DATA 3441: Assignment 2

1. (IPS10-2.2)

(a)

The explanatory variable is the price per load of the detergent because it is independent of a person's rating. The response variable is the rating of the laundry because it is measured on a rated scale, which would make it dependent on the price of detergent.

(b)

The explanatory variable is the day of the week because those are fixed values and cannot change, whereas the response variable is the amount of time spent studying which is continuous.

(c)

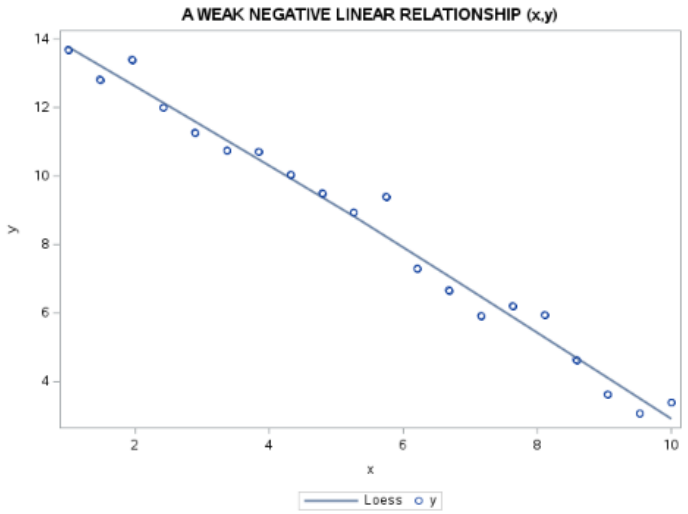
The explanatory variables are the children's age group because their age is fixed and cannot be readily changed (unless there was a longitudinal study and we were recording the years over time), and the response variable is the amount of calcium taken by the children, which is dependent on their age group.

(d)

The explanatory variable is the number of alcoholic drinks consumed by a person because that is something an experimenter can control for, and the response variable is the blood alcohol levels because they are completely dependent on whether or not a person takes a drink or not.

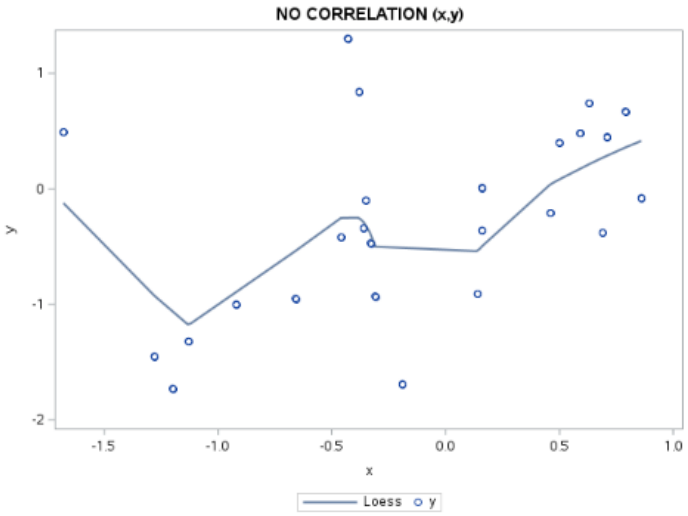
2. (IPS10-2.6)

(a)



[1]

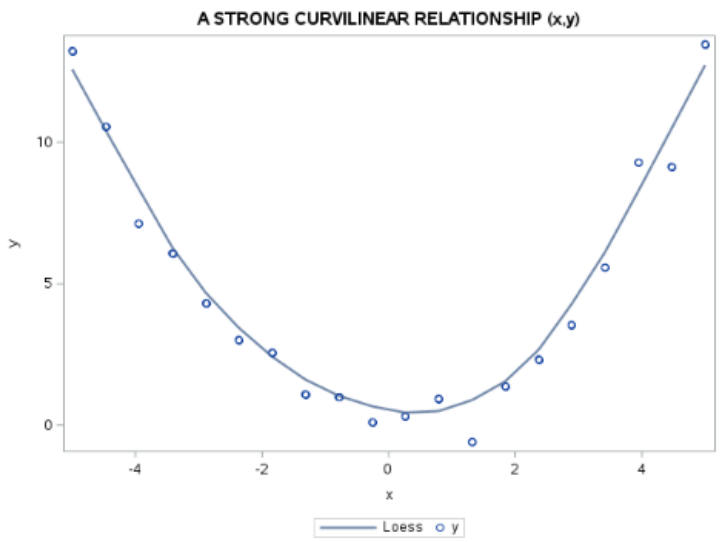
(b)



[2]

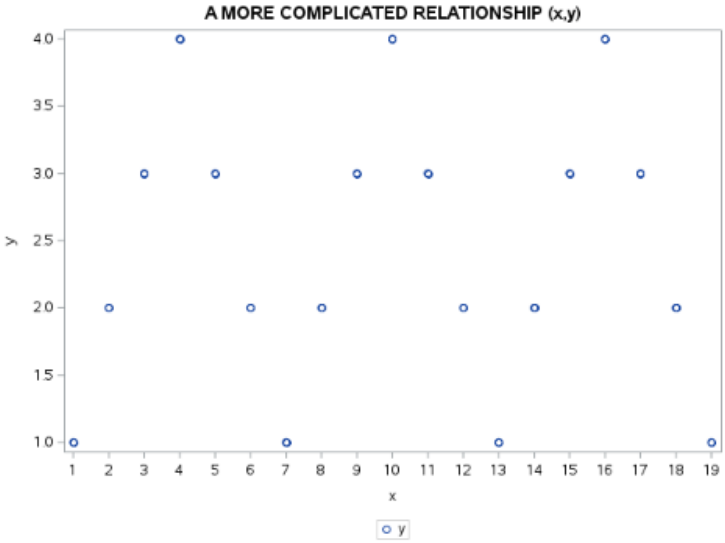
[1] Scatter plot of a weak negative relationship.
[2] Scatter plot of no relationship.

(c)



[3]

(d)



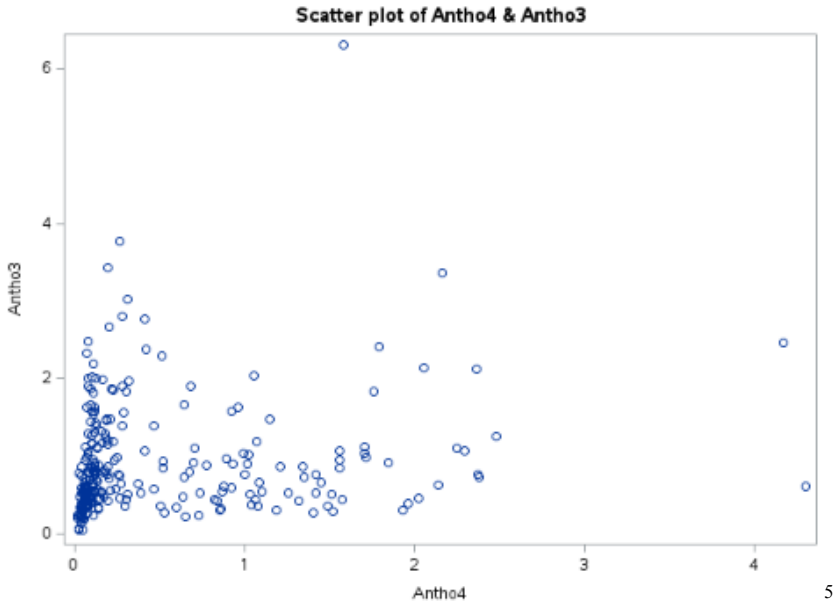
[4]

The data follows a cosine-like pattern, indicating that x and y are occurring at equal intervals. So not much change other than when the period starts and restarts again.

[3] Scatter plot of a strong positive non-linear relationship.
[4] Scatter plot of a complicated relationship.

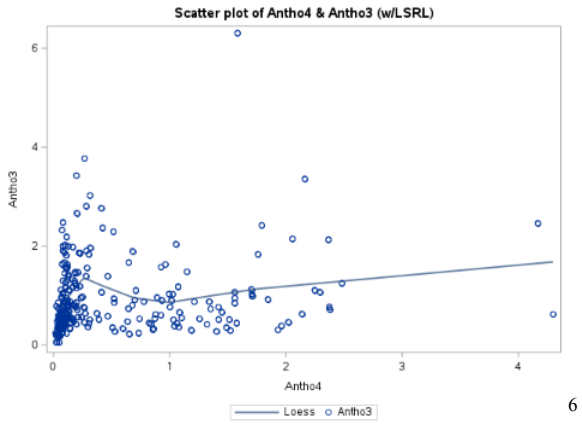
3. (IPS10-2.8)

(a)



[5]

(b)



[6]

The form is a dense cloud, its direction is positive, and I think the strength is weak.

[5] Scatter plot between Antho4 & Antho3.
[6] Scatter plot between Antho4 & Antho3 with a regression line added.

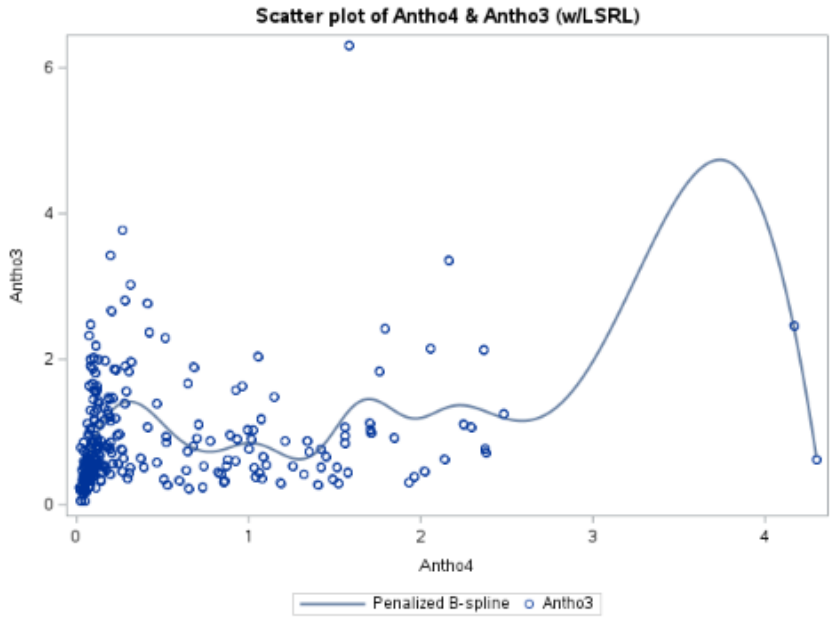
(c)

There are three outliers, one at an extreme y-value of about 6; two others are more than the x-value 4.

(d)

A straight line is not useful. When looking at the regression line, it is erratic at the start of the graph, then it evens out as x approaches 4, so if we were to make a guess about the significance of the relationship between Antho4 and Antho3, it would be difficult.

(e)



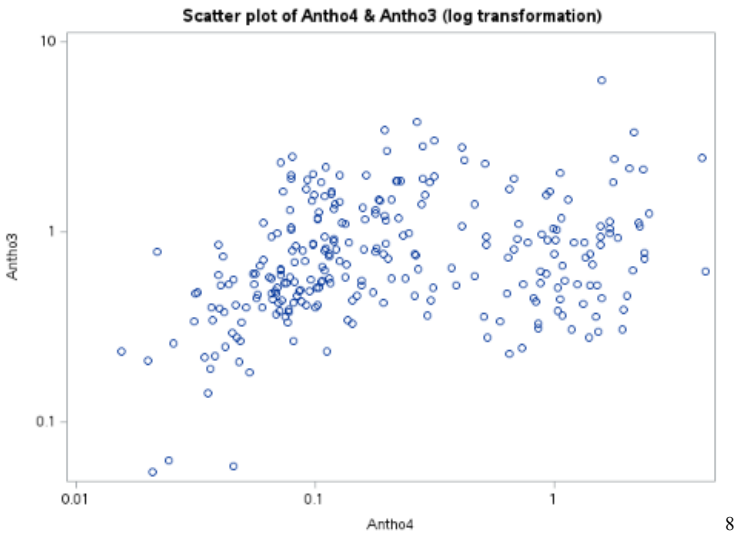
[7]

Interpreting the data is easier with smoothing. I know now that removing the two outliers at x>4 will change the shape, however, I think the probability would be low if we were to look at p-values or correlations.

[7] Scatter plot between Antho4 & Antho3 with a smoothed regression line added.

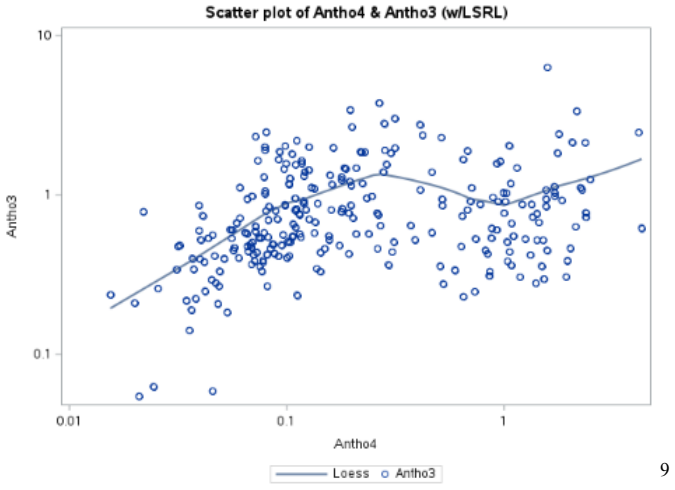
4. (IPS10-2.9)

(a)



[8]

(b)



[9]

The shape is less dense, its direction is positive, and the relationship looks stronger than the un-transformed graph.

[8] Scatter plot between Antho4 & Antho3 with a log transformation.
[9] Scatter plot between Antho4 & Antho3 with a log transformation and a regression line added.

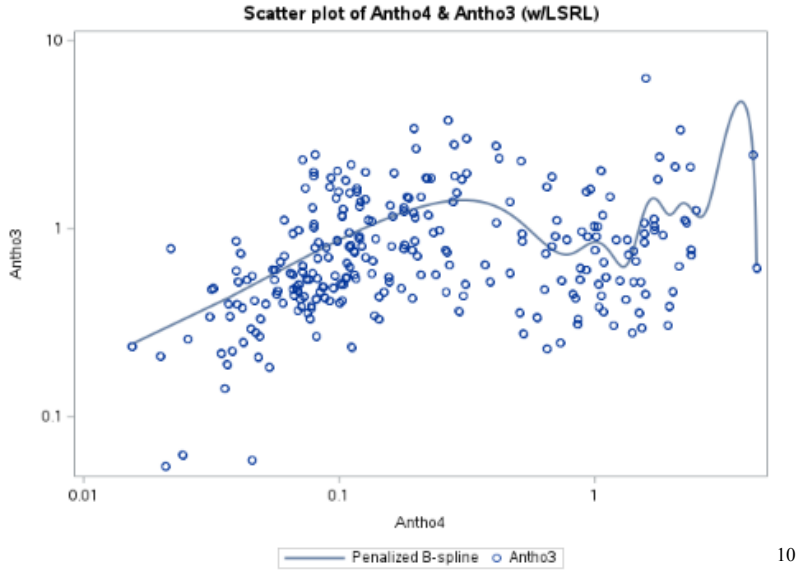
(c)

The outliers are now near the lower left side of the graph. What is unusual about this log transformation is that we can see a concentration of data points around Antho4 0.1 and Antho3 1.

(d)

I think it is not useful to add a straight line to the graph because the shape of the scatter follows different slopes. So, the line we draw won't be linear.

(e)



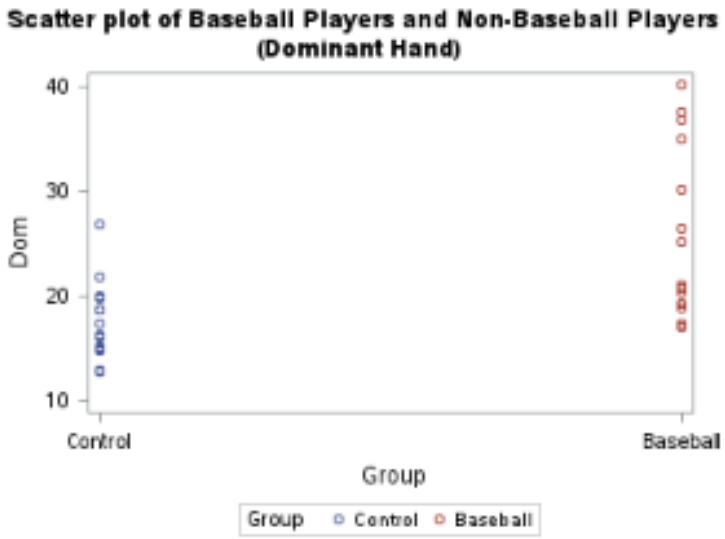
[10]

The transformed data lets us look at the data in a standard way. So now we can make a better assumption for the slope of the line for the interval 0.01 and 0.5; we can't make a great assumption for the intervals after because the relationship starts to become non-linear.

[10] Scatter plot between Antho4 & Antho3 with a log transformation and a smoothed regression line added.

5. (IPS10-2.16)

(a)



[11]

(b)

We can see from the graph that baseball players have a higher Newton meter ratio than non-baseball players, suggesting that they do indeed have stronger bones in their dominant hand. This decreases their risk for developing osteoporosis, compared to the non-baseball players. This confirms the hypothesis that more exercise leads to stronger bones, and in turn, reduces the risk of being diagnosed with osteoporosis.

[11] Scatter plot of Baseball players vs non-Baseball players (Dominant Hand).

6. (IPS10-2.30)

(a)

Correlation of Antho4 and Antho3

The CORR Procedure

2 Variables:	Antho4 Antho3

Simple Statistics

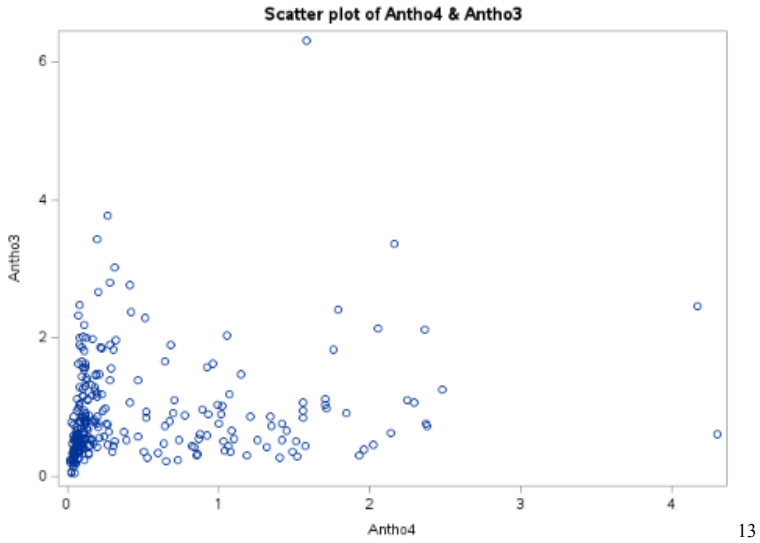
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Antho4	267	0.48370	0.67708	129.14800	0.01540	4.29950
Antho3	267	0.91505	0.72788	244.31900	0.05460	6.31100

Pearson Correlation Coefficients, N = 267
Prob > |r| under H0: Rho=0

	Antho4	Antho3
Antho4	1.00000	0.16324 0.0075
Antho3	0.16324 0.0075	1.00000

[12]

(b)



[13]

[12] Correlation table between Antho4 & Antho3.
[13] Scatter plot between Antho4 & Antho3.

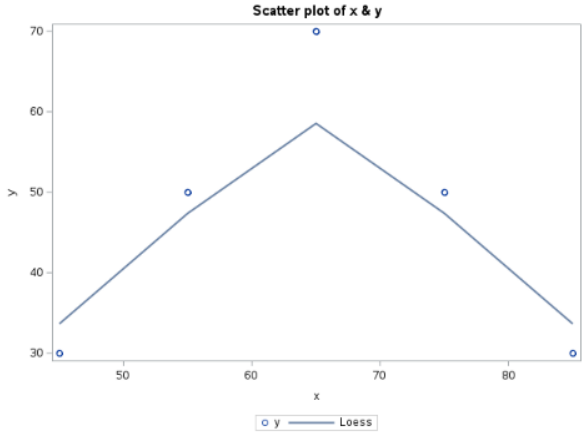
There is a significant relationship, $p < 0.05$, between Anth04 ($M = 0.48$, $STD = 0.67$) and Antho3 ($M = 0.91$, $STD = 0.72$) although it is really weak ($r = 0.16$) and accounts for about 2% of the variance. This correlation is a good numerical summary for the scatter plot because we see a clustering of points around the mean values for both variables.

(c)

The size of the correlation doesn't suggest the amounts are the same because these are predictions, not proven at all. We can look at the coefficient of determination to see that there is only 2% of variance for the data, the pearson's correlation is weak ($r = 0.16$), and the two mean values are not the same for both anthocyanins.

7. (IPS10-1.26)

(a)



[14]

(b)

The relationship is weak and not linear.

(c)

Scatter plot of x & y

The CORR Procedure

2 Variables: x y

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
x	5	65.00000	15.81139	325.00000	45.00000	85.00000
y	5	46.00000	16.73320	230.00000	30.00000	70.00000

Pearson Correlation Coefficients, N = 5
Prob > |r| under H0: Rho=0

	x	y
x	1.00000	0.00000 1.0000
y	0.00000 1.0000	1.00000

[15]

(d)

This exercise emphasizes that drawing a correlation is important to understand the true significance of the relationship between two variables.

[14] Scatter plot between variables x & y.
[15] Correlation table for variables x & y.