# Day17

Olivia Wu

2024-04-30

## Problem 5.8

C

## Problem 5.10

a) The variability within each group changes a lot across groups.

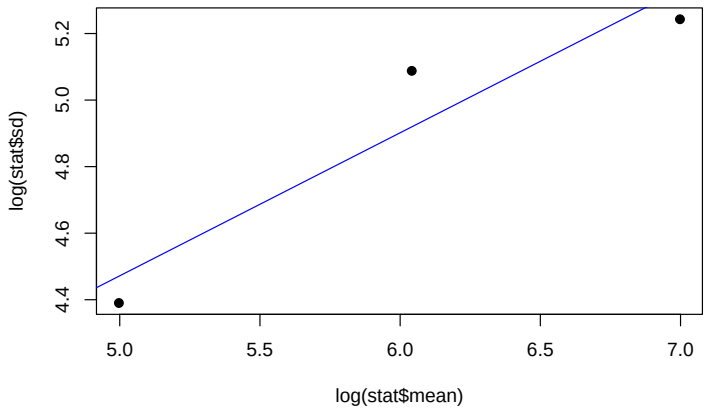
b) The slope tells us which transformation to use after saying $p = 1 - s$.

## Problem 5.43

a) The points look like they could fit on a line, but barely.

b) The slope of the line is 0.4298.

c) $p = 0.5702$, which suggests a square root transformation.

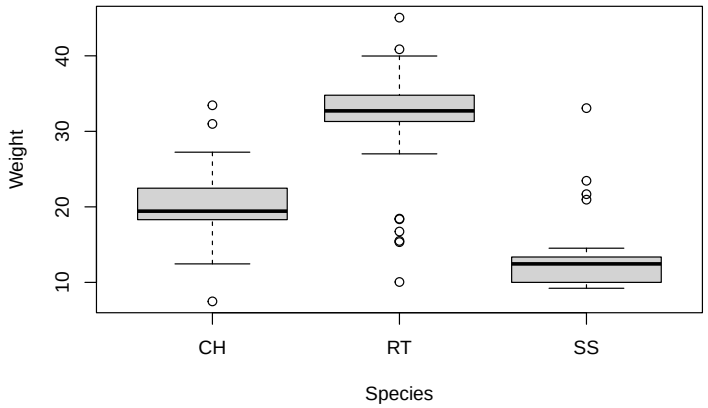



## Problem 5.49

a)

```
##   Species     mean       sd   n
## 1      CH 20.14429 3.860873  70
## 2      RT 32.94032 3.062976 572
## 3      SS 11.92161 2.422181 256
```

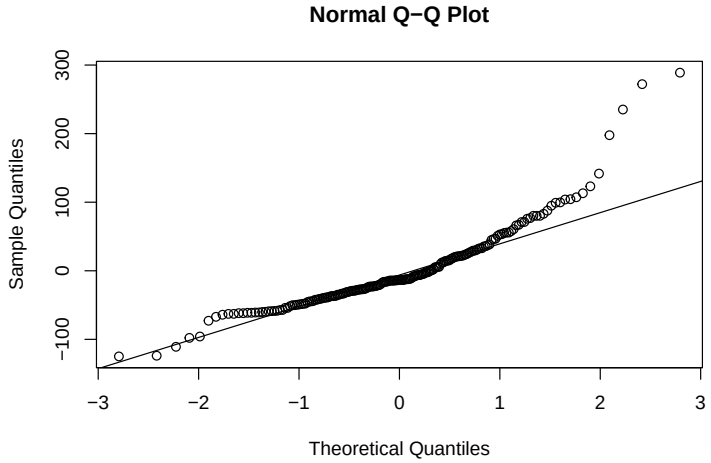
b) The average weight tends to be higher for RT, then CH, then SS.



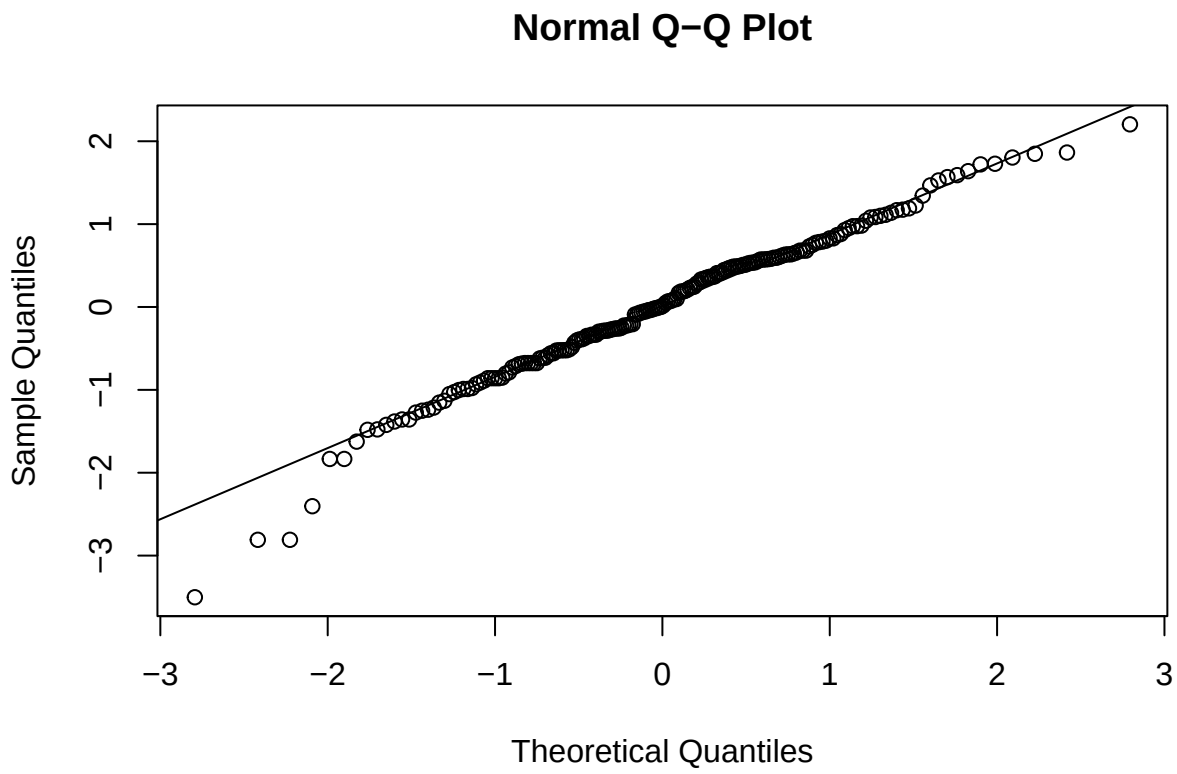

c) The IQR are roughly the same size, suggesting similar variability. There are many outliers for each groups, suggesting a skew. Thus, the normality condition is not met, and an anova test would be inappopriate.

# Problem 5.67

a) The points on the right do not lie close to the line, there seems to be a curved pattern. This fails the normality condition.



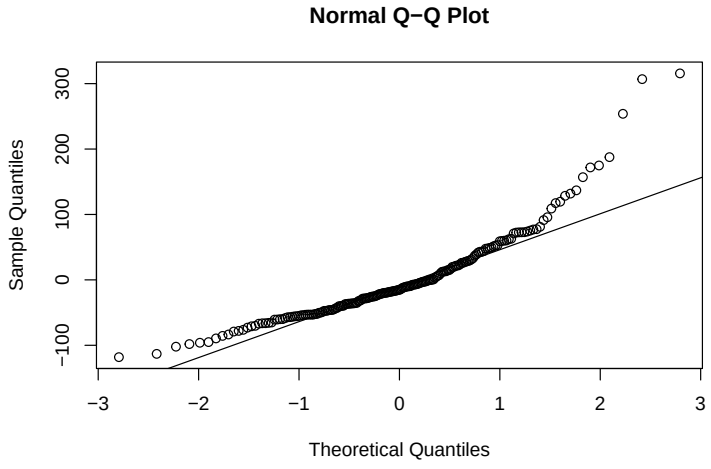

b) The transformed data fits the line much better, and it has better normality.




## Problem 5.69

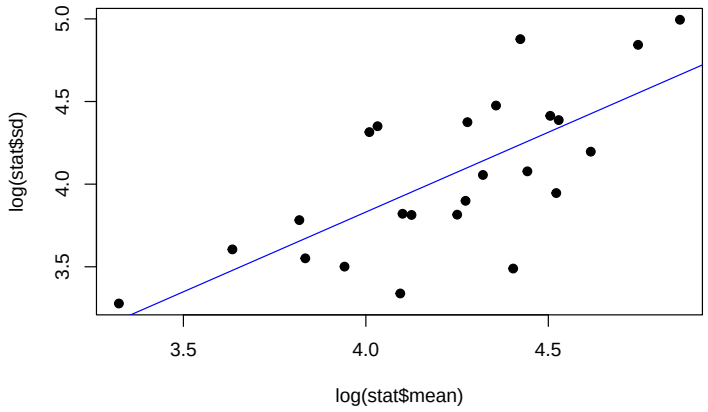
Original data: the qqplot is not linear, so normality fails.

```
model <- aov(Time~Round, data=FanBase)
qqnorm(model$residuals)
qqline(model$residuals)
```



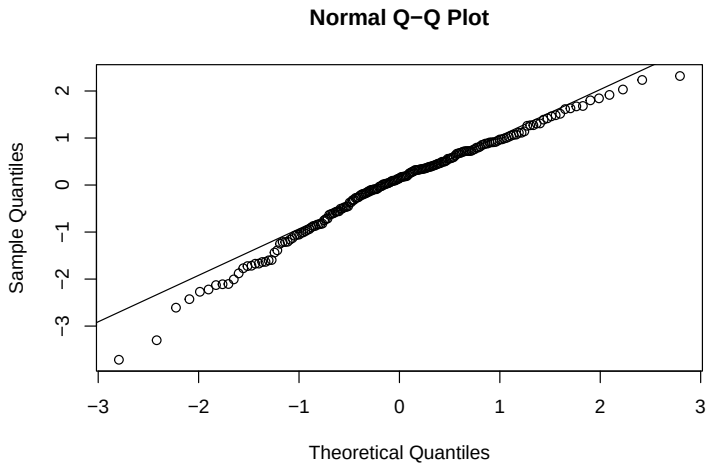
We check log(s) vs. log(ave): the points are roughly linear, and we see that 1-slope $\approx 0.035$, so we proceed to use logarithms to transform the data.

```
stat <- favstats(Time~Round,data=FanBase)[c("Round","mean","sd","n")]
plot(log(stat$sd)~log(stat$mean), pch=19)
abline(lm(log(stat$sd)~log(stat$mean)),col="blue")
```




Transformed data: The points are much more linear, and they meet the normality condition.

```
model <- aov(LnTime~Round, data=FanBase)
qqnorm(model$residuals)
qqline(model$residuals)
```




Anova: $p = 0.713 > 0.05$, so we fail to reject the null hypothesis. There is not enough evidence to suggest that some rounds tend to have significantly longer selection times than other rounds.

```
summary(model)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
```

```
## Round        23  23.82   1.036   0.812  0.713
## Residuals   168 214.30   1.276
```