Lab 8

2024-10-27

Q3

Do you calculate the actual cell means or estimate them based on the code provided?

Q5-7

Do you include "and across all interactions" in the interpretation?

Overview

Factor A: Concentration of copper (0 or 150 ppm) Factor B: Concentration of zinc (0, 750 or 1500 ppm) Response Variable: protein content (μg/tank)

```
library(readr)
minnows <- read_table("minnows.txt",
                      col_names = c("copper", "zinc", "protein"),
                      col_types = cols(copper = col_factor(levels = c("0","150")),
                                       zinc = col_factor(levels = c("0", "750", "1500")))
                      )
```

1.

Using baseline constraints, report the six unique rows in the design matrix X and the parameter vector β for this two-factor experiment.

```
options(contrasts = c("contr.treatment", "contr.treatment"))
baseline.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(baseline.anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## copper       1    234     234   1.809 0.227264
## zinc         2  10234    5117  39.537 0.000351 ***
## copper:zinc  2    288     144   1.113 0.387957
## Residuals    6    777     129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
baseline.anova$coefficients
```

```
##        (Intercept)          copper150            zinc750           zinc1500
##              193.5              -21.0              -26.0              -74.0
##  copper150:zinc750 copper150:zinc1500
##               24.0               12.5
```

```r
model.matrix(baseline.anova)
```

```
##    (Intercept) copper150 zinc750 zinc1500 copper150:zinc750 copper150:zinc1500
## 1            1         0       0        0                 0                 0
## 2            1         0       0        0                 0                 0
## 3            1         0       1        0                 0                 0
## 4            1         0       1        0                 0                 0
## 5            1         0       0        1                 0                 0
## 6            1         0       0        1                 0                 0
## 7            1         1       0        0                 0                 0
## 8            1         1       0        0                 0                 0
## 9            1         1       1        0                 1                 0
## 10           1         1       1        0                 1                 0
## 11           1         1       0        1                 0                 1
## 12           1         1       0        1                 0                 1
## attr(,"assign")
## [1] 0 1 2 2 3 3
## attr(,"contrasts")
## attr(,"contrasts")$copper
## [1] "contr.treatment"
##
## attr(,"contrasts")$zinc
## [1] "contr.treatment"
```

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$$

2.

Using sum-to-zero constraints, report the six unique rows in the design matrix X and the parameter vector β for this two-factor experiment.

```
options(contrasts = c("contr.sum", "contr.sum"))
sumto0.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(sumto0.anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## copper       1    234     234   1.809 0.227264
## zinc         2  10234    5117  39.537 0.000351 ***
## copper:zinc  2    288     144   1.113 0.387957
## Residuals    6    777     129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
sumto0.anova$coefficients
```

```
##   (Intercept)       copper1         zinc1         zinc2 copper1:zinc1
##    155.750000      4.416667     27.250000     13.250000      6.083333
## copper1:zinc2
##     -5.916667
```

```
model.matrix(sumto0.anova)
```

```
##    (Intercept) copper1 zinc1 zinc2 copper1:zinc1 copper1:zinc2
## 1            1       1     1     0             1             0
## 2            1       1     1     0             1             0
## 3            1       1     0     1             0             1
## 4            1       1     0     1             0             1
## 5            1       1    -1    -1            -1            -1
## 6            1       1    -1    -1            -1            -1
## 7            1      -1     1     0            -1             0
## 8            1      -1     1     0            -1             0
## 9            1      -1     0     1             0            -1
## 10           1      -1     0     1             0            -1
## 11           1      -1    -1    -1             1             1
## 12           1      -1    -1    -1             1             1
## attr(,"assign")
## [1] 0 1 2 2 3 3
## attr(,"contrasts")
## attr(,"contrasts")$copper
## [1] "contr.sum"
##
## attr(,"contrasts")$zinc
## [1] "contr.sum"
```

$$X = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{bmatrix}$$

3.

Use the estimated parameters from the R output to calculate the six sample treatment means.

The six sample treatment means is the same as the cell means.

```
options(contrasts = c("contr.sum", "contr.sum"))
summary(lm(protein ~ copper + zinc + copper*zinc, data = minnows))
```

```
##
## Call:
## lm(formula = protein ~ copper + zinc + copper * zinc, data = minnows)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -13.5   -6.0    0.0    6.0   13.5
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    155.750      3.284  47.427 5.89e-09 ***
## copper1          4.417      3.284   1.345  0.22726
## zinc1           27.250      4.644   5.867  0.00108 **
## zinc2           13.250      4.644   2.853  0.02907 *
## copper1:zinc1    6.083      4.644   1.310  0.23816
## copper1:zinc2   -5.917      4.644  -1.274  0.24980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.38 on 6 degrees of freedom
## Multiple R-squared:  0.9327, Adjusted R-squared:  0.8766
## F-statistic: 16.62 on 5 and 6 DF,  p-value: 0.001854
```

```
model <- lm(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(model)
```

```
##
## Call:
## lm(formula = protein ~ copper + zinc + copper * zinc, data = minnows)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -13.5   -6.0    0.0    6.0   13.5
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    155.750      3.284  47.427 5.89e-09 ***
## copper1          4.417      3.284   1.345  0.22726
## zinc1           27.250      4.644   5.867  0.00108 **
## zinc2           13.250      4.644   2.853  0.02907 *
## copper1:zinc1    6.083      4.644   1.310  0.23816
## copper1:zinc2   -5.917      4.644  -1.274  0.24980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 11.38 on 6 degrees of freedom
## Multiple R-squared:  0.9327, Adjusted R-squared:  0.8766
## F-statistic: 16.62 on 5 and 6 DF,  p-value: 0.001854
```

$$\hat{\beta}_{\text{sum to zero}} = \begin{bmatrix} 155.750 \\ 4.417 \\ 27.250 \\ 13.250 \\ 6.083 \\ -5.917 \end{bmatrix} = \begin{bmatrix} \mu \\ \alpha_1 \\ \tau_1 \\ \tau_2 \\ (\alpha\tau)_{11} \\ (\alpha\tau)_{12} \end{bmatrix}$$

```
tapply(X = minnows$protein,
       INDEX = list(minnows$copper,  minnows$zinc),
       FUN = mean)
```

```
##         0   750  1500
## 0   193.5 167.5 119.5
## 150 172.5 170.5 111.0
```

```
aggregate(protein ~ copper + zinc, data = minnows, FUN = mean)
```

```
##   copper zinc protein
## 1      0    0   193.5
## 2    150    0   172.5
## 3      0  750   167.5
## 4    150  750   170.5
## 5      0 1500   119.5
## 6    150 1500   111.0
```

4.

Use the ANOVA table from the R output to conduct the overall F-test for the effects of the six treatments on the minnow protein production. Interpret the results of the test in the context of the study.

```
options(contrasts = c("contr.sum", "contr.sum"))
summary(lm(protein ~ copper + zinc + copper*zinc, data = minnows))
```

```
##
## Call:
## lm(formula = protein ~ copper + zinc + copper * zinc, data = minnows)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -13.5   -6.0    0.0    6.0   13.5
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)    155.750      3.284  47.427 5.89e-09 ***
## copper1          4.417      3.284   1.345  0.22726
## zinc1           27.250      4.644   5.867  0.00108 **
## zinc2           13.250      4.644   2.853  0.02907 *
## copper1:zinc1    6.083      4.644   1.310  0.23816
## copper1:zinc2   -5.917      4.644  -1.274  0.24980
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.38 on 6 degrees of freedom
## Multiple R-squared:  0.9327, Adjusted R-squared:  0.8766
## F-statistic: 16.62 on 5 and 6 DF,  p-value: 0.001854
```

Overall F-statistic is 16.62 with a p-value of 0.001854. The overall F-statistic and associated test is whether there are any differences in the mean protein content across all treatments, i.e. $\mu_{copper0} = \mu_{copper1} = \mu_{zinc0} = \mu_{zinc1} = \mu_{zinc2}$. With a small p-value we have evidence to reject the null hypothesis at the $\alpha = 0.05$ level that all the aforementioned mean protein contents are equal in favor of the alternative hypothesis that at least one mean protein content of a treatment is different from the mean protein content of the other treatments.

5.

Use the ANOVA table from the R output to test for the main effect of copper concentration on the minnow protein production. Assuming there is no significant interaction, interpret the results of the test in the context of the study.

```
sumto0.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(sumto0.anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## copper       1    234     234   1.809 0.227264
## zinc         2  10234    5117  39.537 0.000351 ***
## copper:zinc  2    288     144   1.113 0.387957
## Residuals    6    777     129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Copper (Main Effect): F-value: 1.809 p-value: 0.227 Interpretation: The p-value of 0.227 does not meet our threshold of the $\alpha = 0.05$ level, such that we do not have evidence to reject the null hypothesis that the mean protein content (μg/tank) is the same between between the two copper levels (0 ppm and 150 ppm), when averaging across levels of zinc concentrations.

6.

Use the ANOVA table from the R output to test for the main effect of zinc concentration on the minnow protein production. Assuming there is no significant interaction, interpret the results of the test in the context of the study.

```
sumto0.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(sumto0.anova)
```

```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## copper       1    234     234   1.809 0.227264
## zinc         2  10234    5117  39.537 0.000351 ***
## copper:zinc  2    288     144   1.113 0.387957
## Residuals    6    777     129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Zinc (Main Effect):

F-value: 39.537 p-value: 0.000351 Interpretation: The p-value of 0.000351 meets the threshold of the $\alpha = 0.05$ level, such that we do have evidence to reject the null hypothesis that the mean protein content (μg/tank) is the same between the three zinc levels levels (0, 750 or 1500 ppm), when averaging across levels of copper. This suggests that at least one of the zinc levels (0, 750, or 1500 ppm) has a different mean protein content (μg/tank) in the minnow larvae, when averaging across levels of copper.

7.

Use the ANOVA table from the R output to test for the interaction effect between copper and zinc concentrations on the minnow protein production. Interpret the results of the test in the context of the study.

```
sumto0.anova <- aov(protein ~ copper + zinc + copper*zinc, data = minnows)
summary(sumto0.anova)
```

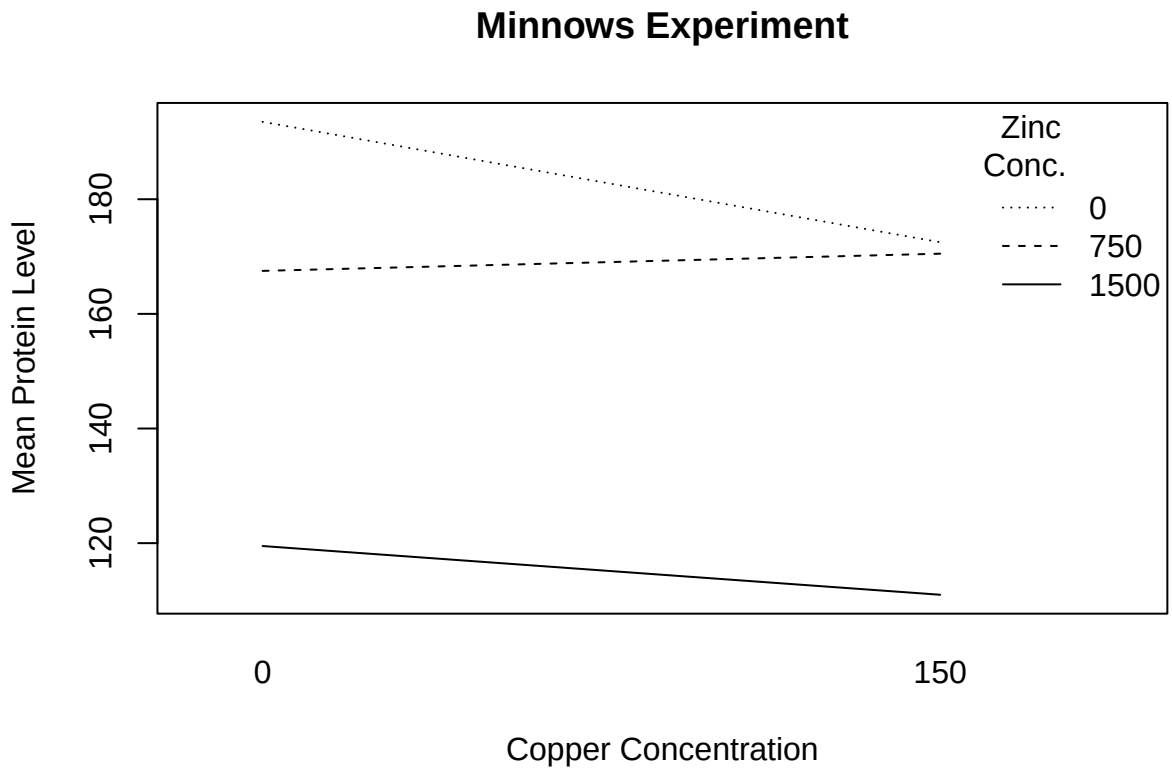
```
##             Df Sum Sq Mean Sq F value   Pr(>F)
## copper       1    234     234   1.809 0.227264
## zinc         2  10234    5117  39.537 0.000351 ***
## copper:zinc  2    288     144   1.113 0.387957
## Residuals    6    777     129
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interaction (Copper $\times$ Zinc): F-value: 1.113 p-value: 0.388 Interpretation: The p-value of 0.388 does not meet our threshold of the $\alpha = 0.05$ level, such that we do not have evidence to reject the null hypothesis that the the mean protein content (μg/tank) of interactions is the same across the pairs of interactions between copper and zinc concentrations.

8.

Study the interaction plot in the R output. Explain why, based on this plot, the interaction is not significant in the model.

```
interaction.plot(x.factor = minnows$copper,
                 trace.factor = minnows$zinc,
                 response = minnows$protein,
  main="Minnows Experiment",
  xlab="Copper Concentration",
  ylab="Mean Protein Level",
  trace.label="Zinc\n Conc.")
```




```
interaction.plot(x.factor = minnows$zinc,
                 trace.factor = minnows$copper,
                 response = minnows$protein,
  main="Minnows Experiment",
  xlab="Zinc Concentration",
  ylab="Mean Protein Level",
  trace.label="Copper\n Conc.")
```

We have two interaction plots: One for "Copper" tracing on "Zinc" and one for "Zinc" tracing on "Copper". The conclusions of the two are similar/consistent, insomuch as we conclude the interaction is not significant in the model.

Overall, what we are looking for is consistency across the different levels, which looks like parallel, non-intersecting lines. Based on the interaction plots above, the lines for different levels of zinc do not strongly deviate from one another across levels of copper (are roughly parallel). Similarly, the lines for different levels of copper do not strongly deviate from one another across levels of zinc; while there is some intersection, I don't believe it is especially concerning since it is consistent in direction and magnitude across the two levels of the copper factor.

Given the above, we do not have overwhelming nor consistent evidence that the interaction is significant in the model.

9.

Use the LSD method to perform all pairwise comparisons of the cell means (simple effects) for the six treatments. Summarize any significant findings in the context of the study.

```
library(emmeans)
```

```
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
```

```
simple.effects <- emmeans(sumto0.anova, c("copper", "zinc"))
pairs(simple.effects, adjust=NULL)
```

```
##  contrast                                estimate   SE df t.ratio p.value
##  copper0 zinc0 - copper150 zinc0             21.0 11.4  6   1.846  0.1144
##  copper0 zinc0 - copper0 zinc750             26.0 11.4  6   2.285  0.0623
##  copper0 zinc0 - copper150 zinc750           23.0 11.4  6   2.022  0.0897
##  copper0 zinc0 - copper0 zinc1500            74.0 11.4  6   6.505  0.0006
##  copper0 zinc0 - copper150 zinc1500          82.5 11.4  6   7.252  0.0003
##  copper150 zinc0 - copper0 zinc750            5.0 11.4  6   0.440  0.6757
##  copper150 zinc0 - copper150 zinc750          2.0 11.4  6   0.176  0.8662
##  copper150 zinc0 - copper0 zinc1500          53.0 11.4  6   4.659  0.0035
##  copper150 zinc0 - copper150 zinc1500        61.5 11.4  6   5.406  0.0017
##  copper0 zinc750 - copper150 zinc750         -3.0 11.4  6  -0.264  0.8008
##  copper0 zinc750 - copper0 zinc1500          48.0 11.4  6   4.219  0.0056
##  copper0 zinc750 - copper150 zinc1500        56.5 11.4  6   4.967  0.0025
##  copper150 zinc750 - copper0 zinc1500        51.0 11.4  6   4.483  0.0042
##  copper150 zinc750 - copper150 zinc1500      59.5 11.4  6   5.230  0.0020
##  copper0 zinc1500 - copper150 zinc1500        8.5 11.4  6   0.747  0.4832
```

Using "significance" of $\alpha = 0.05$, we omit the following pairs: contrast estimate SE df t.ratio p.value copper0 zinc0 - copper150 zinc0 21.0 11.4 6 1.846 0.1144 copper0 zinc0 - copper0 zinc750 26.0 11.4 6 2.285 0.0623 copper0 zinc0 - copper150 zinc750 23.0 11.4 6 2.022 0.0897 copper150 zinc0 - copper0 zinc750 5.0 11.4 6 0.440 0.6757 copper150 zinc0 - copper150 zinc750 2.0 11.4 6 0.176 0.8662 copper0 zinc750 - copper150 zinc750 -3.0 11.4 6 -0.264 0.8008 copper0 zinc1500 - copper150 zinc1500 8.5 11.4 6 0.747 0.4832

The above pairs do not meet the alpha level such that we do not find evidence to reject the null hypothesis. In the above instances, we do not reject the null hypothesis that the mean protein content (μg/tank) for the combination of treatments on the left hand side is the same as the mean protein content (μg/tank) for the combination of treatments on the right hand side (they have a difference of zero).

Meaning for "significant findings": contrast estimate SE df t.ratio p.value copper0 zinc0 - copper0 zinc1500 74.0 11.4 6 6.505 0.0006 copper0 zinc0 - copper150 zinc1500 82.5 11.4 6 7.252 0.0003 copper150 zinc0 - copper0 zinc1500 53.0 11.4 6 4.659 0.0035 copper150 zinc0 - copper150 zinc1500 61.5 11.4 6 5.406 0.0017 copper0 zinc750 - copper0 zinc1500 48.0 11.4 6 4.219 0.0056 copper0 zinc750 - copper150 zinc1500 56.5 11.4 6 4.967 0.0025 copper150 zinc750 - copper0 zinc1500 51.0 11.4 6 4.483 0.0042 copper150 zinc750 - copper150 zinc1500 59.5 11.4 6 5.230 0.0020

The above pairs do meet the alpha level such that we have evidence to reject the null hypothesis at the $\alpha = 0.05$ level. In the above instances, we have evidence to reject the null hypothesis at the alpha level stated that the mean protein content (μg/tank) for the combination of treatments on the left hand side is different from the mean protein content (μg/tank) for the combination of treatments on the right hand side (there is a non-zero difference between the two).

Using this framework, we have evidence to believe the mean protein content (μg/tank) of the following pairwise combinations of treatments are different: copper0 zinc0 - copper0 zinc1500: No copper level changes but an increase of 1500 units of zinc concentration. copper0 zinc0 - copper150 zinc1500: From no zinc and copper concentration, an increase of 150 units for copper concentration in addition to an increase of 1500 units of zinc concentration. copper150 zinc0 - copper0 zinc1500: Copper concentration at 150 and zinc concentration at 0 and lowering copper concentration to 0 while increasing zinc concentration to 1500. copper150 zinc0 - copper150 zinc1500: Copper concentration remains at 150 but zinc concentration increases from 0 to 1500. copper0 zinc750 - copper0 zinc1500: No copper level changes but an increase from 750 to 1500 units of zinc concentration. copper0 zinc750 - copper150 zinc1500: From no copper concentration with 750 zinc concentration to an increase of 150 units for copper concentration in addition to an increase of 750 units of zinc concentration (to 1500 total). copper150 zinc750 - copper0 zinc1500: From 150 copper concentration with 750 zinc concentration to 0 copper concentration (decrease of 150 copper concentration) and an increase of 750 zinc concentration (total 1500 zinc concentration) copper150 zinc750 - copper150 zinc1500: From a copper concentration fo 150 and zinc concentration of 750, keep copper concentration consistent and add an additional 750 zinc concentration (total of 1500 zinc concentration).

Taking the numerous significant pairwise comparisons above, we find the mean protein content (μg/tank) when the zinc concentration increases to 1500 units tends to be different (smaller mean protein content (μg/tank) than what it is being compared to, whether starting from 0 or 750 zinc concentration, and without regard to a particular copper concentration or change in copper concentration).

10.

Use Tukey's HSD method to perform pairwise comparisons of the marginal means for the two copper concentrations (main effect of copper) and the marginal means for the three zinc concentrations (main effect of zinc). Summarize any significant findings in the context of the study.

```
copper.effects <- emmeans(sumto0.anova, "copper")
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
pairs(copper.effects, adjust="tukey")
```

```
##  contrast            estimate   SE df t.ratio p.value
##  copper0 - copper150     8.83 6.57  6   1.345  0.2273
##
## Results are averaged over the levels of: zinc
```

```
zinc.effects <- emmeans(sumto0.anova, "zinc")
```

```
## NOTE: Results may be misleading due to involvement in interactions
```

```
pairs(zinc.effects, adjust="tukey")
```

```
##  contrast           estimate   SE df t.ratio p.value
##  zinc0 - zinc750        14.0 8.04  6   1.740  0.2666
##  zinc0 - zinc1500       67.8 8.04  6   8.422  0.0004
##  zinc750 - zinc1500     53.8 8.04  6   6.682  0.0013
##
## Results are averaged over the levels of: copper
## P value adjustment: tukey method for comparing a family of 3 estimates
```

Within the contest of the study, the following two differences have a statistically significant difference at the $\alpha = 0.05$ level.

Comparing marginal means, we have two results that have a p-value small enough to provide evidence at the $\alpha = 0.05$ level for rejecting the null hypothesis.

contrast estimate SE df t.ratio p.value zinc0 - zinc1500 67.8 8.04 6 8.422 0.0004 zinc750 - zinc1500 53.8 8.04 6 6.682 0.0013

Zinc 0 - Zinc 1500: We have evidence at the $\alpha = 0.05$ level for rejecting the null hypothesis that mean protein content (μg/tank) for the zinc 0 concentrated water is the same as the mean protein content (μg/tank) for the zinc 1500 concentrated water, when averaging across levels of copper. This provides evidence to support the alternative hypothesis that that mean protein content (μg/tank) for the zinc 0 concentrated water is different from the mean protein content (μg/tank) for the zinc 1500 concentrated water, when averaging across levels of copper.

Zinc 750 - Zinc 1500: We have evidence at the $\alpha = 0.05$ level for rejecting the null hypothesis that mean protein content (μg/tank) for the zinc 750 concentrated water is the same as the mean protein content (μg/tank) for the zinc 1500 concentrated water, when averaging across levels of copper. This provides evidence to support the alternative hypothesis that that mean protein content (μg/tank) for the zinc 750 concentrated water is different from the mean protein content (μg/tank) for the zinc 1500 concentrated water, when averaging across levels of copper.

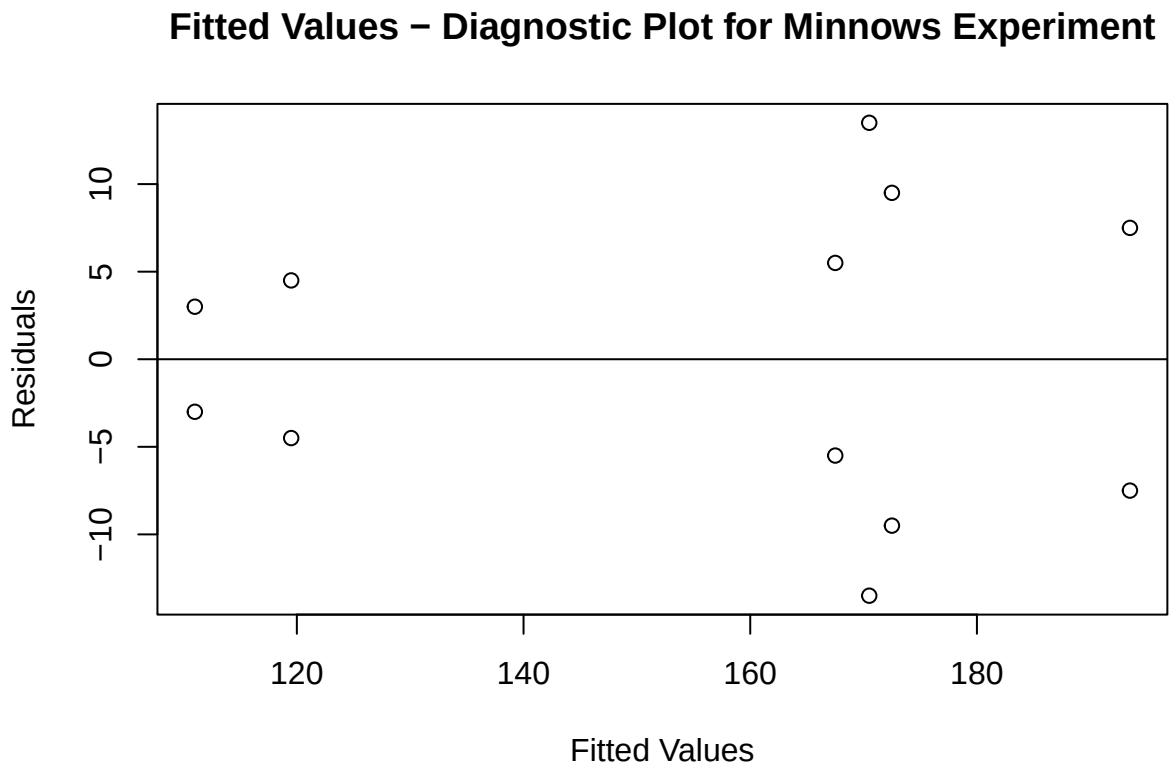
So we overall have evidence of treatment effects for Zinc when increasing zinc concentration in the water to 1500.

11.

Study the plots of the residuals versus: a) predicted value, b) the copper concentrations, and c) the zinc concentrations. Do these plots show any cause for concern?

Since these are separate plots, if any of the below cause concern, then overall we may have cause for concern.

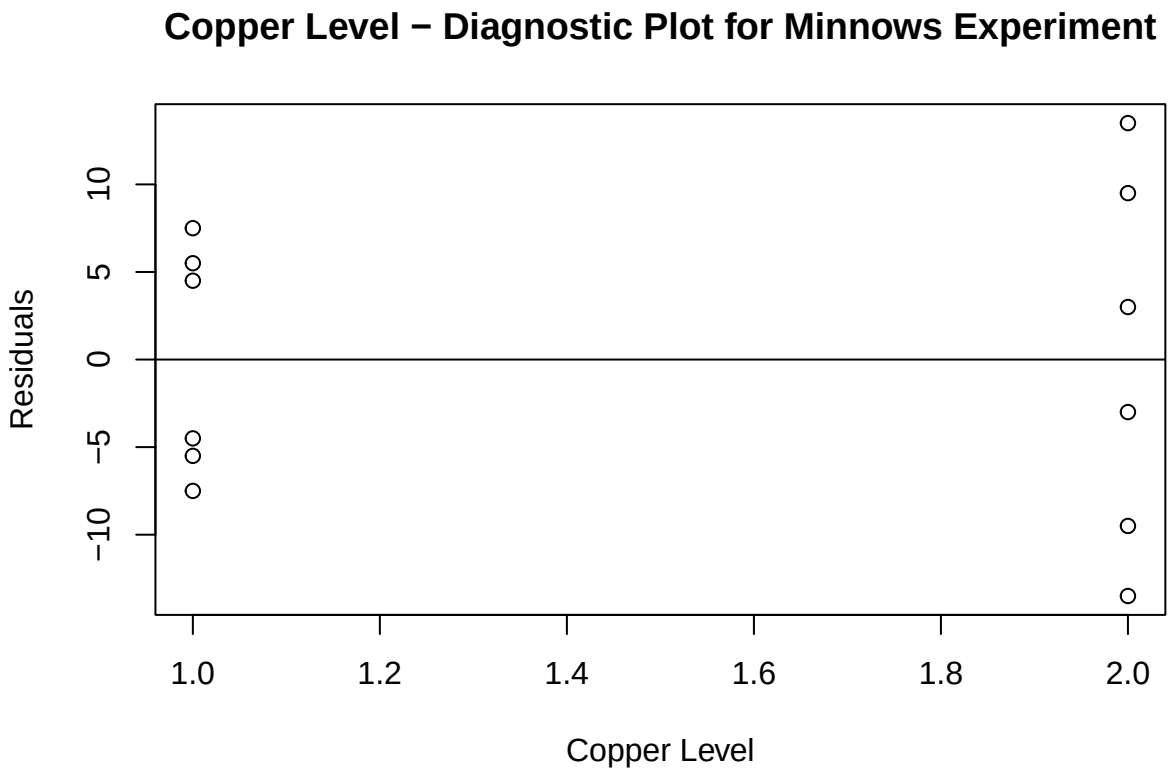
```
plot(sumto0.anova$fitted.values,
     sumto0.anova$residuals,
     main="Fitted Values - Diagnostic Plot for Minnows Experiment",
     xlab="Fitted Values",
     ylab="Residuals")
abline(h = 0)
```



a) predicted value: We overall have a fairly random spread of residuals (both positive and negative), though there is a slight "trumpet" pattern that emerges at we fit larger values of our response variable protein content (μg/tank). We may not in fact have constant variance. However, these values do tend to taper back (decrease their variability) for fitted values greater than 180; it just so happens that the spread/variability of residuals increases a fair amount between fitted values between 160-180.

In addition to potentially nonconstant variance, the data above looks symmetric across y=0. This is another indication that the residuals are not randomly spread across the plot, such that we may have reason to believe the additivity assumption may be violated.

```r
plot(as.numeric(minnows$copper),
     sumto0.anova$residuals,
     main="Copper Level - Diagnostic Plot for Minnows Experiment",
     xlab="Copper Level",
     ylab="Residuals")
abline(h = 0)
```



b) the copper concentrations: Similar to the above, we notice the "spread" or variability of our residuals appear to increase for the larger level (higher concentration) of Copper levels. Though this increase in spread/variability is not dramatic, at least not enough that I would be concerned given it is within the same order of magnitude, there nonetheless is something potentially problematic in the above plot.

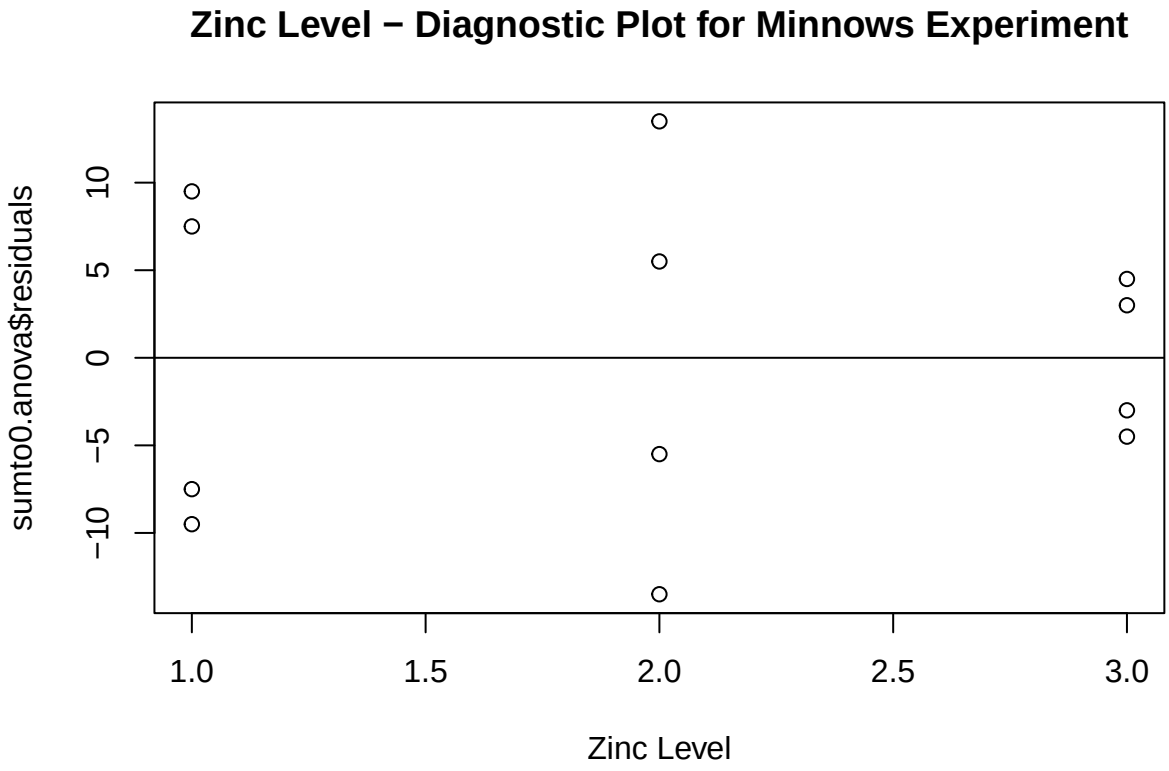
In addition to potentially nonconstant variance, the data above looks symmetric across y=0. This is another indication that the residuals are not randomly spread across the plot, such that we may have reason to believe the additivity assumption may be violated.

```r
plot(as.numeric(minnows$zinc),
     sumto0.anova$residuals,
     main="Zinc Level - Diagnostic Plot for Minnows Experiment",
     xlab="Zinc Level",
     dylab="Residuals")
```

```
## Warning in plot.window(...): "dylab" is not a graphical parameter

## Warning in plot.xy(xy, type, ...): "dylab" is not a graphical parameter
```

```
## Warning in axis(side = side, at = at, labels = labels, ...): "dylab" is not a
## graphical parameter
## Warning in axis(side = side, at = at, labels = labels, ...): "dylab" is not a
## graphical parameter

## Warning in box(...): "dylab" is not a graphical parameter

## Warning in title(...): "dylab" is not a graphical parameter
```

```
abline(h = 0)
```



c) the zinc concentrations: Similar to (a), we observe rather consistent spread/variability of residuals at the tails/extremes of the Zinc levels. However, near the "middle", or near the Zinc concentration levels of "2" or 750 concentration, we observe the residual spread increases a bit. Again, the overall magnitude of the spread of residuals at this particular Zinc level is similar and consistent with the other levels, but it does indicate something potentially problematic.

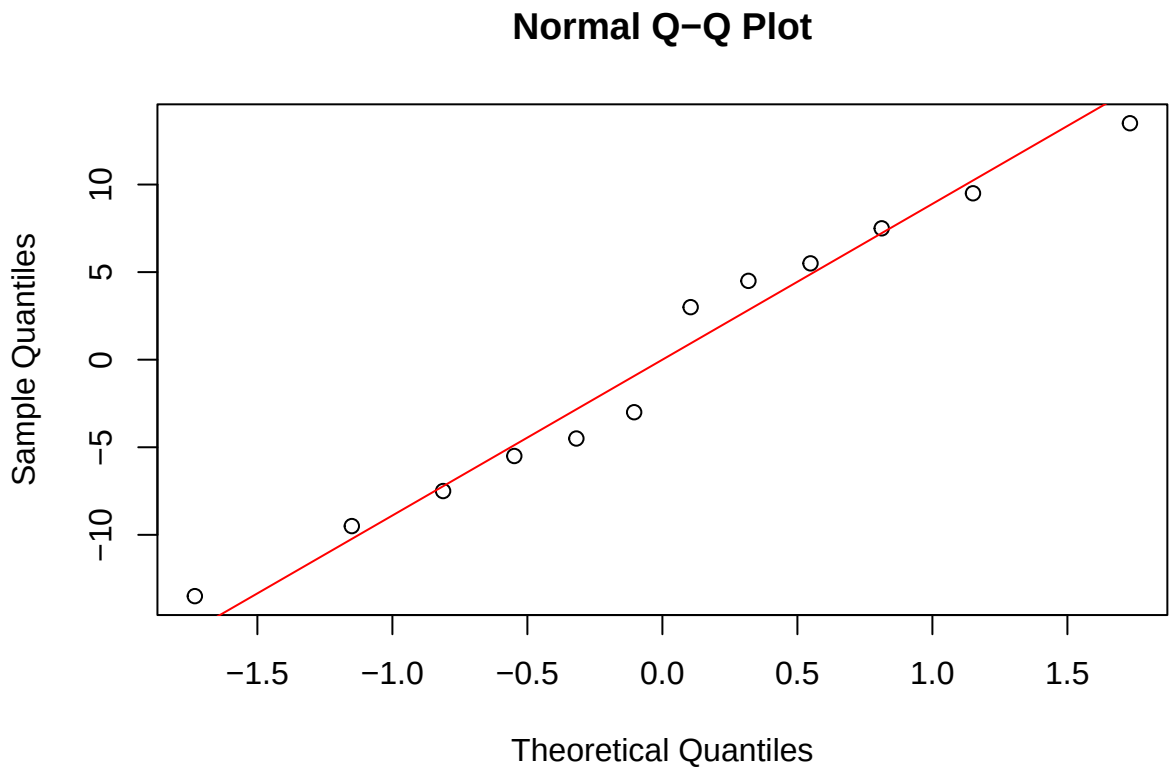
In addition to potentially nonconstant variance, the data above looks symmetric across y=0. This is another indication that the residuals are not randomly spread across the plot, such that we may have reason to believe the additivity assumption may be violated.

Overall, all three residual plots do not appear to have cause for concern regarding the constant variance assumption of the residuals (a component of the $\epsilon \sim N(0, \sigma^2)$ assumption). However, what's more cause for concern is that the residuals do not appear randomly distributed, indicating moreso that tthe addivity assumption may be violated in this experiment.

12.

Study the normal probability plot for the residuals. Is there anything of concern in this plot?

```
qqnorm(sumto0.anova$residuals)
qqline(sumto0.anova$residuals, col="red")
```




```
shapiro.test(sumto0.anova$residuals)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sumto0.anova$residuals
## W = 0.96538, p-value = 0.857
```

```
library(moments)
mean(sumto0.anova$residuals)
```

```
## [1] 2.590159e-16
```

```
median(sumto0.anova$residuals)
```

```
## [1] 3.330669e-15
```

```r
skewness(sumto0.anova$residuals)
```

```
## [1] -1.365054e-16
```

```r
kurtosis(sumto0.anova$residuals)-3
```

```
## [1] -1.171783
```

Reviewing the above QQ Plot to assess normality of residuals: We notice there is a rather subtle "S"-shaped curve to the plot. Despite this, it does appear that the residuals do closely follow along the reference line, with some additional deviations from the reference line near the tails of the quantiles. Overall, it does appear that the residuals are somewhat, or generally, normally distributed but there is some evidence that our assumption of normally distributed residuals may be violated since there are a number of points off from the reference line.

However, we have additional tests to consider and supplement the above QQ Plot interpretation. Namely: We have a large p-value which does not provide evidence for rejecting the null hypothesis that the residuals are normally distributed. Furthermore, summary statistics provide sample values of the mean and median of the residuals being very close (to one another as well as close to zero!) Additionally, when looking at the skewness and (excess) kurtosis of the residuals, we find that the skewness is very close to zero, and that we have slight negative excess kurtosis, or have evidence of residuals being slightly short-tailed. These results are consistent with the above interpretation of the QQ Plot, insomuch as there is not any glaring evidence of violating our assumption of normally distributed residuals.