M9 Problem Set ANOVA for TAs

Pablo E. Gutiérrez-Fonseca

2024-04-02 16:37:11

We conducted an experiment to investigate the effect of different fertilizers on crop yield. Using a controlled environment, we divided a designated cultivation area into three equal plots. Each plot was assigned to one of three fertilizer treatments: Nitrogen, Phosphorus, and Potassium. The application of fertilizers followed recommended dosages to ensure uniformity across plots. Subsequently, we planted the same crop in each plot and maintained consistent agronomic practices throughout the growth period. This setup allowed us to directly assess any disparities in crop yield attributable to the variable under scrutiny, namely the type of fertilizer used.

We hypothesize that the type of fertilizer applied will have a significant influence on crop yield. Specifically, we expect that there will be differences in crop yield among the three fertilizer treatments: Nitrogen, Phosphorus, and Potassium.

#1. Import libraries and load packages

```
library(tidyverse)
library(dplyr)
library(readxl)
```

#2. Importing our data

```
crop.data <- read.csv(file = "crop.data.csv", header = TRUE)
```

#.Normality

```
#normality
by(crop.data$yield, crop.data$fertilizer, shapiro.test)
```

```
## crop.data$fertilizer: Nitrogen
##
##  Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.97914, p-value = 0.7743
##
## ------------------------------------------------------------
## crop.data$fertilizer: Phosphorus
##
##  Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.98329, p-value = 0.8875
```

```
##
## ------------------------------------------------------------
## crop.data$fertilizer: Potasium
##
##  Shapiro-Wilk normality test
##
## data:  dd[x, ]
## W = 0.95878, p-value = 0.2542
```

#Variance

```
#variance
bartlett.test(yield ~ fertilizer, data=crop.data)
```

```
##
##  Bartlett test of homogeneity of variances
##
## data:  yield by fertilizer
## Bartlett's K-squared = 1.0622, df = 2, p-value = 0.5879
```

#3. Run a Simple Linear Regresion

```
one.way <- aov(yield ~ fertilizer, data = crop.data)
summary(one.way)
```

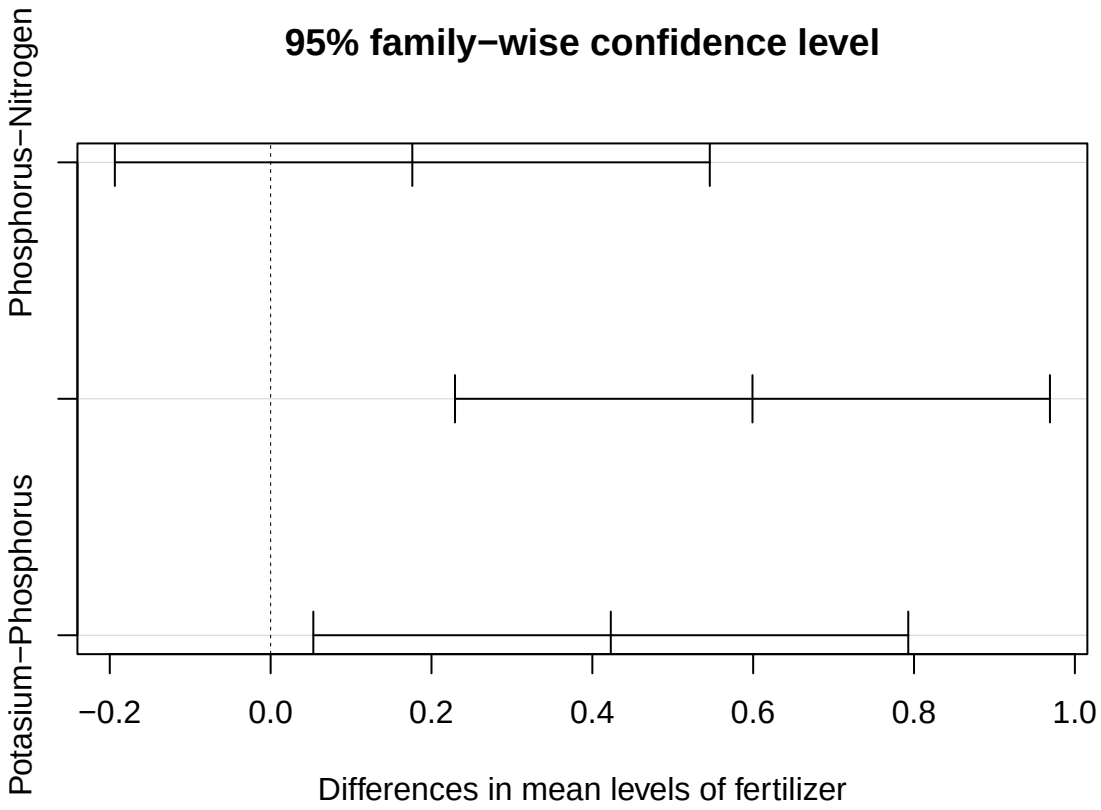
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer   2   6.07  3.0340   7.863  7e-04 ***
## Residuals   93  35.89  0.3859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

#4. Perform a Tukey's Honestly Significant Difference (Tukey's HSD) post-hoc test for pairwise comparisons

```
TukeyHSD(one.way)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = yield ~ fertilizer, data = crop.data)
##
## $fertilizer
##                           diff         lwr       upr     p adj
## Phosphorus-Nitrogen 0.1761687 -0.19371896 0.5460564 0.4954705
## Potasium-Nitrogen   0.5991256  0.22923789 0.9690133 0.0006125
## Potasium-Phosphorus 0.4229568  0.05306916 0.7928445 0.0208735
```

```
plot(TukeyHSD(one.way))
```



#5. Tukey Interpretation. From the post-hoc test results, we see that there are statistically significant differences ($p < 0.05$) between fertilizer groups 3 and 1 and between fertilizer types 3 and 2, but the difference between fertilizer groups 2 and 1 is not statistically significant.

Reference https://www.scribbr.com/statistics/anova-in-r/