M9 Problem Set ANOVA for TAs

Pablo E. Gutiérrez-Fonseca

2024-04-02 17:31:45

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. Are your data normally distributed?
2. Does your data have equal variance?
3. What is the R-squared value for the ANOVA analysis in the context of your hypothesis regarding the influence of different fertilizer types (Nitrogen, Phosphorus, and Potassium) on crop yield?
4. What would the adjusted alpha be for the pairwise comparison when applying the Bonferroni correction to adjust the threshold?
5. Summarize: Write a concise one paragraph summary of this analysis. Remember that any summary should include the following:

a. Statement of the research hypothesis or study objectives.

b. Brief summary of methods (one sentence or less).

c. Statement of the statistical results (including type of test and shorthand: F(Bdf,Wdf)= obtained value, p = 0.xxx, r^2).

d. Description of any differences, if meaningful, along with an interpretation of why these results make sense (or don't make sense).

The aim of our study was to assess the influence of different fertilizers on crop yield, with a focus on Nitrogen, Phosphorus, and Potassium treatments. Using a one-way ANOVA, we found a significant effect of fertilizer type on crop yield ($F(2, 93) = 7.863$, $p < 0.001$, r^2 = 0.1411). Post-hoc analysis with Tukey's HSD revealed that crop yield with Potassium fertilizer was significantly higher than with Nitrogen ($p = 0.0006$) and Phosphorus ($p = 0.0209$) fertilizers. However, there was no significant difference in yield between Nitrogen and Phosphorus treatments ($p = 0.4955$). These results underscore the importance of fertilizer choice, suggesting that Potassium application may lead to superior crop yields compared to other fertilizer types. Such findings emphasize the potential impact of fertilizer management practices on agricultural productivity. The R^2 value of 0.1411 suggests that approximately 14.11% of the variability in crop yield can be explained by the type of fertilizer used, indicating a relatively low but still meaningful effect size.

#1. Import libraries and load packages

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

#2. Importing our data

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

#.Normality

```r
#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

```r
#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

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

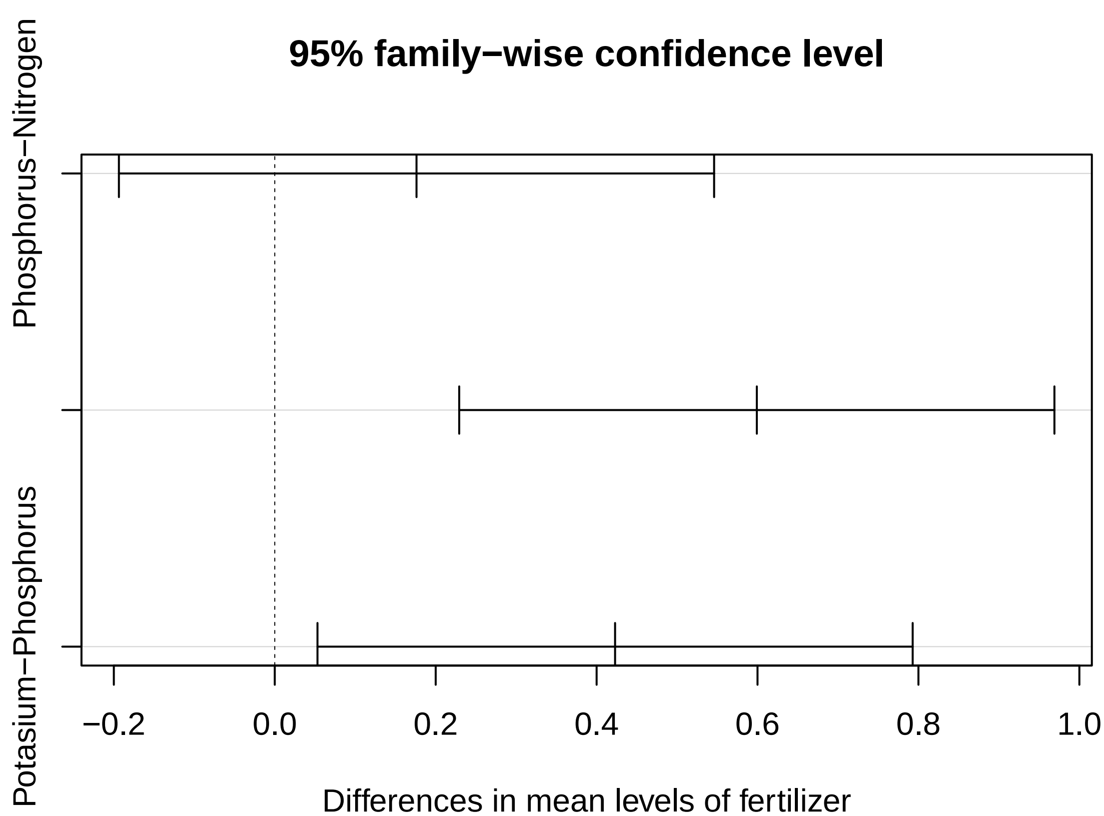
```
##             Df Sum Sq Mean Sq F value Pr(>F)
## fertilizer   2   6.07  3.0340   7.863  7e-04 ***
## Residuals   93  35.89  0.3859
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

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

```r
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
```

```r
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.

```r
#But we don't know WHICH species are different from each other: here is that
pairwise.t.test(crop.data$yield,  crop.data$fertilizer,
                p.adjust.method = "bonferroni")
```

```
##
##  Pairwise comparisons using t tests with pooled SD
##
## data:  crop.data$yield and crop.data$fertilizer
##
##            Nitrogen Phosphorus
## Phosphorus 0.77863  -
## Potasium   0.00063  0.02314
##
## P value adjustment method: bonferroni
```

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