Probability and Statistics: To *p*, or not to *p*?

Module Leader: Dr James Abdey

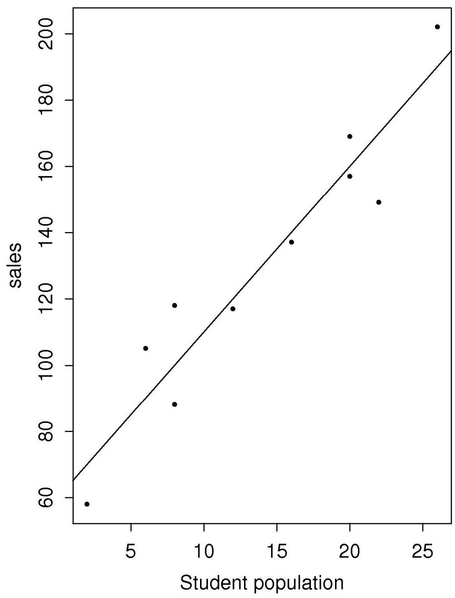
6.3 Linear regression

Linear regression analysis is one of the most frequently-used statistical techniques. It aims to model an explicit relationship between one **dependent variable**, denoted as y, and one or more **regressors** (also called covariates, or independent variables), denoted as $x_1, \ldots, x_p$.

The goal of regression analysis is to **understand** how y depends on $x_1, \ldots, x_p$ and to **predict** or control the unobserved y based on the observed $x_1, \ldots, x_p$. We only consider simple examples with $p = 1$. Interested learners are encouraged to take a full course on regression analysis and/or econometrics.

Example

In a university town, the sales, y, of 10 pizza parlour restaurants are closely related to the student population, x, in their neighbourhoods.



The scatterplot above shows the sales (in thousands of pounds) in a period of three months together with the numbers of students (in thousands) in their neighbourhoods.

We plot y against x, and draw a straight line through the middle of the data points:

$$\boldsymbol{y = \alpha + \beta x + \varepsilon}$$

where ε stands for a **random error term**, α is the intercept and β is the slope of the straight line.

For a given student population, x, the **predicted sales** are:

$$\boldsymbol{\widehat{y} = \alpha + \beta x.}$$

Some other possible examples of y and x are shown in the following table:

y	x
Sales	Price
Weight gain	Protein in diet
Present FTSE 100 index	Past FTSE 100 index
Consumption	Income
Salary	Tenure
Daughter's height	Mother's height

The simple linear regression model

We now present the simple linear regression model. Let the paired observations $(x_1, y_1), \ldots, (x_n, y_n)$ be drawn from the model:

$$\boldsymbol{y_i = \alpha + \beta x_i + \varepsilon_i}$$

where:

$$\mathbf{E}(\varepsilon_i) = \mathbf{0} \quad \textbf{and} \quad \mathbf{Var}(\varepsilon_i) = \sigma^2 > \mathbf{0}.$$

So the model has three parameters: β_0, β_1 and σ^2. In a formal course on regression you would consider the following questions:

- **How to draw a line through data clouds, i.e. how to α and β?**
- **How accurate is the fitted line?**
- **What is the error in predicting a future y?**

Example

We can apply the simple linear regression model to study the relationship between two series of financial returns – a regression of a stock's returns, y, on the returns of an underlying market index, x. This regression model is an example of the **capital asset pricing model (CAPM)**.

Stock returns are defined as:

$$\textbf{return} = \frac{\textbf{current price} - \textbf{previous price}}{\textbf{previous price}} \approx \log\left(\frac{\textbf{current price}}{\textbf{previous price}}\right)$$

when the difference between the two prices is small. Daily prices are definitely not independent. However, daily returns may be seen as a **sequence of uncorrelated random variables**.

The capital asset pricing model (CAPM) is a simple asset pricing model in finance given by:

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

where y_i is a stock return and x_i is a market return at time i.

The **total risk of the stock** is:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\sum_{i=1}^{n}(\widehat{y}_i - \bar{y})^2 + \frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2.$$

The **market-related (or systematic) risk** is:

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

The **firm-specific risk** is:

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2.$$

Some remarks are the following.

i. β measures the market-related (or systematic) risk of the stock.

ii. Market-related risk is unavoidable, while firm-specific risk may be 'diversified away' through *hedging*.

iii. Variance is a simple measure (and one of the most frequently-used) of risk in finance.

So the **'beta' of a stock is a simple measure of the riskiness of that stock** with respect to the market index. By definition, the market index has $\beta = 1$.

If a stock has a beta of 1, then:

$$\text{if the market index} \uparrow \text{ by } 1\%\text{, then the stock} \uparrow \text{ by } 1\%$$

and:

$$\text{if the market index} \downarrow \text{ by } 1\%\text{, then the stock} \downarrow \text{ by } 1\%.$$

If a stock has a beta of 2, then:

$$\text{if the market index} \uparrow \text{ by } 1\%\text{, then the stock} \uparrow \text{ by } 2\%$$

and:

$$\text{if the market index} \downarrow \text{ by } 1\%\text{, then the stock} \downarrow \text{ by } 2\%.$$

If a stock has a beta of 0.5, then:

$$\text{if the market index} \uparrow \text{by } 1\%, \text{ then the stock} \uparrow \text{by } 0.5\%$$

and:

$$\text{if the market index} \downarrow \text{by } 1\%, \text{ then the stock} \downarrow \text{by } 0.5\%.$$

In summary:

$$\textbf{if } \boldsymbol{\beta > 1} \quad \Rightarrow \quad \textbf{risky stocks}$$

as market movements are amplified in the stock's returns, and:

$$\textbf{if } \boldsymbol{\beta < 1} \quad \Rightarrow \quad \textbf{defensive stocks}$$

as market movements are muted in the stock's returns.