

ENV 710: Lab 6

Jiahuan Li

Spring 2023

Contents

Problem 1	2
Problem 2	5
Problem 3	8

Problem 1

Hypothesis

The null hypothesis (H_0) of this test is that there is no linear relationship between mean tree diameter and biomass in the plots.

While the alternative hypothesis (H_1) is that there is a linear relationship between the two variables mentioned above.

Regression model

```
library(here)
tdat <- read.csv(file = here("labs/lab6 linear models/TreePlots_Lab6.csv"), header = TRUE)

lm1 <- lm(AGBH.Mg.ha ~ mDBH.cm, data = tdat)
summary(lm1)
```

```
##
## Call:
## lm(formula = AGBH.Mg.ha ~ mDBH.cm, data = tdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -312.86  -26.12   11.04   53.44  253.23
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -406.186     87.130  -4.662 1.46e-05 ***
## mDBH.cm       29.114       3.713   7.840 3.60e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 91.89 on 70 degrees of freedom
## Multiple R-squared:  0.4675, Adjusted R-squared:  0.4599
## F-statistic: 61.47 on 1 and 70 DF,  p-value: 3.6e-11
```

Result analysis

The ANOVA printout reveals that the F-test for the predictor variable, mean tree diameter, yielded a small p-value ($3.6e-11$), indicating that the regression model accounts for a substantial portion of the variation in the data compared to the null model with only the intercept.

The statistical analysis further confirms the significance of the linear regression model ($R^2 = 0.4599$, $F_{1,70} = 61.47$, $p < 0.001$). The R-squared value of 0.46 demonstrates that the mean tree diameter explains 46% of the variation in plot biomass.

Model Interpretation

The regression results imply that the average tree diameter in forest plots significantly increases plot-level biomass, with a 1 cm increase in diameter resulting in a 29.114 Mg ha⁻¹ increase in plot biomass ($R^2 = 0.4599$, $F_{1,70} = 61.47$, $p < 0.001$).

The equation of the linear regression model is:

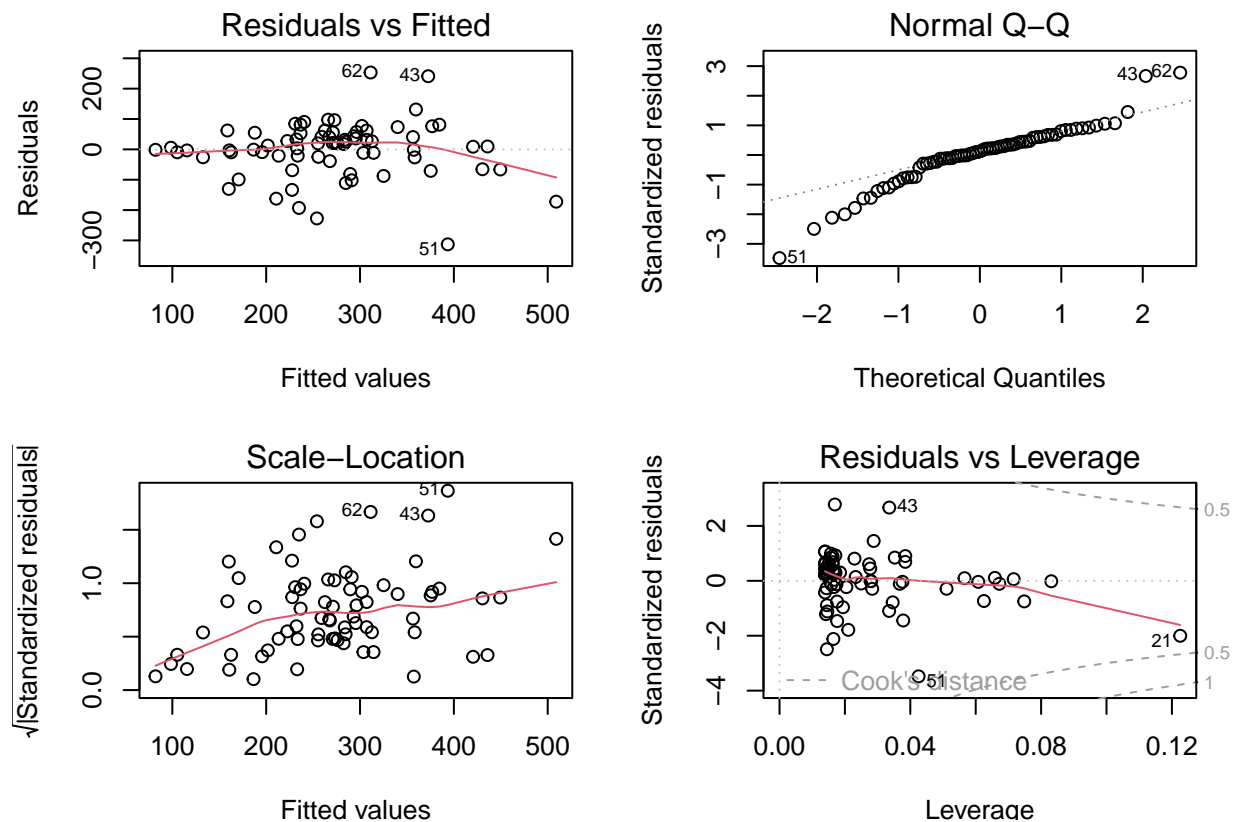
$$\text{biomass} = -406.186 + 29.114 * \text{diameter}$$

The intercept in this model is not interpretable since it pertains to a scenario outside the feasible range of the data.

Model assumptions checking

The assumptions of the linear regression model are checked by examining the residual plots for normality, constant variance, and linearity.

```
par(mfrow=c(2,2), mar = c(3.8, 4, 3, 2))
plot(lm1)
```



In the diagnostic plots for our model,

- Residuals vs. Fitted: looks good as points are randomly scattered around the center line;
- Normal Q-Q: the data generally fits the straight line but with outliers (points 43, 51, 62);
- Scale-Location: points 43, 51 and 62 scattered away from the center suggesting their excessive leverage;
- Residuals vs. Leverage: the potential trouble points 21, 43, and 51 are labelled considering their bigger effects on the parameter estimate compared to other points.

According to the plots, points 43, 51, and 62 could be odd observations. Thus, the analysis could be redone without those points (tree plots) to assess if the inference changes.

```
lm1.c <- with(tdat[-c(43,51,62), ], lm(AGBH.Mg.ha ~ mDBH.cm))
summary(lm1.c)$r.squared
```

```
## [1] 0.5731488
```

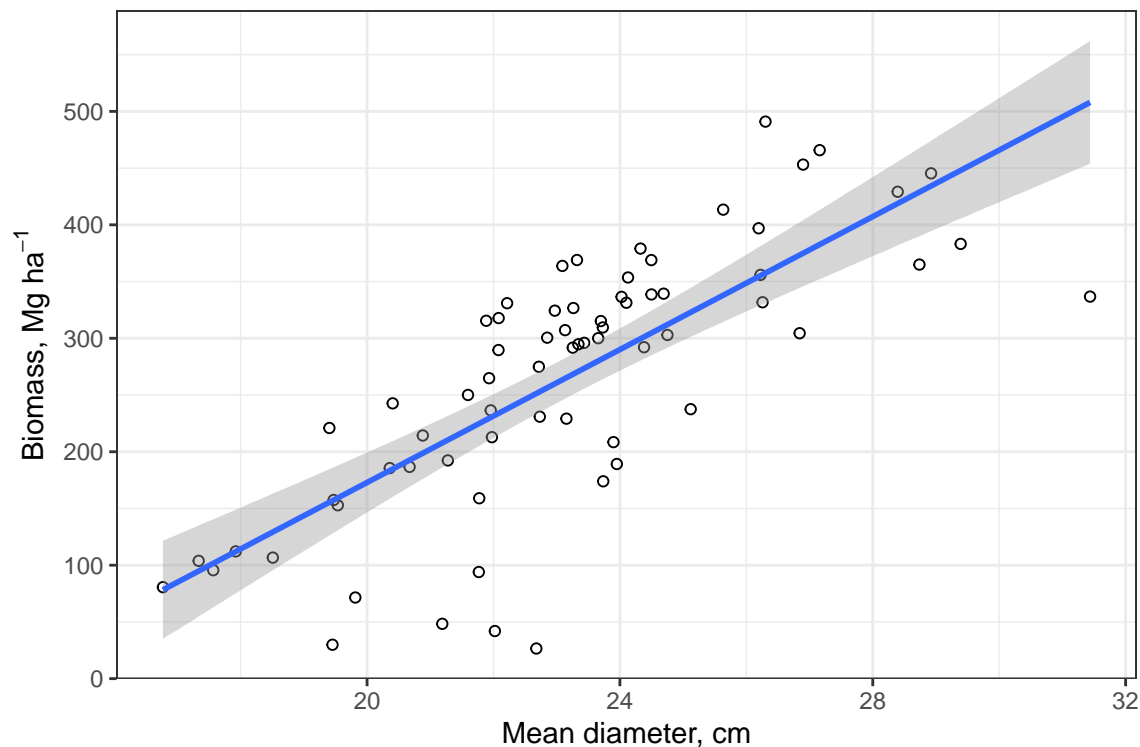
```
coef(lm1.c)
```

```
## (Intercept)      mDBH.cm
## -413.01755    29.29519
```

The adjustment may be appropriate because removing those observations enhances the R^2 value and does not significantly alter the parameter estimates.

Visualization

```
require(ggplot2)
tdat1 <- tdat[-c(43,51,62), ]
ggplot(tdat1, aes(x = mDBH.cm, y = AGBH.Mg.ha)) +
  geom_point(shape = 1) +
  geom_smooth(method = lm) +
  xlab("Mean diameter, cm") +
  ylab(expression(paste("Biomass, Mg ", ha^-1))) +
  theme_bw()
```



Problem 2

Hypothesis

The null hypothesis (H_0) of this test is that there is no linear relationship between mean tree height and mean wood density in the plots.

While the alternative hypothesis (H_1) is that there is a linear relationship between the two variables mentioned above.

Regression model

```
lm2 <- lm(mH.m ~ mWD.g.m3, data = tdat)
summary(lm2)

##
## Call:
## lm(formula = mH.m ~ mWD.g.m3, data = tdat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2022 -0.8029 -0.0336  0.8603  4.5615
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.412      1.548   8.017 1.7e-11 ***
## mWD.g.m3       7.330      2.555   2.869 0.00544 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.428 on 70 degrees of freedom
## Multiple R-squared:  0.1052, Adjusted R-squared:  0.09244
## F-statistic: 8.232 on 1 and 70 DF,  p-value: 0.005438
```

Result analysis

The ANOVA printout reveals that the F-test for the predictor variable, mean wood density, yielded a small p-value (0.005438), indicating that the regression model accounts for a substantial portion of the variation in the data compared to the null model with only the intercept.

The statistical analysis further confirms the significance of the linear regression model ($R^2 = 0.09244$, $F_{1,70} = 8.232$, $p < 0.01$). The R-squared value of 0.092 demonstrates that the mean wood density explains 9.2% of the variation in the tree height.

Model Interpretation

The regression results imply that the average wood density of forest plots significantly increases tree height, with per unit increase resulting in a 7.33 m increase in tree height ($R^2 = 0.09244$, $F_{1,70} = 8.232$, $p < 0.01$).

The equation of the linear regression model is:

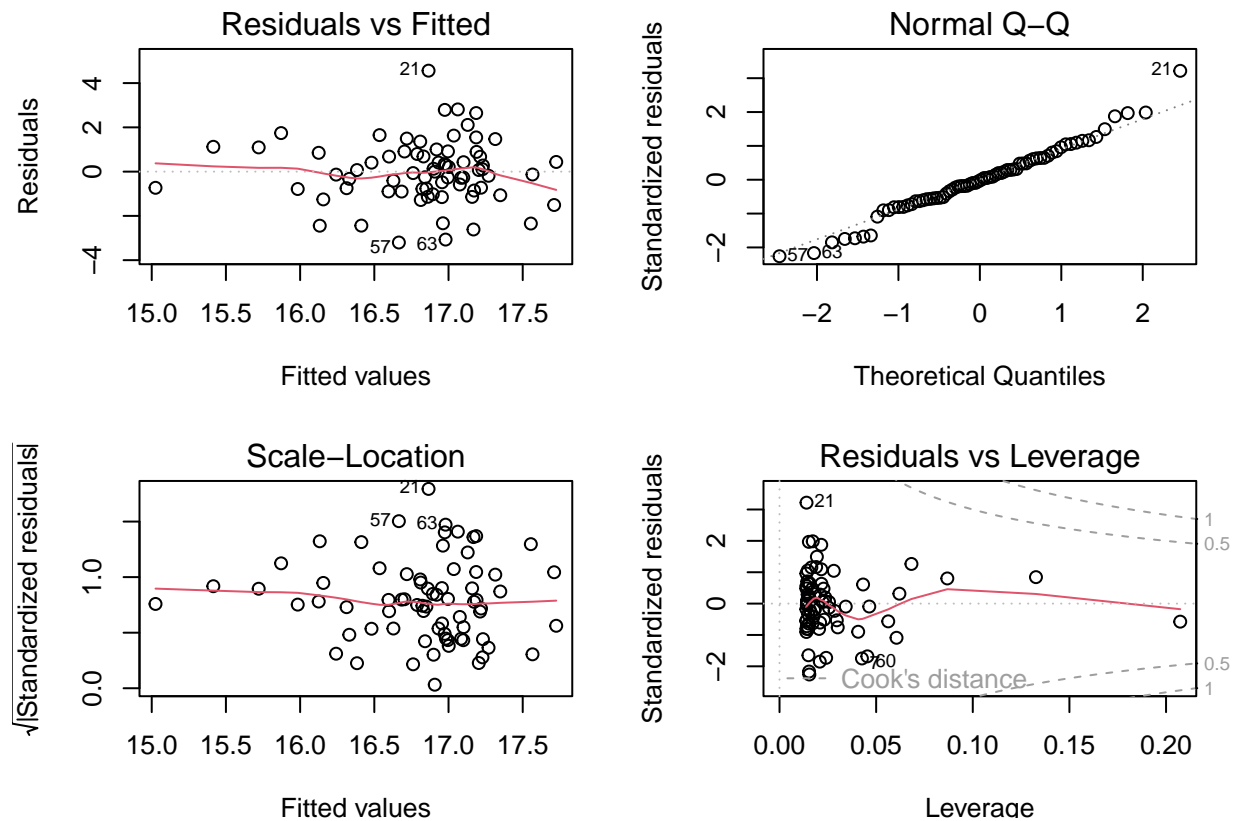
$$height = 12.412 + 7.33 * density$$

The intercept in this model means the height of the trees (12.412m) where the wood density of the plots is 0.

Model assumptions checking

The assumptions of the linear regression model are checked by examining the residual plots for normality, constant variance, and linearity.

```
par(mfrow=c(2,2), mar = c(3.8, 4, 3, 2))
plot(lm2)
```



Similar to the analysis in Problem 1, these plots indicate that points 21, 57, and 63 could be odd observations in this analysis. Thus, the analysis could be redone without those points (tree plots) to assess if the inference changes.

```
lm2.c <- with(tdat[-c(21,57,63), ], lm(mH.m ~ mWD.g.m3))
summary(lm2.c)$r.squared
```

```
## [1] 0.1375224
```

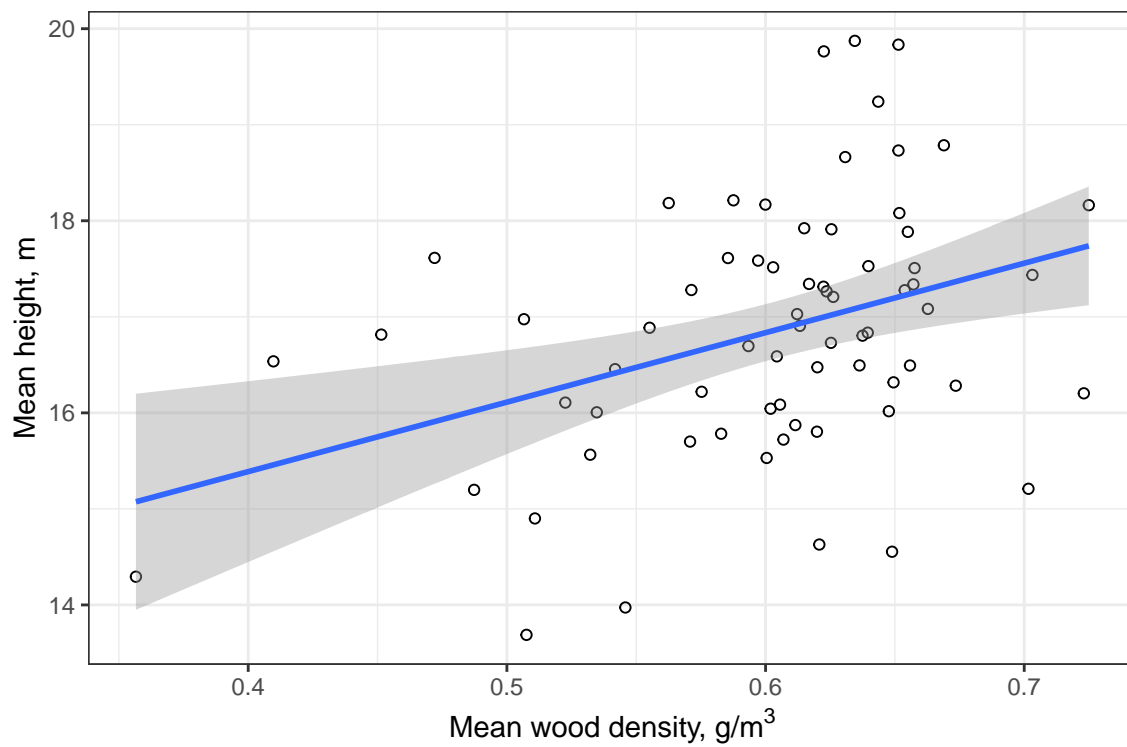
```
coef(lm2.c)
```

```
## (Intercept)    mWD.g.m3
##   12.494815    7.233291
```

The adjustment may be appropriate because removing those observations enhances the R^2 value and does not significantly alter the parameter estimates.

Visualization

```
require(ggplot2)
tdat2 <- tdat[-c(21,57,63), ]
ggplot(tdat2, aes(x = mWD.g.m3, y = mH.m)) +
  geom_point(shape = 1) +
  geom_smooth(method = lm) +
  xlab(expression(paste("Mean wood density, g/", m3))) +
  ylab("Mean height, m") +
  theme_bw()
```



Problem 3

Hypothesis

The null hypothesis (H_0) of this test is that there is no linear relationship between animals' brain weight and body weight.

While the alternative hypothesis (H_1) is that the brain weight increases with body weight for different animals.

Regression model

```
df <- read.table(file = here("labs/lab6 linear models/Brain_Weight_Body_Weight.txt"), header = TRUE)

lm3 <- lm(BrainWeight ~ BodyWeight, data = df)
summary(lm3)
```

```
##
## Call:
## lm(formula = BrainWeight ~ BodyWeight, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1552.25    -8.00    47.36    55.10   1553.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -56.85555   42.97805  -1.323    0.191
## BodyWeight    0.90291    0.04453   20.278 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 323.5 on 60 degrees of freedom
## Multiple R-squared:  0.8727, Adjusted R-squared:  0.8705
## F-statistic: 411.2 on 1 and 60 DF,  p-value: < 2.2e-16
```

Result analysis

The ANOVA printout reveals that the F-test for the predictor variable, body weight, yielded a small p-value ($<2.2e-16$), indicating that the regression model accounts for a substantial portion of the variation in the data compared to the null model with only the intercept.

The statistical analysis further confirms the significance of the linear regression model ($R^2 = 0.8705$, $F_{1,60} = 411.2$, $p < 0.001$). The R-squared value of 0.87 demonstrates that the body weight explains 87% of the variation in the brain weight.

Model Interpretation

The regression results imply that the body weight of animals significantly increases their brain weight, with per unit increase resulting in a 0.90291 kg increase in the brain weight ($R^2 = 0.8705$, $F_{1,60} = 411.2$, $p < 0.001$).

The equation of the linear regression model is:

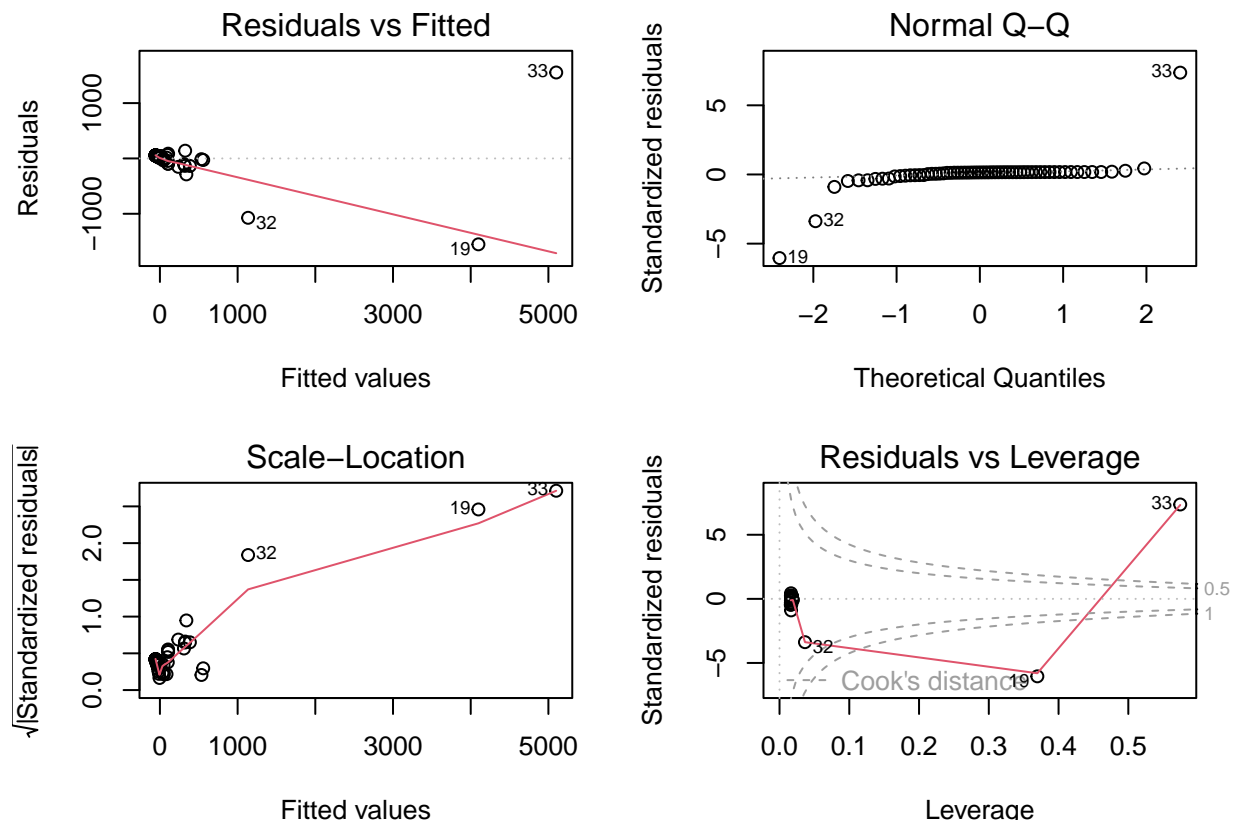
$$height = -56.8555 + 0.90291 * density$$

The intercept in this model is not interpretable since it pertains to a scenario outside the feasible range of the data.

Model assumptions checking

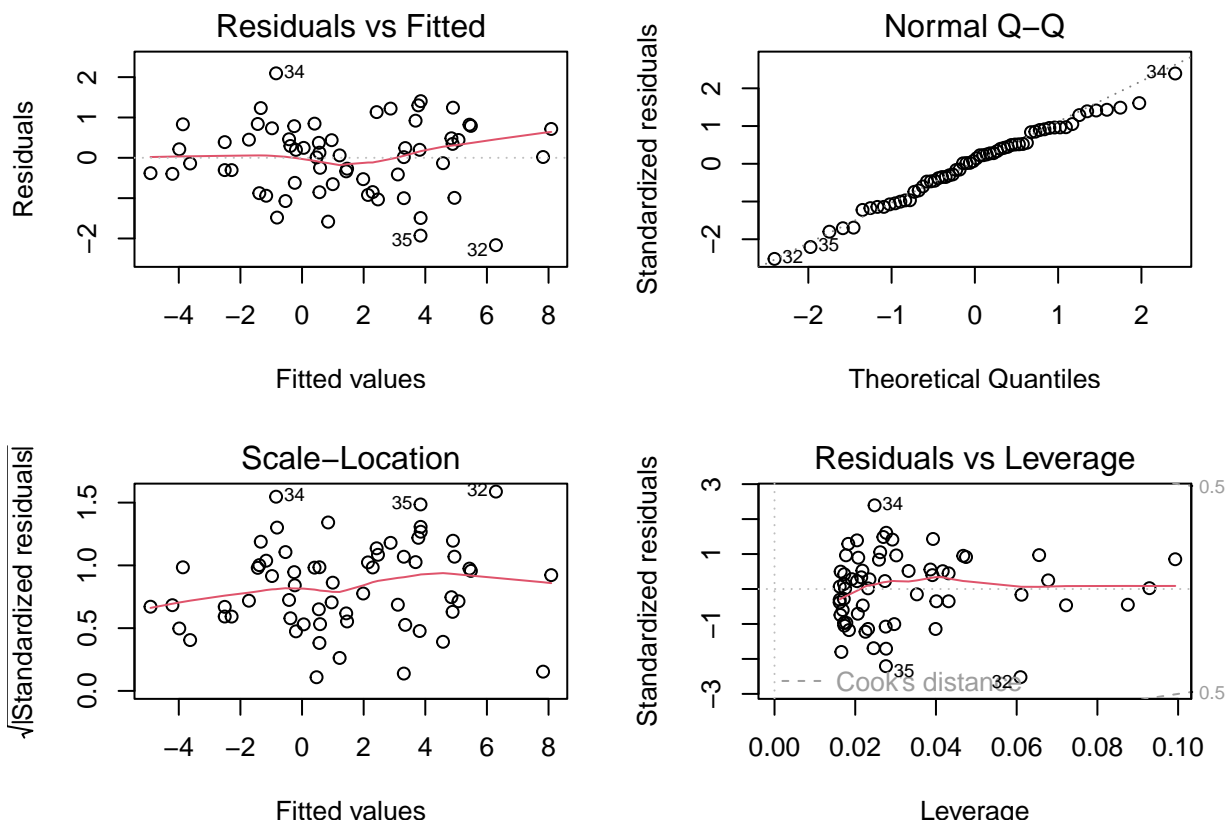
The assumptions of the linear regression model are checked by examining the residual plots for normality, constant variance, and linearity.

```
par(mfrow=c(2,2), mar = c(3.8, 4, 3, 2))
plot(lm3)
```



The plots shown above are not good. The banana shape indicates the potential of log transformation. Also from the common sense, the distribution of animals' weight is not even and has great variations. Thus, instead of viewing some points as outliers, I would rather take a log transformation and make them fit in that model.

```
lm3.log <- lm(log(BrainWeight) ~ log(BodyWeight), data = df)
par(mfrow=c(2,2), mar = c(3.8, 4, 3, 2))
plot(lm3.log)
```



```
summary(lm3.log)$r.squared
```

```
## [1] 0.9207837
```

```
coef(lm3.log)
```

```
##      (Intercept) log(BodyWeight)
##      -2.509072      1.224958
```

The adjustment may be appropriate because the log transformation enhances the R^2 value and does not significantly alter the parameter estimates.

Visualization

```
require(ggplot2)

ggplot(df, aes(x = log(BodyWeight), y = log(BrainWeight))) +
  geom_point(shape = 1) +
  geom_smooth(method = lm) +
  xlab("Log(body weight), kg") +
  ylab("Log(brain weight), kg") +
  theme_bw()
```

