

# MAST30025: Linear Statistical Models

## Week 6 Lab

Questions 1–7 use the ‘sleep’ dataset, which you can download from the course website. This dataset contains (among other things) data on the body weight (kg) and brain weight (g) of 62 mammals. Use the following commands to read the data:

```
> mammals <- read.csv("../data/sleep.csv")
> mammals$BodyWt <- log(mammals$BodyWt)
> mammals$BrainWt <- log(mammals$BrainWt)
```

This creates a data frame, **mammals**, with components (among others) named **BodyWt** and **BrainWt**, then applies a logarithmic transformation to both **BodyWt** and **BrainWt**.

1. Fit a linear model explaining brain weight from body weight, using the **lm** command.  
Display the summary of the fitted model, and then create a scatter plot of the data and superimpose the fitted regression line on it. Does it look like a reasonable fit?  
Use diagnostic plots to assess if the model assumptions are satisfied.
2. Using the fitted model or otherwise, obtain:
  - (a) The least squares estimator of the parameters, **b**;
  - (b) The vector of residuals, **e**;
  - (c) The residual sum of squares,  $SS_{Res}$ ;
  - (d) The regression sum of squares,  $SS_{Reg}$ ;
  - (e) The estimator for the variance of the errors,  $s^2$ ;
  - (f) The standardised residuals;
  - (g) The leverages of the points;
  - (h) The Cook’s distances of the points;
  - (i) 95% confidence intervals for each of the parameters.
3. Find a 95% confidence interval for a mammal weighing 50 kg.
4. Find a 95% prediction interval for a mammal weighing 50 kg.
5. Find and draw a 95% joint confidence region for the parameters.
6. Test the following hypotheses, using the **anova** function.
  - (a)  $H_0 : \beta = 0$
  - (b)  $H_0 : \beta_1 = 0$
  - (c)  $H_0 : \beta_0 = 0$
  - (d)  $H_0 : \beta = (2, 1)$
7. By visualising the raw data, justify the use of a double logarithmic transformation. Write down the final model for the (untransformed) brain weight vs. body weight.
8. In this question we consider the hypothesis  $H_0 : \beta = \beta^*$ . The test statistic for this hypothesis is

$$\frac{(\mathbf{b} - \beta^*)^T X^T X (\mathbf{b} - \beta^*) / p}{SS_{Res} / (n - p)}.$$

(a) Show that

$$(\mathbf{b} - \boldsymbol{\beta}^*)^T X^T X (\mathbf{b} - \boldsymbol{\beta}^*) = (\mathbf{y} - X\boldsymbol{\beta}^*)^T (\mathbf{y} - X\boldsymbol{\beta}^*) - (\mathbf{y} - X\mathbf{b})^T (\mathbf{y} - X\mathbf{b}).$$

That is, it is the  $SS_{Res}$  for the null model minus the  $SS_{Res}$  for the full model.

Also show that

$$(\mathbf{b} - \boldsymbol{\beta}^*)^T X^T X (\mathbf{b} - \boldsymbol{\beta}^*) \neq \mathbf{y}^T X (X^T X)^{-1} X^T \mathbf{y} - \boldsymbol{\beta}^{*T} X^T X \boldsymbol{\beta}^*.$$

That is, in this case we can not write it as the  $SS_{Reg}$  for the full model minus the  $SS_{Reg}$  for the model under  $H_0$ .

(b) Show directly that  $(\mathbf{b} - \boldsymbol{\beta}^*)^T X^T X (\mathbf{b} - \boldsymbol{\beta}^*)$  and  $SS_{Res}$  are independent, that is without using our existing results that  $\mathbf{b}$  and  $SS_{Res}$  are independent.

Hint: set  $\mathbf{q} = \mathbf{y} - X\boldsymbol{\beta}^*$  then

i. Show that  $(\mathbf{b} - \boldsymbol{\beta}^*)^T X^T X (\mathbf{b} - \boldsymbol{\beta}^*) = \mathbf{q}^T X (X^T X)^{-1} X^T \mathbf{q}$ .

ii. Show that  $SS_{Res} = \mathbf{q}^T [I - X (X^T X)^{-1} X^T] \mathbf{q}$  and hence that these two quadratic forms are independent.