Augusta University: STAT 7630

Applied Linear Models

Dustin Pluta

2024 JAN 11

Lecture 2

- Review of Confidence Intervals
- One-way Analysis of Variance (ANOVA)

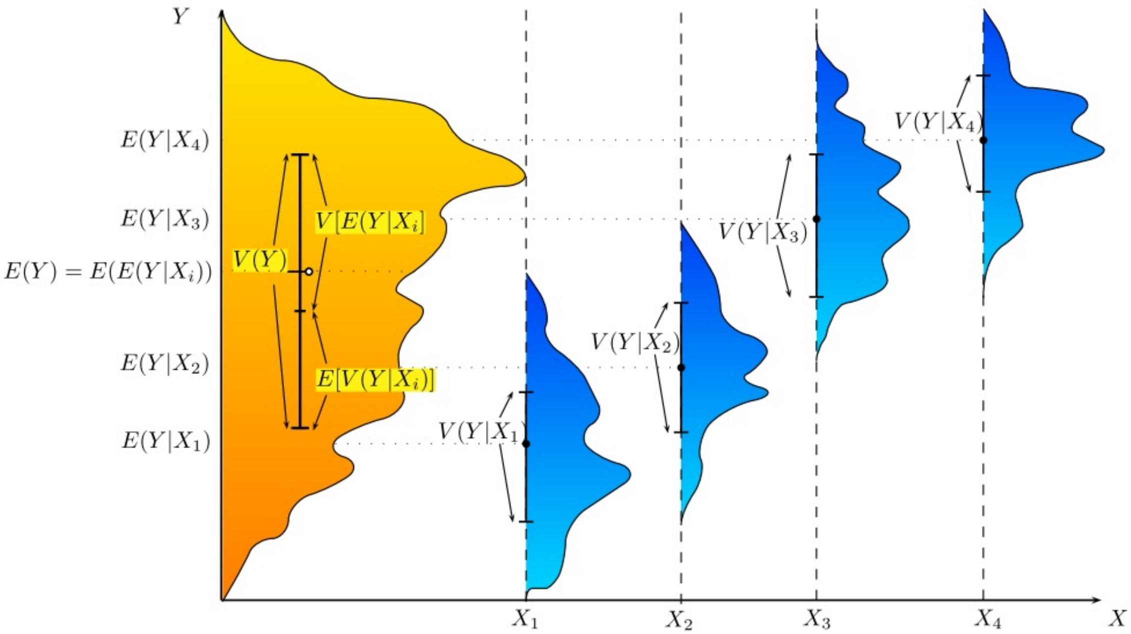


Figure 1: ANOVA : Fair fit

Two-sample z-test

Let $X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \dots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$.

We wish to test

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

If σ^2 is **known**, we can use the two-sample z-test, which has test statistic

$$T(X) = \frac{\overline{X} - \overline{Y}}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \overset{H_0}{\sim} \mathcal{N}(0, 1).$$

Two-sample t-test

Let $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\mu_X, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{iid}{\sim} \mathcal{N}(\mu_Y, \sigma^2)$.

We wish to test

$$H_0 : \mu_X = \mu_Y$$
$$H_1 : \mu_X \neq \mu_Y$$

If σ^2 is **unknown**, we instead use the two-sample t-test, with pooled variance estimator s_p calculated as

$$s_p^2 = \frac{(n-1)s_X^2 + (m-1)s_Y^2}{m+n-2}.$$

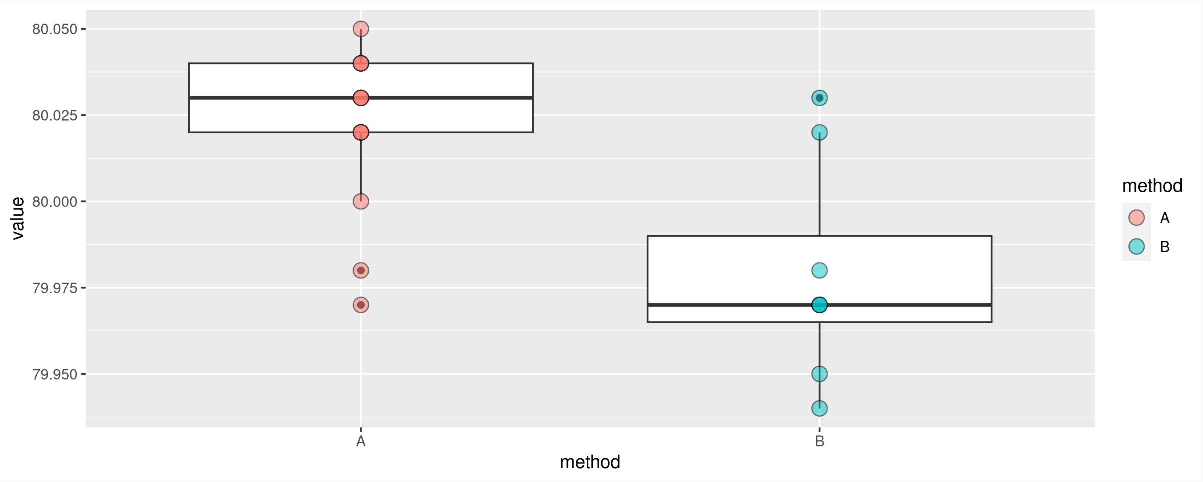
The test statistic is

$$T(X) = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{\frac{1}{n} + \frac{1}{m}}} \overset{H_0}{\sim} t(m+n-2).$$

Two-sample t-test

Example A (Rice, p423)

Two methods, A and B, were used in a determination of the latent heat of fusion of ice (Matrella 1963). The investigators wanted to find out by how much the methods differed. The data give the change in total heat from ice at $-0.72^{\circ}\mathrm{C}$ to water $0^{\circ}\mathrm{C}$ in calories per gram of mass:



Two-sample t-test

Example A

```
A ← c(79.98, 80.04, 80.02, 80.04, 80.03, 80.03, 80.04, 79.97, 80.05, 80.03, 80.02, 80.00,
B ← c(80.02, 79.94, 79.98, 79.97, 79.97, 80.03, 79.95, 79.97)
n ← length(A)
m ← length(B)
dat ← data.frame(value = c(A, B), method = c(rep("A", n), rep("B", m)))
dat
```

```
##    value method
## 1  79.98      A
## 2  80.04      A
## 3  80.02      A
## 4  80.04      A
## 5  80.03      A
## 6  80.03      A
## 7  80.04      A
## 8  79.97      A
## 9  80.05      A
## 10 80.03      A
## 11 80.02      A
## 12 80.00      A
## 13 80.02      A
## 14 80.02      B
```

Two-sample t-test

Example A

```
library(dplyr)
dat_summary <- dat %>% group_by(method) %>%
  summarize(mean = mean(value), sd = sd(value))
```

method	mean	sd
A	80.02077	0.0239658
B	79.97875	0.0313676

Two-sample t-test

Example A

```
s_p <- sqrt(((n - 1) * dat_summary$sd[1]^2 +
          (m - 1) * dat_summary$sd[2]^2) / (n + m - 2))
print(s_p)

## [1] 0.02693052

test_stat <- (mean(A) - mean(B)) / (s_p * sqrt(1 / n + 1 / m))
print(test_stat)

## [1] 3.472245

## [1] 0.002551004
```

Two-sample t-test

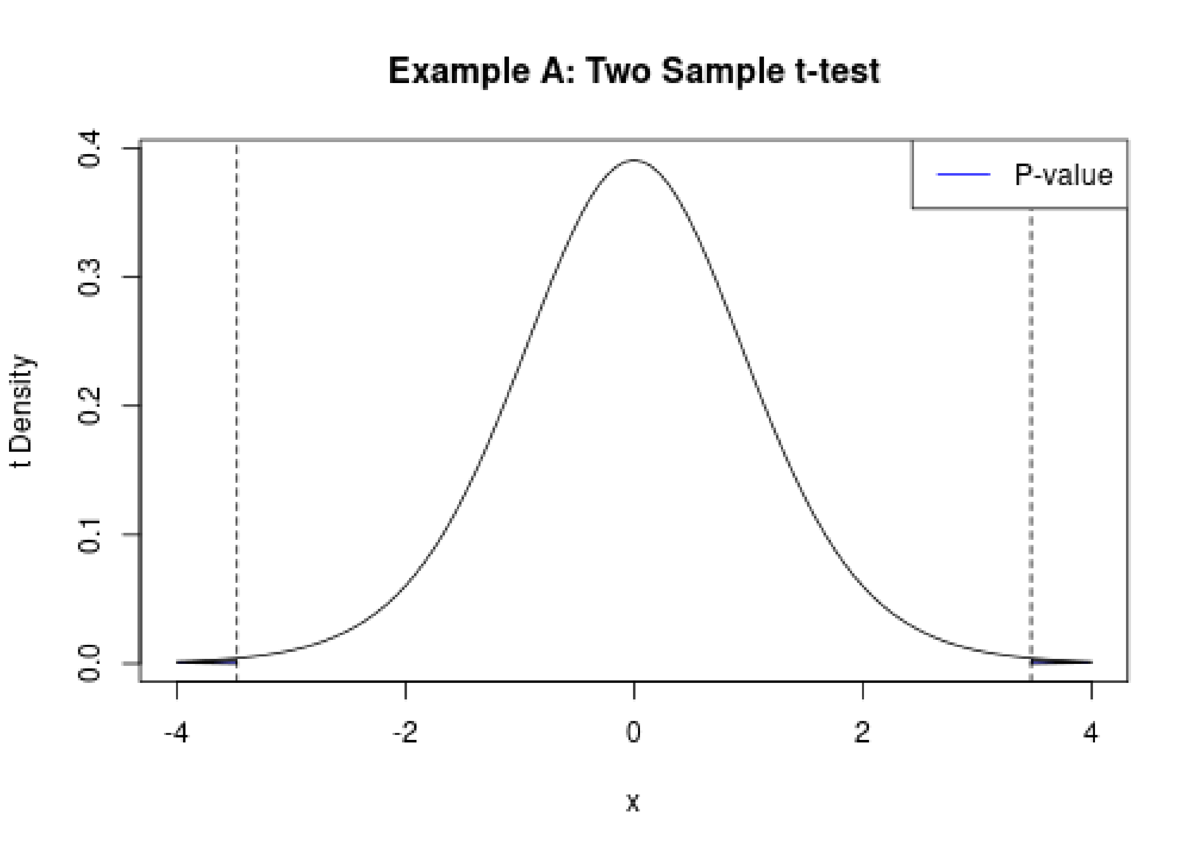
Example A

Using the built-in `R` function:

```
##
##     Two Sample t-test
##
## data:  A and B
## t = 3.4722, df = 19, p-value = 0.002551
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01669058 0.06734788
## sample estimates:
## mean of x mean of y
##  80.02077  79.97875
```

Two-sample t-test

Example A

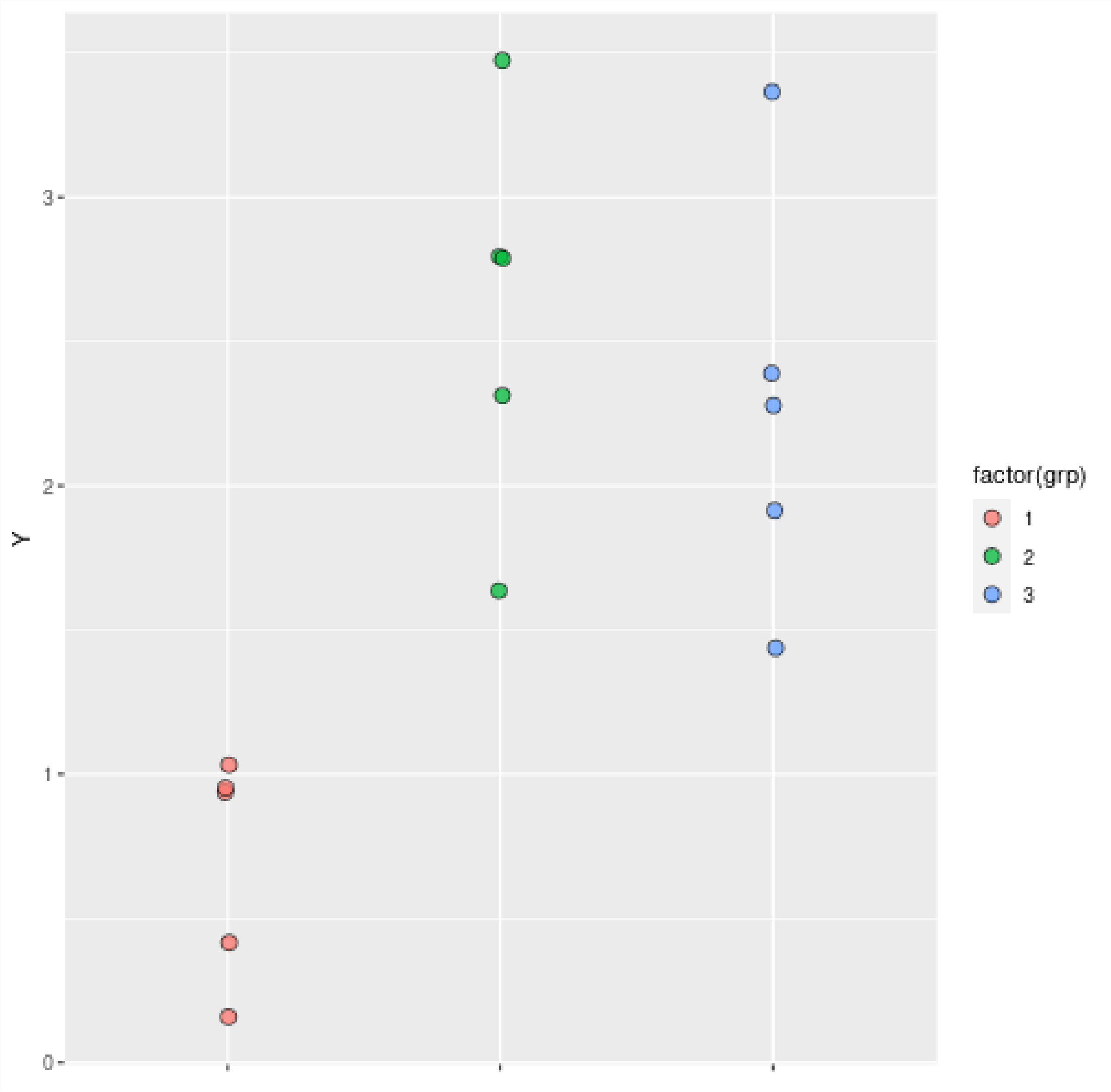


One-way ANOVA

```
I ← 3
J ← 5
mu ← 1
alpha1 ← -0.5
alpha2 ← 2
alpha3 ← 1
sigma ← 0.75

set.seed(12345)
Y1 ← mu + alpha1 + rnorm(J, 0, sigma)
Y2 ← mu + alpha2 + rnorm(J, 0, sigma)
Y3 ← mu + alpha3 + rnorm(J, 0, sigma)
```

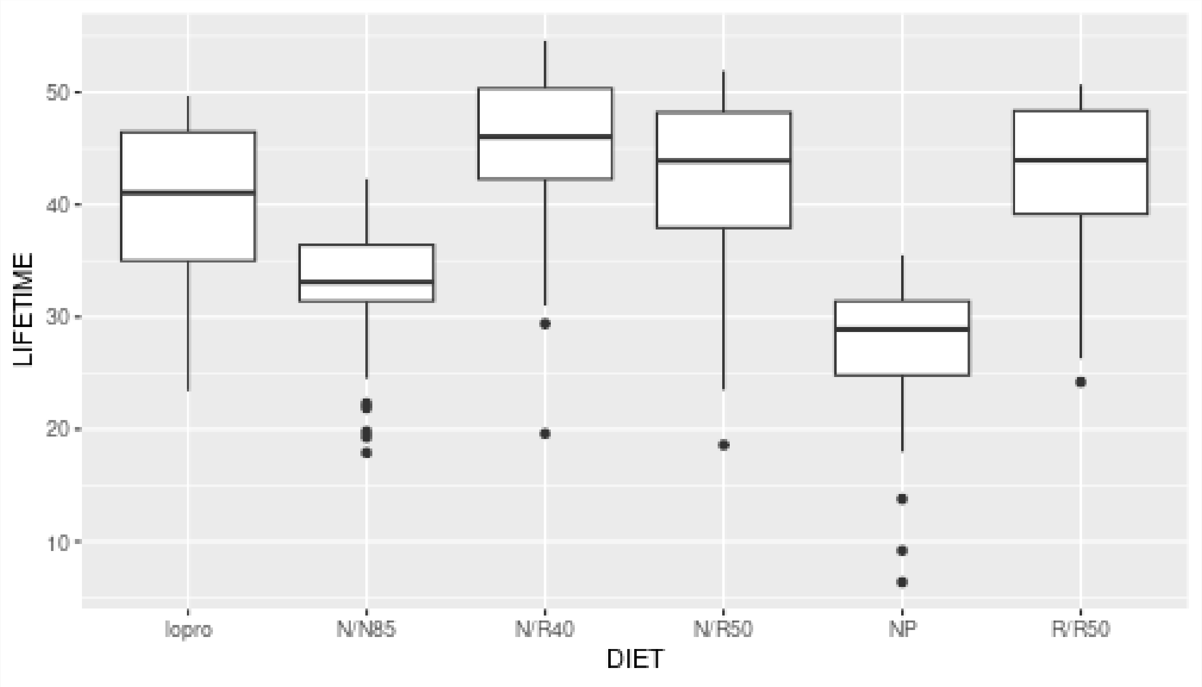
One-way ANOVA



One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

It is hypothesized that animals on restricted calorie diets have longer lifespans on average. A study by Weindruch et al. examined the effects of 6 different types of diets on lifespans in female mice.



One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

We wish to test whether there is a significant difference in the average lifespans across the treatment groups.

$$\begin{aligned} H_0 &: \text{All group mean lifespans are equal.} \\ H_1 &: \text{At least one group has a different mean lifespan.} \end{aligned}$$

One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

DIET	mean	sd
N/N85	32.69123	5.125297
N/R40	45.11667	6.703406
N/R50	42.29718	7.768195
NP	27.40204	6.133701
R/R50	42.88571	6.683152
lopro	39.68571	6.991695

One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

```
SSTOT <- sum((dat$LIFETIME - mean(dat$LIFETIME))^2)
SSTOT
```

```
## [1] 28031.36
```

```
SSW <- 0
for (subj in 1:n) {
  SSW <- SSW + (dat$LIFETIME[subj] - dat_summary$mean[which(dat_summary$DIET == dat$DIET[su
}
SSW
```

```
## [1] 15297.42
```

```
SSB <- SSTOT - SSW
SSB
```

```
## [1] 12733.94
```

One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

```
MSB ← SSB / (I - 1)
MSB
```

```
## [1] 2546.788
```

```
MSE ← SSW / (n - I)
MSE
```

```
## [1] 44.59888
```

One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

F statistic

```
test_stat <- MSB / MSE
test_stat
```

```
## [1] 57.10431
```

P-value

```
P_val <- pf(test_stat, df1 = I - 1, n - I, lower.tail = FALSE)
P_val
```

```
## [1] 4.111744e-43
```

One-way ANOVA

Diet and Longevity Study (12.5.33, Rice)

We wish to test whether there is a significant difference in the average lifespans across the treatment groups.

```
fit <- lm(LIFETIME ~ DIET, data = dat)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: LIFETIME
##            Df Sum Sq Mean Sq F value    Pr(>F)
## DIET        5  12734  2546.8  57.104 < 2.2e-16 ***
## Residuals 343  15297    44.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

One-way ANOVA

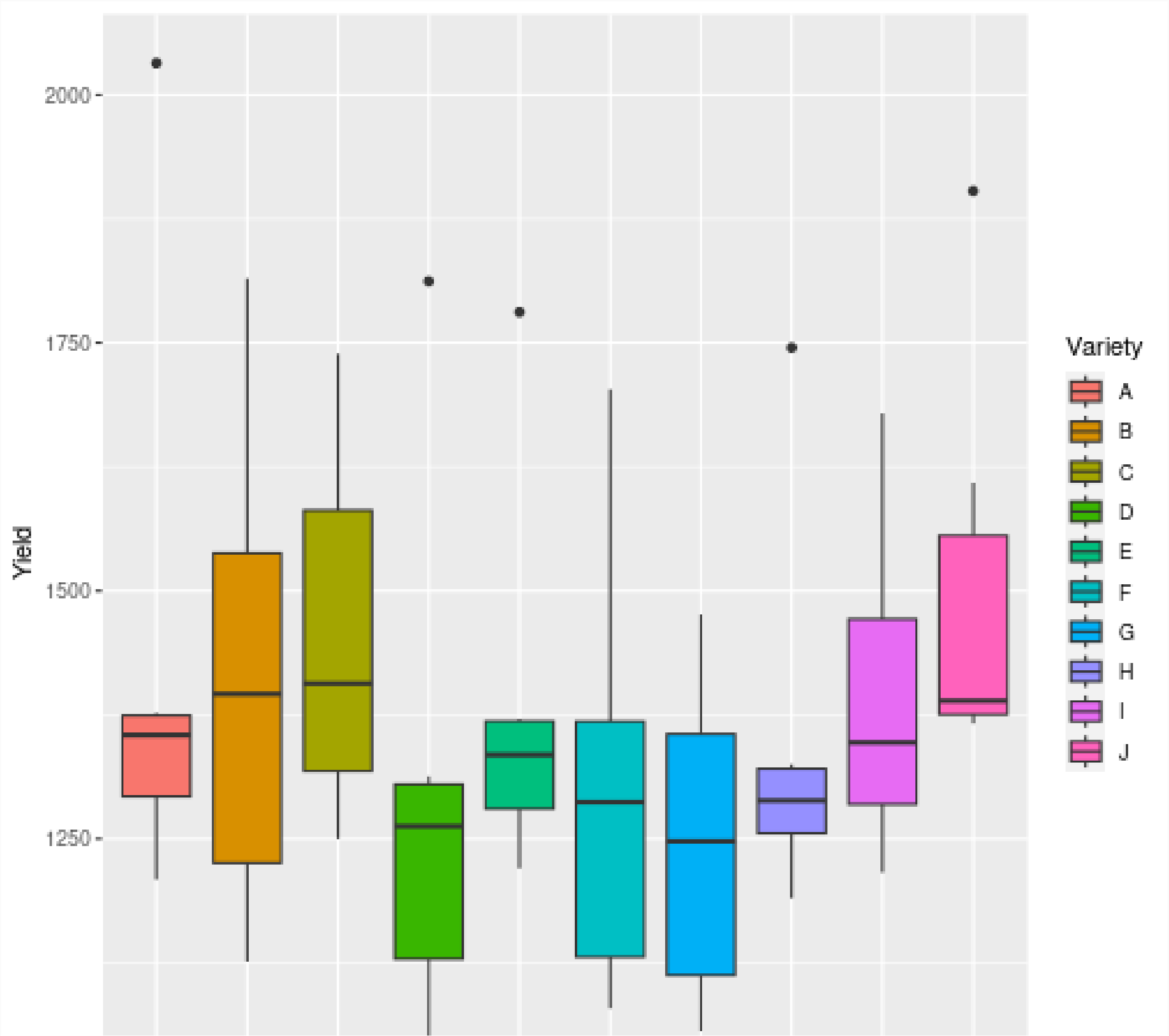
Linseed Experiment

```
## Rows: 60 Columns: 3
## ── Column specification ────────────────────────────────────────────────────────
## Delimiter: ","
## chr (1): Variety
## dbl (2): Plot, Yield
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
## # A tibble: 10 × 4
##    Variety     n  mean    sd
##    <chr>   <int> <dbl> <dbl>
##  1 A           6 1434.  300.
##  2 B           6 1414.  255.
##  3 C           6 1455   191.
##  4 D           6 1297.  274.
##  5 E           6 1385.  202.
##  6 F           6 1305   230.
##  7 G           6 1248.  167.
##  8 H           6 1348   200.
##  9 I           6 1394.  171.
## 10 J           6 1505.  216.
```

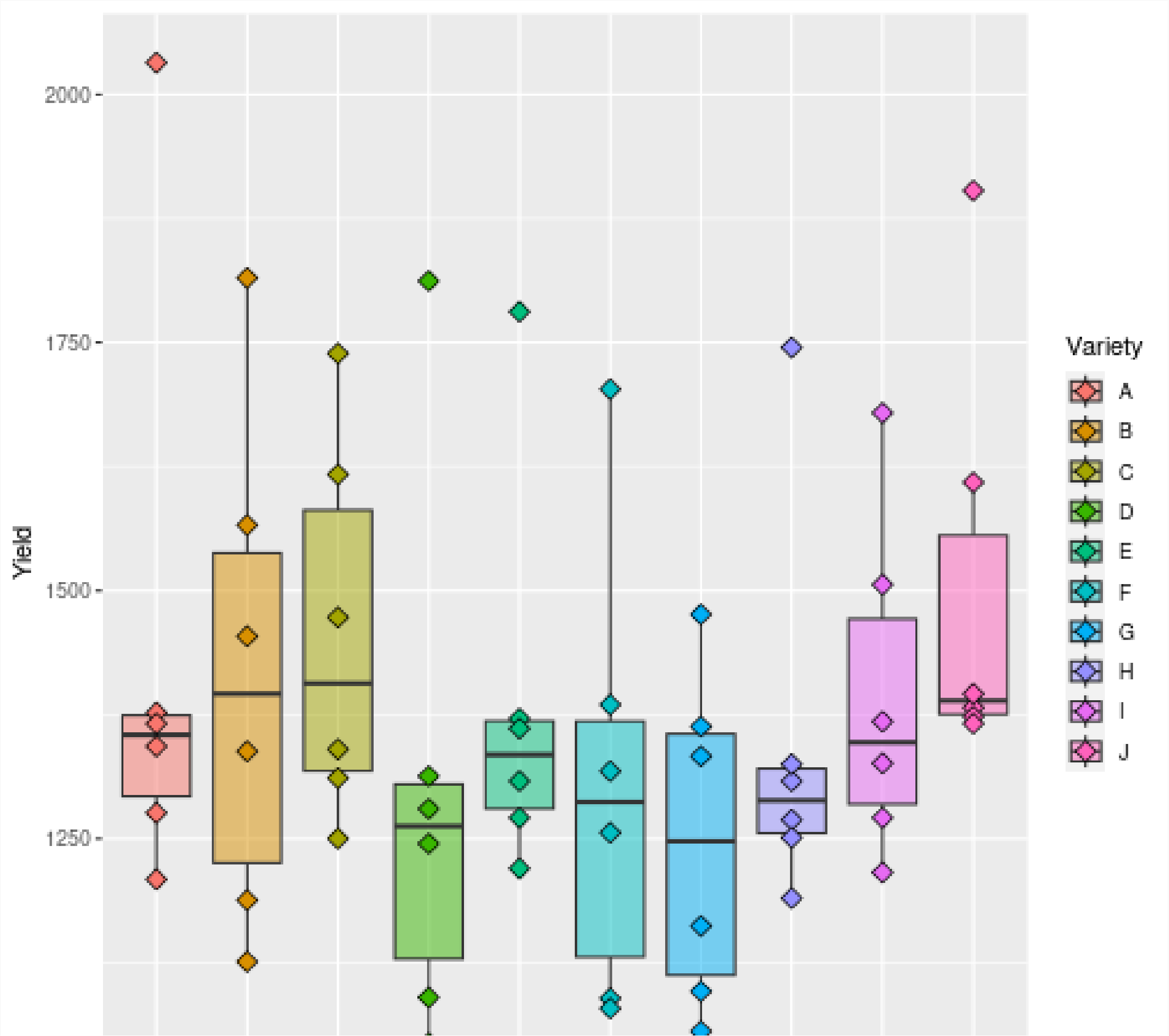
One-way ANOVA

Linseed Experiment



One-way ANOVA

Linseed Experiment



One-way ANOVA

Linseed Experiment

```
fit <- lm(Yield ~ Variety, data = dat)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: Yield
##           Df  Sum Sq Mean Sq F value Pr(>F)
## Variety    9  339032   37670  0.7471 0.6644
## Residuals 50 2521036   50421
```