Guided Question Set 5 Solutions

1)

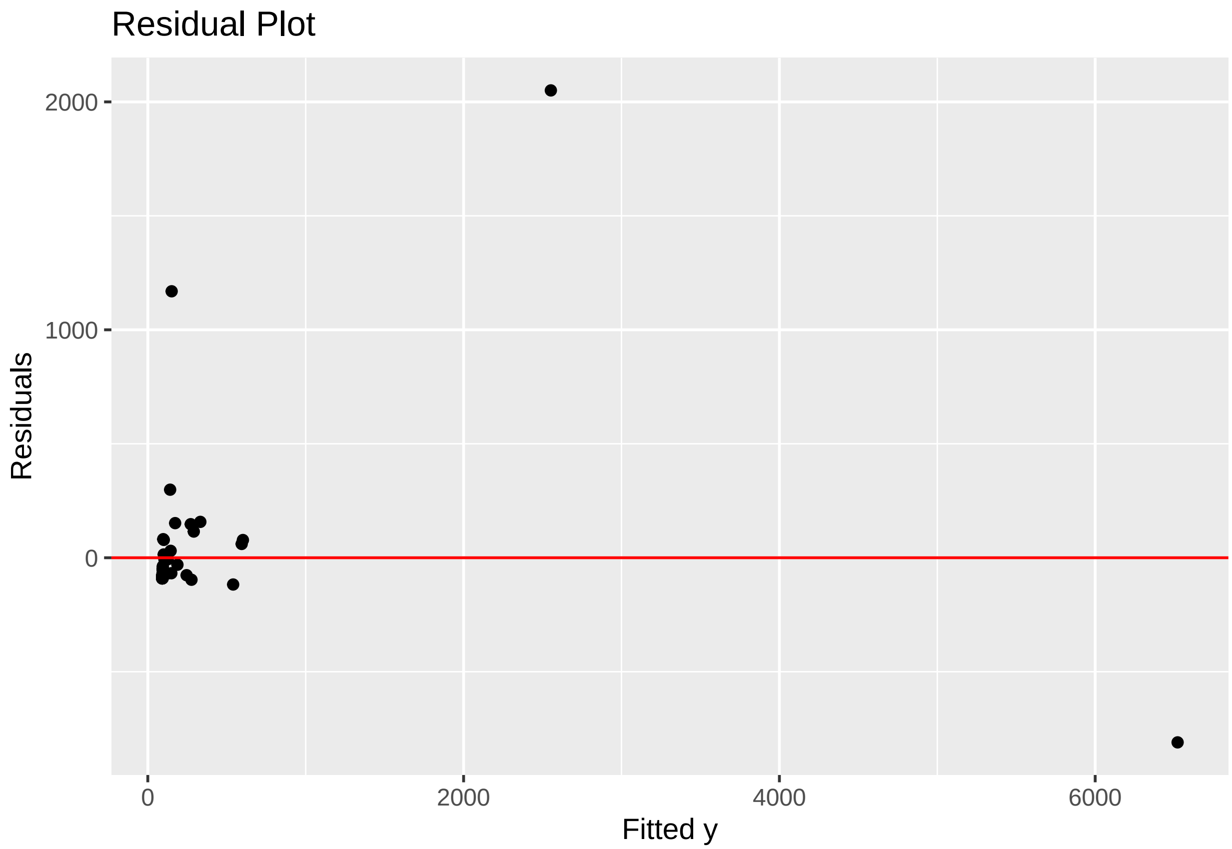
```
## `geom_smooth()` using formula 'y ~ x'
```

Generally, there appears to be an increasing association between body weight and brain weight of land mammals. The heavier the mammal, the heavier the brain. However, the relationship may be more logarithmic rather than linear. There are a couple of mammals with heavy body weights (around 2500kg and 6500kg) that make this distinction between a logarithmic or linear relationship difficult.

So assumption 1, that the relationship is linear, may not be met.

It is difficult to assess assumption 2, that the variance is constant, is difficult to assess with this scatterplot (due to the two heavy mammals).

2)

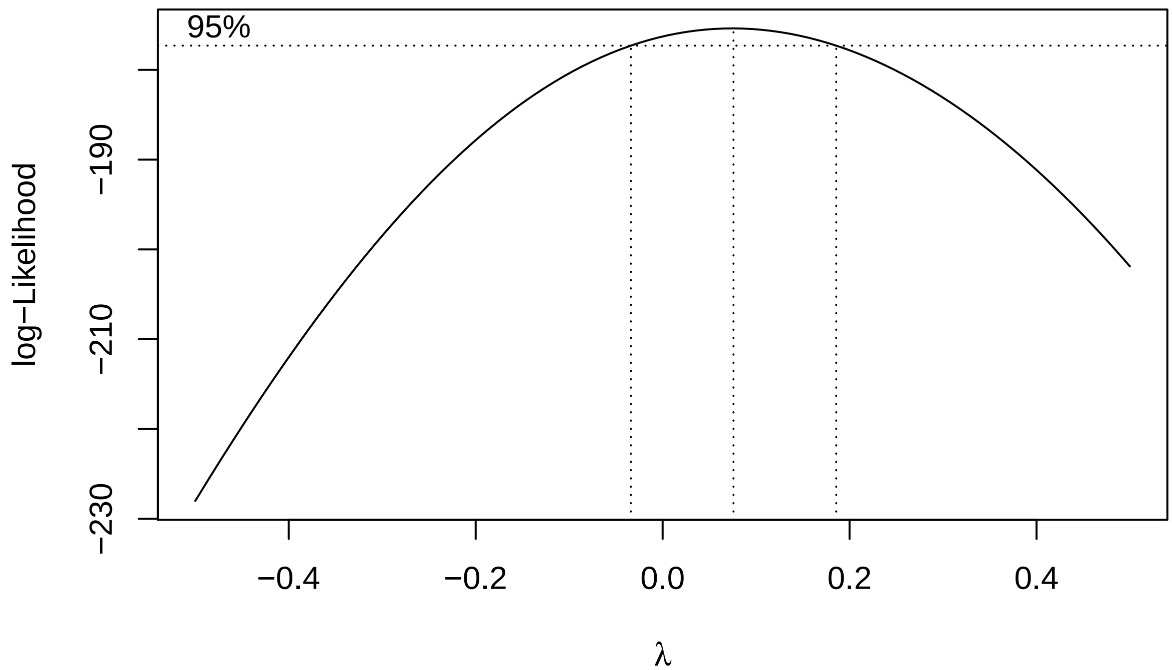


Similar to the scatterplot, this residual plot is a little difficult to assess due to the presence of two observations with large fitted values (probably the 2 heavy mammals identified earlier).

3)

We should think about transforming the predictor, since the linearity of the relationship is questionable. The plots were difficult to evaluate if the constant variance assumption was met.

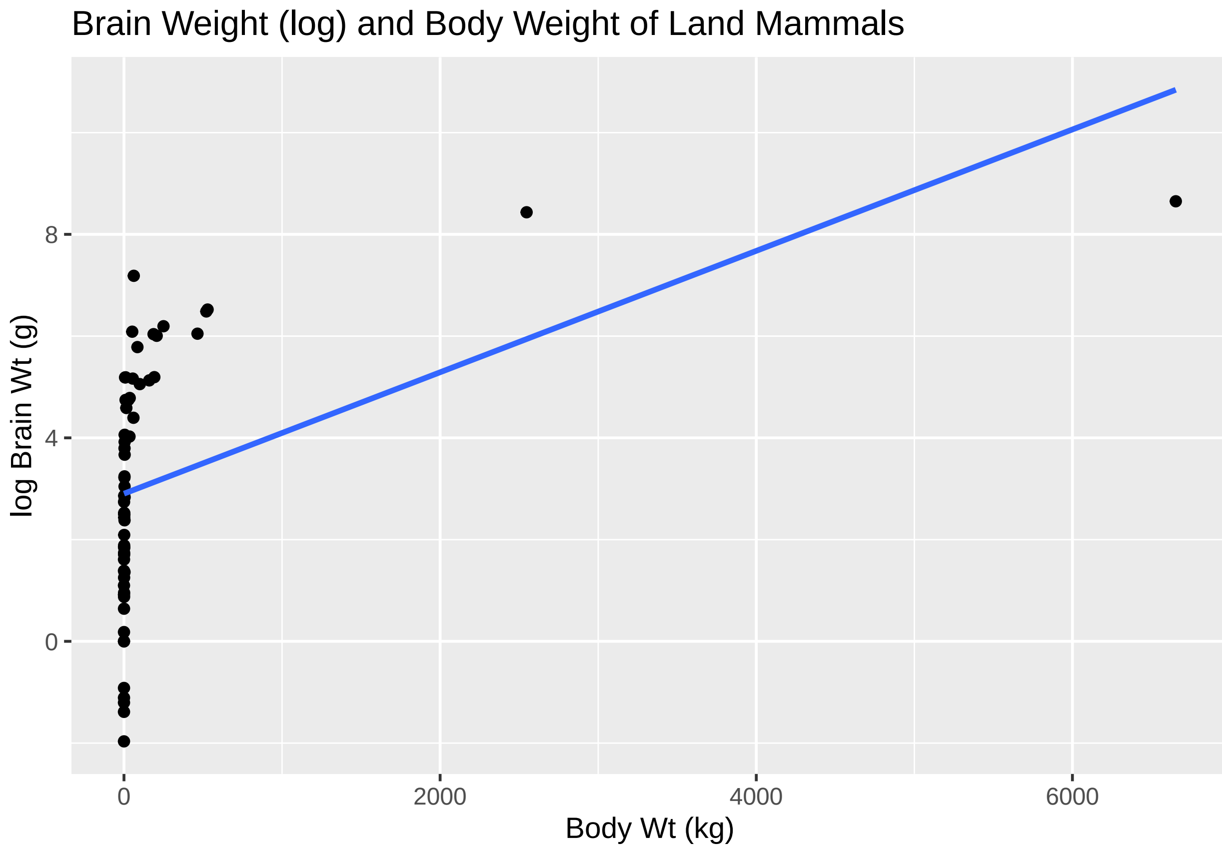
4)



Based on the Box Cox plot, a log transformation on the response variable should be performed, so $y^* = \log(y)$.

5)

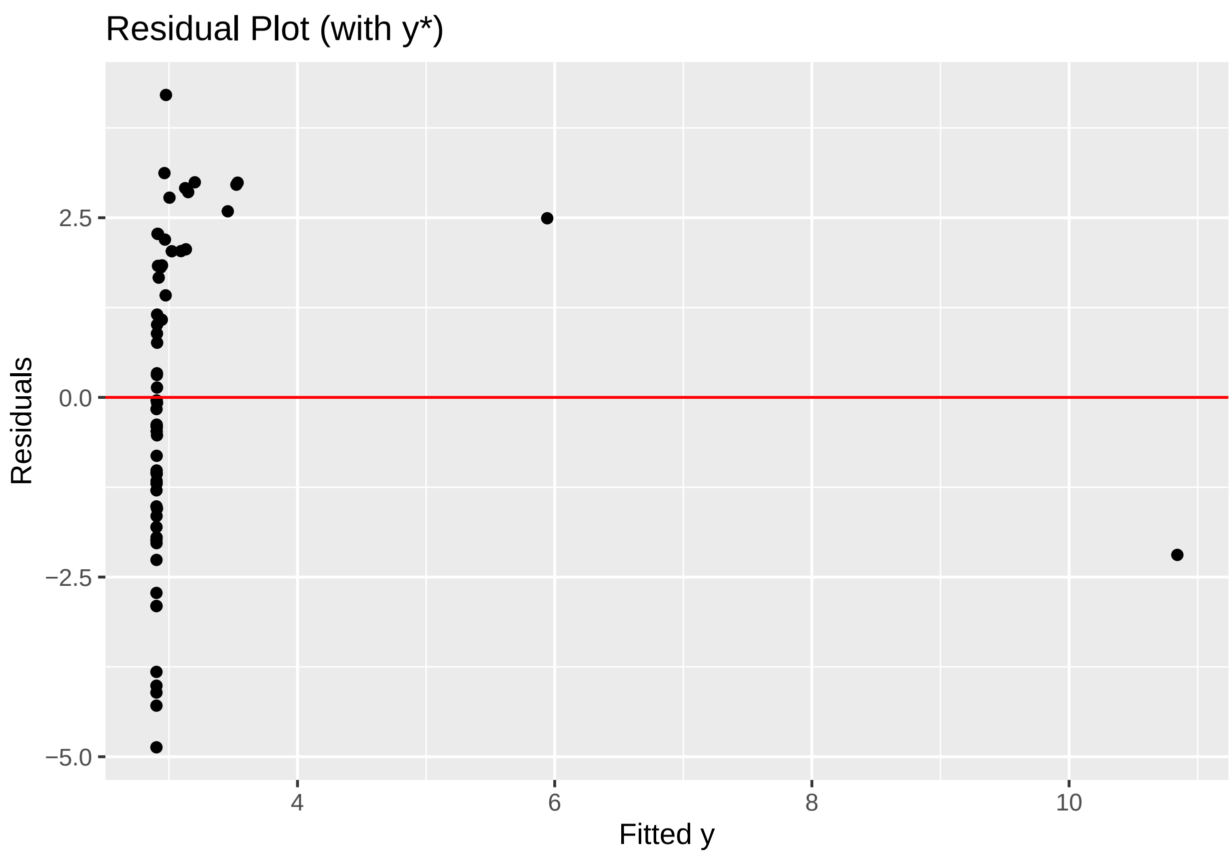
```
## `geom_smooth()` using formula 'y ~ x'
```



The relationship between y^* and x does not appear to be linear. The relationship appears logarithmic.

Again, the constant variance assumption is difficult to evaluate with this scatterplot.

6)



The mean of the residuals do not appear to be 0 when fitted y is great than 2 and below 10. This indicates that the relationship between y^* and x is not linear.

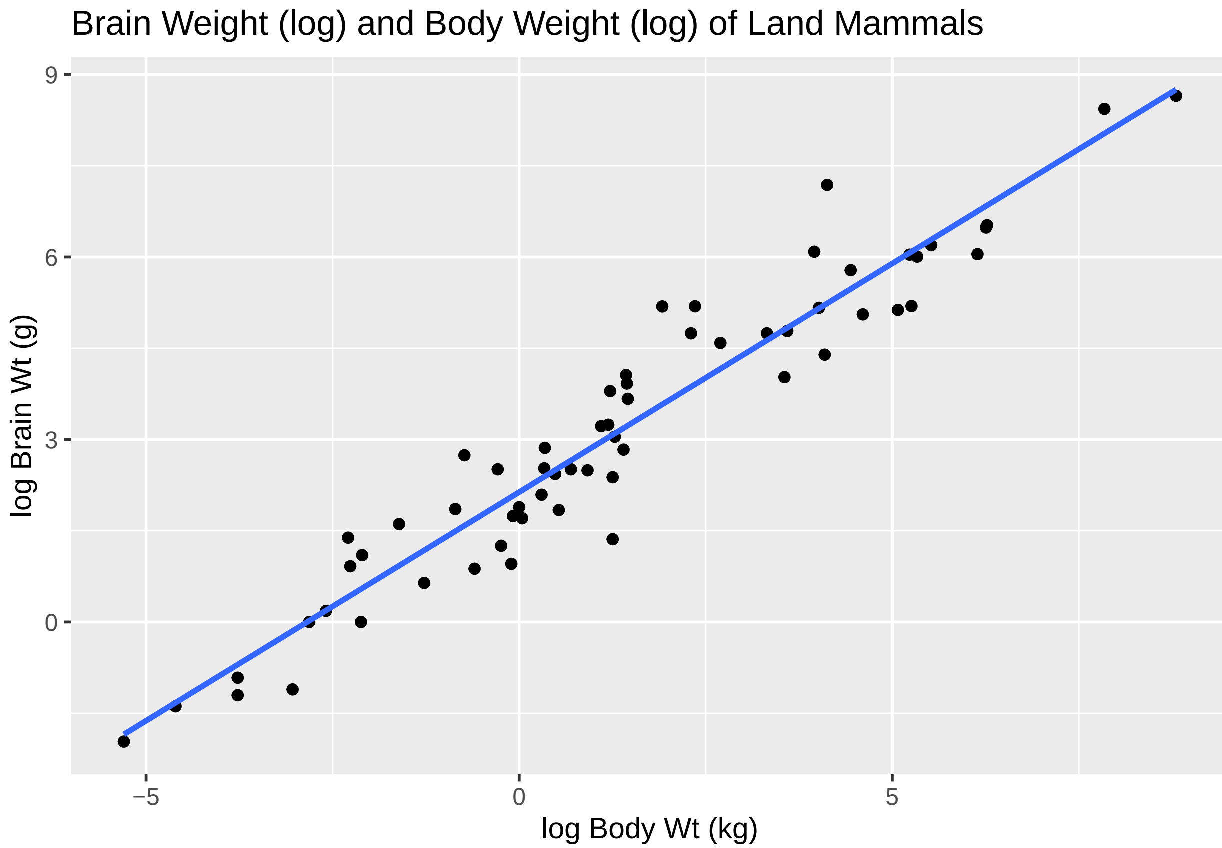
Again, it is quite difficult to assess if the variance is constant.

7)

Since we already performed the transformation on the response variable based on the Box Cox plot, we consider only transforming the predictor.

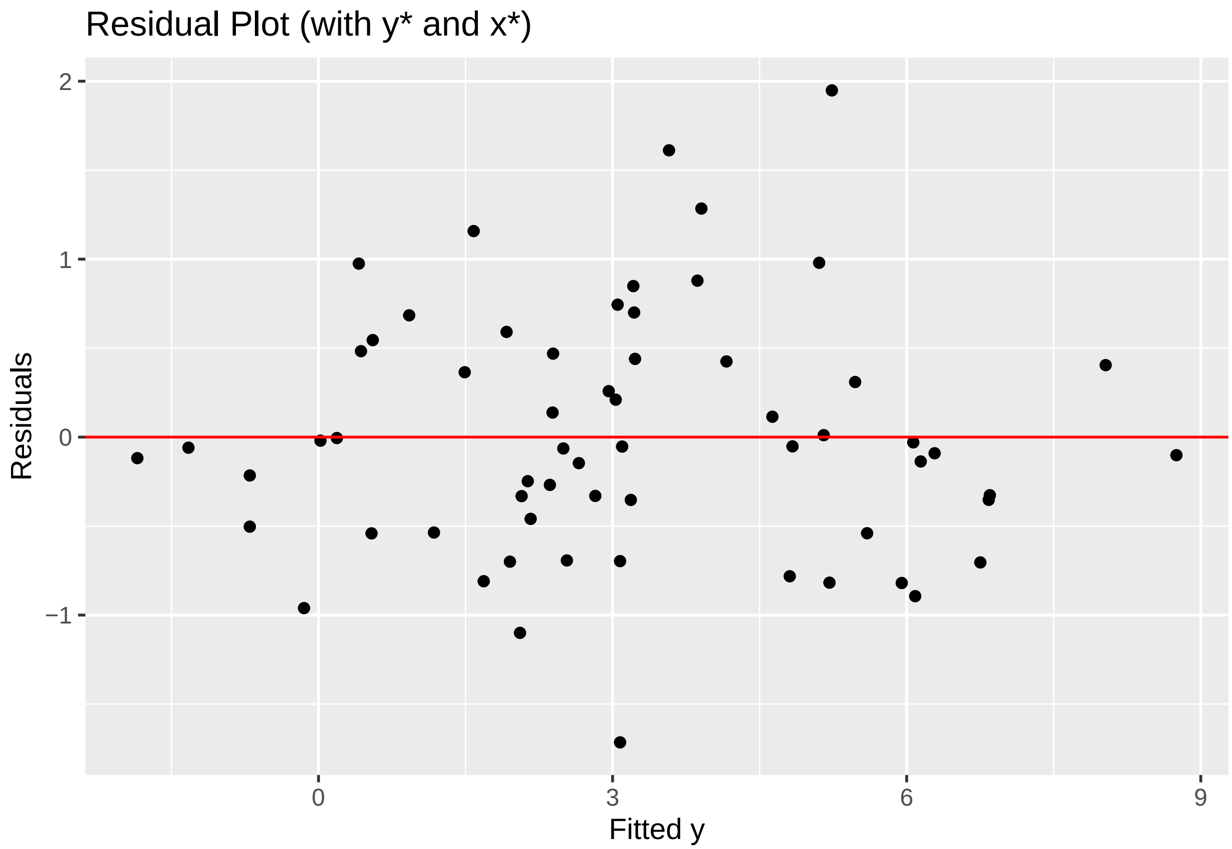
Based on the scattterplot in part 5, we should perform a log transformation on the predictor, so $x^* = \log(x)$.

```
## `geom_smooth()` using formula 'y ~ x'
```



Looking at the scatterplot of y^* against x^*, the relationship appears to be linear (and positive). The constant variance assumption appears to be reasonably met as we don't see the variance increasing or decreasing.

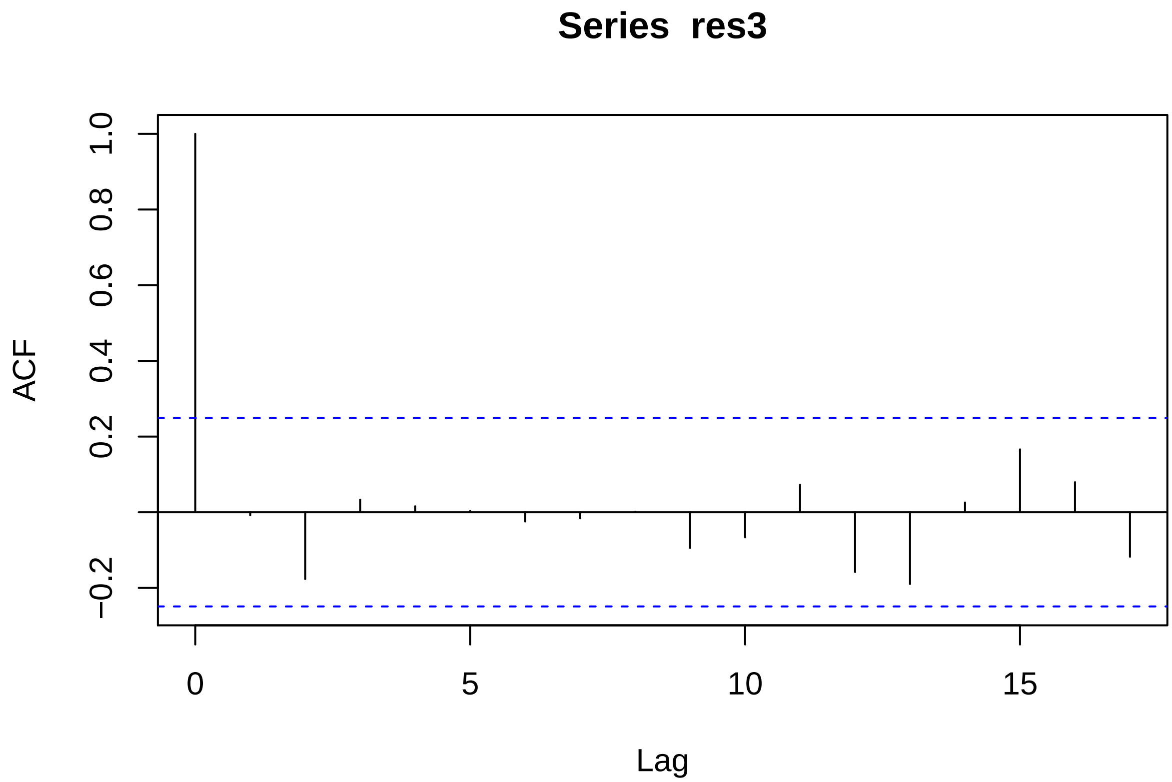
8)



We see huge improvements in the residual plot of y^* against x^*. The residuals are generally evenly scattered on both sides of the x-axis, so the assumption that the errors have 0 mean is met.

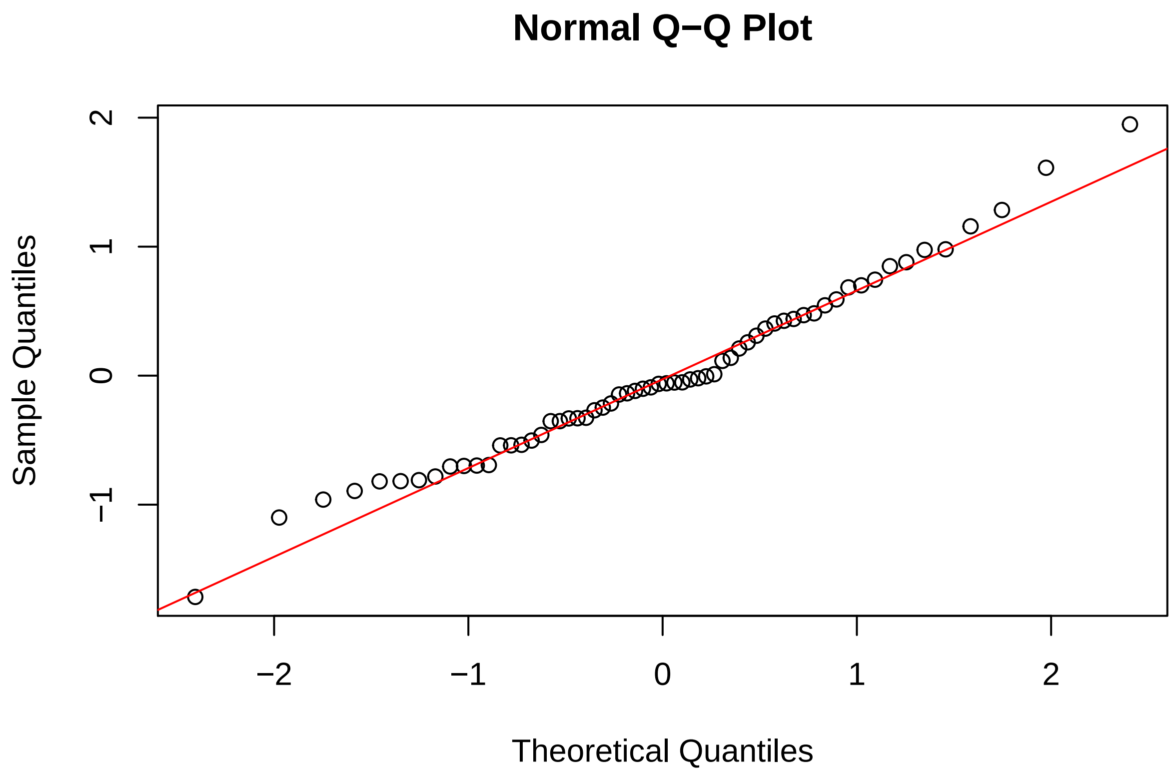
The spread of the residuals also is fairly constant as we move across the residual plot. We do not see the variance increasing or decreasing. So the constant variance assumption for the errors is met.

9)



Based on the ACF plot, the residuals are uncorrelated, so we don't have evidence that the errors are dependent.

10)



The plots fall closely to the line, so the residuals follow a normal distribution.

11)

```
summary(result3)
```

```
##
## Call:
## lm(formula = log.brain ~ log.body, data = Data)
##
## 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 ***
## log.body     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
```

We have $y^* = 2.13 + 0.75x^*$, where $y^* = \log(y), x^* = \log(x)$.

Since both variables were log transformed, we interpret the slope of 0.75 as, for a 1% increase in body weight, the weight of the brain increases by approximately 0.75%.

We note that based on the residual plot, ACF plot of residuals, and QQ plot of residuals in parts 8, 9, and 10, the assumptions for this regression model are met.