Assignment 3

STAT3373

Isaac Shoebottom

Oct 2nd, 2025

```
library(tidyverse)
library(broom)
library(car)
library(effectsize)
library(knitr)
library(kableExtra)
```

```
# Create the fertilizer dataset
fertilizer_data <- data.frame(
  yield = c(4.2, 4.5, 4.1, 4.8, 4.3,  # Fertilizer A
            5.1, 5.3, 4.9, 5.2, 5.0,  # Fertilizer B
            3.8, 4.0, 3.6, 4.2, 3.9,  # Fertilizer C
            5.5, 5.8, 5.4, 5.7, 5.6), # Fertilizer D
  fertilizer = factor(rep(c("A", "B", "C", "D"), each = 5))
)

kable(fertilizer_data, caption = "Tomato Yield Data") %>%
  kable_styling(bootstrap_options = c("striped", "hover"))
```

Table 1: Tomato Yield Data

yield	fertilizer
4.2	A
4.5	A
4.1	A
4.8	A
4.3	A
5.1	B
5.3	B
4.9	B
5.2	B
5.0	B
3.8	C
4.0	C
3.6	C
4.2	C
3.9	C
5.5	D
5.8	D

5.4	D
5.7	D
5.6	D

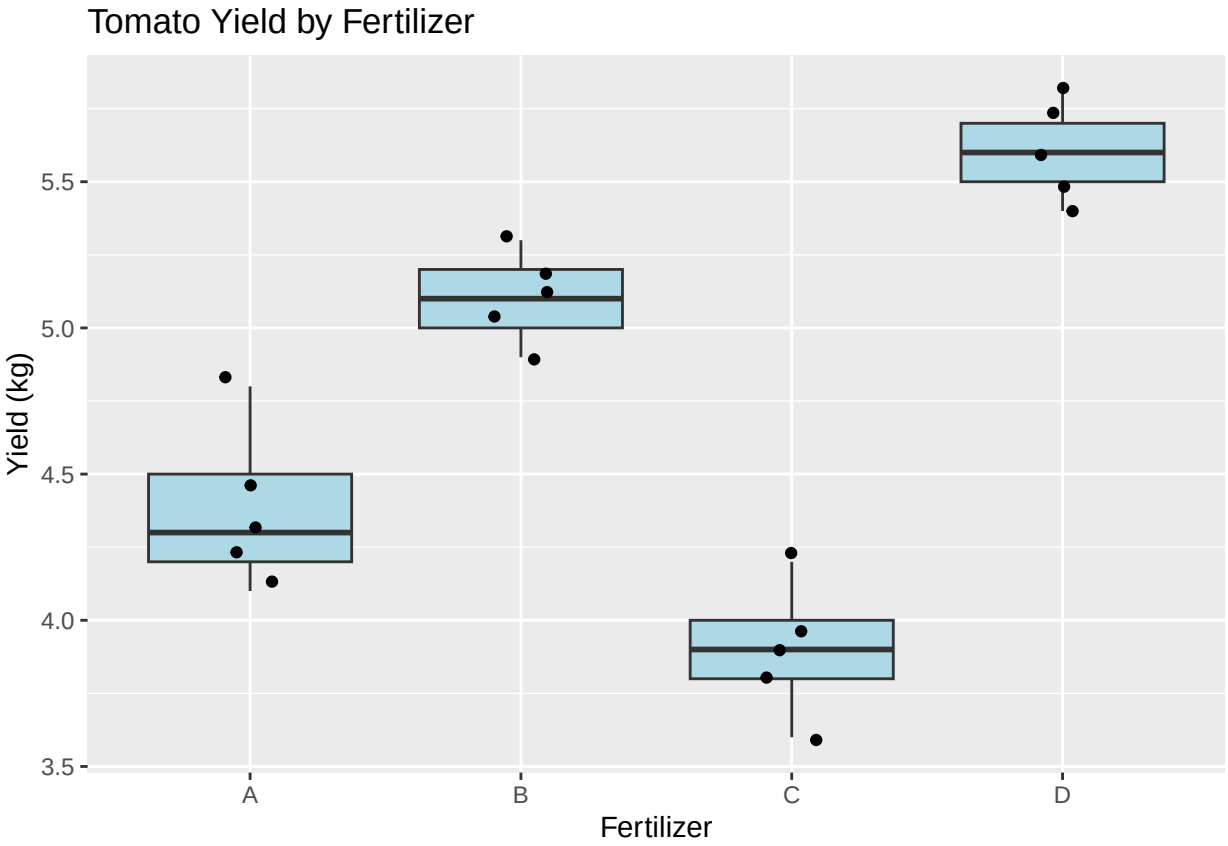
```
fertilizer_data %>%
  group_by(fertilizer) %>%
  summarise(
    mean = mean(yield),
    sd = sd(yield),
    n = n()
  )
```

```
## # A tibble: 4 x 4
##   fertilizer  mean    sd     n
##   <fct>      <dbl> <dbl> <int>
## 1 A           4.38 0.277     5
## 2 B           5.1  0.158     5
## 3 C           3.9  0.224     5
## 4 D           5.6  0.158     5
```

Question 1

a)

```
ggplot(fertilizer_data, aes(x = fertilizer, y = yield)) +
  geom_boxplot(fill = "lightblue") +
  geom_jitter(width = 0.1) +
  labs(title = "Tomato Yield by Fertilizer",
       y = "Yield (kg)",
       x = "Fertilizer")
```



Analysis:

- Fertilizer D appears to produce the highest yields.
- Fertilizer C appears lowest.
- Variability is similar across groups.

b)

```
fertilizer_aov <- aov(yield ~ fertilizer, data = fertilizer_data)
summary(fertilizer_aov)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## fertilizer   3  8.522  2.8405   64.19 3.86e-09 ***
## Residuals   16  0.708  0.0443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value < 0.05, reject H_0. There is a statistically significant difference in mean yield among fertilizers

Question 2

a)

```
teaching_data <- tibble(
  score = c(78,82,75,80,77,83,79,
            85,88,84,87,86,90,83,
```

```
            81,79,83,85,82,78,80),
  method = factor(rep(c("Traditional", "Interactive", "Online"), each = 7))
)
```

b)

```
teaching_aov <- aov(score ~ method, data = teaching_data)
summary(teaching_aov)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## method       2  182.0   91.00   14.05 0.000211 ***
## Residuals   18  116.6    6.48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value < 0.05, reject H_0, significant differences exist.

Question 3

```
supplier_data <- tibble(
  strength = c(245,250,248,252,249,247,
               240,238,242,241,239,243,
               255,258,254,257,256,253,
               248,246,250,249,247,251),
  supplier = factor(rep(paste("Supplier", 1:4), each = 6))
)

supplier_aov <- aov(strength ~ supplier, data = supplier_data)
summary(supplier_aov)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## supplier     3  676.5   225.5      55 7.66e-10 ***
## Residuals   20   82.0     4.1
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

p-value < 0.01, reject H_0, significant differences exist