# Lab 5

Grace Randall

2024-02-22

Question 1

The null hypothesis in this test is that there is no relationship between the mean tree diameter and the amount of biomass. The alternative hypothesis is that biomass increases with increasing mean tree diameter.

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

summary(lm1)
```

the regression model has a p-value of 3.6e-11 which means we can reject the null hypothesis of no relationship. This suggests there is a positive relationship between tree diameter and biomass. However, the low adjusted R-squared value of 0.46 suggests that much of the variability is still not explained.

```r
ggplot(treedata, aes(x = mDBH.cm, y = AGBH.Mg.ha)) +
 geom_point(shape = 1) +
  geom_smooth(method = lm) +
   xlab("Mean Tree Diameter, cm") +
   ylab(expression(paste("Biomass, Mg ", ha^-1))) +
  theme_bw()
```
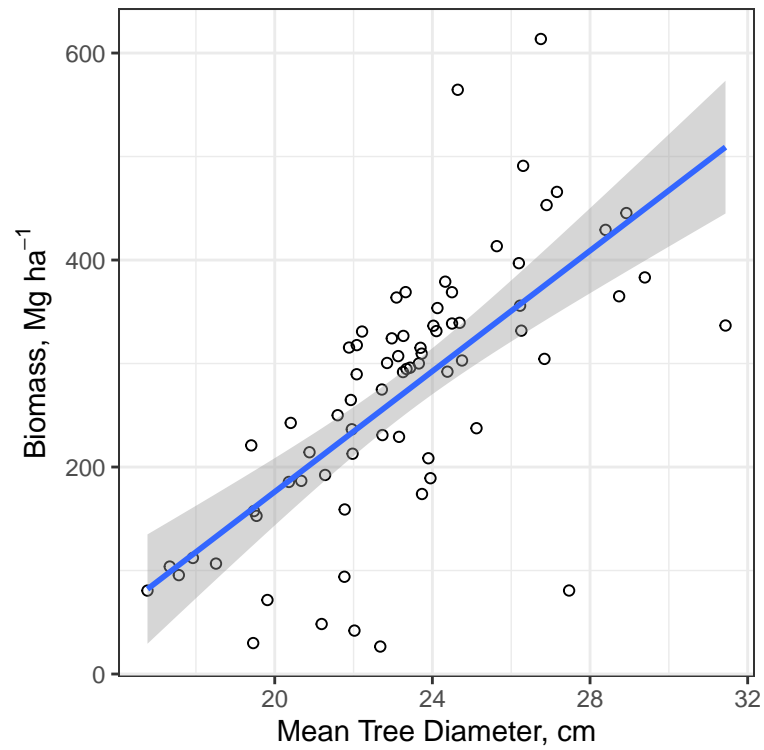
this linear model has a slope of about 29 which shows that as the average tree diameter increases by 1 cm, the biomass increases by 29 Mg ha^{-1}.

```r
par(mfrow=c(2,2))
plot(lm1)
```
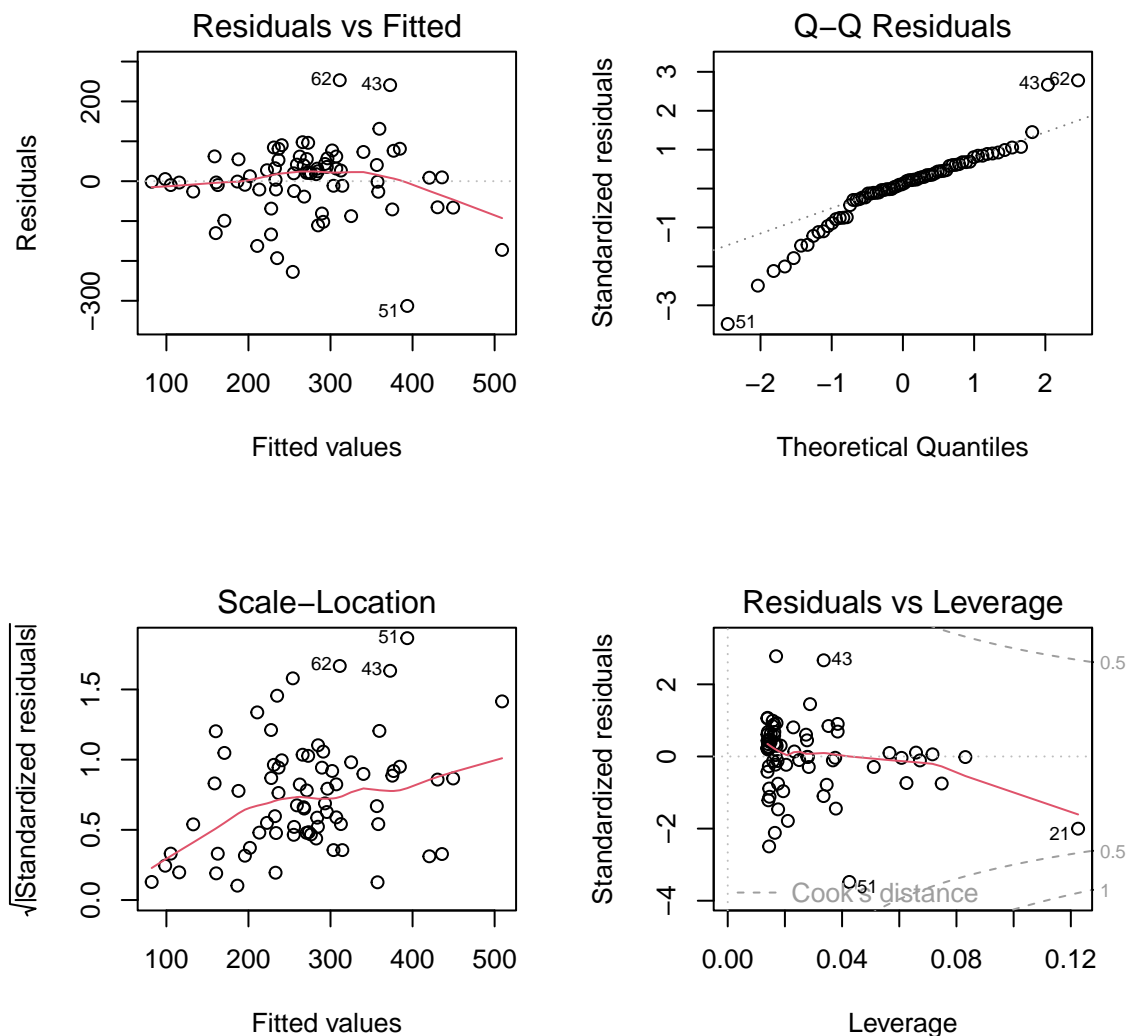
The assumption of normal distribution was checked by considering the plots of the traits of the residuals above. The points in the first graph are scattered around 0 which supports my assumption of normal distribution. the qq plot follows along the line as expected. In the last graph none of the points are outside of the cook's distance limits which suggests that there are no points that need to be removed. However it is possible the regression would be better without points 43, 62 and 51.

Question 2

The null hypothesis in this test is that there is no relationship between the mean wood density and the mean tree height. The alternative hypothesis is that mean tree height increases with increasing mean wood density.

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

summary(lm2)
```

The regression model has a p-value of 0.005 which means we can reject the null hypothesis of no relationship, even if this is not as strong as in question 1. This suggests there is a positive relationship between tree height and average wood density. However, the extremely low adjusted R-squared value of 0.09 suggests that much

of the variability is still not explained under this model.

```
ggplot(treedata, aes(x = mWD.g.m3, y = mH.m)) +
 geom_point(shape = 1) +
  geom_smooth(method = lm) +
  xlab(expression(paste("Mean Wood Density, g ", m^-3))) +
  ylab("Mean Tree Height, m") +
  theme_bw()
```
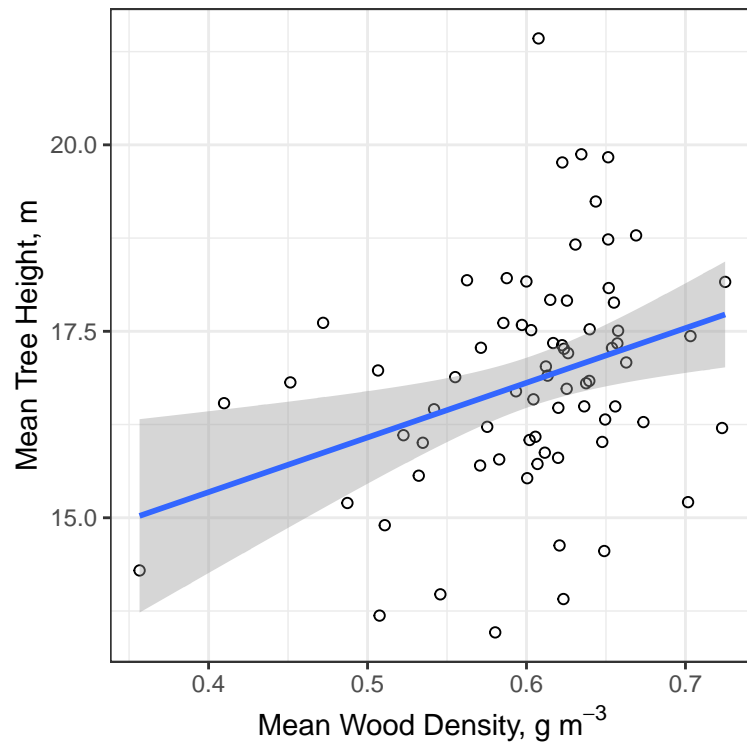

this linear model has a slope of about 7.3 which suggests that as the average wood density increases by 1 g m^{-3}, average tree height increases by 7.3 m.

```
par(mfrow=c(2,2))
plot(lm2)
```
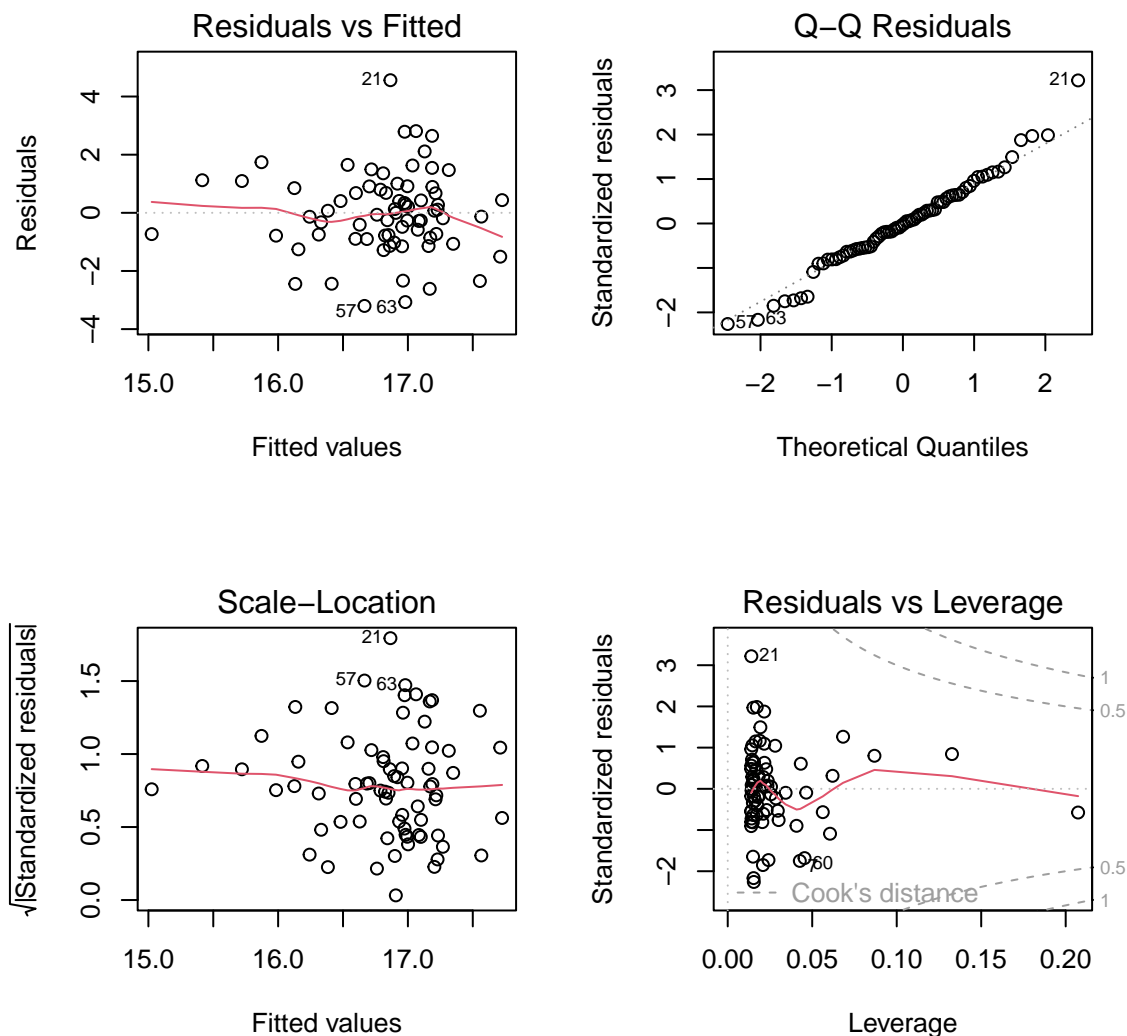
The assumption of normal distribution was checked by considering the plots of the traits of the residuals above. The points in the first graph are scattered around 0 which supports my assumption of normal distribution. the qq plot follows along the line as expected. In the last graph none of the points are outside of the cook's distance limits which suggests that there are no points that need to be removed. However it is possible the regression would be better without points 21, 63 and 57.

Question 3

The null hypothesis in this test is that there is no relationship between the an animals body weight and brain weight. The alternative hypothesis is that log of an animals brain weight increases with increasing log of body weight.

```
lm3 <- lm(log(Brain_Weight) ~ log(Body_Weight), data = bbw)

summary(lm3)
```

the regression model has a p-value of 2.2e-16 which means we can confidently reject the null hypothesis of no

relationship. This suggests there is a clear positive relationship between body weight and brain weight. the relatively high adjusted R-squared value of 0.919 suggests that most of the variability is explained under this model.

```
ggplot(bbw, aes(x = log(Body_Weight), y = log(Brain_Weight))) +
 geom_point(shape = 1) +
  geom_smooth(method = lm) +
  xlab("log of Body Weight (kg) ") +
  ylab("log of Brain Weight (kg)") +
  theme_bw()
```
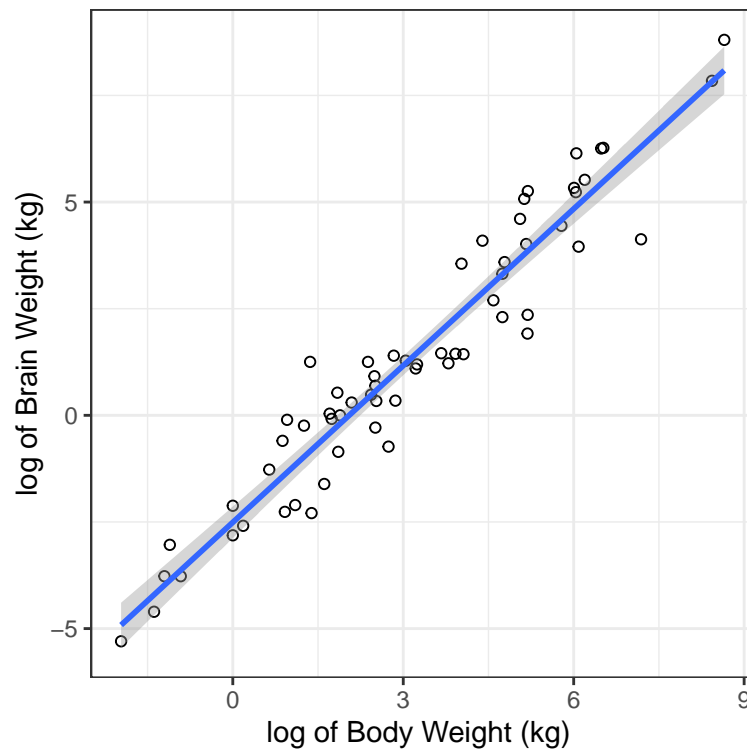

This linear model has a slope of about 1.22 which suggests that as the log of body weight increases by 1 logkg, the log of brain weight increases by 1.22 logkg.

```
par(mfrow=c(2,2))
plot(lm3)
```
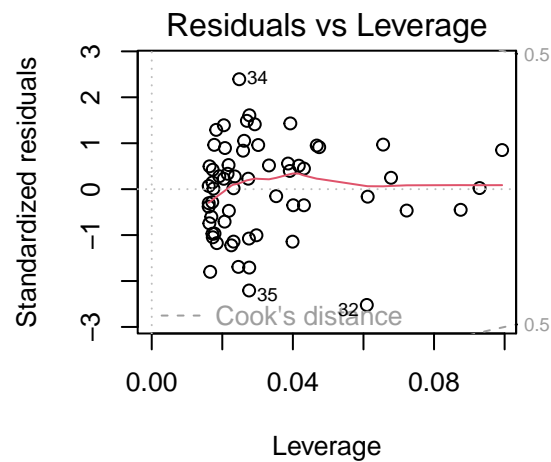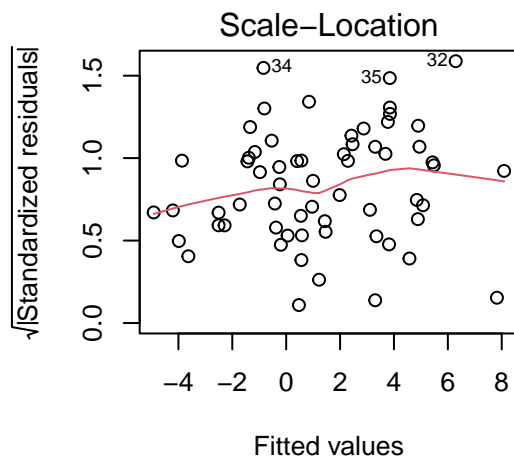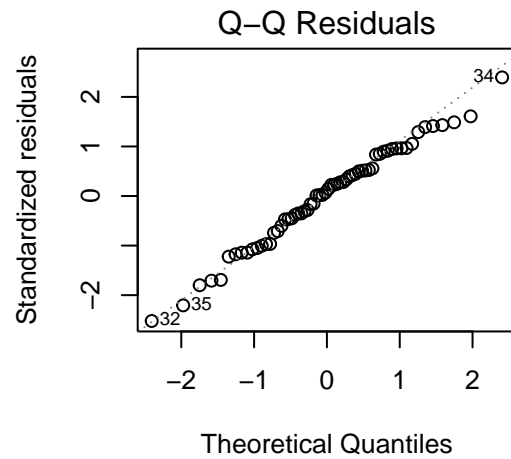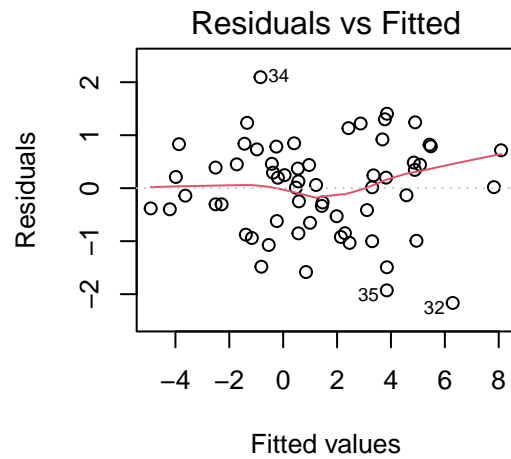
Residuals vs Fitted

Q–Q Residuals

Scale–Location

Residuals vs Leverage

The assumption of normal distribution was checked by considering the plots of the traits of the residuals above. The points in the first graph are scattered around 0 which supports my assumption of normal distribution. the qq plot follows along the line as expected. In the last graph none of the points are outside of the cook's distance limits which suggests that there are no points that need to be removed. However, it is possible the regression would be better without points 35, 34 and 32.