Exercise 2: Crops

Exercise 2: Crops

```
crops <- read.table("../data/crops.txt", header = TRUE)
crops_frame <- data.frame(crops)
yes_filter <- data.frame(match = c("yes"))
crops_frame$Related <- c(0, 1)[(crops_frame$Related %in% yes_filter) + 1]
crops_frame$County <- factor(crops_frame$County)
crops_frame$Related <- factor(crops_frame$Related)

county_related_anova <- lm(Crops ~ Related * County, data = crops_frame)
```

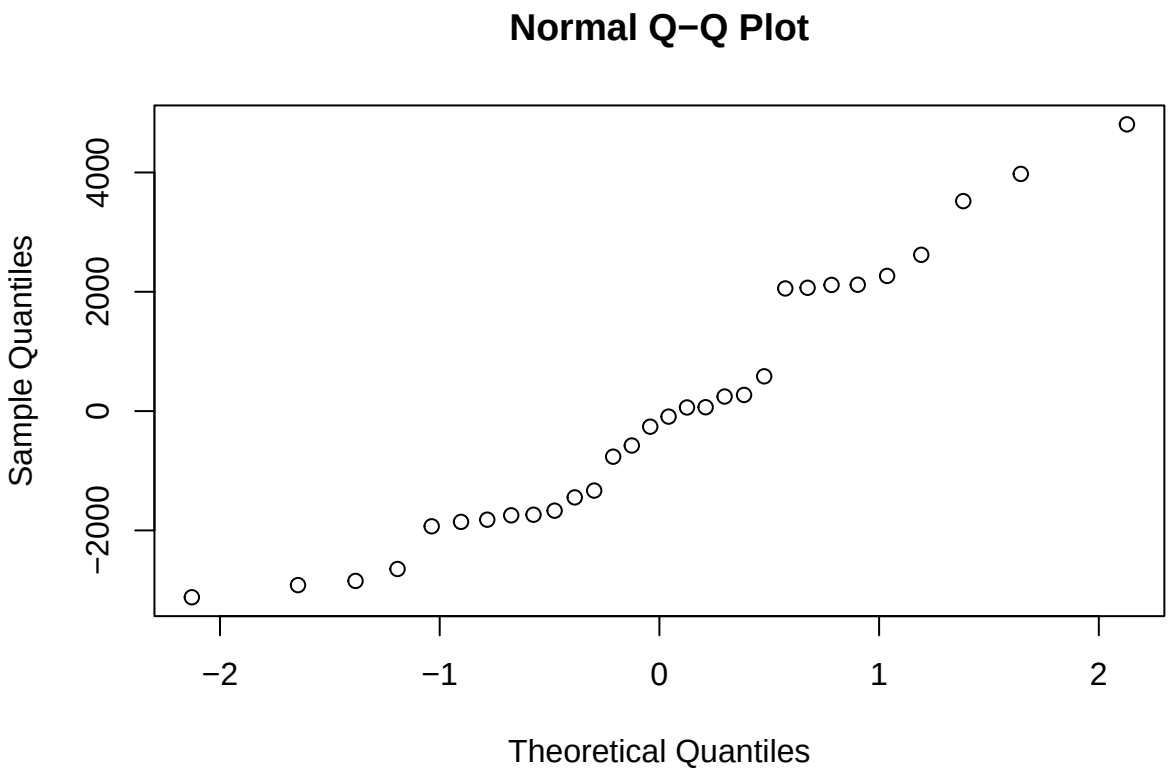
This data comes from a sample of farms from three counties in Iowa. We want to know how the factors of the county and whether the farmer is related to the landlord of the farmland is related to the total crop yield of the farms. Firstly, we perform a two-way ANOVA on the Count, Related and Crops columns, which gives

```
anova(county_related_anova)
```

```
## Analysis of Variance Table
##
## Response: Crops
##                Df    Sum Sq Mean Sq F value Pr(>F)
## Related         1   2378957 2378957  0.4113 0.5274
## County          2   8841441 4420721  0.7644 0.4766
## Related:County  2   1497573  748786  0.1295 0.8792
## Residuals      24 138805865 5783578
```

The p-value here is given by the `Pr(>F)` column. For the Related and County factors separately, the p-value is not below 0.05, which means that a linear relation between these factors and the crop yield cannot be conclusively established. This is also true for the interaction between County and Related. As the two-way ANOVA assumes the data is normally distributed, we have to asses the normality of the data.

```
qqnorm(residuals(county_related_anova))
```



This Q-Q plot of the residuals of the ANOVA places the point in roughly a straight line, which implies that this data is normally distributed. # `{r} plot(fitted(county_related_anova), residuals(county_related_anova)) #` This plot shows how the spread of the residuals is roughly equal for all values. This implies that the underlying data is normally distributed.