

Water Quality Testing

Isa D'Amario-McShane

February 21, 2025

Escherichia coli (E. coli) is a bacterium commonly found in the intestines of humans and animals, where most strains are harmless and play a role in digestion. However, the presence of E. coli in water is a major public health concern, as it serves as an indicator of fecal contamination and the potential presence of harmful pathogens. Certain strains can cause serious illnesses, including diarrhea, urinary tract infections, and even kidney failure. Contaminated water sources—whether from agricultural runoff, sewage leaks, or stormwater overflows—pose risks not only to drinking water supplies but also to recreational waters and agricultural irrigation. Ingestion of water containing pathogenic E. coli can lead to widespread outbreaks, highlighting the importance of routine water quality monitoring and treatment. Effective detection methods, such as the Colilert test, help identify E. coli contamination, allowing for timely interventions to protect public health and prevent the spread of waterborne diseases.

So this week in class we took water samples from three different places across campus - pHake lake, Seal Court (my group) and Pomona Bioswale - to test it for exactly these reasons.

Skipping this section

```
# mlh None of this is my code, so I am going to take it offline...because I don't understand it...:-) t
data = read.csv(data.csv)

# view(data) # MLH I don't know what this is supposed to do, but it doesn't work for me so I commented
head(data)
colnames(data)
summary(data)

unique(data$Dilution)
sum(is.na(data$Dilution))
sum(is.na(data$Yellow_Large_Wells))
sum(is.na(data$Yellow_Small_Wells))
sum(is.na(data$Fluorescence_Large_Wells))

library(dplyr)
colnames(data) <- gsub(" ", "_", colnames(data))

MPN_data.csv = "/home/idjm2024/Water_Quality_Ecoli_NTM./Ecoli and Total Coliform Data Entry - Data Entry
MPN_data <- read.csv(MPN_data.csv)
```

Starting Here

NOTE: my guide says to download the clean up version...

```

# MLH not my path, you can remove comment.

# MPN_data.csv = "/home/idjm2024/Water_Quality_Ecoli_NTM./Ecoli and Total Coliform Data Entry - Data En

MPN_data.csv = "/home/mwl04747/RTricks/06_Ecoli_NTM/Ecoli and Total Coliform Data Entry - Data Export (
MPN_data <- read.csv(MPN_data.csv)

MPN_data <- MPN_data[!grepl("Control", MPN_data$Site_ID),]
MPN_data$TC_MPN = MPN_data$TotalColiform
MPN_data$EC_MPN = MPN_data$Ecoli

MPN_data$TC_MPN[MPN_data$Dilution=="Diluted, 1:10"] <-
MPN_data$TotalColiform[MPN_data$Dilution=="Diluted, 1:10"] *10
MPN_data$Ecoli <- as.numeric(MPN_data$Ecoli)
MPN_data$Ecoli[is.na(MPN_data$Ecoli)] <- 0
MPN_data$EC_MPN[MPN_data$Dilution == "Diluted, 1:10"] <- MPN_data$Ecoli[MPN_data$Dilution == "Diluted, 1:10"]
head(MPN_data)

```

```

##           Team      Site_ID Date      Dilution TotalColiform Ecoli TC_MPN
## 3      Salix      Seal Court      None          658.6    0.0  658.6
## 4      Salix      Seal Court      None          368.1    1.0  368.1
## 5      Salix      Seal Court      Diluted, 1:10      79.4    0.0  794.0
## 6      Salix      Seal Court      Diluted, 1:10     129.6    0.0 1296.0
## 9  Ceanothus  Pomona Bioswale      None          133.4    2.0  133.4
## 10 Ceanothus  Pomona Bioswale      None          533.5    9.8  533.5
##      EC_MPN
## 3      0.0
## 4      1.0
## 5      0.0
## 6      0.0
## 9      2.0
## 10     9.8

```

Not run... not sure what the goals are here, but I think not needed.

```

head(MPN_data$Collection.Location) # MLH -- not sure your goal here
unique(MPN_data$Collection.Location) # good info, but not needed...
# sum(is.na(MPN_data$Collection.Location)) # I have code to remove that...

MPN_data <- MPN_data[!is.na(MPN_data$Collection.Location) & MPN_data$Collection.Location != "", ]
MPN_data$Site_ID <- as.factor(MPN_data$Collection.Location)
levels(MPN_data$Site_ID)

MPN_data$Site_ID <- MPN_data$`Collection.Location`

```

back to good stuff...

```

aov_result <- aov(EC_MPN ~ Site_ID, data = MPN_data)
summary(aov_result)

```

```

##           Df Sum Sq Mean Sq F value    Pr(>F)

```

```
## Site_ID      4  517.5  129.38   8.193 0.00104 **
## Residuals    15  236.9   15.79
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 6 observations deleted due to missingness
```

```
aov_result_TC <- aov(TC_MPN ~ Site_ID, data = MPN_data)
summary(aov_result_TC)
```

```
##              Df    Sum Sq Mean Sq F value Pr(>F)
## Site_ID      4 27337227 6834307   1.692  0.204
## Residuals    15 60594759 4039651
## 6 observations deleted due to missingness
```

```
Df Sum Sq Mean Sq F value Pr(>F) Site_ID 2 114316 57158 5.095 0.0331* Residuals 9 100971 11219 —
Signif. codes:  0 ' ' 0.0001 ' ' 0.01 ' ' 0.05 ' ' 0.1 ' ' 1
```

Since the $\text{Pr}(>F)$ value (p-value) is less than 0.05 (as it's 0.03), it means that there is statistically significant evidence to reject the null hypothesis. The null hypothesis for the ANOVA test is that the means of the groups (in this case, the Site_ID locations) are all equal. A p-value less than 0.05 suggests that at least one of the sites has a significantly different mean for the EC_MPN (or TC_MPN) value, implying the locations are not the same with respect to the measured MPN (microbial contamination level).