One-Way ANOVA: Effect of Fertilizer on Crop Yield

Statisticians World

August 22, 2024

Load Libraries

```
library(ggplot2)
library(ggpubr)
library(tidyverse)
library(broom)
library(rstatix)
```

Import Data

```
# Set working directory (update this path as needed)
setwd("C:/Users/O&1/OneDrive/Documents/R-Youtube")

# Load the dataset
my_data <- read.csv("crop.data.csv")

# View summary
summary(my_data)
```

```
##     density        block        fertilizer     yield
##  Min.   :1.0   Min.   :1.00   Min.   :1    Min.   :175.4
##  1st Qu.:1.0   1st Qu.:1.75   1st Qu.:1    1st Qu.:176.5
##  Median :1.5   Median :2.50   Median :2    Median :177.1
##  Mean   :1.5   Mean   :2.50   Mean   :2    Mean   :177.0
##  3rd Qu.:2.0   3rd Qu.:3.25   3rd Qu.:3    3rd Qu.:177.4
##  Max.   :2.0   Max.   :4.00   Max.   :3    Max.   :179.1
```

One-Way ANOVA

We test if different fertilizers result in significantly different crop yields.

Model Fit

```
oneway <- aov(yield ~ factor(fertilizer), data = my_data)
summary(oneway)
```

```
##                    Df Sum Sq Mean Sq F value Pr(>F)
## factor(fertilizer)  2   6.07  3.0340   7.863  7e-04 ***
## Residuals          93  35.89  0.3859
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Post-Hoc Test (Tukey HSD)

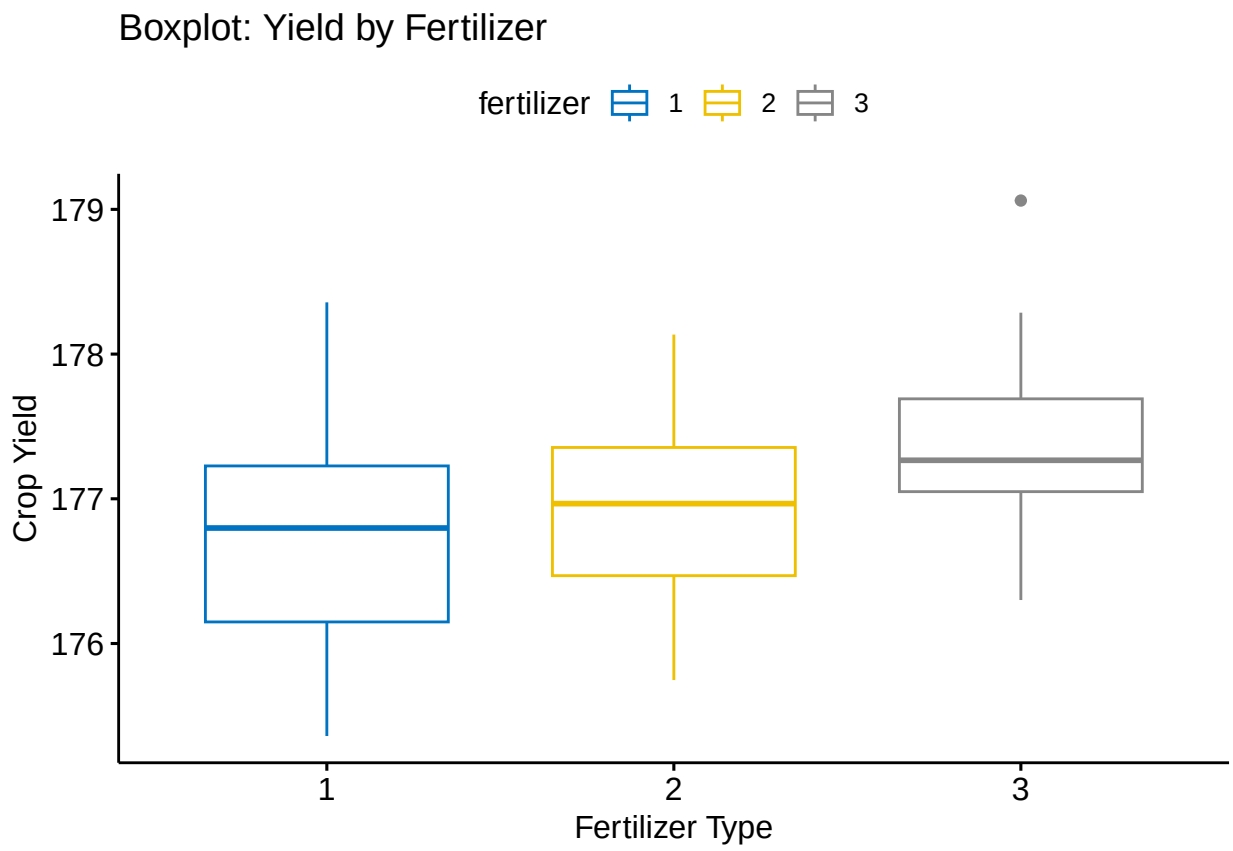
```
TukeyHSD(oneway)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = yield ~ factor(fertilizer), data = my_data)
##
## $`factor(fertilizer)`
##          diff         lwr       upr     p adj
## 2-1 0.1761687 -0.19371896 0.5460564 0.4954705
## 3-1 0.5991256  0.22923789 0.9690133 0.0006125
## 3-2 0.4229569  0.05306916 0.7928445 0.0208735
```

Assumptions Checking

Outliers (Boxplot)

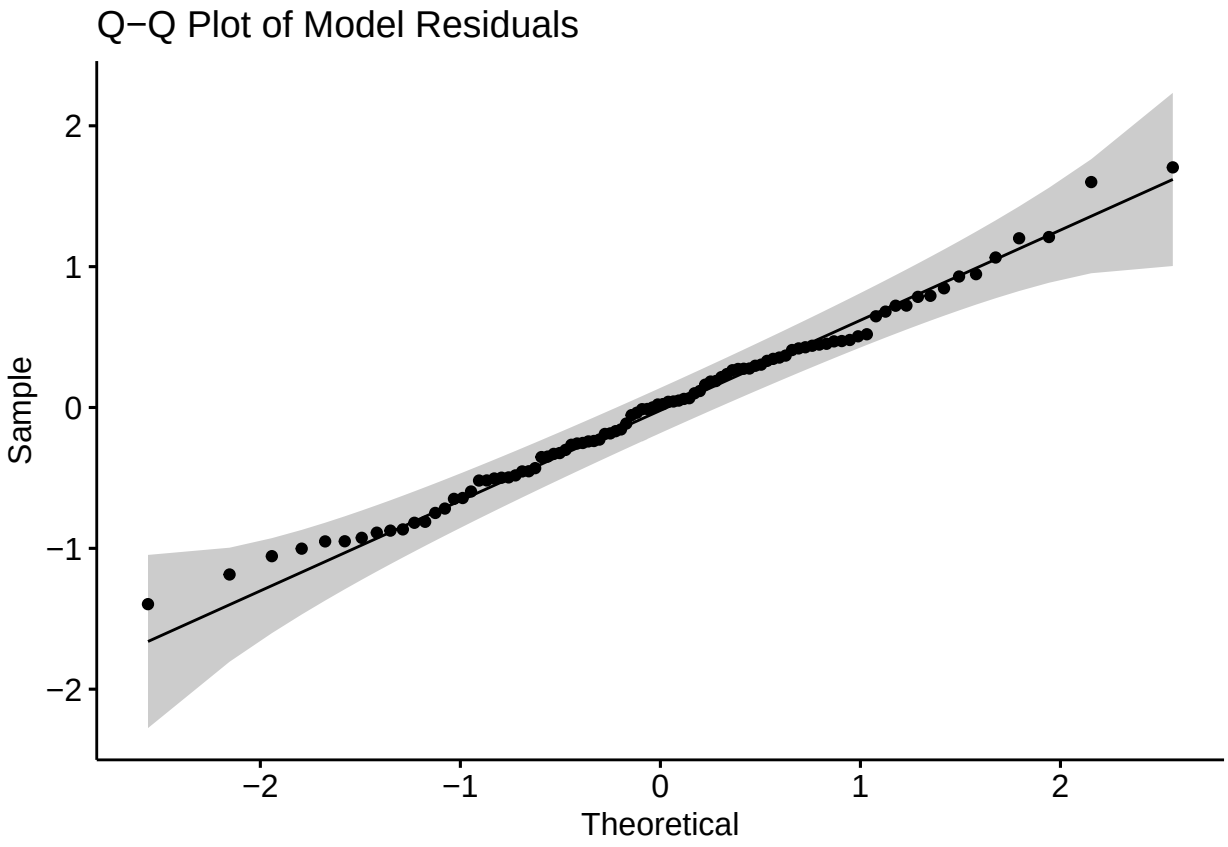
```
ggboxplot(my_data, x = "fertilizer", y = "yield",
          color = "fertilizer", palette = "jco",
          ylab = "Crop Yield", xlab = "Fertilizer Type",
          title = "Boxplot: Yield by Fertilizer")
```



Normality of Residuals

Q-Q Plot of Residuals

```
model <- lm(yield ~ factor(fertilizer), data = my_data)
ggqqplot(residuals(model), title = "Q-Q Plot of Model Residuals")
```



Shapiro-Wilk Test on Residuals

```
shapiro_test(residuals(model))
```

```
## # A tibble: 1 x 3
##   variable         statistic p.value
##   <chr>                <dbl>   <dbl>
## 1 residuals(model)     0.991   0.759
```

Shapiro-Wilk Test by Group

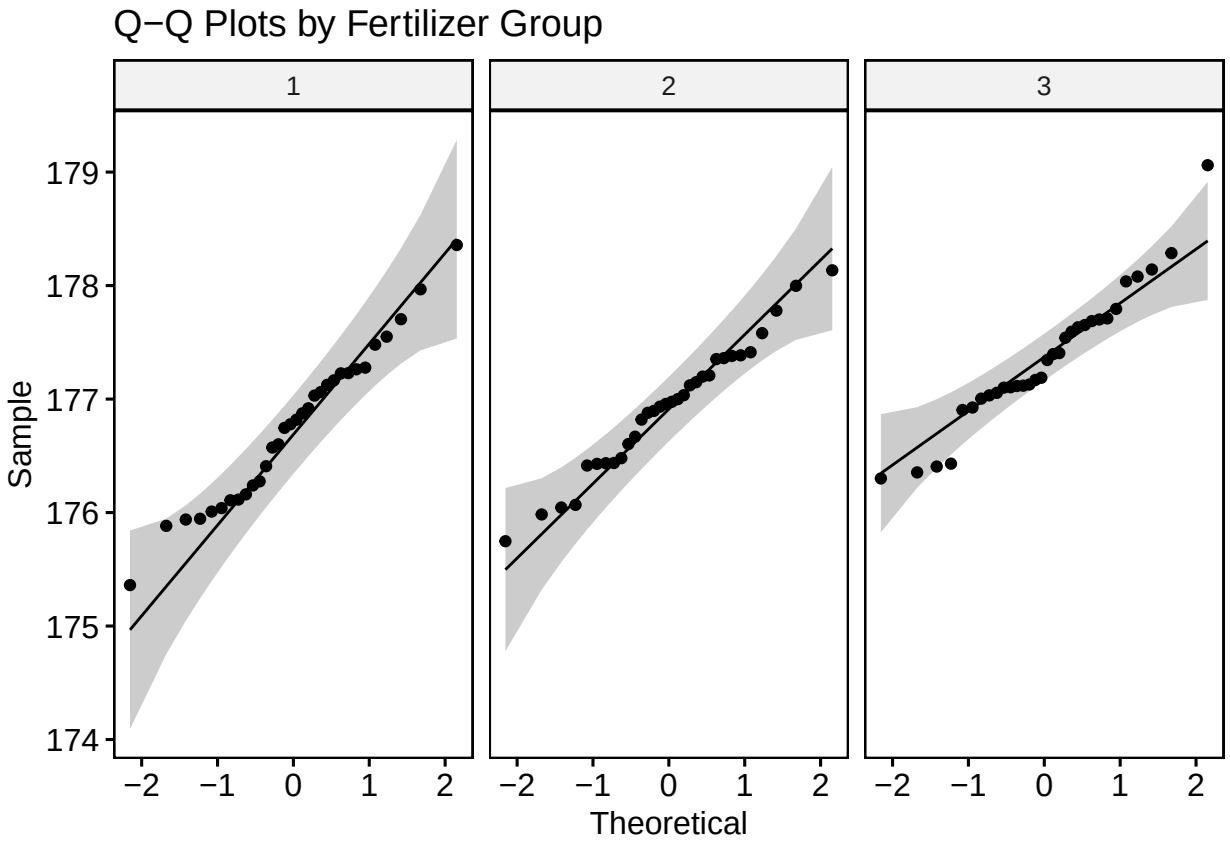
```
my_data %>%
  group_by(fertilizer) %>%
  shapiro_test(yield)
```

```
## # A tibble: 3 x 4
##   fertilizer variable statistic     p
##        <int> <chr>        <dbl> <dbl>
## 1          1 yield        0.979 0.774
## 2          2 yield        0.983 0.887
```

```
## 3          3 yield        0.959 0.254
```

Group-Wise Q-Q Plots

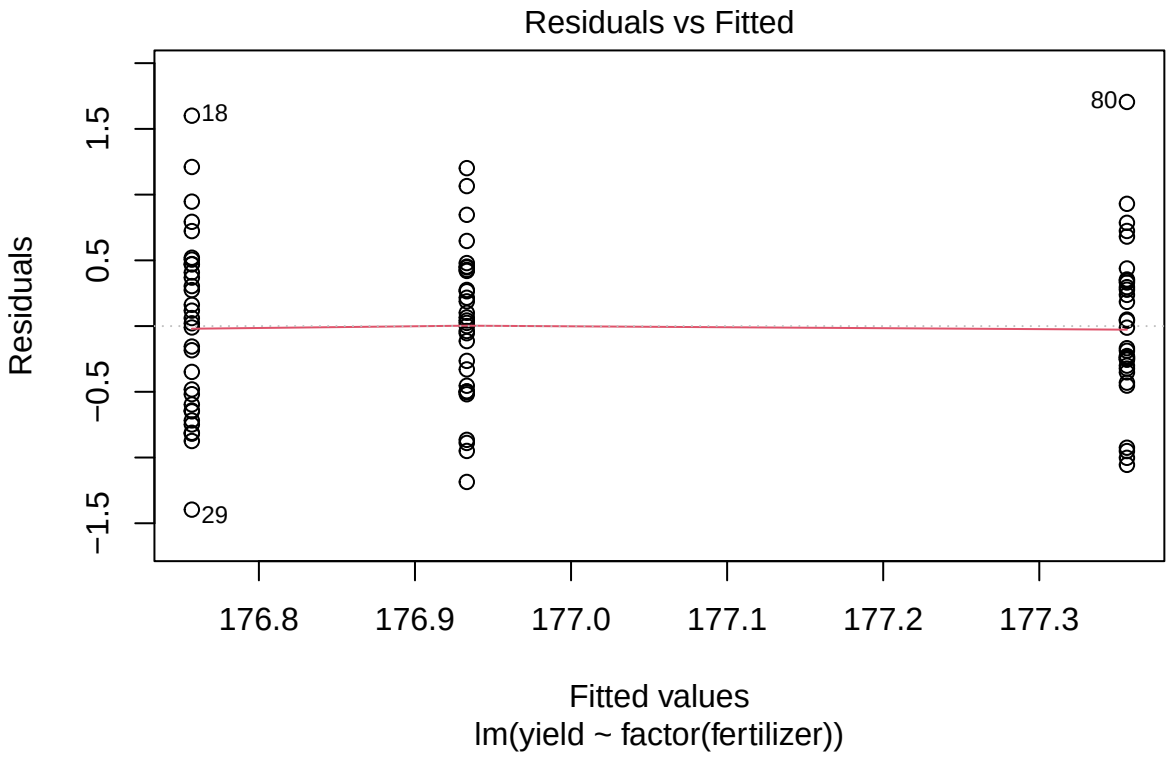
```
ggqqplot(my_data, "yield", facet.by = "fertilizer", title = "Q-Q Plots by Fertilizer Group")
```



Homogeneity of Variances

Residual Plot

```
plot(model, which = 1)
```



Levene's Test

```
my_data %>%
  levene_test(yield ~ factor(fertilizer))
```

```
## # A tibble: 1 x 4
##     df1   df2 statistic     p
##   <int> <int>     <dbl> <dbl>
## 1     2    93     0.847 0.432
```

Conclusion

This analysis used a one-way ANOVA to test for yield differences across fertilizer groups. Diagnostic checks confirmed key assumptions (normality, homogeneity). The post-hoc Tukey test identified which specific fertilizer types differed significantly in mean yield.

Created by Statisticians World — August 22, 2024