Case Study 2

Emily, Ruth, Minh

Ecological researchers conducted a study to examine the effects of sulfamerazine concentration on hemoglobin in the blood of (rainbow) trout. Trout were randomly assigned to one of four tanks with each tank containing a different concentration of sulfamerazine (0, 5, 10, or 15 grams per 100 pounds of fish). After 35 days, 10 trout from each tank were randomly selected and their hemoglobin level (grams per 100 ml) was measured.

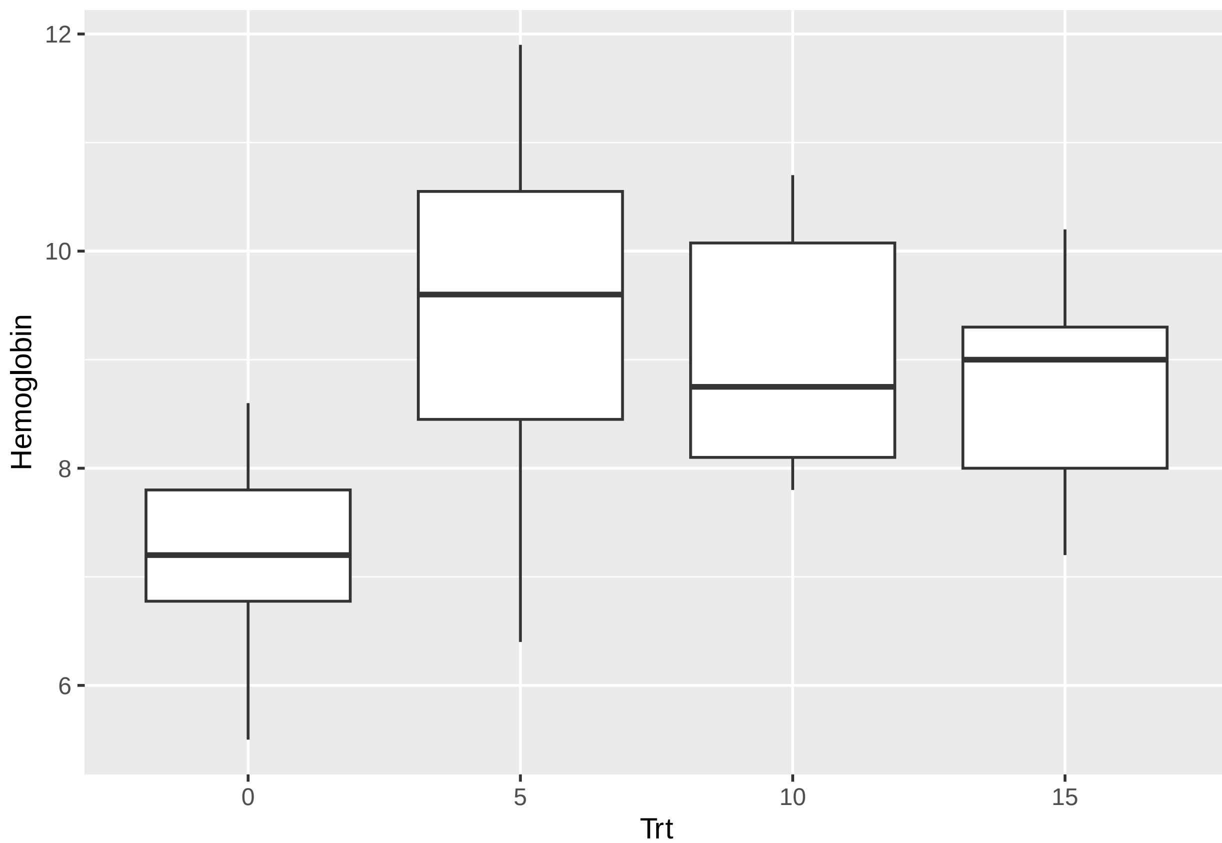
```r
#load data set
trout <- read_csv("./trout.csv")
#convert Trt to factor
trout <- trout %>% mutate(Trt = as.factor(Trt))
```

1. [5pts] Using the appropriate graphical display and summary statistics, describe the relationship between sulframerazine concentration and hemoglobin levels.

```r
favstats(Hemoglobin~Trt, data = trout) %>%
  mutate(IQR = Q3-Q1)
```

```
##   Trt min    Q1 median     Q3  max mean       sd  n missing   IQR
## 1   0 5.5 6.775   7.20  7.800  8.6 7.20 1.018714 10       0 1.025
## 2   5 6.4 8.450   9.60 10.550 11.9 9.33 1.716618 10       0 2.100
## 3  10 7.8 8.100   8.75 10.075 10.7 9.03 1.135341 10       0 1.975
## 4  15 7.2 8.000   9.00  9.300 10.2 8.69 1.000500 10       0 1.300
```

```r
trout %>%
  ggplot(aes(x = Trt, y = Hemoglobin)) +
  geom_boxplot()
```



Answer: The distribution of hemoglobin for fish given sulfamerazine in concentrations of 0g, 5g, 10g, or 15g per 100 pounds of fish vary in shape with the 0g and 5g group being relatively symmetric while the 10g group is right skewed and the 15g group is left skewed. When looking at the centers of these distributions, the median hemoglobin level of the 5g group is highest, with a value of 9.60 g/mL, while the 0g group had the lowest median hemoglobin level, with a value of 7.20 g/mL (10g: 8.75 g/mL, 15g: 9.00 g/mL). When looking at the spread of these groups, the IQR was greatest for the 5g group, with an IQR of 2.10 g/mL, while the 0g group had the smallest IQR, with a value of 1.025 g/mL (10g: 1.975 g/mL, 15g: 1.30 g/mL). There were no possible outliers in the data.

2. [3pts] The cell means model for these data is $Hemoglobin_{ij} = \mu + \alpha_i + \epsilon_{ij}$.

a. [1pt] Define what i represents. Make sure to write in context and provide the values for i.

Answer: i represents the treatment group that a fish is in, with each i value representing a different concentration of sulfamerazine in grams per 100 pounds of fish. The possible i values in this experiment are 0g, 5g, 10g, and 15g per 100 pounds of fish.

b. [1pt] Define what j represents. Make sure to write in context.

Answer: j represents which fish in a specific treatment group is being referred to. The possible j values in this experiment range from 1-10 because there are 10 fish in each tank/treatment group that are randomly selected to be measured.

c. [1pt] Define what α_3 represents. Make sure to write in context.

Answer α_3 represents the difference in means between the 3rd treatment group and the baseline group, which in this experiment, is the 0g group. The 3rd treatment group is the 15g group, so α_3 equals the difference between 15g group's average hemoglobin level and the 0g group's average hemoglobin level.

3. [4pts] The first question of interest is whether average hemoglobin levels differ for at least one concentration of sulframerazine.

a. [1pt] Set up the hypotheses of interest.

Answer: $H_o : \mu_{0g} = \mu_{5g} = \mu_{10g} = \mu_{15g}$ H_a : **At least one** μ_i **differs with** $i = 0g, 5g, 10g, 15g$

b. [1pt] Run the appropriate analysis to test these hypotheses and report the appropriate test statistic, df (if needed), and p-value.

```
trout_model <- lm(Hemoglobin~Trt, data = trout)
summary(trout_model)
```

```
##
## Call:
## lm(formula = Hemoglobin ~ Trt, data = trout)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.930 -0.930 -0.010  0.725  2.570
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.2000     0.3961  18.179  < 2e-16 ***
## Trt5          2.1300     0.5601   3.803 0.000534 ***
## Trt10         1.8300     0.5601   3.267 0.002391 **
## Trt15         1.4900     0.5601   2.660 0.011588 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.252 on 36 degrees of freedom
## Multiple R-squared:  0.3219, Adjusted R-squared:  0.2654
## F-statistic: 5.696 on 3 and 36 DF,  p-value: 0.002685
```

```
anova(trout_model)
```

```
## Analysis of Variance Table
##
## Response: Hemoglobin
##           Df Sum Sq Mean Sq F value   Pr(>F)
## Trt        3 26.803  8.9343  5.6955 0.002685 **
## Residuals 36 56.471  1.5686
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Answer: F = 5.696, df = 3 and 36, and p value = 0.003

c. [2pts] Provide a conclusion for this test making sure to clearly explain what the results mean.

Answer: We reject the null hypothesis and conclude that there is strong evidence that the average hemoglobin level of rainbow trout with varying sulfamerazine concentrations (0g, 5g, 10g, 15g per 100 pounds of fish) differs (F = 5.696, df = 3 and 36, p = 0.003).

4. [3pts] Before we get too far in our analysis, we need to check whether the conditions for this analysis are reasonably satisfied (Note: in the notes we did this in a different order. In practice it is recommended to check assumptions early in an analysis. If any changes need to be made it is better to do them early instead of having to redo all your work). For each of the assumption, explain if it is reasonably met referencing appropriate plots/numerical values as needed.

a. [1pt] Independence

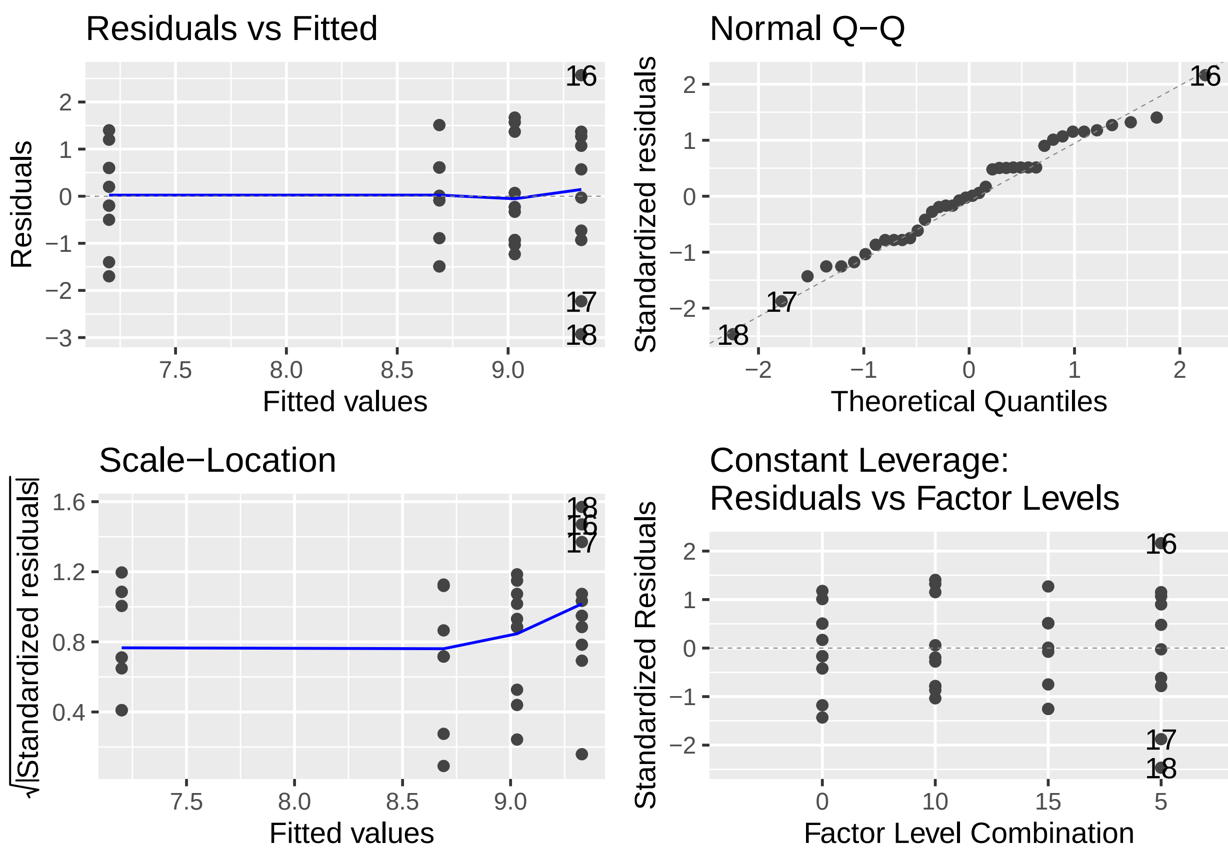
Answer: There is no indication that the fish were measured more than once or that there was any cluster effect, but the experimental unit is not equal to the observational unit, so the independence assumption is not reasonably satisfied. The experimental unit is tanks of fish because each treatment was applied to the whole tank, but the observational unit was individual fish because the fish were randomly picked from the tank.

b. [1pt] Equal variance

Answer: The equal variances assumption is reasonably satisfied because when dividing the largest standard deviation by the smallest standard deviation, the ratio is less than 2 (1.717/1.001 = 1.715).

c. [1pt] Normality

```
autoplot(trout_model)
```




```
1.716618/1.000500
```

```
## [1] 1.71576
```

Answer: The normality assumption is reasonably satisfied because when looking at the QQ plot, there is little deviation from the dashed line, especially on the edges, and the sample size is 40, which is enough to compensate for this deviation.

5. [2pts] During data collection, the researchers realized that one of the trout was a steelhead trout which can look similar to rainbow trout. Because of this, the researchers did not measure hemoglobin levels of this fish and instead they randomly sampled another fish from this sulframerazine group. Do you agree or disagree with this decision? Explain.

Answer: We agree with this decision because the population of interest for this experiment was rainbow trout, so the steelhead trout was not part of the population of interest, so it is appropriate to remove it. Also, in order to replace this fish, the replacement fish should be measured at the same time as the other fish and there must be more than 10 fish in the tank so sample size is not affected and fish are not counted twice.

6. [3pts] Regardless of what you may have found in Question 4, let's proceed assuming the assumptions are reasonably satisfied. One part of this study is to compare average hemoglobin levels among the four sulframerazine levels. Using the appropriate tools (e.g. HTs, CIs, plots) summarize these results for the researchers. Make sure to clearly state what methods you used and why.

```r
confint(trout_model)
```

```
##                 2.5 %   97.5 %
## (Intercept) 6.396752 8.003248
## Trt5        0.994036 3.265964
## Trt10       0.694036 2.965964
## Trt15       0.354036 2.625964
```

Answer: With 95% confidence, the average hemoglobin level for fish in tanks with 0g of sulfamerazine is between 6.397 and 8.003 g/mL of hemoglobin. The average hemoglobin level for fish in tanks with 5g of sulfamerazine is between 0.994 and 3.266 g/mL greater than the 0g group, and the average hemoglobin for fish in 10g tanks is between 0.694 and 2.966 g/mL greater than the 0g group, and the average hemoglobin level for fish in 15g tanks is between 0.354 and 2.626 g/mL greater than the tank with 0g of sulfamerazine. There is strong evidence of a difference in the average hemoglobin level between the different treatment groups with different concentrations of sulfamerazine, but further analysis is required to determine which groups differ.

7. [4pts] The researchers are specifically interested in comparing hemoglobin levels between trout that received no sulframerazine to those that did.

a. [1pt] Set up the linear combination of interest to the researchers using correct statistical notation.

Answer: $\gamma = \mu_{0g} - (\mu_{5g} + \mu_{10g} + \mu_{15g})/3$

b. [3pts] Obtain a point estimate and 95% confidence interval for the linear combination of interest. Interpret the confidence interval in context and in a way that is easily understandable.

```r
trout_means <- emmeans(trout_model,~Trt)
levels(trout$Trt)
```

```
## [1] "0"  "5"  "10" "15"
```

```r
confint(contrast(trout_means,list(NvR=c(1, -(1/3), -(1/3), -(1/3)))))
```

```
##  contrast estimate    SE df lower.CL upper.CL
##  NvR         -1.82 0.457 36    -2.74   -0.889
##
## Confidence level used: 0.95
```

Answer: With 95% confidence, the average difference between fish in a tank with no sulfamerazine is between 0.889 and 2.74 g/mL lower than fish in tanks with sulfamerazine, with a point estimate of 1.82 g/mL less for fish with no sulfamerazine.

8. [3pts] The researchers would like to rerun this experiment and have come to you for help. You explain to them that a more appropriate way to design this experiment is to use multiple tanks of trout and randomly assign sulframerazine level to the tanks. The response would then be average hemoglobin level of the fish in the tank.

a. [1pt] Explain why this would be a better experimental design than the original.

Answer: If the experimental design was changed to this way, the independence assumption would be reasonably satisfied because the experimental unit and observational unit would both be the tank of fish. This would allow for more accurate analysis.

b. [1pt] The researchers have the resources to have 4 tanks for each sulframerazine concentration. They believe that average hemoglobin levels will be lowest with no sulframerazine at 7.25 grams/100 ml and highest for fish exposed to a 5 gram sulframerazine concentration at 9.5 grams/100 ml. Hemoglobin levels are expected to be slightly lower in the 10 and 15 gram concentration groups at 9.1 and 8.7 grams/100 ml, respectively. The researchers will use a significance level of 5% and would like to achieve a power of 80%. Is this possible given their sample size?

```
means <- c(7.25, 9.5, 9.1, 8.7)
var(means)
```

```
## [1] 0.9622917
```

```
power.anova.test(groups = 4, n = 4, sig.level = 0.05, power = NULL, within.var = 1.5686, between.var = 0
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 4
##               n = 4
##     between.var = 0.9622917
##      within.var = 1.5686
##       sig.level = 0.05
##           power = 0.4683812
##
## NOTE: n is number in each group
```

Answer: It is not possible to get 80% power using only 4 tanks per treatment, as the estimated power would only be 46.83%. More tanks must be used per treatment.

c. [1pt] How many tanks per treatment would be required for the researchers to achieve 80% power?

```
power.anova.test(groups = 4, n = NULL, sig.level = 0.05, power = 0.8, within.var = 1.5686, between.var =
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##          groups = 4
##               n = 6.982471
##     between.var = 0.9622917
##      within.var = 1.5686
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

Answer: There would need to be 7 tanks per treatment to allow for 80% power. '