

Homework3

PART 1:

Keough and Raimondi looked at the effects of four biofilm types on recruitment of serpulid larvae (see Logan page 268 for more info). Is recruitment different among the biofilm types?

```
# 1. Import the Keough dataframe  
library(readxl)  
library(car)
```

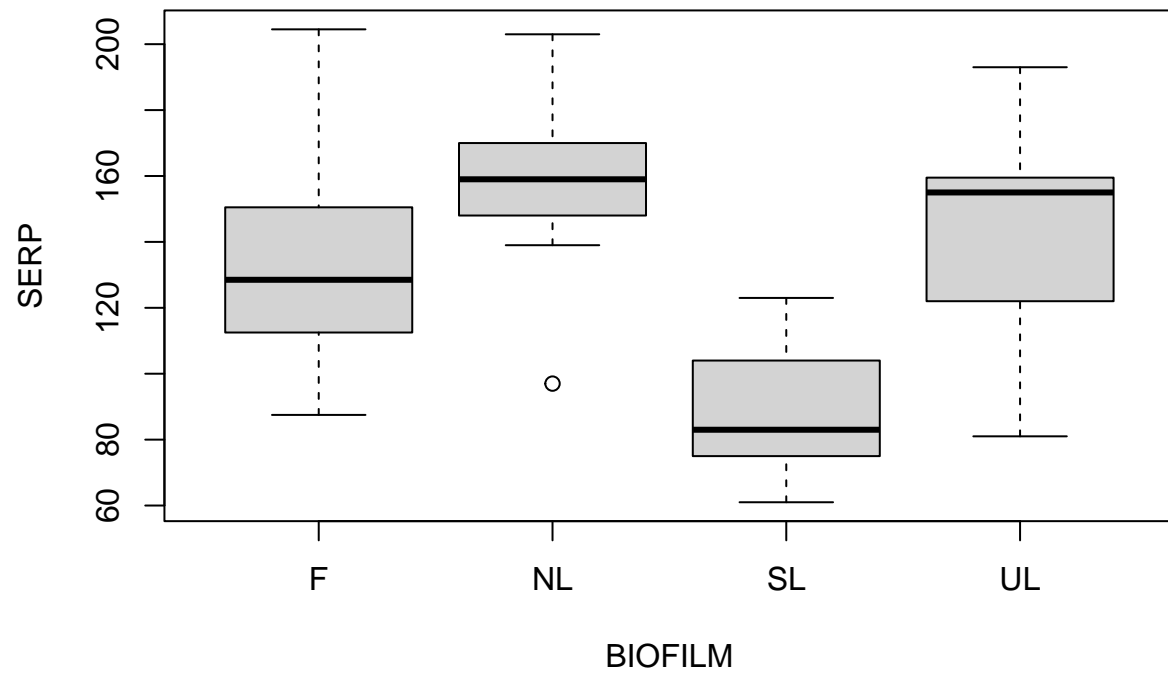
```
## Loading required package: carData
```

2. Understand how to formulate the Null and Alternate hypotheses

H0: The mean recruitment of serpulid larvae is equal across all four biofilm types.

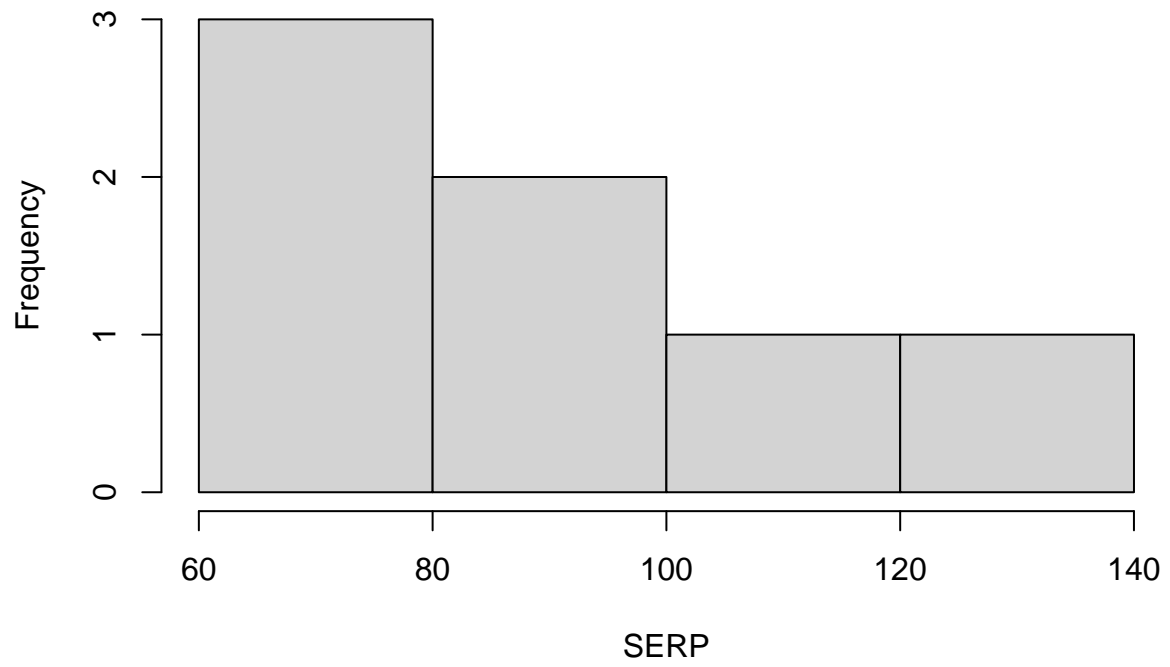
H1: At least one biofilm type has a mean recruitment that differs from the others.

```
# Read data  
data <- read_excel('/Users/kanoalindiwe/Downloads/Projects/playground/R/Quantitative Ecology/Datasets/k  
  
# 3. Check assumptions of ANOVA with boxplots and histogram and normality test  
# Test for normality  
boxplot(SERP ~ BIOFILM, data = data)
```

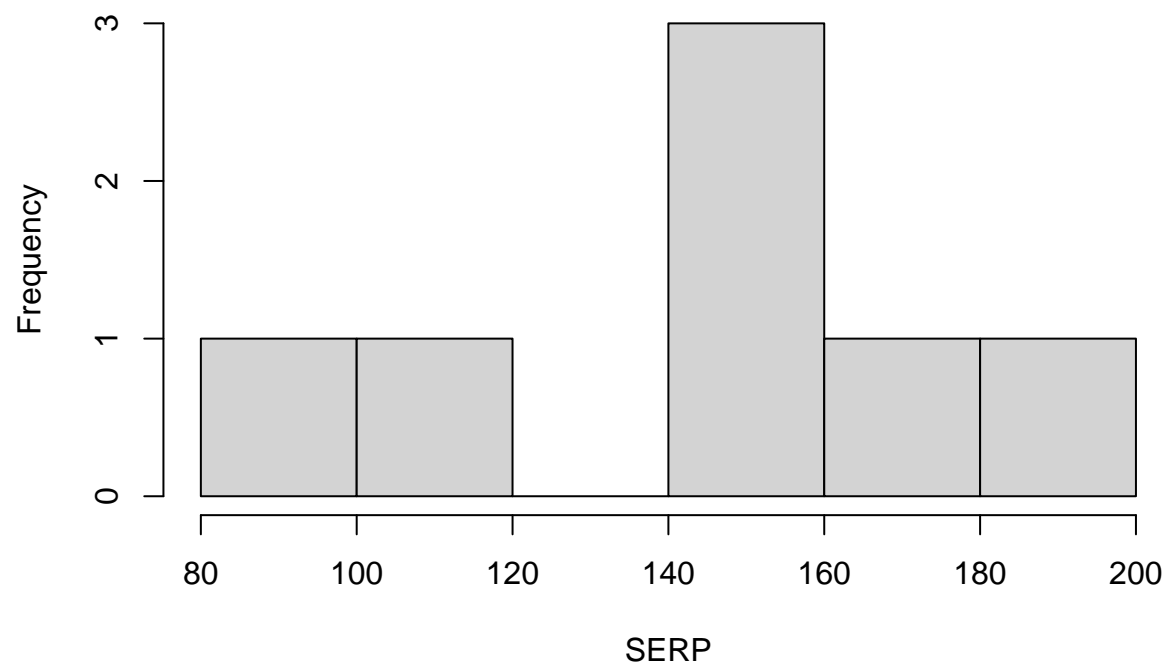


```
for (g in unique(data$BIOFILM)) {  
  hist(data$SERP[data$BIOFILM == g],  
    main = paste("BIOFILM =", g),  
    xlab = "SERP")  
}
```

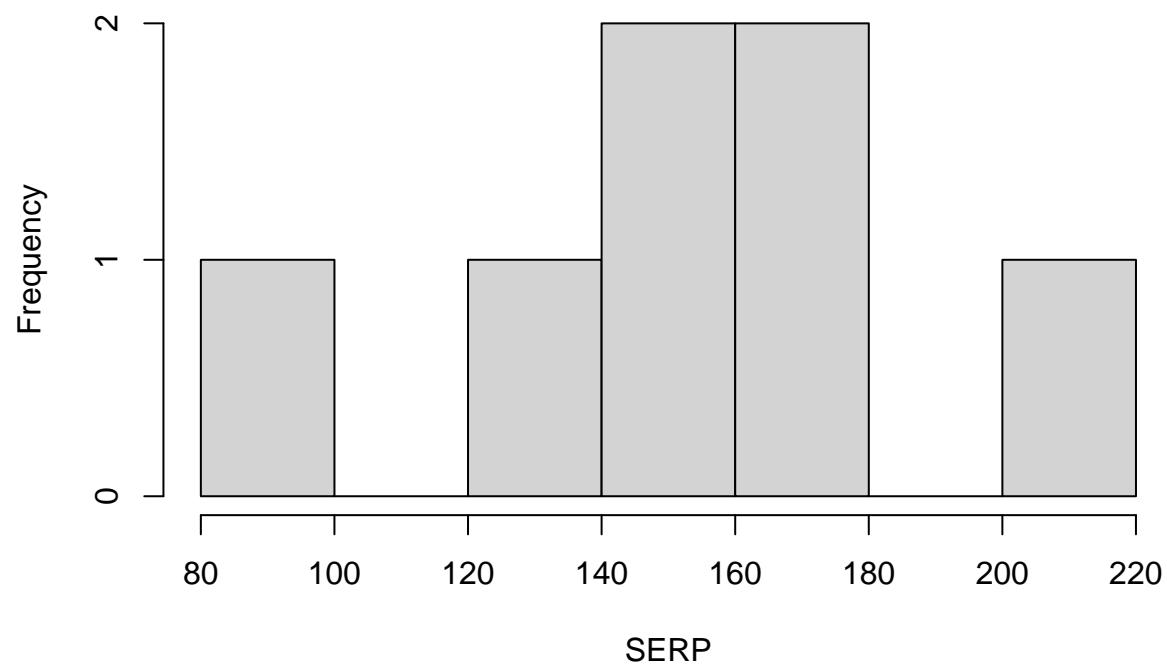
BIOFILM = SL

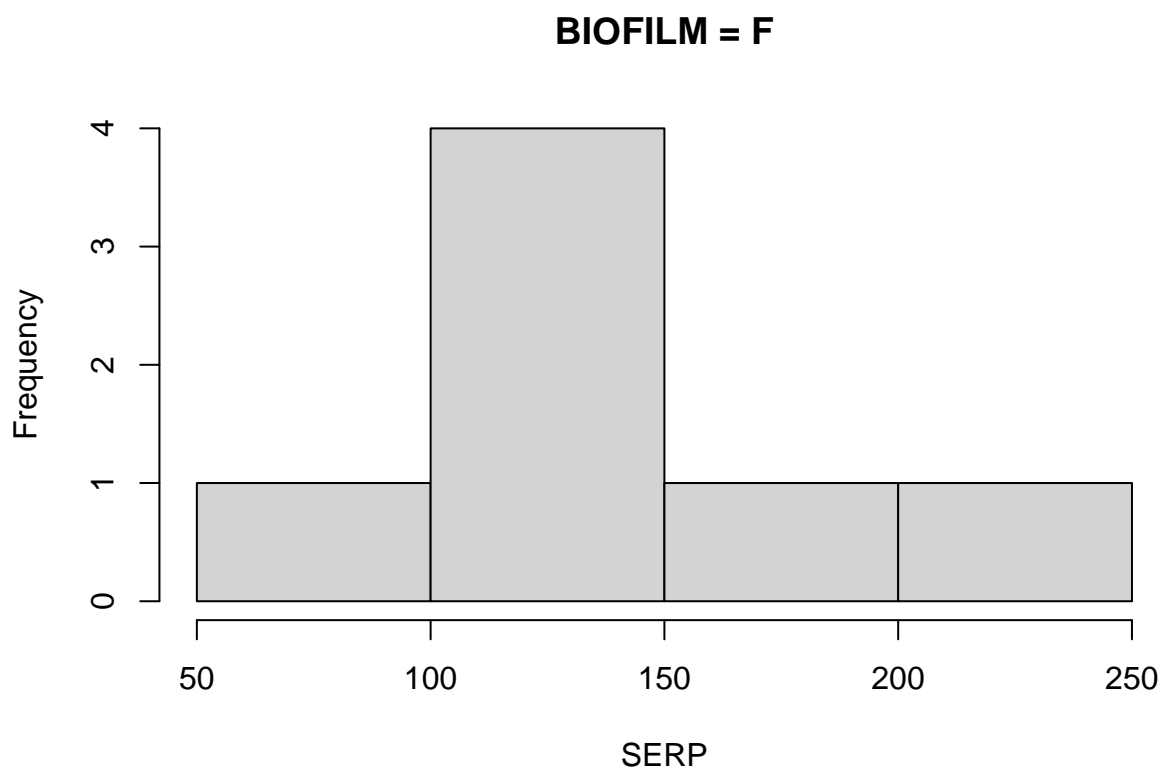


BIOFILM = UL



BIOFILM = NL



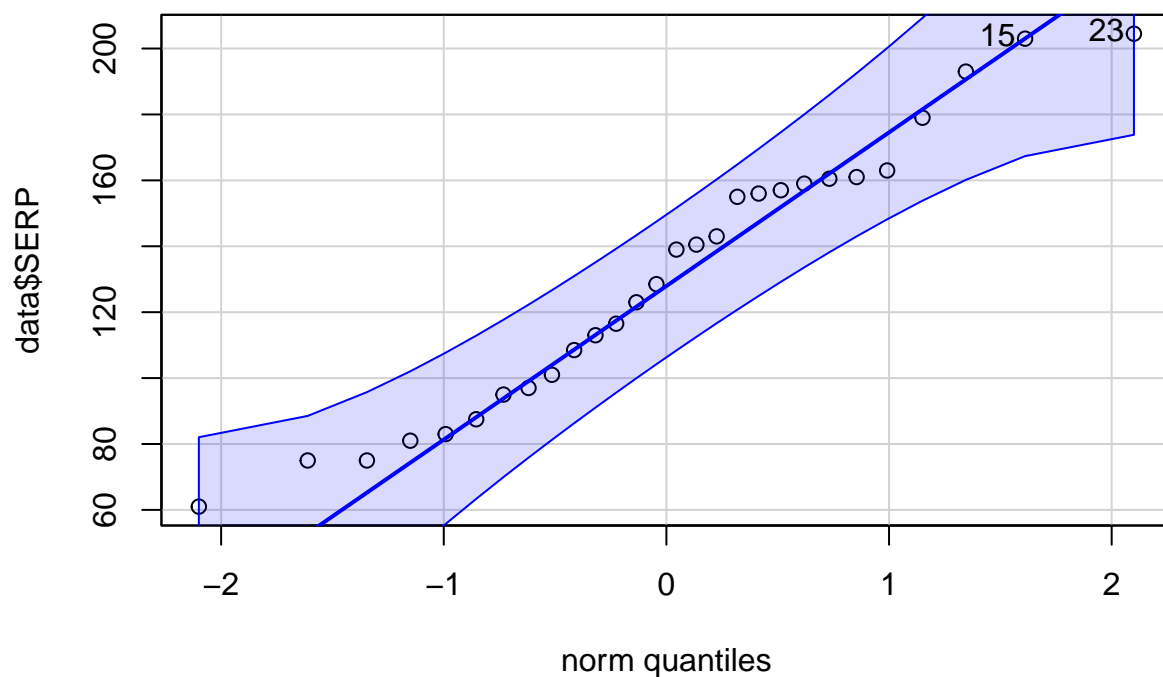


```
by(data$SERP, data$BIOFILM, function(x) shapiro.test(as.numeric(na.omit(x))))
```

```
## data$BIOFILM: F
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.957, p-value = 0.7926
##
## -----
## data$BIOFILM: NL
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.94763, p-value = 0.7081
##
## -----
## data$BIOFILM: SL
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.94173, p-value = 0.6543
##
```

```
## -----
## data$BIOFILM: UL
##
## Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.92786, p-value = 0.5329
```

```
qqPlot(data$SERP)
```



```
## [1] 23 15
```

```
# 4. Check assumptions for violations of unequal variances
# Test for equal variances
leveneTest(SERP ~ BIOFILM, data = data)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

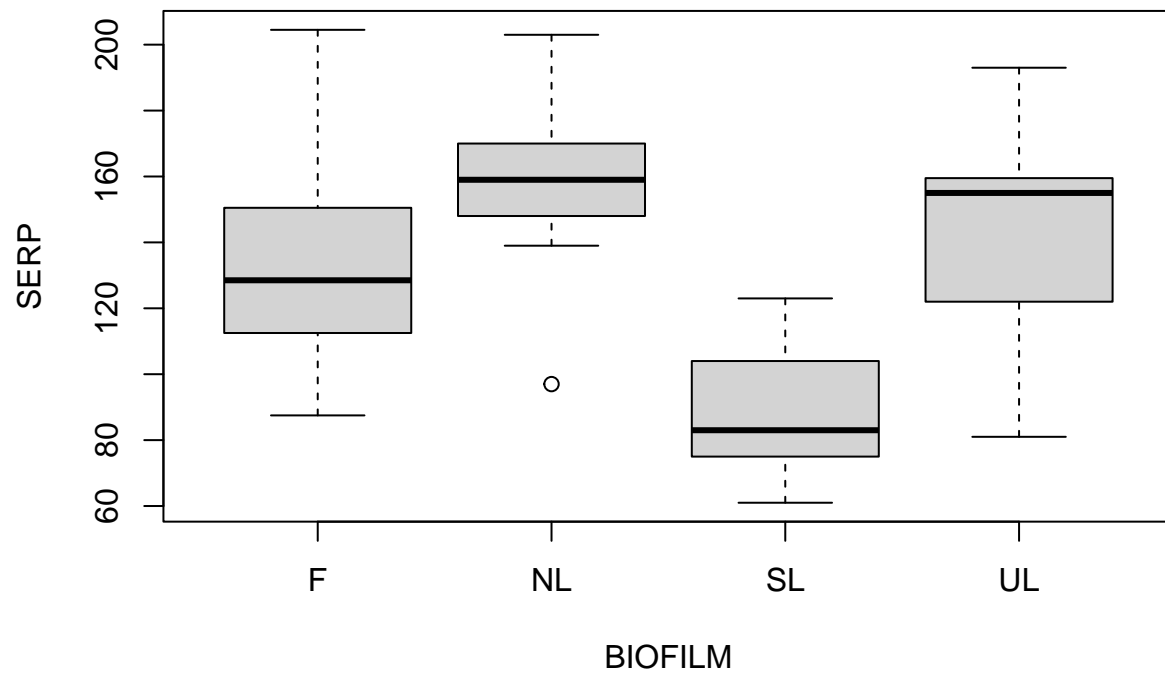
```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.2964 0.8276
##      24
```

```

# Log for normality
data$log_SERP <- log(data$SERP)

# 5. Transform data to improve normality and heteroscedasticity. Recheck assumptions.
# Test for normality again
boxplot(SERP ~ BIOFILM, data = data)

```

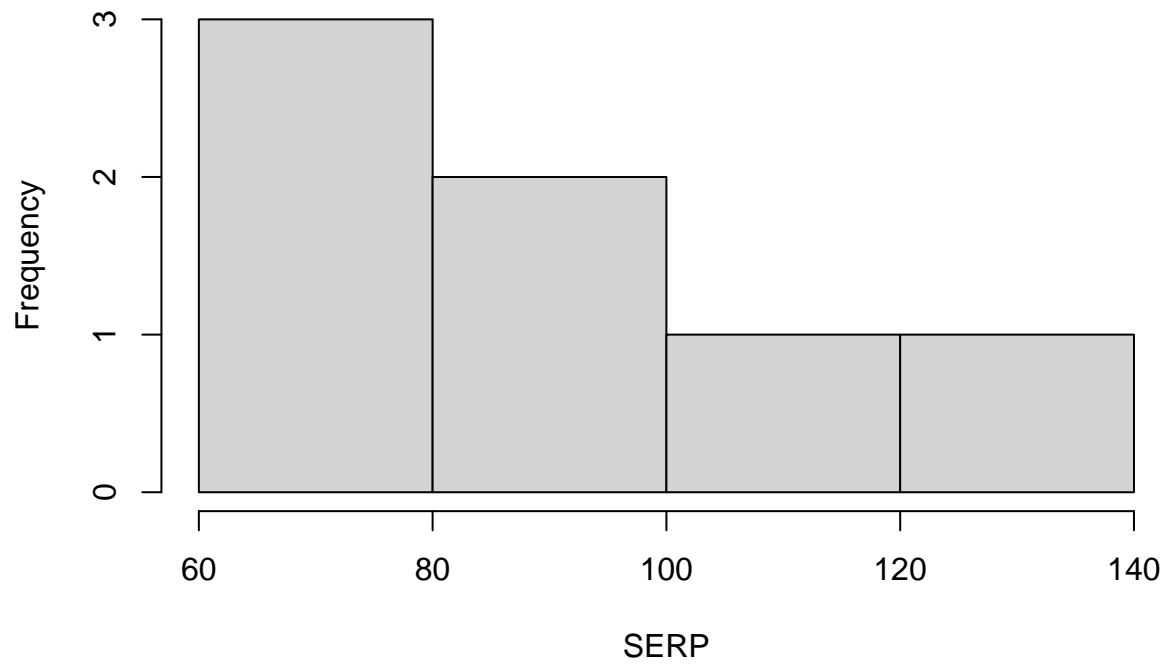


```

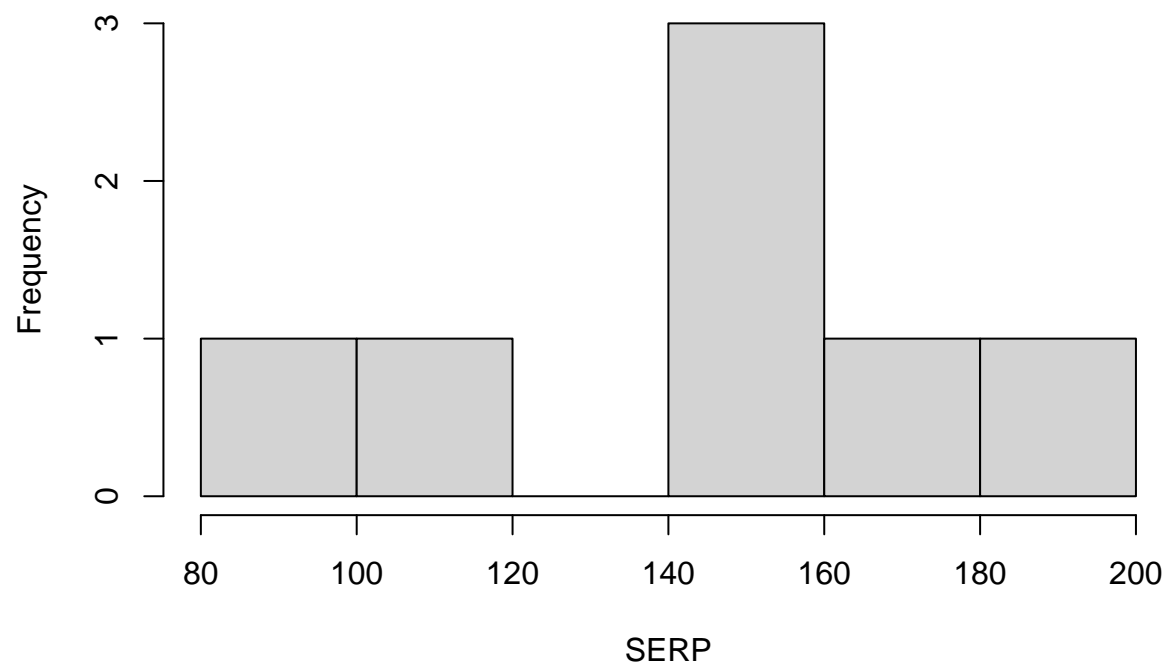
for (g in unique(data$BIOFILM)) {
  hist(data$SERP[data$BIOFILM == g],
    main = paste("BIOFILM =", g),
    xlab = "SERP")
}

```

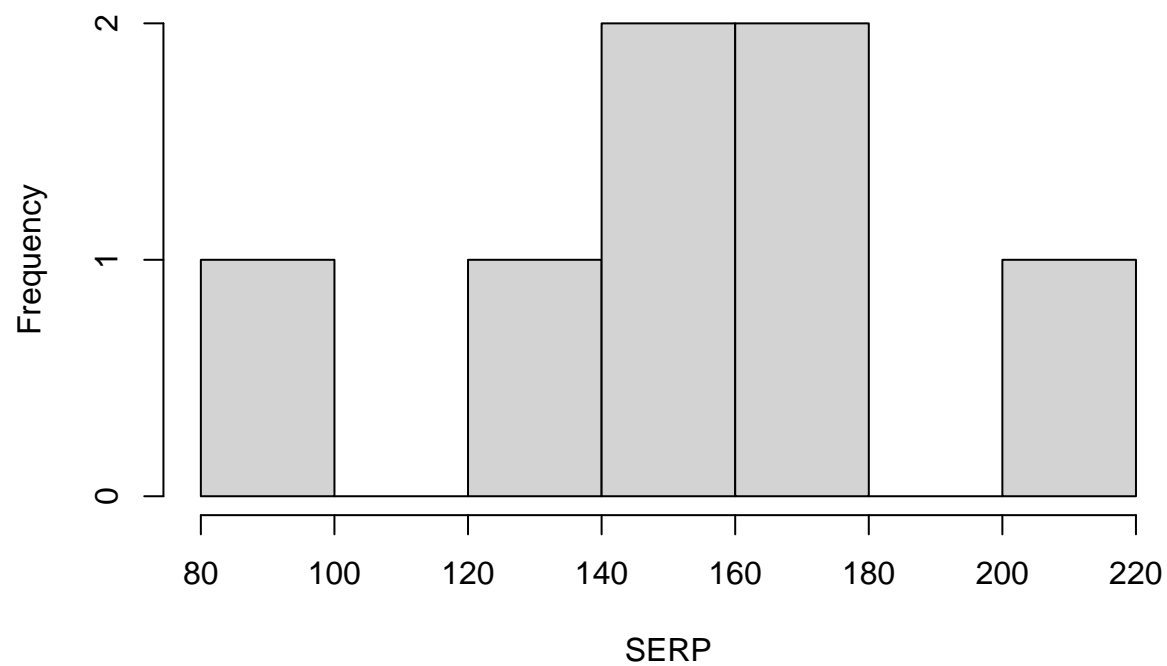

BIOFILM = SL

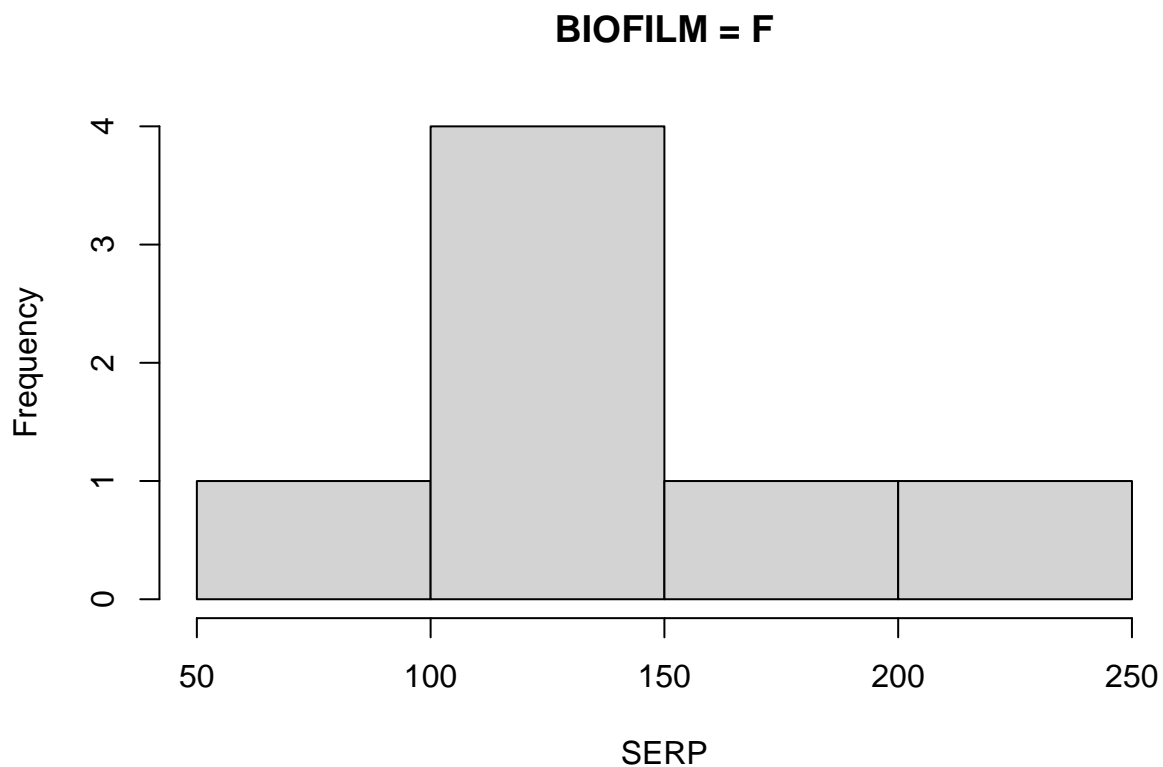


BIOFILM = UL



BIOFILM = NL



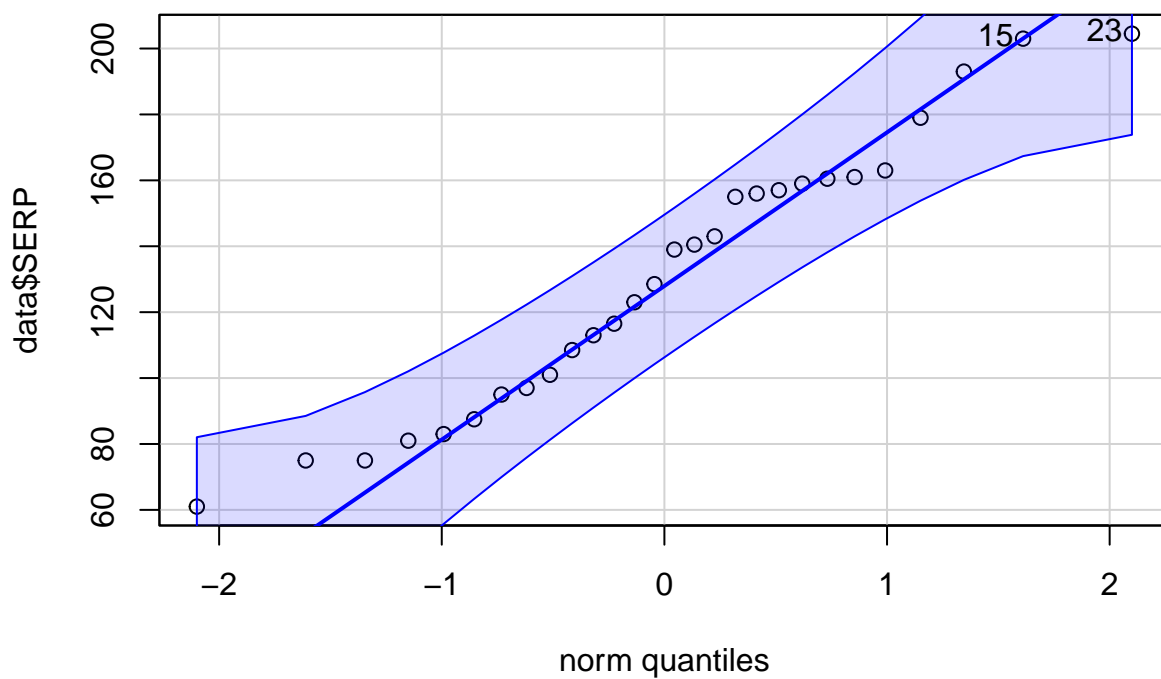


```
by(data$SERP, data$BIOFILM, function(x) shapiro.test(as.numeric(na.omit(x))))
```

```
## data$BIOFILM: F
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.957, p-value = 0.7926
##
## -----
## data$BIOFILM: NL
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.94763, p-value = 0.7081
##
## -----
## data$BIOFILM: SL
##
##  Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.94173, p-value = 0.6543
##
```

```
## -----
## data$BIOFILM: UL
##
## Shapiro-Wilk normality test
##
## data:  as.numeric(na.omit(x))
## W = 0.92786, p-value = 0.5329
```

```
qqPlot(data$SERP)
```



```
## [1] 23 15
```

```
# Test for equal variances again
leveneTest(SERP ~ BIOFILM, data = data)
```

```
## Warning in leveneTest.default(y = y, group = group, ...): group coerced to
## factor.
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group 3  0.2964 0.8276
##      24
```

```
# 6. Run ANOVA with transformed data
# Run ANOVA
anova <- aov(log_SERP ~ BIOFILM, data = data)
summary(anova)
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## BIOFILM        3  1.278   0.4260    6.006 0.00334 **
## Residuals     24  1.702   0.0709
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

7. Provide a 1-2 sentence summary of the outcome, and report results (df, F, and P values). Do you accept or reject the NULL?

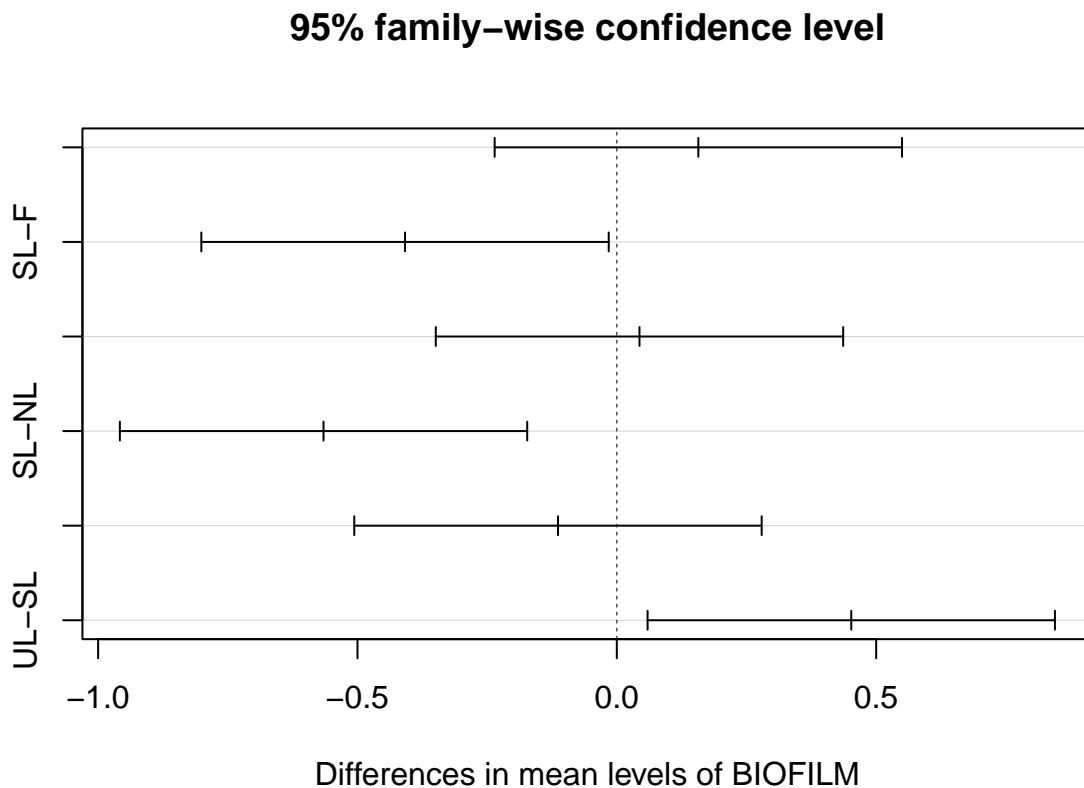
The results of the ANOVA indicate that we can accept our alternative hypothesis that at least one biofilm type has a mean recruitment that differs from the others. This is supported by these results: $F(3, 24) = 6.006$, $p = 0.00334$.

8. Run a post Hoc Tukey's test (if results were significant). Include the Tukey results statement in report format (as a summary or table) for the Tukey's results as well as a Tukey's plot showing 95% CI's

```
# Hoc Tukey's
tuk = TukeyHSD(anova)
tuk
```

```
##      Tukey multiple comparisons of means
##      95% family-wise confidence level
##
## Fit: aov(formula = log_SERP ~ BIOFILM, data = data)
##
## $BIOFILM
##              diff              lwr              upr              p adj
## NL-F      0.1571504 -0.23554812  0.54984884  0.6905958
## SL-F     -0.4083055 -0.80100402 -0.01560706  0.0394724
## UL-F      0.0436939 -0.34900458  0.43639237  0.9897078
## SL-NL    -0.5654559 -0.95815438 -0.17275742  0.0029670
## UL-NL    -0.1134565 -0.50615494  0.27924201  0.8550855
## UL-SL     0.4519994  0.05930096  0.84469791  0.0198663
```

```
plot(tuk)
```



Given the acceptance of the alternative hypothesis in the ANOVA, we run Tukeys post hoc which shows SL-F, SL-NL, and UL-SL have statistically different means while the others do not.

PART 2:

Now examine the dataset above using an appropriate non parametric alternative to the One Way ANOVA.

1. State appropriate null and alternative hypotheses

H0: The median recruitment of serpulid larvae is equal across all four biofilm types.

H1: At least one biofilm type has a median recruitment that differs from the others.

2. Run the non-parametric test and if appropriate, run a non-parametric pairwise comparison (using the Bonferroni adjustment).

```
kruskal.test(log_SERP ~ BIOFILM, data = data)
```

```
##
## Kruskal-Wallis rank sum test
##
## data: log_SERP by BIOFILM
## Kruskal-Wallis chi-squared = 11.042, df = 3, p-value = 0.0115
```

```
pairwise.wilcox.test(data$log_SERP, data$BIOFILM, p.adjust.method = "bonferroni")
```

```
## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties
## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties
## Warning in wilcox.test.default(xi, xj, paired = paired, ...): cannot compute
## exact p-value with ties

##
## Pairwise comparisons using Wilcoxon rank sum exact test
##
## data: data$log_SERP and data$BIOFILM
##
##      F      NL      SL
## NL 1.000 -      -
## SL 0.128 0.029 -
## UL 1.000 1.000 0.128
##
## P value adjustment method: bonferroni
```

3. Provide a 1-2 sentence summary of the outcome, and report results (df, F, and P values). Do you accept or reject the NULL?

The results of the non-parametric Kruskal-Wallis test indicate post hoc is necessary. The post hoc shows that one mean differ between the groups, which is NL-SL, indicating we can accept our alternative hypothesis that at least one biofilm type has a median recruitment that differs from the others.