

[Home](#) » [Forecasting: principles and practice](#) » [Simple regression](#) » 4.2 Least squares estimation

4.2 Least squares estimation

In practice, of course, we have a collection of observations but we do not know the values of β_0 and β_1 . These need to be estimated from the data. We call this “fitting a line through the data”.

There are many possible choices for β_0 and β_1 , each choice giving a different line. The least squares principle provides a way of choosing β_0 and β_1 effectively by minimizing the sum of the squared errors. That is, we choose the values of β_0 and β_1 that minimize

$$\sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2.$$

Using mathematical calculus, it can be shown that the resulting **least squares estimators** are

$$\hat{\beta}_1 = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$

and

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$

where $\bar{x}$ is the average of the x observations and $\bar{y}$ is the average of the y observations. The *estimated* line is known as the “regression line” and is shown in Figure 4.2.

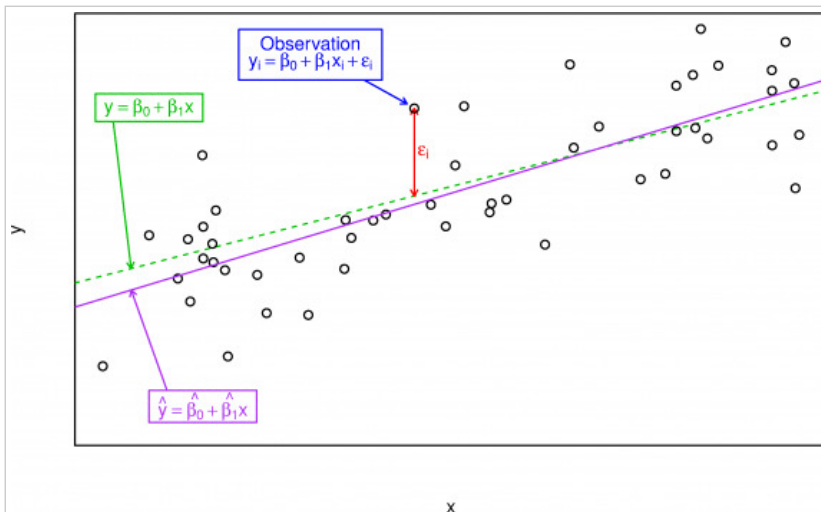


Figure 4.2: Estimated regression line for a random sample of size N .

We imagine that there is a “true” line denoted by $y = \beta_0 + \beta_1 x$ (shown as the dashed green line in Figure 4.2, but we do not know β_0 and β_1 so we cannot use this line for forecasting. Therefore we obtain estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ from the observed data to give the “regression line” (the solid purple line in Figure 4.2).

The regression line is used for forecasting. For each value of x , we can forecast a corresponding value of y using $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$.

Book information



[About this book](#)

[Feedback on this book](#)

[Buy a printed copy](#)

Rob J Hyndman

George Athanasopoulos

Forecasting: principles and practice

- ▶ [Getting started](#)
- ▶ [The forecaster's toolbox](#)
- ▶ [Judgmental forecasts](#)
- ▼ [Simple regression](#)
 - [The simple linear model](#)
 - [Least squares estimation](#)
 - [Regression and correlation](#)
 - [Evaluating the regression model](#)
 - [Forecasting with regression](#)
 - [Statistical inference](#)
 - [Non-linear functional forms](#)
 - [Regression with time series data](#)
 - [Summary of notation and terminology](#)
 - [Exercises](#)
 - [Further reading](#)
- ▶ [Multiple regression](#)
- ▶ [Time series decomposition](#)
- ▶ [Exponential smoothing](#)
- ▶ [ARIMA models](#)
- ▶ [Advanced forecasting methods](#)
 - [Data](#)
 - [Using R](#)
 - ▶ [Resources](#)
 - [Reviews](#)

Fitted values and residuals

The forecast values of y obtained from the observed x values are called “fitted values”. We write these as $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, for $i = 1, \dots, N$. Each $\hat{y}_i$ is the point on the regression line corresponding to observation x_i .

The difference between the observed y values and the corresponding fitted values are the “residuals”:

$$e_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i.$$

The residuals have some useful properties including the following two:

$$\sum_{i=1}^N e_i = 0 \quad \text{and} \quad \sum_{i=1}^N x_i e_i = 0.$$

As a result of these properties, it is clear that the average of the residuals is zero, and that the correlation between the residuals and the observations for the predictor variable is also zero.

[◀ 4.1 The simple linear model](#)

[up](#)

[4.3 Regression and correlation ▶](#)
