

ANOVA Example — Treatment Systems and *E. coli*

Your Name

2025-10-04

Introduction

Water-treatment systems are designed to reduce microbial contamination in wastewater. In this study, we compare **four different water-management systems** for their effect on *E. coli* concentration (log CFU/100 mL) after treatment:

1. Raw Sewage
2. Treatment Plant
3. Treatment Wetland
4. Plant + Wetland

Each treatment was replicated 4 times (16 observations total).

Research Question

Do mean *E. coli* levels (log CFU/100 mL) differ among the four treatment systems?

Hypotheses

- **Null hypothesis (H):**

$$\mu_{Raw} = \mu_{Plant} = \mu_{Wetland} = \mu_{Plant+Wetland}$$

- **Alternative hypothesis (H):**

At least one group mean differs.

We will use a **one-way ANOVA** to test for differences among the treatment means.

Data

The data are simulated log-transformed *E. coli* concentrations for illustration.

ID	Treatment	Ecoli_LogCFU
1	Raw_Sewage	6.9
2	Raw_Sewage	6.7
3	Raw_Sewage	6.8

ID	Treatment	Ecoli_LogCFU
4	Raw_Sewage	7.0
5	Treatment_Plant	3.7
6	Treatment_Plant	3.9
7	Treatment_Plant	3.6
8	Treatment_Plant	3.8
9	Treatment_Wetland	2.6
10	Treatment_Wetland	2.4
11	Treatment_Wetland	2.7
12	Treatment_Wetland	2.5
13	Plant_and_Wetland	1.9
14	Plant_and_Wetland	2.0
15	Plant_and_Wetland	1.8
16	Plant_and_Wetland	2.1

ANOVA Analysis

Create the dataset

```

ID <- 1:16
Treatment <- c(rep("Raw_Sewage", 4),
               rep("Treatment_Plant", 4),
               rep("Treatment_Wetland", 4),
               rep("Plant_and_Wetland", 4))
Ecoli_LogCFU <- c(6.9, 6.7, 6.8, 7.0,
                   3.7, 3.9, 3.6, 3.8,
                   2.6, 2.4, 2.7, 2.5,
                   1.9, 2.0, 1.8, 2.1)

data <- data.frame(ID, Treatment, Ecoli_LogCFU)

# Display first few rows
head(data)

##   ID      Treatment Ecoli_LogCFU
## 1  1      Raw_Sewage     6.9
## 2  2      Raw_Sewage     6.7
## 3  3      Raw_Sewage     6.8
## 4  4      Raw_Sewage     7.0
## 5  5 Treatment_Plant    3.7
## 6  6 Treatment_Plant    3.9

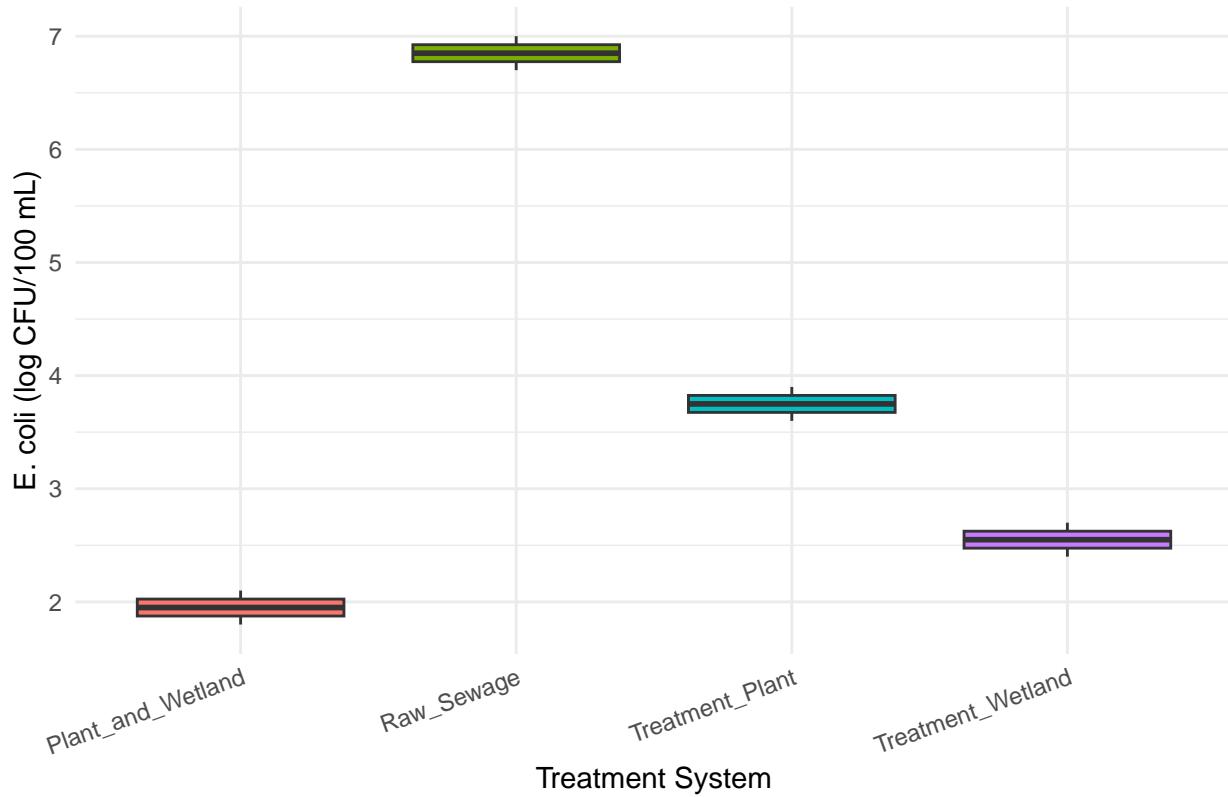
library(ggplot2)

ggplot(data, aes(x = Treatment, y = Ecoli_LogCFU, fill = Treatment)) +
  geom_boxplot() +
  labs(title = "E. coli (log CFU/100 mL) Across Treatment Systems",
       x = "Treatment System",
       y = "E. coli (log CFU/100 mL)") +
  theme_minimal() +

```

```
theme(legend.position = "none",
      axis.text.x = element_text(angle = 20, hjust = 1))
```

E. coli (log CFU/100 mL) Across Treatment Systems



```
anova_model <- aov(Ecoli_LogCFU ~ Treatment, data = data)
summary(anova_model)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## Treatment    3  57.15  19.050     1143 5.27e-15 ***
## Residuals   12   0.20   0.017
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```