Lab 8

2024-10-27

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
```

```
baseline.anova$coefficients
```

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

```
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"
```

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

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

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"
```

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

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

The dataset's structure confirms the full factorial design and the coding remains suitable for the two-factor analysis.

3.

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

This question is to calculate the six sample treatment means or the cell means.

```
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
```

```
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
```

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 there is no significant difference in the mean protein content (μg/tank) between the two copper levels (0 ppm and 150 ppm).

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 there is no significant difference in the mean protein content (μg/tank) between the three zinc levels levels (0, 750 or 1500 ppm). 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.

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
```

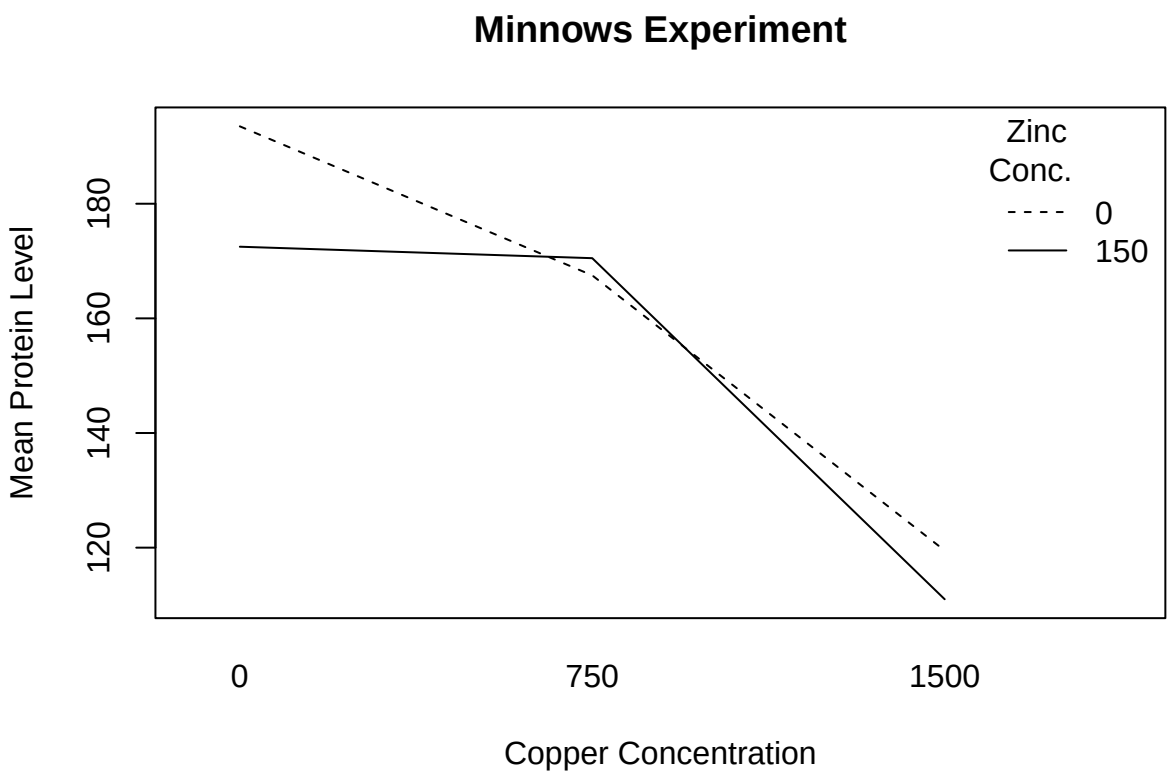
Interaction (Copper $\times$ Zinc): F-value: 1.113 p-value: 0.388 Interpretation: The p-value of 0.388 does not meet our threshold of the $\alpha = 0.05$ level, such that we do not have evidence to reject the null hypothesis that the interaction between copper and zinc concentrations does not significantly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

8.

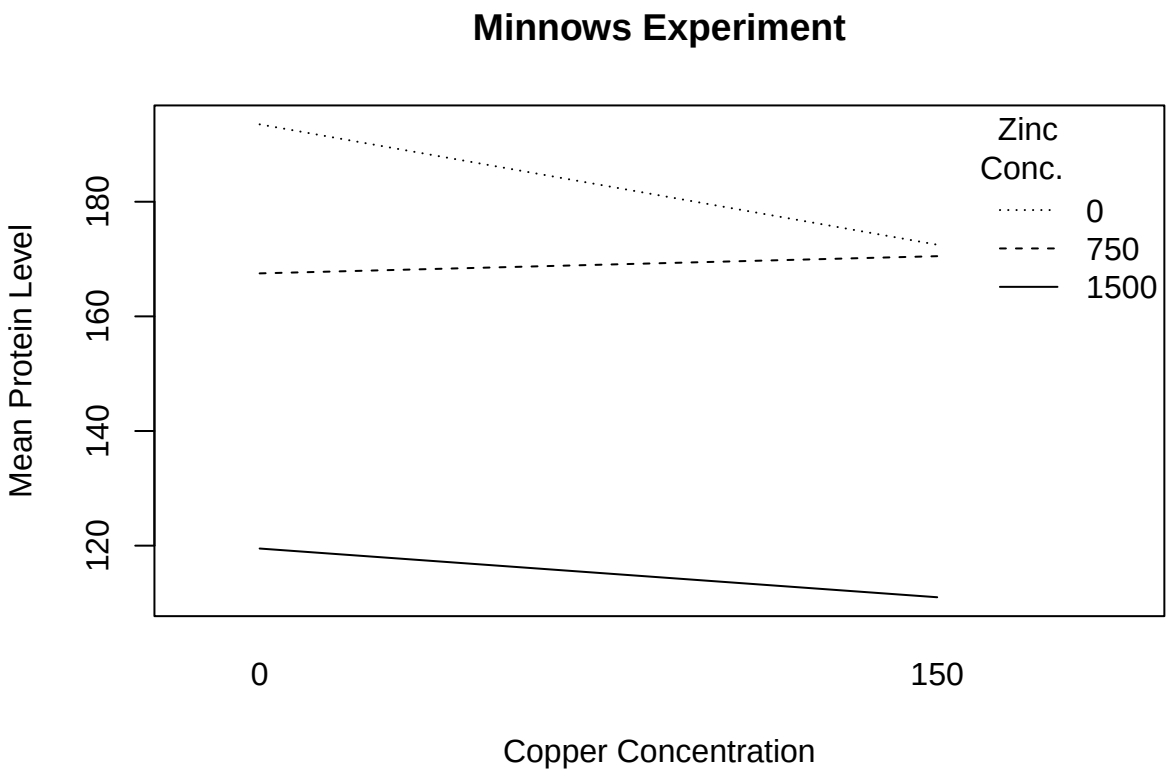
Study the interaction plot in the R output. Explain why, based on this plot, the interaction is not significant in the model.

```
interaction.plot(minnows$copper, minnows$zinc, minnows$protein, main="Minnows Experiment", xlab="Zinc C
```


```
interaction.plot(minnows$zinc, minnows$copper, minnows$protein, main="Minnows Experiment", xlab="Copper
```




```
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.")
```

Based on the interaction plots shown, it appears that the lines for different levels of zinc do not strongly deviate from being roughly parallel. In an interaction plot, if the lines representing different factor levels (in this case, copper and zinc) are nearly parallel, it indicates that there is little to no interaction effect.

Parallel lines suggest that the effect of one factor (e.g., zinc) on the response variable (protein content) is consistent across the levels of the other factor (copper). In other words, changing the copper concentration does not significantly alter the effect that zinc concentration has on the mean protein content.

When the interaction is significant, we would expect to see the lines crossing or diverging noticeably, indicating that the effect of zinc on protein content depends on the level of copper.

Therefore, the interaction is not significant in the model because the interaction plot indicates that the effect of zinc concentration on mean protein content remains relatively consistent regardless of the copper level. This visual evidence aligns with the statistical result showing a non-significant interaction term in the ANOVA table.

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. In the above instances, we have evidence to reject the null hypothesis at the alpha level 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 copper0 zinc0 - copper150 zinc1500: A simultaneous change of 150 units for copper with an increase of 1500 units of zinc copper150 zinc0 - copper0 zinc1500: copper150 zinc0 - copper150 zinc1500: copper0 zinc750 - copper0 zinc1500: copper0 zinc750 - copper150 zinc1500: copper150 zinc750 - copper0 zinc1500: copper150 zinc750 - copper150 zinc1500:

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
```

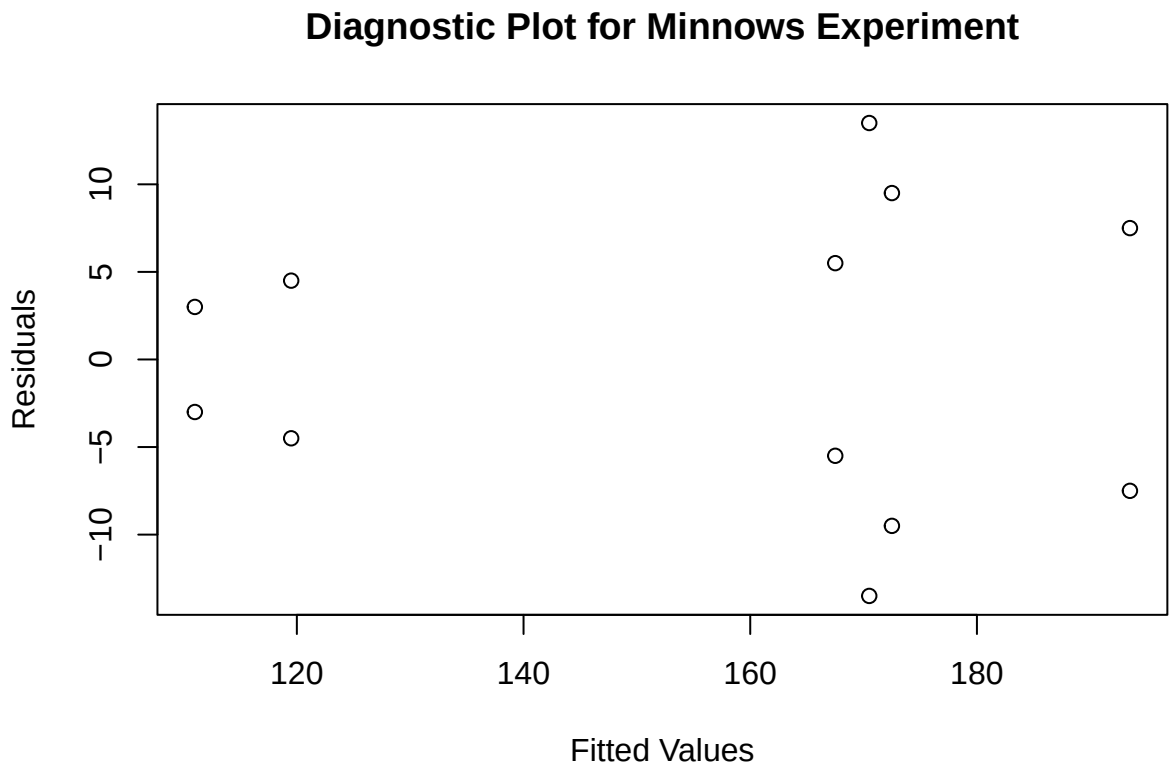
Within the contest of the study, the following two differences have a statistically significant difference at the $\alpha = 0.05$ level.

zinc0 - zinc1500 67.8 8.04 6 8.422 0.0004 zinc750 - zinc1500 53.8 8.04 6 6.682 0.0013

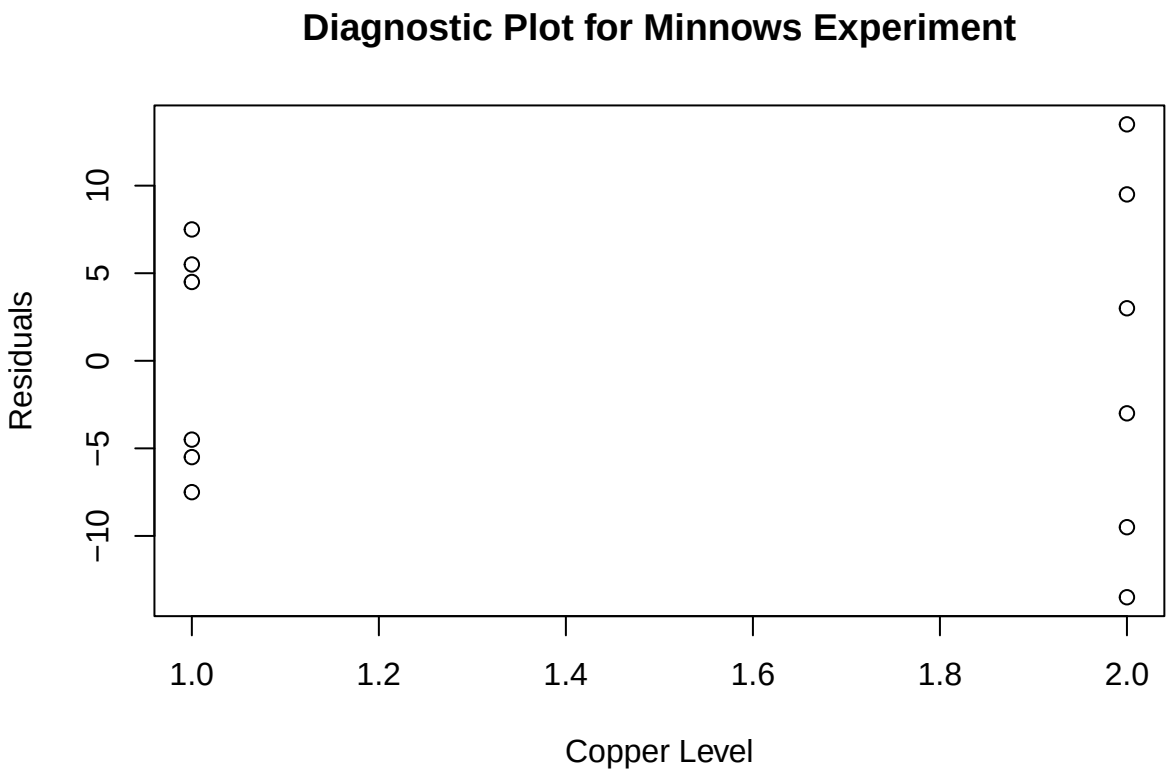
11.

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?

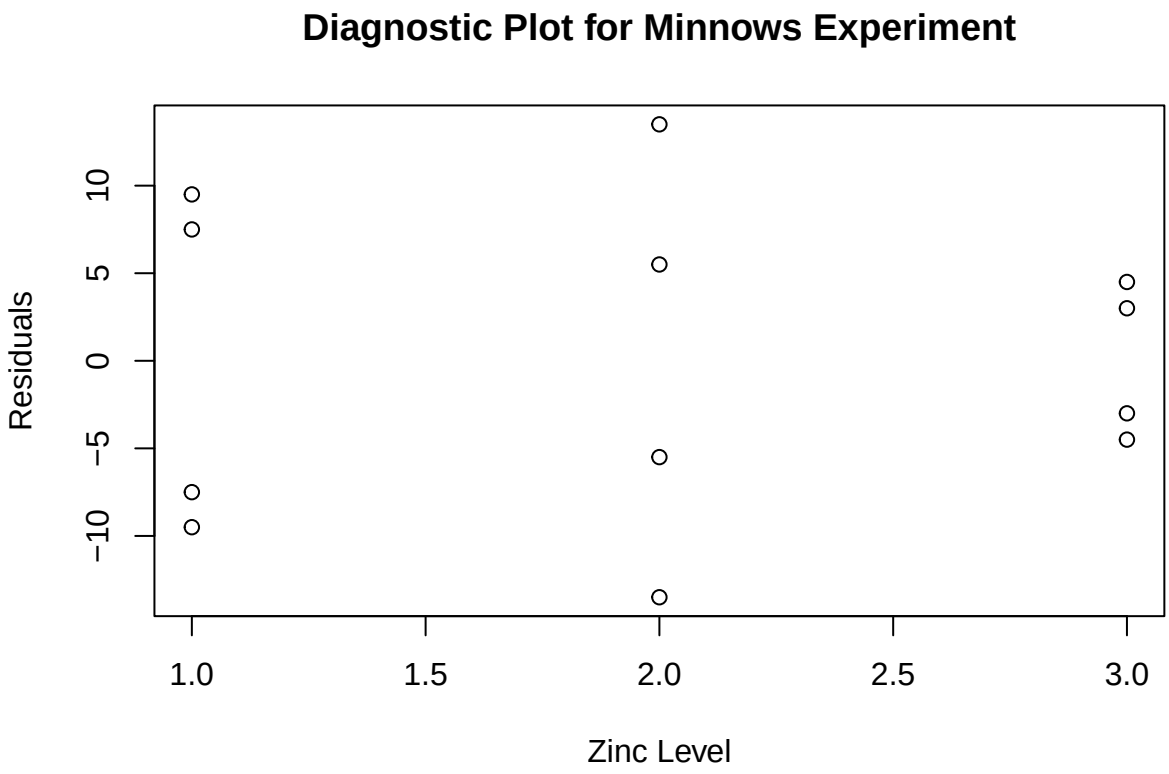
```
plot(sumto0.anova$fitted.values, sumto0.anova$residuals,
main="Diagnostic Plot for Minnows Experiment",
xlab="Fitted Values", ylab="Residuals")
```




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



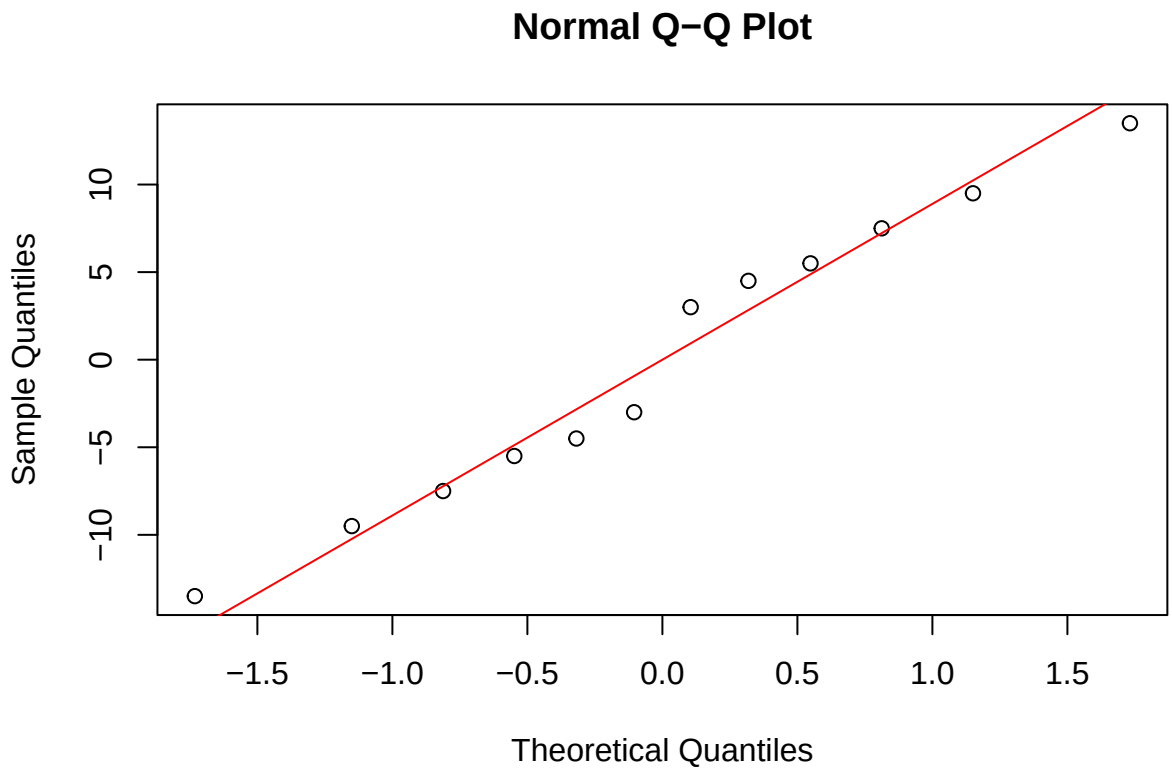

```
plot(as.numeric(minnows$zinc), sumto0.anova$residuals,
main="Diagnostic Plot for Minnows Experiment",
xlab="Zinc Level", ylab="Residuals")
```


Diagnostic Plot for Minnows Experiment
Residuals
10
5
0
−5
−10
1.0
1.5
2.0
2.5
3.0
Zinc Level

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
```