Level M Regression Models

Lecture 12

Colette Mair

School of Mathematics and Statistics

University of Glasgow

Inference

Introduction

- Now we will focus on the construction interval estimates for various parameters of our models

Confidence Interval

A 100c% confidence interval estimate for $\mathbf{b}^T\beta$ with confidence c is

$$\mathbf{b}^T\hat{\beta} \pm t\left(n-p; \frac{1+c}{2}\right)\sqrt{\frac{RSS}{n-p}(\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b})}.$$

Prediction Interval

A 100c% prediction interval for a future observation y_f is

$$\mathbf{b}^T\hat{\beta} \pm t\left(n-p; \frac{1+c}{2}\right)\sqrt{\frac{RSS}{n-p}(1+\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b})}.$$

Confidence c

- In this course, we will assume $c = 0.95$ unless otherwise specified.

95% Confidence Interval

A 95% confidence interval estimate for $\mathbf{b}^T\beta$ is

$$\mathbf{b}^T\hat{\beta} \pm t\,(n-p; 0.975)\,\sqrt{\frac{RSS}{n-p}(\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b})}.$$

95% Prediction Interval

A 95% prediction interval for a future observation y_f is

$$\mathbf{b}^T \hat{\beta} \pm t\,(n-p; 0.975)\,\sqrt{\frac{RSS}{n-p}(1 + \mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b})}.$$

Illusration

$$y_i = \alpha + \beta x_i \quad i = 1, \dots, n.$$

A 95% confidence interval estimate for β is

- $\beta^T = (\alpha \quad \beta)$
- $\mathbf{b}^T = (0 \quad 1)$

$\mathbf{b}^T \beta$ is

$$\mathbf{b}^T \hat{\beta} \pm t\,(n-p; 0.975)\,\sqrt{\frac{RSS}{n-p}(\mathbf{b}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{b})}.$$

Example

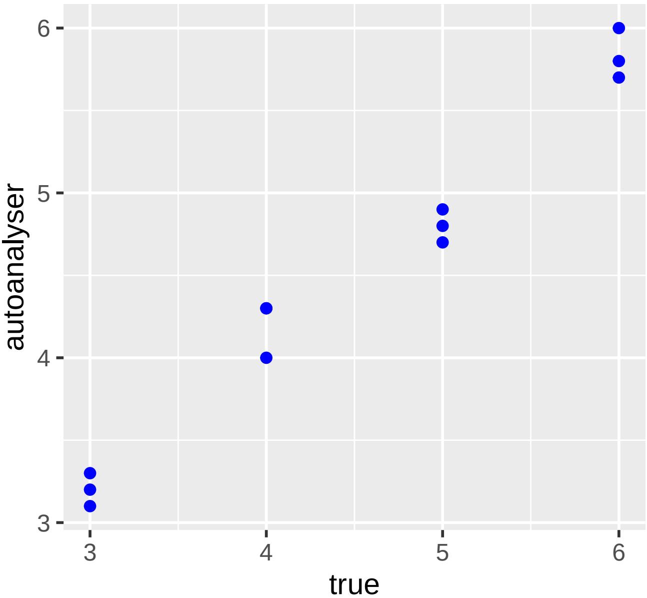
Autoanalyser

- Blood plasma concentrations are usually measured using a lengthy laboratory process.
- The autoanalyser is regularly tested to see if it is performing properly.
- A method using an autoanalyser is often used
- On this occasion, 12 measurements have been made on samples of known concentration (3 replicates at each of 4 concentrations).

Autoanalyser

- Construct a 95% C.I. for the population mean autoanalyser concentration when the true concentration is 6 units and interpret the interval.
- Construct a 95% prediction interval for the autoanalyser concentration when the true concentration is 6 units and interpret the interval.

Autoanalyser



Autoanalyser

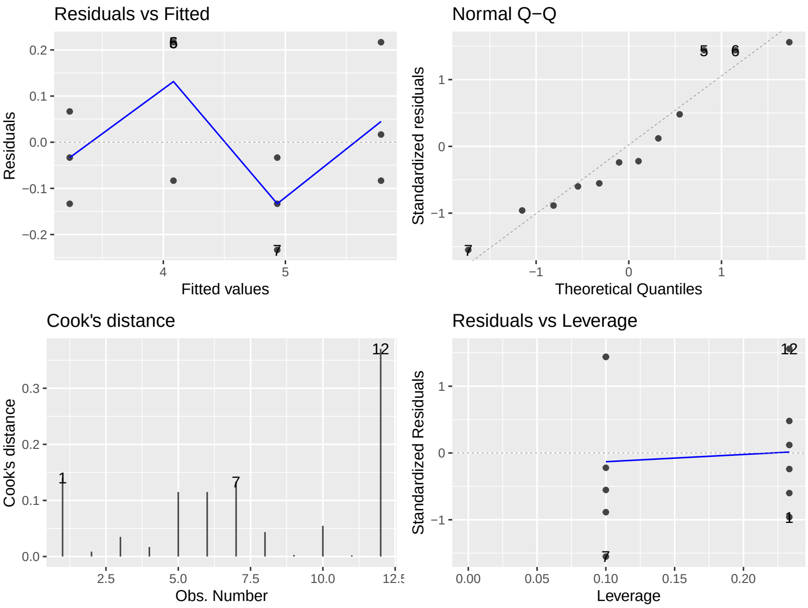
$$\text{autoanalyser}_i = \beta_0 + \beta_1 \text{true}_i + \epsilon_i, \quad i = 1, \ldots, 12$$

Autoanalyser

Simple linear model `R` output

term	estimate	std.error	statistic	p.value
(Intercept)	0.68	0.19	3.60	0.00487
true	0.85	0.04	20.75	0.00000

Autoanalyser



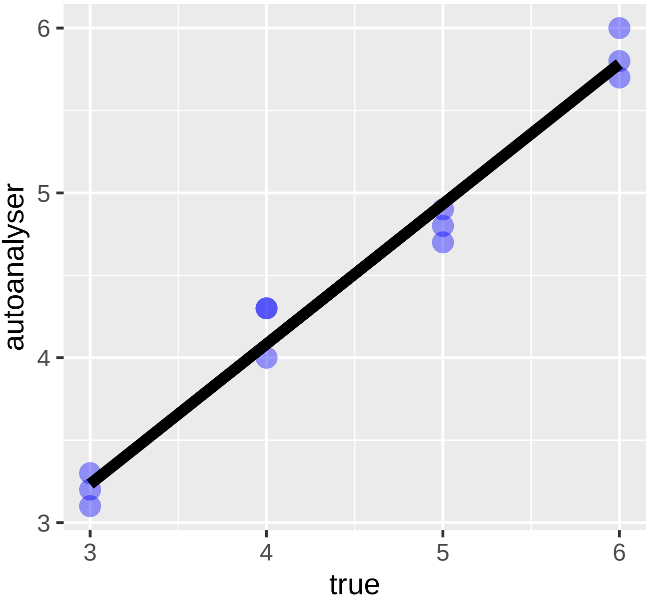
Autoanalyser

$$
\begin{aligned}
\mathbf{b}^T\hat{\beta} &= \hat{\alpha} + 6 \times \hat{\beta} \\
&= 0.683 + 0.85 \times 6 \\
&= 5.783 \\
RSS &= 0.252 \\
n &= 12
\end{aligned}
$$

t(10;0.975) $= 2.228$

$$
(\mathbf{X}^T\mathbf{X})^{-1} = \frac{1}{180}\begin{pmatrix} 258 & -54 \\ -54 & 12 \end{pmatrix}
$$

Autoanalyser



Autoanalyser

- Construct a 95% C.I. for the population mean autoanalyser concentration when the true concentration is 6 units and interpret the interval.
- Construct a 95% prediction interval for the autoanalyser concentration when the true concentration is 6 units and interpret the interval.

95% C.I.

95% prediction interval

Trees

- The dataset refers to the volume (cubic feet) and diameter (inches) (at 54 inches above the ground) and height (feet) for a sample of 31 black cherry trees in the Allegheny National Forest Pennsylvania.
- The data were collected in order to and an estimate for the volume of a tree (and therefore for the timber yield), given its height and diameter. A

Trees

Our full model for the trees data is

$$E(Y) = \alpha + \beta x_1 + \gamma x_2$$

where Y denotes log(volume), x_1 denotes log(diameter) and x_2 denotes log(height) of 31 trees.

Trees

The fitted model produced:

$$\hat{\beta} = \begin{pmatrix} -6.632 \\ 1.983 \\ 1.117 \end{pmatrix}$$

$$RSS = 0.1855$$

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{pmatrix} 96.5721 & 3.1393 & -24.1651 \\ 3.1393 & 0.8495 & -1.2275 \\ -24.1651 & -1.2275 & 6.3099 \end{pmatrix}$$

Trees

Construct 95% confidence intervals for β and γ to test their significance.

95% C.I. for β

95% C.I. for γ