## Exercises 2 solutions

1. (a) $\mathrm{E}(\mathbf{Z}) = \mathbf{A}\mathrm{E}(\mathbf{Y}) + \mathbf{c}$ and $\mathrm{var}(\mathbf{Z}) = \mathbf{A}\mathbf{V}(\mathbf{Y})\mathbf{A}^{\mathrm{T}}$ (Rice section 14.4.1, theorems A and B).

   (b) From equation (2.10) in Section 2.1.3 of the notes:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{Y}.$$

   So apply the results in part (a) with $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}, \mathbf{V}(\mathbf{Y}) = \mathbf{V}, \mathbf{A} = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}$ and $\mathbf{c} = \mathbf{0}$. Then, a little algebra gives

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \ \text{ and } \ \mathrm{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{X})^{-1}.$$

   *(Try this if you haven't done so already.)*

2. (a) The normal equations are $\mathbf{X}^{\mathrm{T}}\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^{\mathrm{T}}\mathbf{Y}$. Using $\sum(x_i - \overline{x}) = 0$, obtain

$$\mathbf{X} = \left[\begin{array}{ll} 1 & x_1 - \overline{x} \\ 1 & x_2 - \overline{x} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & x_N - \overline{x} \end{array}\right], \ \mathbf{X}^{\mathrm{T}}\mathbf{X} = \left[\begin{array}{ll} N & 0 \\ 0 & C_{xx} \end{array}\right], \ \mathbf{X}^{\mathrm{T}}\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^N Y_i \\ \sum_{i=1}^N (x_i - \overline{x})Y_i \end{pmatrix}, \ \hat{\boldsymbol{\beta}} = \begin{pmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \end{pmatrix}.$$

   This yields

$$N\hat{\alpha}_0 = \sum_{i=1}^N Y_i, \qquad \sum_{i=1}^N (x_i - \overline{x})^2 \hat{\alpha}_1 = \sum_{i=1}^N (x_i - \overline{x})Y_i.$$

   Hence

$$\hat{\alpha}_0 = \overline{Y} \text{ and } \hat{\alpha}_1 = \frac{\sum_{i=1}^N (x_i - \overline{x})Y_i}{\sum_{i=1}^N (x_i - \overline{x})^2}.$$

   (b) From the above,

$$\mathbf{V}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} = \sigma^2 \left[\begin{array}{ll} N^{-1} & 0 \\ 0 & C_{xx}^{-1} \end{array}\right].$$

   (Note that $\mathrm{cov}(\hat{\alpha}_0, \hat{\alpha}_1) = 0$ in this version of the simple linear regression model.)

   (c)

$$\mathrm{RSS} = \sum_{i=1}^N [Y_i - \hat{\alpha}_0 - \hat{\alpha}_1(x_i - \overline{x})]^2 = \sum_{i=1}^N [Y_i - \overline{Y} - \frac{C_{xY}}{C_{xx}}(x_i - \overline{x})]^2 = C_{YY} - 2\frac{C_{xY}^2}{C_{xx}} + \frac{C_{xY}^2}{C_{xx}} = C_{YY} - \frac{C_{xY}^2}{C_{xx}}.$$

   (Note that in these calculations, $\hat{\alpha}_1$ is expressed as $C_{xY}/C_{xx}$ because $C_{xY} = \sum(x_i - \overline{x})(Y_i - \overline{Y}) = \sum(x_i - \overline{x})Y_i$.)

   (d) The parameters in the two models are related as follows:

$$\beta_0 = \alpha_0 - \alpha_1\overline{x}, \qquad \beta_1 = \alpha_1.$$

   - Hence
$$\hat{\beta}_1 = \frac{C_{xy}}{C_{xx}} = 2.7642, \ \hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x} = 2.6662.$$
   - $\mathrm{RSS} = C_{yy} - \hat{\beta}_1 C_{xy} = 0.57858$,
     $\hat{\sigma}^2 = \mathrm{RSS}/8 = 0.072323$, Residual standard error $\hat{\sigma} = 0.2689$.

- $\text{var}(\hat{\beta}_1) = \text{var}(\hat{\alpha}_1)$ ,
  $\text{cov}(\hat{\beta}_0, \hat{\beta}_1) = \text{cov}(\hat{\alpha}_0 - \hat{\alpha}_1\overline{x}, \hat{\alpha}_1) = \text{cov}(\hat{\alpha}_0, \hat{\alpha}_1) - \overline{x}\text{var}(\hat{\alpha}_1) = -\overline{x}\text{var}(\hat{\alpha}_1)$,
  $\text{var}(\hat{\beta}_0) = \text{var}(\hat{\alpha}_0 - \hat{\alpha}_1\overline{x}) = \text{var}(\hat{\alpha}_0) - 2\overline{x}\text{cov}(\hat{\alpha}_0, \hat{\alpha}_1) + \overline{x}^2\text{var}(\hat{\alpha}_1) = \text{var}(\hat{\alpha}_0) + \overline{x}^2\text{var}(\hat{\alpha}_1)$.
  From above, $\text{var}(\hat{\alpha}_0) = \sigma^2/N, \text{var}(\hat{\alpha}_1) = \sigma^2/C_{xx}$. Then use these results with the estimate of $\sigma^2$ to verify numerically that the estimates of the above quantities are found to be as in the R output.

3. (a)
   $\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} \pm t_{N-p,\frac{1}{2}\alpha}\sqrt{\hat{\sigma}^2 v}$.
   (b)
   $\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} - e_{N+1}$ is a sum of normally distributed rv, therefore normally distributed, $\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}}$ and $e_{N+1}$ are independent, therefore the variance of their sum is the sum of their variances, so $\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} - e_{N+1} \sim \mathcal{N}(\mathbf{x}_{N+1}^T\boldsymbol{\beta}, \sigma^2 v + \sigma^2)$. Since $\frac{RSS}{\sigma^2} = \frac{(N-p)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{N-p}$ and the $t_{N-p}$-distribution is defined as the distribution of $\frac{Z}{\sqrt{S/(N-p)}}$, where $Z \sim \mathcal{N}(0,1)$ and $S \sim \chi^2_{N-p}$, $\frac{\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} - \mathbf{x}_{N+1}^T\boldsymbol{\beta} - e_{N+1}}{\sqrt{\hat{\sigma}^2(v+1)}} \sim t_{N-p}$ follows from $\frac{\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} - \mathbf{x}_{N+1}^T\boldsymbol{\beta} - e_{N+1}}{\sqrt{\sigma^2(v+1)}} \sim \mathcal{N}(0,1)$.
   (c)
   $\mathbf{x}_{N+1}^T\hat{\boldsymbol{\beta}} \pm t_{N-p,\frac{1}{2}\alpha}\sqrt{\hat{\sigma}^2(v+1)}$.

4. Here is some R-code and the summary results:

```
treevol <- read.table("treevol.dat",header=TRUE)
logHT <- log(treevol$HT)
logD16 <- log(treevol$D16)
logVOL <- log(treevol$VOL)
pairs(cbind(logVOL,logHT,logD16))
lmtreelog <- lm(logVOL ~ logHT + logD16)
summary(lmtreelog)
# > summary(lmtreelog)
#
# Call:
# lm(formula = logVOL ~ logHT + logD16)
#
# Residuals:
#       Min        1Q    Median        3Q       Max
# -0.069602 -0.040922  0.002851  0.024690  0.106917
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)  -5.6243     1.0605  -5.304 5.83e-05 ***
# logHT         1.0771     0.2532   4.255 0.000534 ***
# logD16        1.8321     0.1034  17.713 2.16e-12 ***
# ---
# Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
#
# Residual standard error: 0.04976 on 17 degrees of freedom
# Multiple R-Squared: 0.9671,   Adjusted R-squared: 0.9632
# F-statistic: 249.9 on 2 and 17 DF,  p-value: 2.485e-13


par(mfrow=c(2,2))
plot(lmtreelog)
# In some more recent R-versions the plot.lm default has
# changed and you get the same plots by
# plot(lmtreelog,which=1:4,add.smooth=FALSE)
```

The model looks generally very similar to the model fitting VOL by D16 and HT. The $R^2$ is higher for the logarithmic model than for the nonlogarithmic model, and it is even higher than for the original model that included DBH. This indicates that in terms of prediction strength it is better to use logarithms. The residual plots look still well. Point 1 is an even stronger leverage point this time.

Prediction:

```
newtree <- data.frame(HT=100,D16=10)
lognewtree <- data.frame(logHT=log(newtree$HT),logD16=log(newtree$D16))

predict(lmtreelog,lognewtree,interval=c("confidence"))
#       fit     lwr      upr
#1 3.554553 3.46666 3.642446

# Don't need to specify level=0.95 because you can find out from
# ?predict.lm that this is the default value anyway.

predict(lmtreelog,lognewtree,interval=c("prediction"))
#       fit     lwr      upr
#1 3.554553 3.41764 3.691466

# Much larger than the confidence interval because it includes
# the random variation.

exp(predict(lmtreelog,lognewtree,interval=c("prediction")))
#       fit      lwr      upr
#1 34.97218 30.49736 40.10359

# Need to take exp to have a prediction interval for VOL, not logVOL.

lmtree <- lm(VOL ~ D16+HT,data=treevol)
predict(lmtree,newtree,interval=c("prediction"))
#       fit      lwr      upr
#1 35.88012 27.23681 44.52344

# Much larger, so much less precise!
```

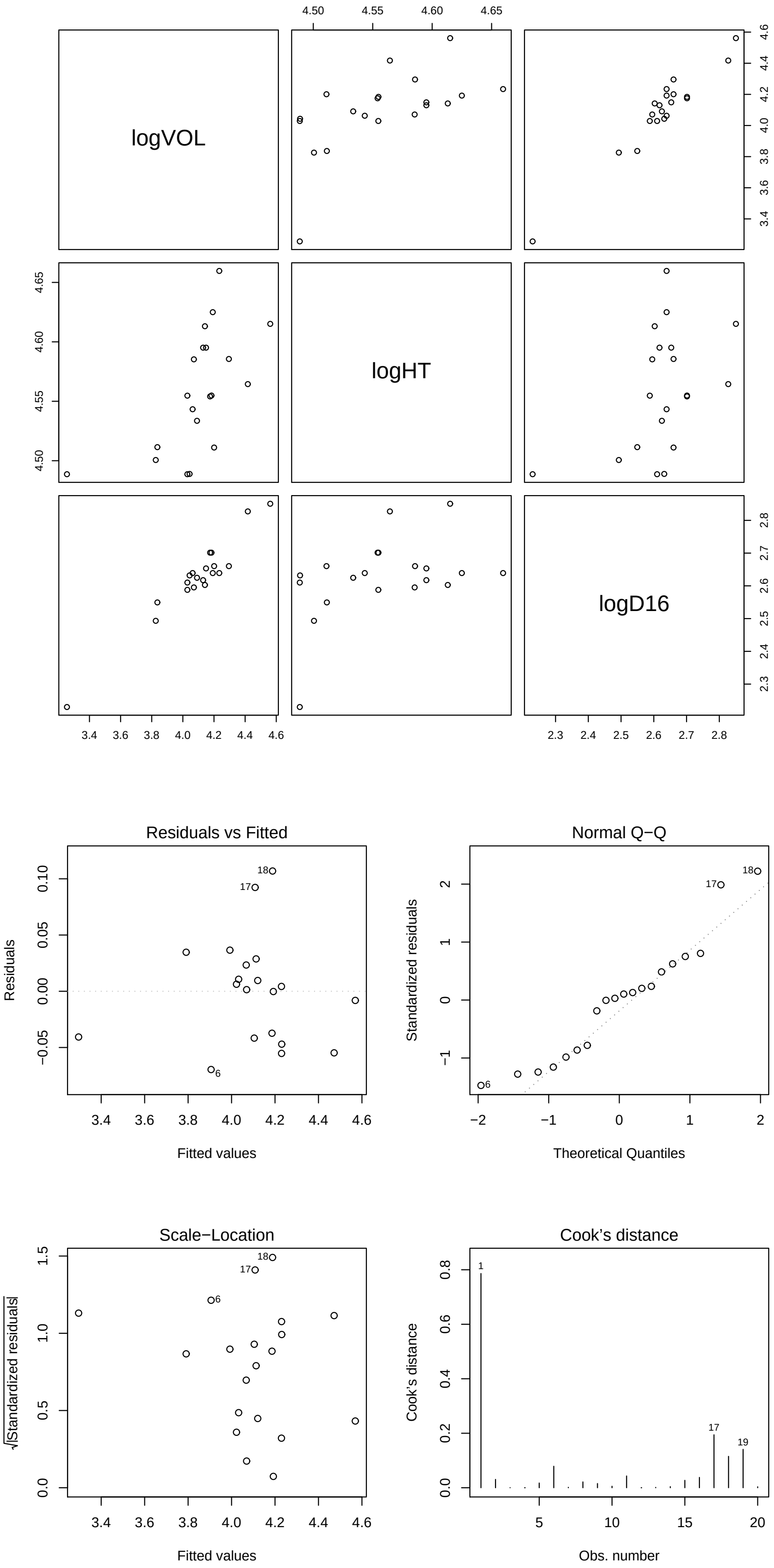

Figure 1: Matrix plot and residual plots for logarithmic treevol data.