r_{xy} is the correlation between the x_i's and the y_i's, we get

$$\beta = \frac{\sigma_x\sigma_y r_{xy}}{\sigma_x^2} = r_{xy}\frac{\sigma_y}{\sigma_x}$$

Hence the estimate for β is

$$\hat{\beta} = r_{xy}\frac{\sigma_y}{\sigma_x} \tag{5}$$

Comments

- Note that the right hand side of equation (4) is equivalent to $-2\sum_{i=1}^n e_i x_i$, so setting $\frac{\partial \text{RSS}}{\partial \beta}$ to 0 implies that

$$\sum_{i=1}^n e_i x_i = 0 \tag{6}$$

- Note that the population factor $\frac{1}{n}$ has been used in all formulas (for example, for covariance, standard deviation, etc.).
- Note that $\hat{\beta}$ is proportional to the correlation r_{xy}.

The critical point is a local minimum

In order to conclude that the critical point $(\hat{\alpha}, \hat{\beta})$ is a local minimum for RSS, it is sufficient to show that the Jacobian of RSS at $(\hat{\alpha}, \hat{\beta})$ is a positive definite 2×2 matrix.

From (1) it follows that $\frac{\partial \text{RSS}}{\partial \alpha} = 2n\alpha + 2n\beta\bar{x} - 2n\bar{y}$ and from (4), it follows that $\frac{\partial \text{RSS}}{\partial \beta} = 2n\alpha\bar{x} + 2\beta\sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i y_i$. Therefore the Jacobian of RSS at $(\hat{\alpha}, \hat{\beta})$ is the matrix

$$\begin{pmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2\sum_{i=1}^n x_i^2 \end{pmatrix}$$

The matrix is positive definite if and only if two conditions are satisfied: (i) the (1,1) entry of the matrix is positive; and (ii) the determinant of the Jacobian is positive. Condition (i) is satisfied because the (1,1) entry of the Jacobian is $2n$, which is positive. Condition (ii) is also satisfied because the determinant is $4(n\sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2)$, which is equivalent to $4\sum_{i=1}^n\sum_{j=1}^n (x_i - x_j)^2$, which is positive.

The model

Using (2) and (5), the model $\hat{y} = \hat{\alpha} + \hat{\beta}x$ can be written as

$$\hat{y} = \bar{y} + r_{xy}\frac{\sigma_y}{\sigma_x}(x - \bar{x}) \tag{7}$$

or

$$\frac{\hat{y}-\bar{y}}{\sigma_y} = r_{xy}\frac{x-\bar{x}}{\sigma_x} \tag{8}$$

Regression to the mean

From either (7) or (8), it follows that

$$\sigma_{\hat{y}} = |r_{xy}|\sigma_y \leq \sigma_y \tag{9}$$

Inequality (9) is the essence of the phenomenon that is commonly known as 'regression to the mean'.

Summing up

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = r_{xy}\frac{\sigma_y}{\sigma_x}$$

$$\hat{y} = \bar{y} + r_{xy}\frac{\sigma_y}{\sigma_x}(x-\bar{x})$$

Mean squared error (MSE)

$$\begin{aligned}
\text{MSE} &= \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{y}_i)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\bar{y}-r_{xy}\frac{\sigma_y}{\sigma_x}(x_i-\bar{x})\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left((y_i-\bar{y})-r_{xy}\frac{\sigma_y}{\sigma_x}(x_i-\bar{x})\right)^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_i-\bar{y})^2 + \frac{1}{n}\sum_{i=1}^{n}\left(r_{xy}\frac{\sigma_y}{\sigma_x}(x_i-\bar{x})\right)^2 - \frac{1}{n}\cdot 2r_{xy}\frac{\sigma_y}{\sigma_x}\sum_{i=1}^{n}(y_i-\bar{y})(x_i-\bar{x}) \\
&= \sigma_y^2 + r_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}\sigma_x^2 - 2r_{xy}\frac{\sigma_y}{\sigma_x}\sigma_x\sigma_y r_{xy} \\
&= \sigma_y^2 + r_{xy}^2\sigma_y^2 - 2r_{xy}^2\sigma_y^2 \\
&= \sigma_y^2(1-r_{xy}^2)
\end{aligned}$$

Sums of squares (residual, explainable, total)

$$\text{RSS} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$
$$\text{ESS} = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$
$$\text{TSS} = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

Theorem TSS = RSS + ESS

The geometric interpretation of the theorem

In $\mathbb{R}^n$, consider the plane $\mathcal{P}$ spanned by $[1 \quad 1 \quad \cdots \quad 1]^T$ and $[x_1 \quad x_2 \quad \cdots \quad x_n]^T$. Let C be the point $(\bar{y}, \ldots, \bar{y})$, which lies on the line generated by $[1 \quad 1 \quad \cdots \quad 1]^T$ in $\mathcal{P}$. If P is the point $(y_1, y_2, \ldots, y_n)$ and $\hat{P}$ is the point $(\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n)$, it follows that $\hat{P}$ is the projection of the point P on $\mathcal{P}$. The vector from P to $\hat{P}$ is perpendicular to the vector from $\hat{P}$ to C, so the triangle with vertices C, P, and $\hat{P}$ is a right triangle with a right angle at $\hat{P}$. The theorem is equivalent to the Pythagorean theorem applied to the right triangle $CP\hat{P}$.

Proof of TSS = RSS + ESS Recall that $\sum_{i=1}^{n} e_i = 0$ (equation 3) and $\sum_{i=1}^{n} e_i x_i = 0$ (equation 6).

$$\begin{aligned}
\text{TSS} &= \sum_{i=1}^{n} (y_i - \bar{y})^2 \\
&= \sum_{i=1}^{n} ((y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}))^2 \\
&= \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 + 2\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) \\
&= \text{RSS} + \text{ESS} + 2\boxed{\sum_{i=1}^{n} (y_i - \hat{y}_i)(\hat{y}_i - \bar{y})}
\end{aligned}$$

To complete the proof, it's sufficient to prove that the boxed expression is 0.

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y}) &= \sum_{i=1}^{n} e_i(\hat{y}_i-\bar{y}) \\
&= \sum_{i=1}^{n} e_i r_{xy}\frac{\sigma_y}{\sigma_x}(x_i-\bar{x}) \quad \text{using equation (7)} \\
&= r_{xy}\frac{\sigma_y}{\sigma_x}(\sum_{i=1}^{n} e_i x_i - \bar{x}\sum_{i=1}^{n} e_i) \\
&= r_{xy}\frac{\sigma_y}{\sigma_x}(0-\bar{x}\cdot 0) \\
&= 0
\end{aligned}$$

Definition $R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$

Theorem $R^2 = r_{xy}^2$

Proof By definition, $\text{TSS} = n\sigma_y^2$.

$$\begin{aligned}
\text{ESS} &= \sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2 \\
&= \sum_{i=1}^{n} r_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}(x_i-\bar{x})^2 \quad \text{using equation (7)} \\
&= r_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}(\sum_{i=1}^{n}(x_i-\bar{x})^2) \\
&= r_{xy}^2\frac{\sigma_y^2}{\sigma_x^2}\cdot n\sigma_x^2 \\
&= r_{xy}^2 n\sigma_y^2
\end{aligned}$$

Therefore, $R^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{r_{xy}^2 n\sigma_y^2}{n\sigma_y^2} = r_{xy}^2$.

Theorem $R^2 = r_{y\hat{y}}^2$

Proof sketch

- Show that the mean of $x_i - \bar{x}$ is 0.
- Use equation (7) to show that the mean of $\hat{y}_i - \bar{y}$ is 0, which implies that the mean of $\hat{y}_i$ is $\bar{y}$.
- Use the bilinearity of covariance and equation (7) to show that $\sigma_{\hat{y}}^2 = r_{xy}^2\sigma_y^2$.
- Use the bilinearity of covariance to show that $\text{Cov}(\hat{y}, y) = r_{xy}\frac{\sigma_y}{\sigma_x}\text{Cov}(x, y)$.
- Use the definition of correlation in terms of covariance and variance to conclude that $r_{y\hat{y}}^2 = r_{xy}^2$.

Reference

Ordinary least squares, https://en.wikipedia.org/wiki/Ordinary_least_squares