

620-371: Linear Models

Practice Class 7

21st April, 2009

In this practice class, we shall analyse a large(ish) dataset. Go to the 'Datasets' section of the 620-371 website and download the 'sleep' dataset. This dataset contains (among other things) data on the body weight (kg) and brain weight (g) of 62 mammals. Use the following function to read the data:

```
mammals <- read.table("sleep.txt", header=TRUE)
```

This creates a data frame, `mammals`, with components (among others) named `BodyWt` and `BrainWt`.

1. This data needs a logarithmic transformation. Apply `log` to both `BodyWt` and `BrainWt`. You will have to logarithmise each component individually as the entire data frame contains some non-numeric entries, which cannot be log'ed.

Solution:

```
> mammals$BodyWt <- log(mammals$BodyWt)
> mammals$BrainWt <- log(mammals$BrainWt)
```

2. We want to fit a linear model explaining brain weight from body weight. Fit a `lm` model to the data. Display the summary of it.

Solution:

```
> model <- lm(BrainWt ~ BodyWt, data = mammals)
> summary(model)
```

Call:

```
lm(formula = BrainWt ~ BodyWt, data = mammals)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-1.71550	-0.49228	-0.06162	0.43597	1.94829

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.13479	0.09604	22.23	<2e-16 ***
BodyWt	0.75169	0.02846	26.41	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

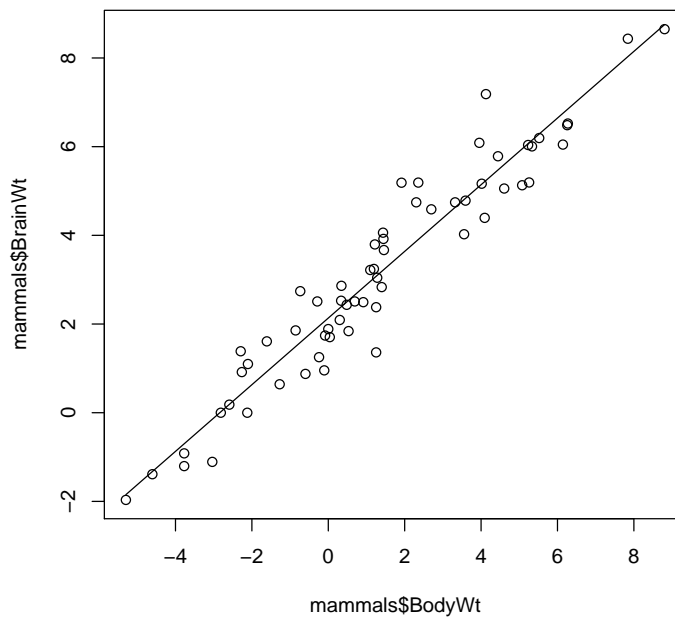
Residual standard error: 0.6943 on 60 degrees of freedom

Multiple R-squared: 0.9208, Adjusted R-squared: 0.9195
F-statistic: 697.4 on 1 and 60 DF, p-value: < 2.2e-16

3. Create a scatter plot of the data and superimpose the fitted regression line on it.

Solution:

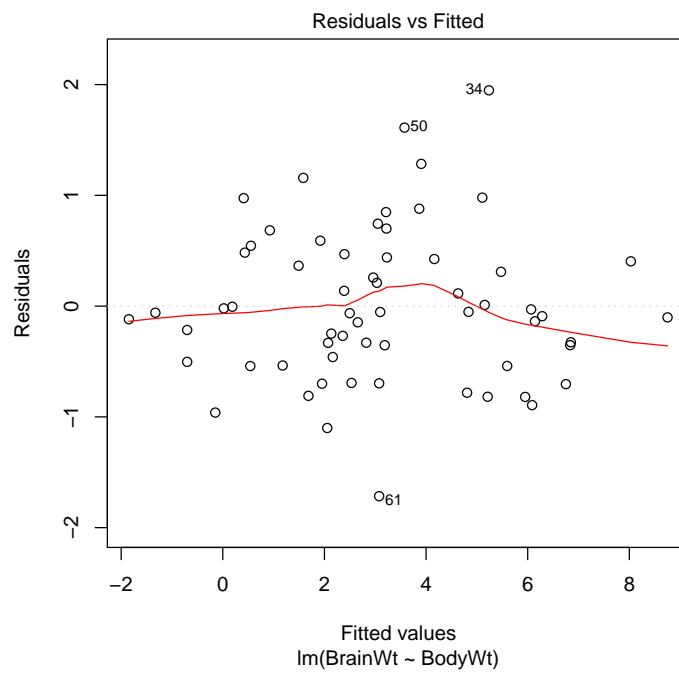
```
> plot(mammals$BodyWt, mammals$BrainWt)
> curve(model$coefficients[1] + model$coefficients[2] * x, add = TRUE,
+       color = "red")
```



4. Create diagnostic plots and look at them. Are the model assumptions satisfied?

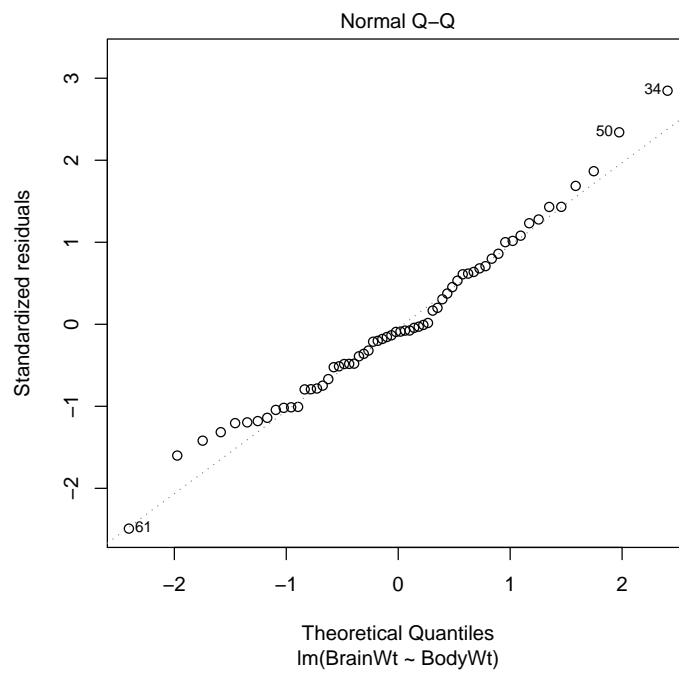
Solution:

```
> plot(model, which = 1)
```



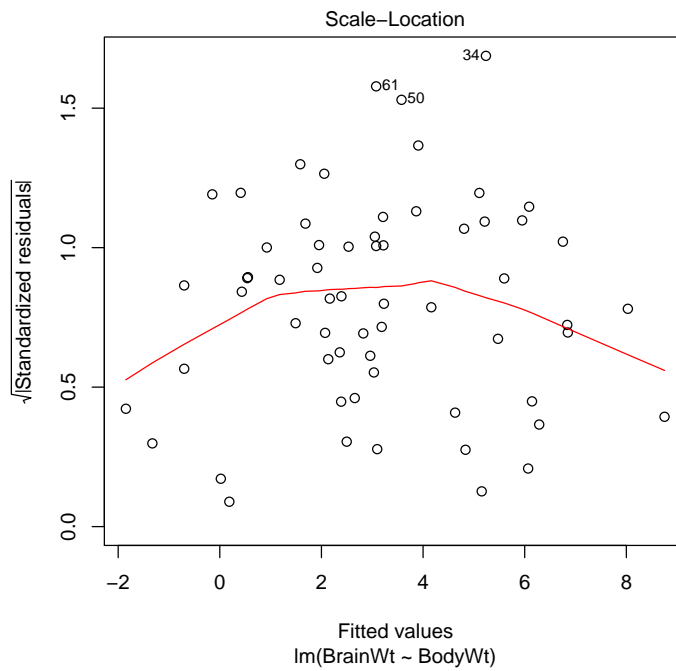
The residuals show a slight trend toward negativity as the fitted values increase. This may indicate a nonlinear trend.

```
> plot(model, which = 2)
```



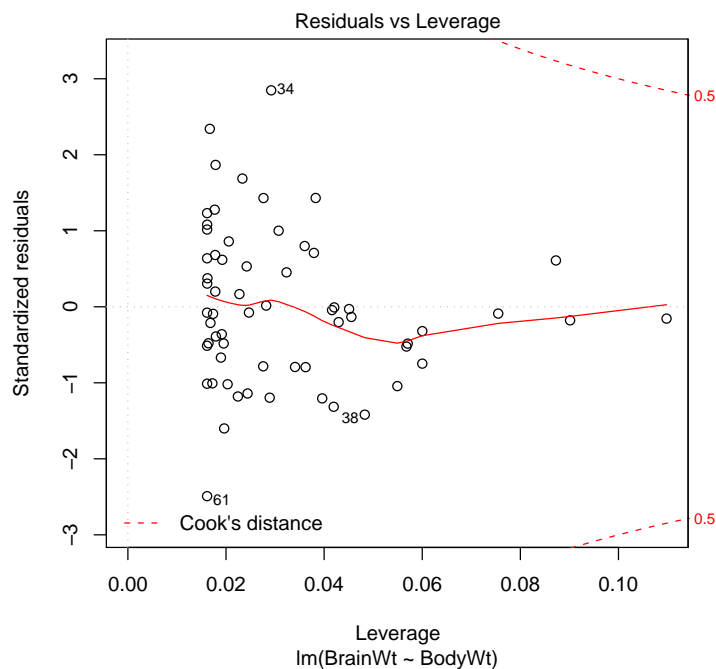
The Q-Q plot looks reasonably linear.

```
> plot(model, which = 3)
```



The standardised residuals get smaller on both sides of the plot. This is not ideal, but the lack of a definite trend makes it difficult to correct.

```
> plot(model, which = 5)
```



This plot is fine.

5. Calculate:

(a) The least squares estimator of the parameters, \mathbf{b} ;

Solution:

```
> model$coefficients
(Intercept)      BodyWt
  2.134789      0.751686
```

(b) The vector of residuals, \mathbf{e} ;

Solution:

The first few residuals are:

```
> str(model$residuals)
Named num [1:62] -0.102 -0.248  0.744 -0.332  0.404 ...
- attr(*, "names")= chr [1:62] "1" "2" "3" "4" ...
```

(c) The residual sum of squares, SS_{Res} ;

Solution:

```
> deviance(model)
[1] 28.92271
```

(d) The regression sum of squares, SS_{Reg} ; and

Solution:

```
> t(mammals$BrainWt) %*% mammals$BrainWt - deviance(model)
      [,1]
[1,] 947.5602
```

(e) The estimator for the variance of the errors, s^2 .

Solution:

```
> deviance(model)/model$df.residual
[1] 0.4820452
```

6. Find a 95% prediction interval for a particular mammal weighing 50 kg.

Solution:

```
> predict(model, data.frame(BodyWt = log(50)), interval = "prediction",
+   level = 0.95)

      fit      lwr      upr
1 5.075401 3.667797 6.483006
```

7. Test for model adequacy — $H_0 : \beta = 0$.

Solution:

```
> basemodel <- lm(BrainWt ~ 0, data = mammals)
> anova(basemodel, model)
```

Analysis of Variance Table

```
Model 1: BrainWt ~ 0
Model 2: BrainWt ~ BodyWt
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1     62 976.48
2     60  28.92  2    947.56 982.85 < 2.2e-16 ***
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We reject the null hypothesis at a very strong level.

8. Test for model adequacy with a possibly nonzero intercept, $H_0 : \beta_1 = 0$, using the `anova` function.

Solution:

```
> basemodel <- lm(BrainWt ~ 1, data = mammals)
> anova(basemodel, model)
```

Analysis of Variance Table

```
Model 1: BrainWt ~ 1
Model 2: BrainWt ~ BodyWt
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      61 365.11
2      60  28.92  1    336.19 697.42 < 2.2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

We again reject the null hypothesis.

9. Test the hypothesis $H_0 : \beta_0 = 0$, using the `anova` function.

Solution:

```
> basemodel <- lm(BrainWt ~ 0 + BodyWt, data = mammals)
> anova(basemodel, model)
```

Analysis of Variance Table

```
Model 1: BrainWt ~ 0 + BodyWt
Model 2: BrainWt ~ BodyWt
  Res.Df  RSS Df Sum of Sq    F    Pr(>F)
1      61 267.079
2      60  28.923  1    238.157 494.05 < 2.2e-16 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Again, we reject the null hypothesis.