STAT 331 - Applied Linear Models

Fantastic Models and How to Abuse Them

Jose Luis Avilez

Faculty of Mathematics

University of Waterloo

Chapter 1

Introduction

Definition 1.1. We define a **statistical model** as an equation

$$y = \mu + \epsilon$$

where μ is a **deterministic** component and ϵ is a **stochastic** component (or noise).

Definition 1.2. A **response** variable is denoted Y and its values are $(y_1, \ldots, y_n)$; an **independent** variable is denoted X and its values are $(x_1, \ldots, x_n)$; the **regression slope** is denoted β; the **noise** term is denoted ϵ; the regression equation is then given by

$$Y = \beta X + \epsilon$$

Definition 1.3. To emphasise that the model applies to each potential experiment, we index using our dataset (i.e. $\{(x_i, y_i)\}_{i=1,\ldots,n}$ are data points) to say

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

Definition 1.4. We say that the noise is exhibits **homoscedasticity** if each ϵ_i has equal variance. **Heteroscedasticity** means they have unequal variances.

Definition 1.5. In a **simple linear model** there is only one explanatory variable and we make the following assumptions for the error term ϵ:

1. ϵ_i is normally distributed for each i
2. $\mathrm{E}(\epsilon_i) = 0$, for $i = 1, 2, \ldots, n$
3. $\mathrm{Var}(\epsilon_i) = \sigma^2$
4. ϵ_i and ϵ_j are independent random variables for $i \neq j$

Theorem 1.6. *In a simple linear model, if we take x_i to be deterministic and each y_i as a random variable, $\mathrm{E}(y_i) = \beta_0 + \beta_1 x_i$.*

Proof. Trivial.

Definition 1.7. We define a **general linear model**[1] as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_p x_p$$

Note that it has multiple independent variables. A more efficient way to write this is in matrix form

$$\vec{y} = X\vec{\beta} + \vec{\epsilon}$$

Except, no sane person puts those funny hats on top of their vectors, so we shall simply write $y = X\beta + \epsilon$ where X is the design matrix. Note it has a column of 1s to multiply out the constant β_0 term.

Definition 1.8. We say that a model is **"parsimonious"** if it is "economic" and has "low complexity". We use inverted commas since these are not well-defined mathematical constructs.

[1] Not to be confused with **generalised**.

Chapter 2

Simple Linear Regression

For this chapter, we explore the consequences of Definition 1.5 and how to test their assumptions.

To obtain estimates of the parameters in a simple linear model we have two available methods: (i) **maximum likelihood estimation**, and (ii) **least squares estimate**. The former requires distributional assumptions; the latter does not.

Theorem 2.1. *For a simple linear model, the maximum likelihood estimators are given by* $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ *and* $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$

Proof. Given that the y_i are independent, we have that the likelihood function is

$$L(\beta_0, \beta_1, \sigma^2) = f(y_1, \ldots, y_n|\beta_0, \beta_1, \sigma^2) = \prod_i^n f(y_i|\beta_0, \beta_1, \sigma^2)$$

Under the normality assumption for y_i, we then have

$$f(y_i|\beta_0, \beta_1, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_1)^2\right)$$

Thus, the log-likelihood function is given by

$$l(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_1)^2$$

The remainder of the result follows from maximising the log-likelihood for the parameters. We show the computation in an upcoming Theorem. ∎

Definition 2.2. We say that $\hat{\beta}_0$ and $\hat{\beta}_1$ are least squares estimates if they minimise the equation

$$S(\beta_0, \beta_1) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$$

Theorem 2.3. *The least squares estimates are equal to the maximum likelihood estimates*[1].

Proof. Taking partial derivatives with respect to the parameters, we obtain,

$$\frac{\partial S}{\partial \beta_0} = -2\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)$$
$$\frac{\partial S}{\partial \beta_1} = -2\sum_{i=1}^n x_i(y_i - \beta_0 - \beta_1 x_i)$$

[1] Proofs for this theorem can be seen in Lectures 1 and 4 of Shalizi's notes

To maximise the parameters, we set the partial derivatives to zero. It is easy to see that the first expression is minimised when $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}$. Minimising the second expression requires a bit more algebraic mumbo-jumbo.

$$\begin{aligned}
0 &= \sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) \\
&= \sum_{i=1}^{n}(x_i y_i) - \beta_0 \sum_{i=1}^{n} x_i - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}(\bar{y} - \beta_1\bar{x}) - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&= \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} - n\beta_1\bar{x}^2 - \beta_1 \sum_{i=1}^{n} x_i^2 \\
&\iff \\
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n}(x_i y_i) - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 + n\bar{x}^2} \\
&= \frac{S_{xy}}{S_{xx}}
\end{aligned}$$

Ta-da! ■

Definition 2.4. The following two equations are called **normal equations**:

$$n\hat{\beta}_0 + \left(\sum x_i\right)\hat{\beta}_1 = \sum y_i \tag{2.1}$$

$$\left(\sum x_i\right)\hat{\beta}_0 + \left(\sum x_i^2\right)\hat{\beta}_1 = \sum x_i y_i \tag{2.2}$$

Definition 2.5. The **residual**, e_i, of the fitted value at x_i is $e_i = y_i - \hat{\mu}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$.

Theorem 2.6. *In a regression line fitted by the least squares estimate procedure, the following are facts about residuals:*

1. $\sum e_i = 0$
2. $\sum e_i x_i = 0$
3. $\sum \hat{\mu}_i e_i = 0$

Proof. Follows from the minimisation procedure used in Theorem 2.3. ■

Theorem 2.7. *The maximum likelihood estimate of σ^2 is $\hat{\sigma^2} = \frac{S(\hat{\beta}_0, \hat{\beta}_1)}{n}$.*

Proof. Exercise. ■

Theorem 2.8. *The estimated value of σ^2 using the least squares estimate method is*

$$S^2 = \frac{S(\hat{\beta}_0, \hat{\beta}_1)}{n-2}$$

*We call this the least square error and it has $n-2$ degrees of freedom. In R, the summary output for a linear model is the **residual standard error**, which is simply $S = \sqrt{S^2}$.*

Proof. Exercise.

Theorem 2.9. *The mean squared error, S^2 is an unbiased estimate for σ^2. That is, $\mathrm{E}\left(S^2\right) = \sigma^2$.*

Proof. Exercise. ■

Theorem 2.10. *The estimators $\hat{\beta}_0, \hat{\beta}_1$ is unbiased; that is* $\mathrm{E}\left[\hat{\beta_{0,1}}\right] = \beta_{0,1}$

Proof. We can write

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n c_i y_i$$

where $c_i = \frac{x_i - \bar{x}}{S_{xx}}$. Thus,

$$\mathrm{E}\left[\hat{\beta}_1\right] = \mathrm{E}\left[\sum c_i y_i\right] = \sum c_i \mathrm{E}\left[y_i\right] = \sum c_i \mathrm{E}\left[\beta_0 + \beta_1 x_i\right] = E[\beta_0] \sum c_i + \beta_1 \sum c_i \mathrm{E}\left[x_i\right] = \beta_1 \frac{S_{xx}}{S_{xx}} = \beta_1$$

Likewise,

$$\mathrm{E}\left[\hat{\beta}_0\right] = \mathrm{E}\left[y_i - \hat{\beta}_1 x_i\right] = \bar{y} - \beta_1 \bar{x} = \beta_0$$

■

Theorem 2.11. *The estimator $\hat{\mu}$ is an unbiased estimate for μ and S^2 is an unbiased estimator for σ^2.*

Proof. The first follows easily from Theorem 2.10. The second estimator requires finding a pivotal quantity which follows a chi-squared distribution with $n - 2$ degrees of freedom. I'll provide details later. ■

Theorem 2.12. *The following are the variances for the estimators:*

1. $\mathrm{Var}\left(\hat{\beta}_1\right) = \frac{\sigma^2}{S_{xx}}$
2. $\mathrm{Var}\left(\hat{\beta}_0\right) = \sigma^2\left[\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right]$
3. $\mathrm{Var}\left(\hat{\mu_0}\right) = \sigma^2\left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]$

Proof. The first two follow by our usual variance formulas. The third point requires writing

$$\mathrm{Var}\left(\hat{\mu_0}\right) = \mathrm{Var}\left(\bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}\right) = \mathrm{Var}\left(\bar{y} + \hat{\beta}_1 (x_0 - \bar{x})\right)$$

and using the independence[2] of $\bar{y}$ and $\hat{\beta}_1$.

[2] The professor claimed this. I am not entirely convinced... I'll check this at a later date.