

Generalised Regression Models

GRM: Examples — Residual Analysis

Semester 1, 2022–2023

Example 1: Forbes' Data

The file `Forbes.txt` contains values of the barometric pressure (in inches of mercury) and the boiling point of water (in °F) at 17 locations in Scotland and the Alps, as given in Numerical Example 1.4. The following R commands and output are for the regression of pressure on boiling point.

```
Forbes.dat <- read.table('Forbes.txt',header=T)
attach(Forbes.dat)
plot(Boiling_point,Pressure)      # scatterplot

fit1 <- lm(Pressure~Boiling_point)
summary(fit1)
anova(fit1)
par(mfrow=c(2,2))
plot(fit1)
```

Call:

```
lm(formula = Pressure ~ Boiling_point)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-0.23856	-0.13020	-0.08163	0.11475	0.64652

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-81.018815	2.021023	-40.09	<2e-16 ***
Boiling_point	0.522690	0.009954	52.51	<2e-16 ***

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

Residual standard error: 0.2293 on 15 degrees of freedom

Multiple R-squared: 0.9946, Adjusted R-squared: 0.9942

F-statistic: 2757 on 1 and 15 DF, p-value: < 2.2e-16

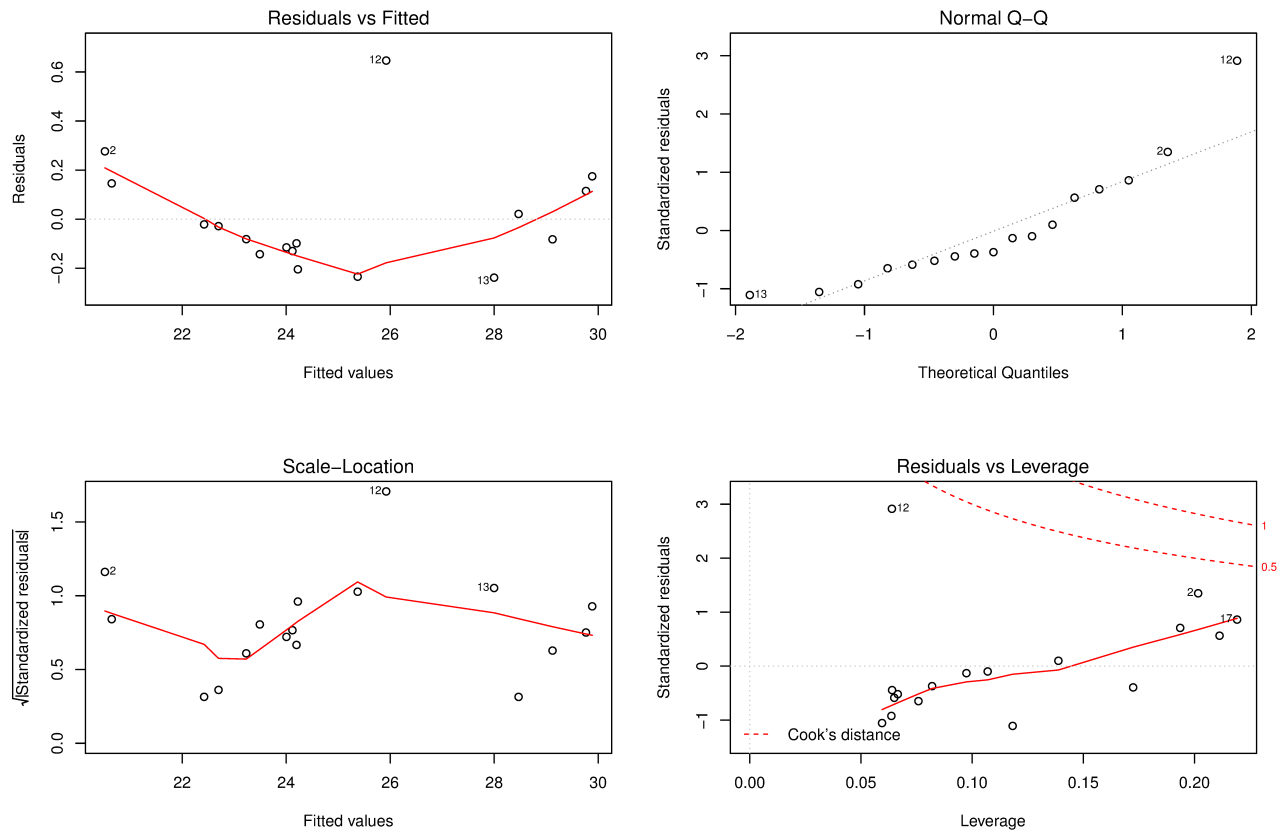
Analysis of Variance Table

Response: Pressure

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Boiling_point	1	144.968	144.968	2757.3	< 2.2e-16 ***
Residuals	15	0.789	0.053		

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

The command `plot(fit1)` produces the following plots of residuals.



The corresponding commands, output and plots are given below for the regression of the logarithm (to base 10) of pressure on boiling point.

```
fit2 <- lm(log10(Pressure)~Boiling_point)
summary(fit2)
anova(fit2)
plot(fit2)
```

```
Call:
lm(formula = log10(Pressure) ~ Boiling_point)
```

```
Residuals:
      Min       1Q   Median       3Q      Max
-0.0029788 -0.0011862 -0.0006242  0.0000616  0.0135358
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.4208817  0.0331809  -12.68 2.02e-09 ***
Boiling_point  0.0089528  0.0001634   54.78 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

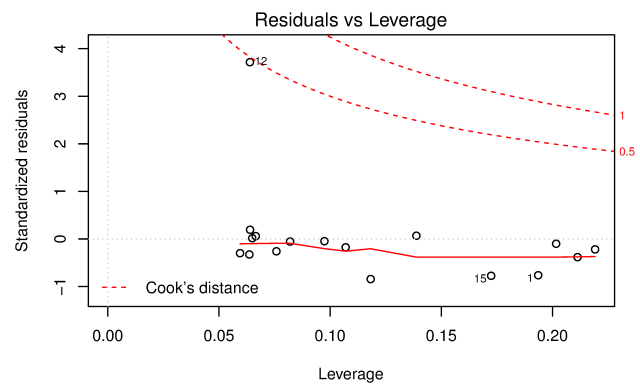
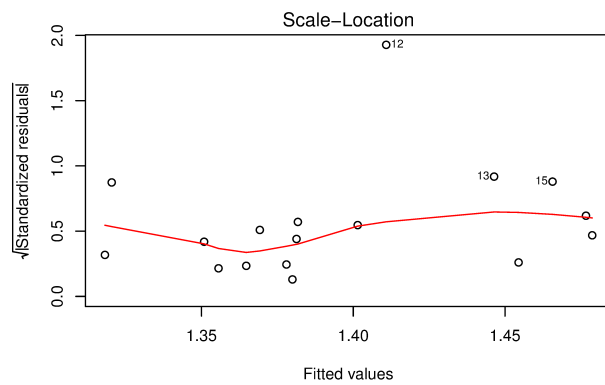
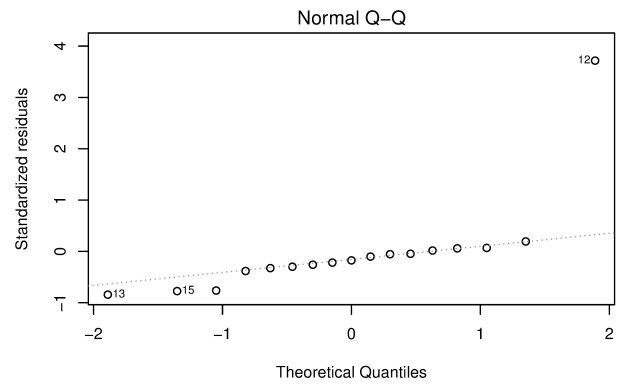
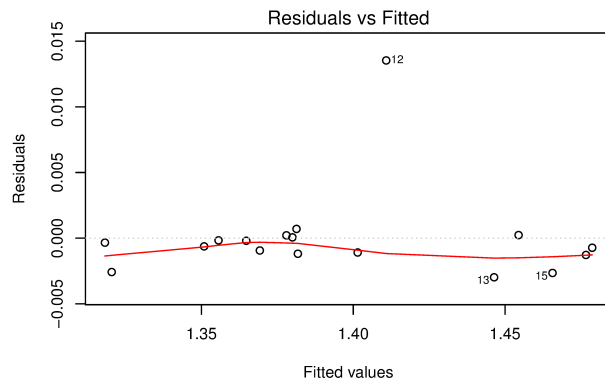
```
Residual standard error: 0.003765 on 15 degrees of freedom
Multiple R-squared:  0.995, Adjusted R-squared:  0.9947
F-statistic: 3001 on 1 and 15 DF, p-value: < 2.2e-16
```

Analysis of Variance Table

Response: log10(Pressure)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Boiling_point	1	0.042530	0.042530	3001	< 2.2e-16 ***
Residuals	15	0.000213	0.000014		

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

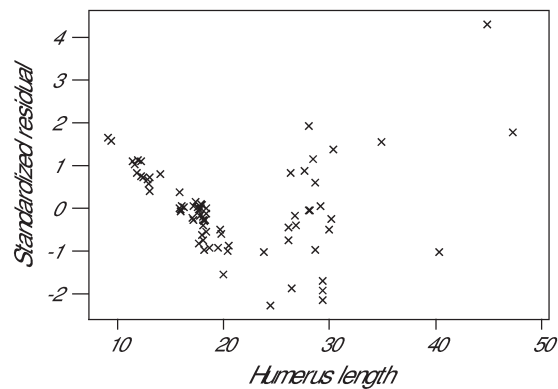


Example 2: Bird Weight Data

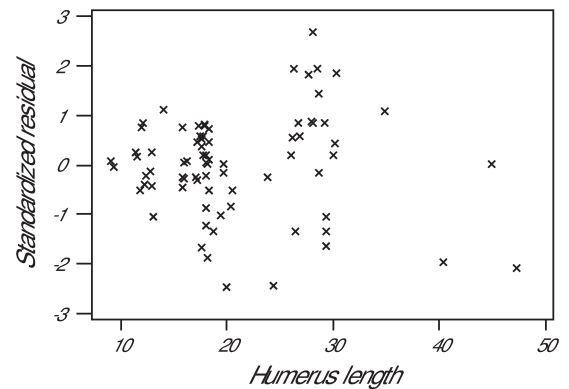
Numerical Example 1.5 concerns the body weights of 78 birds (in g) and the lengths of their humerus bones (in mm); these are held in BirdWt.txt. The following plots are of the standardized residuals against the explanatory variables from the regressions of

- (i) body weight on humerus length,
- (ii) the cube root of body weight on humerus length, and
- (iii) log body weight on log humerus length.¹

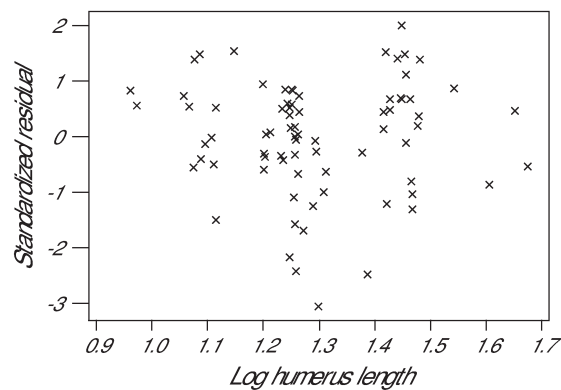
Plot of standardized residuals for body weight against humerus length



Plot of standardized residuals for cube root of body weight against humerus length



Plot of standardized residuals for log body weight against log humerus length



¹Here we use logs to base 10 in the analysis, but note that the pattern of points would be the same if using natural logarithms.