University of Toronto Mississauga

STA305 H5S - Winter 2020

Instructor: Dr. Luai Al Labadi

Final Project

Due on: April 8th, 2020

Last Name: Mansoor

First Name: Sarah

Student Number: 1004183251

EVALUATION

Total	Mark Earned
20	

Table of Contents:

Introduction

In this project I will cover 2 case studies. The first studies the levels of mercury in different lakes in Maine to investigate which characteristics of lakes lead to higher levels of mercury in the fish. I will discuss if lakes in Maine are of high concern, if the presence of dams increases the level of mercury, if mercury levels differ by lake type, and if mercury levels differ by lake types with dams and without dams. The second studies how vitamins affect weight gain in animals and if their caloric intake has anything to do with their weight gain. I will discuss if vitamins increase weight gain, if their caloric intake increases weight gain, and if the combination of the two increases their weight gain.

CASE STUDY I. Are the Fish Safe to Eat?

Lake: Name of the lake
Mercury: the mercury level in the fish in parts per million (ppm)
Dam: whether a dam is present or not
Type: 1. 'oligotrophic', 2. 'eutrophic', 3. 'mesotrophic'

Questions:

a) When looking at the boxplot of the data before [1] removing outliers and missing data we can see that there are 2 missing points in the data as well as 6 outliers. I had removed these outliers in the excel sheet and had the new data points. These can be seen in the boxplot of the data after [2] where the data is more spread apart and covers a lot more area now. There are still some outliers in this new boxplot, but they aren't of too much concern since they are still relatively close to the third quartile.

b) **Experimental unit(s):** A lake
Population(s): All lakes
Sample(s): 120 lakes in Maine
Variables: Whether a *Dam* is present or not (explanatory & categorical), Type of lake (explanatory & categorical), Level of mercury (response & quantitative)
Factor(s): Dam, Type of lake
Levels: *Dam* – There **is** a dam, There **is not** a dam

[1] Appendix # 1.A # Visual 1
[2] Appendix # 1.A # Visual 2

Type of lake – **'oligotrophic'** – sustains fish based on its vegetation and oxygen, **'eutrophic'** – has few fish, **'mesotrophic'** – in between oligotrophic and eutrophic

Treatments:

1. There is a dam & 'oligotrophic'
2. There is not a dam & 'oligotrophic'
3. There is a dam & 'eutrophic'
4. There is not a dam & 'eutrophic'
5. There is a dam & 'mesotrophic'
6. There is not a dam & 'mesotrophic'

Experimental or observational: This study is observational since there is no control or causation that the researches can conduct. They only observe what is already present.

Balanced or unbalanced:

Type 1: 17, Type 2: 52, Type 3: 43, Dam: 66, No dam: 46

Since each level of all the factors have different number of samples, the design is unbalanced.

Crossed or nested: Crossed, the design is complete since each treatment is available in our sample.

Treatment 1: 11, Treatment 2: 6, Treatment 3: 29, Treatment 4: 23, Treatment 5: 26, Treatment 6: 17

Since each treatment has a different number of samples, the design is unbalanced. But since each treatment

Confounding Variable(s): How much water is in the dam (this could differ by lake). How big the lake is (this could differ by lake). We could take these things into account and measure them and construct blocking variables.

c) Let's take a look at our first question of interest: are mercury levels in Maine high enough to be of concern? To test this, I will use the t.test in R. As seen in my code [3] I tested mercury levels with $H_o: \mu \leq 0.5 \; vs \; H_a: \mu > 0.5$. The p-value result was 0.982 so we fail to reject H_o with $\propto = 0.05$. This suggests that the mercury levels in Maine are not higher than 0.5 ppm since our test yields $\mu \leq 0.5$. If we look at the mean of mercury levels for all the lakes sampled, I get 0.4466607 < 0.5, this also suggests that the mercury levels are not high enough to be of concern.

d) Now we will take a look at the 2nd question of interest: Are high levels of mercury in fish related to the presence of dams? To test this question, I will conduct a test on $H_o: \beta_{dam} = 0 \; vs \; H_a: \beta_{dam} \neq 0$. From my code [4], I created a model of the response variable (mercury) in relation to the explanatory variable dam. I then used a One-Way ANOVA to find the p-value associated to the levels of mercury to the presence of dams. This resulted in: p-value

[3] Appendix # 1.C
[4] Appendix # 1.D

= 0.05549 which is > 0.05, the alpha value. I can now fail to reject H_o and say that $\beta_{dam} = 0$ which means that the presence of dams does not result in higher levels of mercury. Since we FTR H_o a post-hoc test is not necessary. Therefore, industries that benefit from dams should not be concerned.

e) The 3rd question of interest is if mercury levels vary by lake type. To test this question, I will conduct a test on $H_o: \beta_{lake\ type} = 0\ vs\ H_a: \beta_{lake\ type} \neq 0$. From my code [5], I created a model of the response variable (mercury) in relation to the explanatory variable lake types. I then used a completely randomized design to test this and used the ANOVA method to find a p-value = 0.0112 < 0.05, the alpha value. This suggests that we can reject H_o and conclude that lake types do have an influence on the levels of mercury in fish. The question now is how is that influence depending on the lake type. To test this, I conducted a post-hoc test using Tukeys HSD method. This tested the pairwise comparisons of the lake types to see which comparison was most significant. From the output [6], I can see that the most significant comparison is that of Lake Type 2 and Lake Type 1. This comparison has a p-value of 0.009 < 0.05. The other comparisons had a p-value > 0.05. This suggests that lake type 2 and lake type 1 have the most significant difference in the levels of mercury that are in the fish in those lakes. This would make sense since lake type 1 is more natural than lake type 2, and lake type 3 is a combination of the two.

f) Now I will discuss our last question of interest: do mercury levels vary by lake type differently for lakes with dams than for lakes without dams. $H_o: \beta_{lake\ type(s)\ with\ dam} = \beta_{lake\ type(s)\ without\ dam} = 0\ vs. H_a: \beta_{lake\ type(s)\ with\ dam} \neq 0$ or $\beta_{lake\ type(s)\ without\ dam} \neq 0$. From my code[7], I created a model that took the response variable (mercury) in relation to all the treatments of the two factors (lake types and dams). Conducting a Two-Way ANOVA results in a p-value of Dam : Type = 0.09908 which is > 0.05, the alpha value. This suggests that there is no interaction effect between the two factors. This means that mercury levels do not vary by lake type differently for lakes with dams and for lakes without dams.

g) If we take a look at the 3rd question of interest again, we can use dams as a blocking factor, this is because the dam's factor can be looked at as a nuisance variable, meaning it might have some effect on the response. This way we can control and reduce experimental error. From my code [8] , I created two models. One was a completely randomized (CR) design model with only the lake type factor and the other was a randomized complete block design

[5] Appendix # 1.E
[6] Appendix # 1.E
[7] Appendix # 1.F
[8] Appendix # 1.G

(RCB). Using the ANOVA method, I found that the p-value of lake types in the CR design is 0.0112 which is of significance and the p-value of lake types in the RCB design is 0.1465 which is also of significance. Blocking gave slightly more power since the variation of lake type is more than expected. If I look at the MSE of the two designs, I can also notice that blocking decreased the MSE and resulted in a more powerful hypothesis test.

$$\widehat{RE} = \frac{(df_{RCB}+1)(df_{CR}+3)}{(df_{RCB}+3)(df_{CR}+1)} \cdot \frac{(b-1)MS_B + b(a-1)MSE_{RCB}}{(ab-1)MSE_{RCB}}$$

$$= \frac{(108+1)(109+3)}{(108+3)(109+1)} \cdot \frac{(2-1)0.41762 + 2(3-1)0.06366}{(3\cdot 2-1)0.06366} = 0.998362 \cdot \frac{0.67226}{0.3182}$$

$$= 2.109235$$

Since the ratio is greater than 1, the CR has a larger error variance and the block design is more efficient.

CASE STUDY II. Vitamin May Affect the Weight Gain?

a) **Experimental unit(s):** A laboratory animal
Population(s): All animals
Sample(s): 20 laboratory animals
Variables: Vitamin supplement (explanatory, categorical), Calorie intake (explanatory, quantitative), weight gain (response, quantitative)
Factor(s): Vitamin supplement, calorie intake
Levels:

Vitamin supplement –

A: 5, B: 5, C: 5, D: 5

Treatments:

Vitamin supplement –

A: 5, B: 5, C: 5, D: 5

Experimental or observational: Experimental, there is causation
Balanced or unbalanced: Balanced, since every level has equal amount of samples
Crossed or nested: Nested, since the factors for each treatment do not intersect. The design is complete and balanced since there is an equal number of samples for all vitamin levels and treatments.
Confounding variables: The health of the laboratory animal could affect their weight gain (whether they are sick or not), the age could affect, how much water the animal drinks. We could take these things into account and create blocking variables to control these variables.

b) Parallel planes model dummy coding (supplement A as reference group)

$y_{ij} = \beta_0 + \beta_1 x + \beta_2 I_{B,j} + \beta_3 I_{C,j} + \beta_4 I_{D,j} + \varepsilon_{i,j}$

$y_{ij} : response$

$x : calorie\ intake$

$\beta_0 : mean\ of\ Group\ A$
$\beta_1 : effect\ of\ caloric\ intake\ on\ weight\ gain$
$\beta_2 : difference\ between\ mean\ of\ Group\ B\ and\ Group\ A$
$I_{B,j} : indicator\ of\ Group\ B\ (whether\ jth\ item\ is\ in\ Group\ B\ or\ not)$
$\beta_3 : difference\ between\ mean\ of\ Group\ C\ and\ Group\ A$
$I_{C,j} : indicator\ of\ Group\ C\ (whether\ jth\ item\ is\ in\ Group\ C\ or\ not)$
$\beta_4 : difference\ between\ mean\ of\ Group\ D\ and\ Group\ A$
$I_{D,j} : indicator\ of\ Group\ C\ (whether\ jth\ item\ is\ in\ Group\ D\ or\ not)$
$\varepsilon_{i,j} : error\ term\ for\ the\ jth\ item\ in\ the\ ith\ group$

Vitamin	$I_{B,j}$	$I_{C,j}$	$I_{D,j}$	$\beta_0 + \beta_1 x + \beta_2 I_{B,j} + \beta_3 I_{C,j} + \beta_4 I_{D,j}$
A	0	0	0	$\beta_0 + \beta_1 x$
B	1	0	0	$(\beta_0 + \beta_2) + \beta_1 x$
C	0	1	0	$(\beta_0 + \beta_3) + \beta_1 x$
D	0	0	1	$(\beta_0 + \beta_4) + \beta_1 x$

c) in terms of betas what null hypothesis can test the following:

i. $H_o: \beta_2 = \beta_3 = \beta_4 = 0$

ii. $H_o: \beta_1 = 0$

iii. $H_o: \beta_2 = \beta_3 + \beta_4$

d) parallel planes model but allowing for the effect of caloric intake on weight gain to vary by types of vitamin supplements. Use dummy coding again.

$y_{ij} = \beta_0 + \beta_1 x + \beta_2 I_{B,j} + \beta_3 x I_{B,j} + \beta_4 I_{C,j} + \beta_5 x I_{C,j} + \beta_6 I_{D,j} + \beta_7 x I_{D,j} + \varepsilon_{i,j}$
$y_{ij} : response$
$x : caloric\ intake$
$\beta_0 : mean\ of\ Group\ A$
$\beta_1 : effect\ of\ caloric\ intake\ on\ weight\ gain$
$\beta_2 : difference\ between\ mean\ of\ Group\ B\ and\ Group\ A$
$I_{B,j} : indicator\ of\ Group\ B\ (whether\ jth\ item\ is\ in\ Group\ B\ or\ not)$
$\beta_3 : effect\ of\ caloric\ intake\ on\ weight\ gain\ by\ Vitamin\ B$
$\beta_4 : difference\ between\ mean\ of\ Group\ C\ and\ Group\ A$
$I_{C,j} : indicator\ of\ Group\ C\ (whether\ jth\ item\ is\ in\ Group\ C\ or\ not)$
$\beta_5 : effect\ of\ caloric\ intake\ on\ weight\ gain\ by\ Vitamin\ C$
$\beta_6 : difference\ between\ mean\ of\ Group\ D\ and\ Group\ A$
$I_{D,j} : indicator\ of\ Group\ C\ (whether\ jth\ item\ is\ in\ Group\ D\ or\ not)$
$\beta_7 : effect\ of\ caloric\ intake\ on\ weight\ gain\ by\ Vitamin\ D$
$\varepsilon_{i,j} : error\ term\ for\ the\ jth\ item\ in\ the\ ith\ group$

Vitamin	$I_{B,j}$	$I_{C,j}$	$I_{D,j}$	$\beta_0 + \beta_1 x + \beta_2 I_{B,j} + \beta_3 x I_{B,j} + \beta_4 I_{C,j} + \beta_5 x I_{C,j} + \beta_6 I_{D,j} + \beta_7 x I_{D,j}$
A	0	0	0	$\beta_0 + \beta_1 x$
B	1	0	0	$(\beta_0 + \beta_2) + (\beta_1 + \beta_3)x$
C	0	1	0	$(\beta_0 + \beta_4) + (\beta_1 + \beta_5)x$
D	0	0	1	$(\beta_0 + \beta_6) + (\beta_1 + \beta_7)x$

e) Using d. in terms of betas what null hypothesis would you test for the following:

i. $H_o: \beta_3 = \beta_5 = \beta_7 = 0$

ii. $H_o: \beta_3 = \beta_5 = \beta_7 = 0$

iii. $H_o: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

f) The assumptions of ANCOVA is that there is a linear relationship between the response Y and the covariate x, the slopes of the regression lines in the groups are the same or if we can fit a model with interactions. From the regression lines in my code [9], I can see that the interaction lines are all linear and they do intersect at some points. If we conduct a test on the interactions, which is also in my code [10] to test the homogeneity, I find that the p-value = 0.7223 which suggests that there is homogeneity and the slopes are the same. Thus, the assumptions of ANCOVA are satisfied. I also conducted an equal variances test [11] on each of the groups. For this I had 4 tests: $H_0: {\sigma^2}_A = {\sigma^2}_{cov}\ vs. H_0: {\sigma^2}_A \neq {\sigma^2}_{cov}$, $H_0: {\sigma^2}_B = {\sigma^2}_{cov}\ vs. H_0: {\sigma^2}_B \neq {\sigma^2}_{cov}, H_0: {\sigma^2}_C = {\sigma^2}_{cov}\ vs. H_0: {\sigma^2}_C \neq {\sigma^2}_{cov}, H_0: {\sigma^2}_D = {\sigma^2}_{cov}\ vs. H_0: {\sigma^2}_D \neq {\sigma^2}_{cov}$. From the var.test of each of these, none of the p-values were less than the significance level. This means that we fail to reject the null hypothesis for all 4 tests and can conclude that the variances are equal. Thus, another ANCOVA assumption is satisfied.

g) Now I will test whether supplements influence weight gain or not. Using a one-way ANOVA, from my code [12], I created a model that took the response variable (weight gain) in relation to the vitamins. I will test this using, $H_o: \beta_1 = \beta_2 = \beta_3 = 0\ vs. H_a: at\ least\ one\ not\ equal\ to\ 0$. From this test, I find a p-value of 0.1728 > 0.05, the alpha value, so we cannot reject the null hypothesis and claim that there is no influence from vitamin supplements on weight gain in animals.

h) Now we can test is supplements influence weight gain, with the addition of the caloric intake. For this, I created [13] a model that took the response variable (weight gain) in relation

[9] Appendix # 2.F # Visual 3
[10] Appendix # 2.F
[11] Appendix # 2.F
[12] Appendix # 2.G
[13] Appendix # 2.H

to the vitamins and the covariate (caloric intake). I test this using: $H_o: \beta_2 = \beta_3 = \beta_4 = 0 \; vs. H_a: at\ least\ one\ not\ equal\ to\ 0$. Using the ANCOVA method, I find that the p-value = 0.005352 < 0.05. This suggests that vitamins, with the covariate of caloric intake, does actually influence the weight gain of the animal. This means that it is important to look at not only what vitamin the animal is receiving but also how much their caloric intake is. This is interesting since in part (g), my test resulted in me concluding that weight gain is not influenced by vitamin supplements, whereas now, with a blocking covariate variable, there is in fact very strong evidence that vitamins do influence weight gain.

i) Since we know that vitamins with the animal's caloric intake does influence the weight gain in the animal, we should also test the significance of the covariate. I can do this by using the same model as before, but now looking at the summary[14] of the model and the caloric variable. The summary yields a p-value of 0.00203 for my caloric variable, which is less than the significance level. This means that the caloric intake covariate is very significant in determining the weight gain of the animal. This was also seen when I compared the results from part (h) to part (g). Without the caloric intake covariate, the vitamins did not seem significant at all for the influence of weight gain, but when this covariate was looked at, we found that vitamins are significant in influencing weight gain. This can also be used to show that there is interaction between the covariate and the factor and that (even according to the regression plot) there will be different slopes for each vitamin since I had tested that the beta values will in fact not be 0, which also suggests that they will not be equal either.

j) If we look at the non-parallel model that was seen in part (d), we can also us this to test whether supplements do influence weight gain or not. For this part, I created[15] a model that took the response variable (weight gain) in relation to the vitamins with every covariate (caloric intake). I used the ANCOVA method to test: $H_o: \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0 \; vs. H_a: at\ least\ one\ not\ equal\ to\ 0$. From this I found that the vitamin variable had a p-value = 0.01182 < 0.05, thus we can reject the null hypothesis and conclude that vitamins do have an influence on weight gain. This is also significant since this was the same conclusion I had for part (h) and (i). From either the parallel-linear model or the non-parallel-linear model, we can conclude that the covariate has an effect on the weight gain of animals as well as the vitamin supplement that the animal receives.

Conclusion

In this project I covered 2 case studies. The first studied the levels of mercury in different lakes in Maine to investigate which characteristics of lakes lead to higher levels of mercury in the fish. I found that lakes in Maine do not have levels of mercury enough to be of high concern, the

[14] Appendix # 2.I
[15] Appendix # 2.J

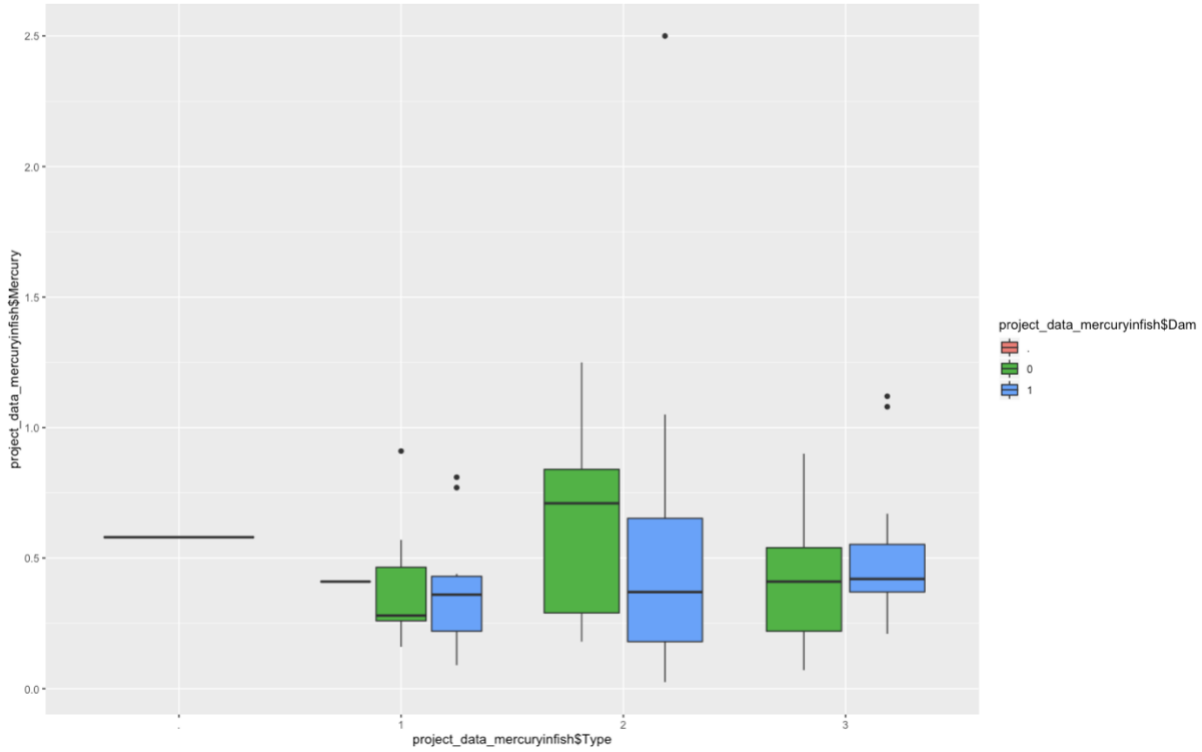
presence of dams does not increase the level of mercury in fishes, mercury levels do differ by lake type, and mercury levels do not differ by lake types with dams and without dams. The second studied how vitamins affect weight gain in animals and if their caloric intake has anything to do with their weight gain. I found that vitamins do not increase weight gain when looked at on their own, the animal's caloric intake can increase weight gain, and the combination of the two definitely increases their weight gain as I saw with multiple tests in this study.

Appendix

 # Course: STA305
 # Final Project: # Case Study 1.A
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> require(ggplot2)
> ggplot(project_data_mercuryinfish, aes(x = project_data_mercuryinfish$Type, y =
project_data_mercuryinfish$Mercury
+                                , fill = project_data_mercuryinfish$Dam)) + geom_boxplot()
```

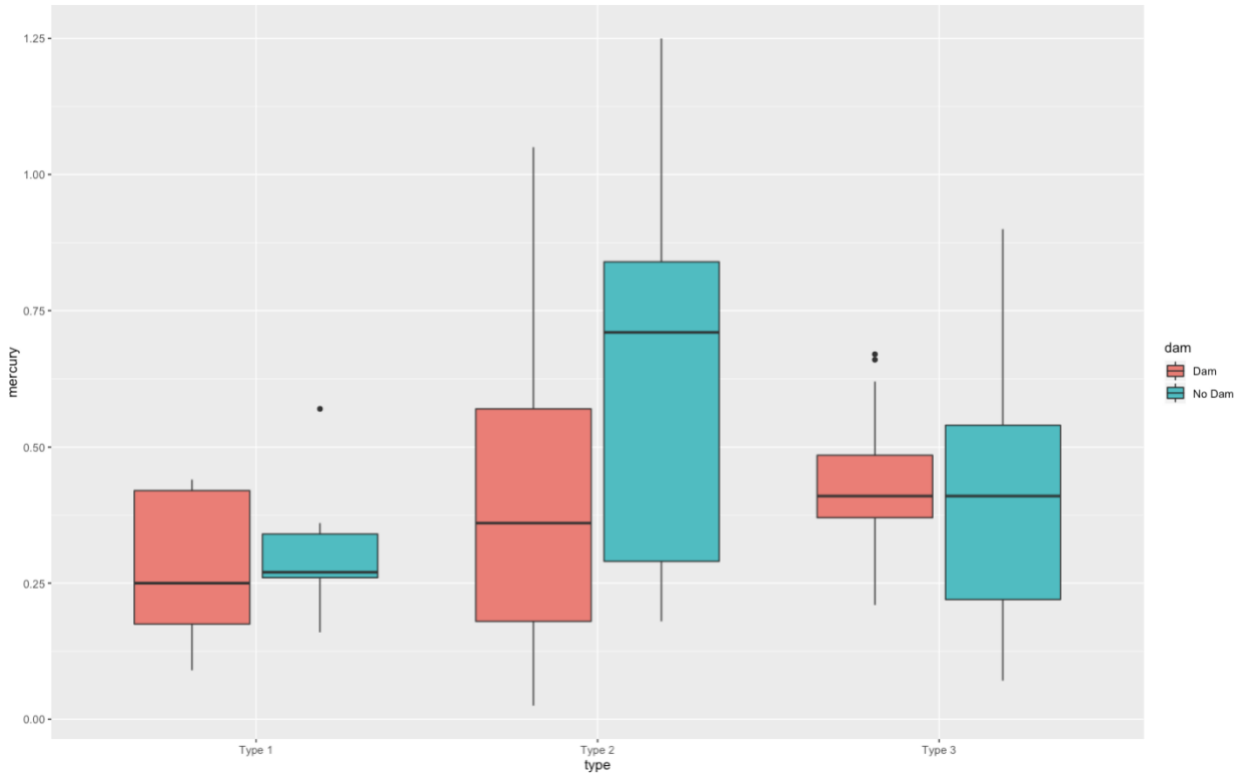
 # Visual 1




```
> y <- Mercury_edited
> a <- c(rep("Type 1", 6), rep("Type 2", 23), rep("Type 3", 17), rep("Type 1", 11),
+       rep("Type 2", 29), rep("Type 3", 26))
> b <- c(rep("No Dam", 46), rep("Dam", 66))
> lakes <- data.frame(mercury = y, type = a, dam = b)
```

> ggplot(lakes, aes(x = type, y - mercury, fill = dam)) + geom_boxplot()

 # Visual 2



 # Course: STA305
 # Final Project: # Case Study 1.C
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

> t.test(y, data = lakes, mu = 0.5, alternative = 'greater', paired = FALSE, var.equal = FALSE)

One Sample t-test

data: y
t = -2.1223, df = 111, p-value = 0.982
alternative hypothesis: true mean is greater than 0.5
95 percent confidence interval:
0.4049731 Inf
sample estimates:
mean of x
0.4466607

> mean(y)
[1] 0.4466607

 # Course: STA305
 # Final Project: # Case Study 1.D
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model_1 <- lm(mercury ~ dam, data = lakes)
> anova(model1)
Analysis of Variance Table

Response: mercury
           Df  Sum Sq  Mean Sq  F value  Pr(>F)
dam         1  0.2870  0.287047  3.7427  0.05549 .
Residuals 115  8.8198  0.076694
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

 # Course: STA305
 # Final Project: # Case Study 1.E
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model_2 <- lm(mercury ~ type, data = lakes)
> anova(model_2)
Analysis of Variance Table

Response: mercury
            Df  Sum Sq  Mean Sq  F value  Pr(>F)
type         2  0.6212  0.310616  4.6819  0.0112 *
Residuals  109  7.2315  0.066344
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(aov(model_2), conf.level = 0.95)
      Tukey multiple comparisons of means
        95% family-wise confidence level
Fit: aov(formula = model_2)
$type
                          diff         lwr         upr        p adj
Type 2-Type 1  0.21655543  0.04556431  0.38754655  0.0090547
Type 3-Type 1  0.13544049 -0.03990398  0.31078497  0.1629587
Type 3-Type 2 -0.08111494 -0.20726894  0.04503907  0.2819582
```

 # Course: STA305

 # Final Project: # Case Study 1.F
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model_3 <- lm(mercury ~ dam * type, data = lakes)
> anova(model_3)
Analysis of Variance Table

Response: mercury
          Df Sum Sq Mean Sq F value  Pr(>F)
dam        1 0.4176 0.41762  6.7253 0.01085 *
type       2 0.5593 0.27965  4.5034 0.01327 *
dam:type   2 0.2935 0.14674  2.3630 0.09908 .
Residuals 106 6.5823 0.06210
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

 # Course: STA305
 # Final Project: # Case Study 1.G
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

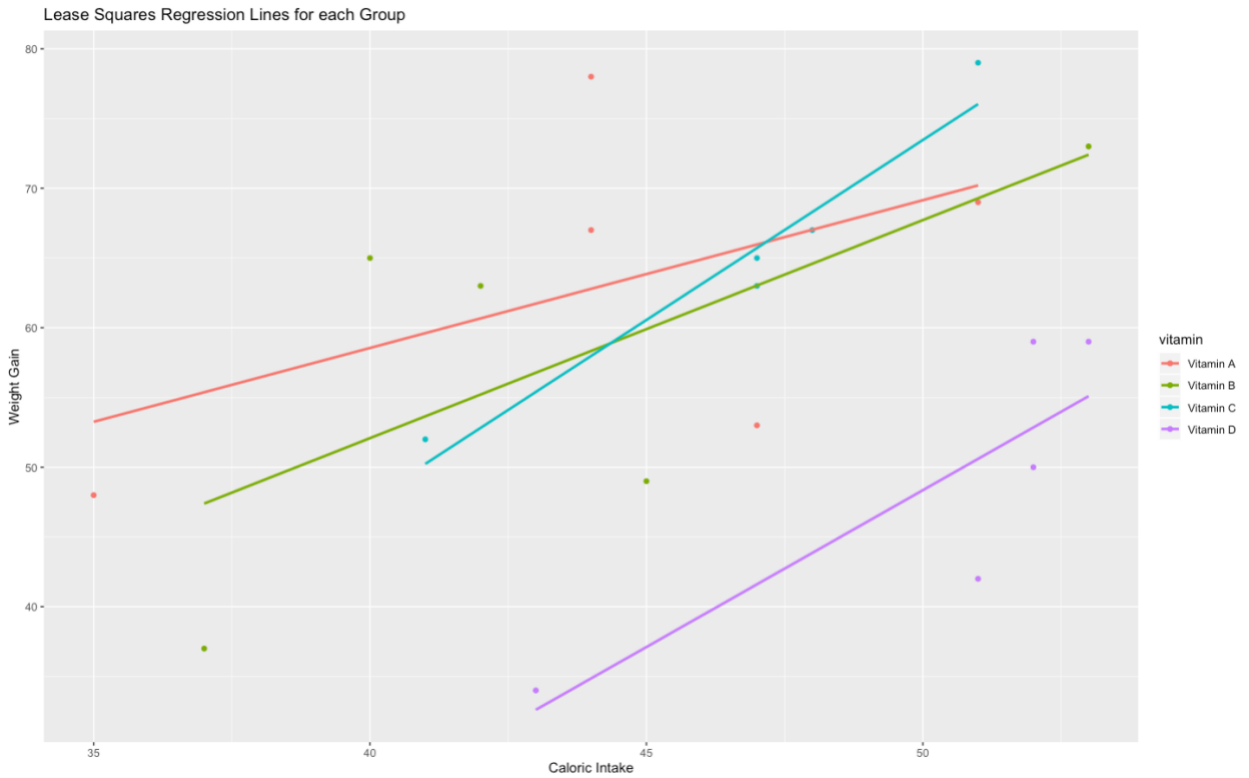
```
> modelCR <- lm(mercury ~ type, data = lakes)
> anova(modelCR)
Analysis of Variance Table

Response: mercury
          Df Sum Sq  Mean Sq F value Pr(>F)
type       2 0.6212 0.310616  4.6819 0.0112 *
Residuals 109 7.2315 0.066344
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> modelRCB <- lm(mercury ~ dam + type, data = lakes)
> anova(modelRCB)
Analysis of Variance Table

Response: mercury
          Df  Sum Sq  Mean Sq  F value   Pr(>F)
dam        1  0.4176  0.41762  6.5597  0.01181 *
type       2  0.5593  0.27965  4.3926  0.01465 *
Residuals 108 6.8758 0.06366
---
```

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

 # Course: STA305
 # Final Project: # Case Study 2.F
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> y <- y_II
> a <- factor(rep(c("Vitamin A", "Vitamin B", "Vitamin C", "Vitamin D"), each = 5))
> x <- x_II
> animals <- data.frame(weight_gain = y, vitamin = a, caloric = x)
> require(ggplot2)
> qplot(caloric, weight_gain, data = animals, colour = vitamin, main = "Lease Squares
+     Regression Lines for each Group", xlab = "Caloric Intake", ylab = "Weight Gain") +
+     geom_smooth(method = "lm", se = FALSE)
```

 # Visual 3




```
> homo <- aov(weight_gain ~ caloric * vitamin, data = animals)
> summary(homo)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
caloric	1	391.1	391.1	4.430	0.0571 .
vitamin	3	1501.0	500.3	5.667	0.0118 *
caloric:vitamin	3	119.1	39.7	0.449	0.7223

```
Residuals        12  1059.6     88.3
---
  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> y1 <- y[0:5]
> y2 <- y[6:10]
> y3 <- y[11:15]
> y4 <- y[16:20]
> x1 <- x[0:5]
> x2 <- x[6:10]
> x3 <- x[11:15]
> x4 <- x[16:20]
> var.test(x1, y1)

        F test to compare two variances

data:  x1 and y1
F = 0.23056, num df = 4, denom df = 4, p-value = 0.1843
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.02400584 2.21446636
sample estimates:
ratio of variances
         0.2305648
> var.test(x2, y2)

        F test to compare two variances

data:  x2 and y2
F = 0.1783, num df = 4, denom df = 4, p-value = 0.1235
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01856398 1.71247115
sample estimates:
ratio of variances
         0.1782983
> var.test(x3, y3)

        F test to compare two variances

data:  x3 and y3
```

```
F = 0.14163, num df = 4, denom df = 4, p-value = 0.08471
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01474626 1.36029822
sample estimates:
ratio of variances
        0.1416309
> var.test(x4, y4)

        F test to compare two variances

data:  x4 and y4
F = 0.14069, num df = 4, denom df = 4, p-value = 0.08377
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01464838 1.35126916
sample estimates:
ratio of variances
        0.1406908
```

 # Course: STA305
 # Final Project: # Case Study 2.G
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model1 <- lm(weight_gain ~ vitamin, data = animals)
> anova(model1)
Analysis of Variance Table

Response: weight_gain
           Df Sum Sq Mean Sq F value Pr(>F)
vitamin     3  802.0  267.33  1.8853 0.1728
Residuals 16 2268.8  141.80
```

 # Course: STA305
 # Final Project: # Case Study 2.H
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model2 <- lm(weight_gain ~ caloric + vitamin, data = animals)
> anova(model2)
Analysis of Variance Table
```

```
Response: weight_gain
            Df  Sum Sq Mean Sq  F value  Pr(>F)
caloric      1   391.13  391.13   4.9778  0.041361 *
vitamin      3  1501.05  500.35   6.3678  0.005352 **
Residuals 15  1178.62   78.57
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

 # Course: STA305
 # Final Project: # Case Study 2.I
 # Last Name: Mansoor, First Name: Sarah
 # St. # : 1004183251

```
> model2 <- lm(weight_gain ~ caloric + vitamin, data = animals)
> summary(model2)

Call:
lm(formula = weight_gain ~ caloric + vitamin, data = animals)

Residuals:
    Min      1Q  Median      3Q     Max
-14.5792  -4.0661  -0.3448   5.9484  15.3271

Coefficients:
                  Estimate   Std.Error  t value  Pr(>|t|)
(Intercept)       -9.2859    19.8073  -0.469  0.64595
caloric            1.6354     0.4391   3.725  0.00203 **
vitaminVitamin B  -4.2917     5.6172  -0.764  0.45671
vitaminVitamin C  -2.0521     5.7213  -0.359  0.72483
vitaminVitamin D -24.0126     6.1943  -3.877  0.00149 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 8.864 on 15 degrees of freedom
Multiple R-squared:  0.6162,  Adjusted R-squared:  0.5138
F-statistic:  6.02 on 4 and 15 DF,  p-value: 0.004273
```

 # Course: STA305
 # Final Project: # Case Study 2.J
 # Last Name: Mansoor, First Name: Sarah

 # St. # : 1004183251

```
> model3 <- lm(weight_gain ~ caloric * vitamin, data = animals)
> anova(model3)
Analysis of Variance Table

Response: weight_gain
                Df  Sum Sq Mean Sq F value  Pr(>F)
caloric          1  391.13  391.13  4.4297 0.05707 .
vitamin          3 1501.05  500.35  5.6667 0.01182 *
caloric:vitamin  3  119.06   39.69  0.4495 0.72229
Residuals       12 1059.56   88.30
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```