

## Question 1

(a) a scatter plot of the dependent and independent variable with a line added (using curve) representing best fit least squares model (See Figure 1)

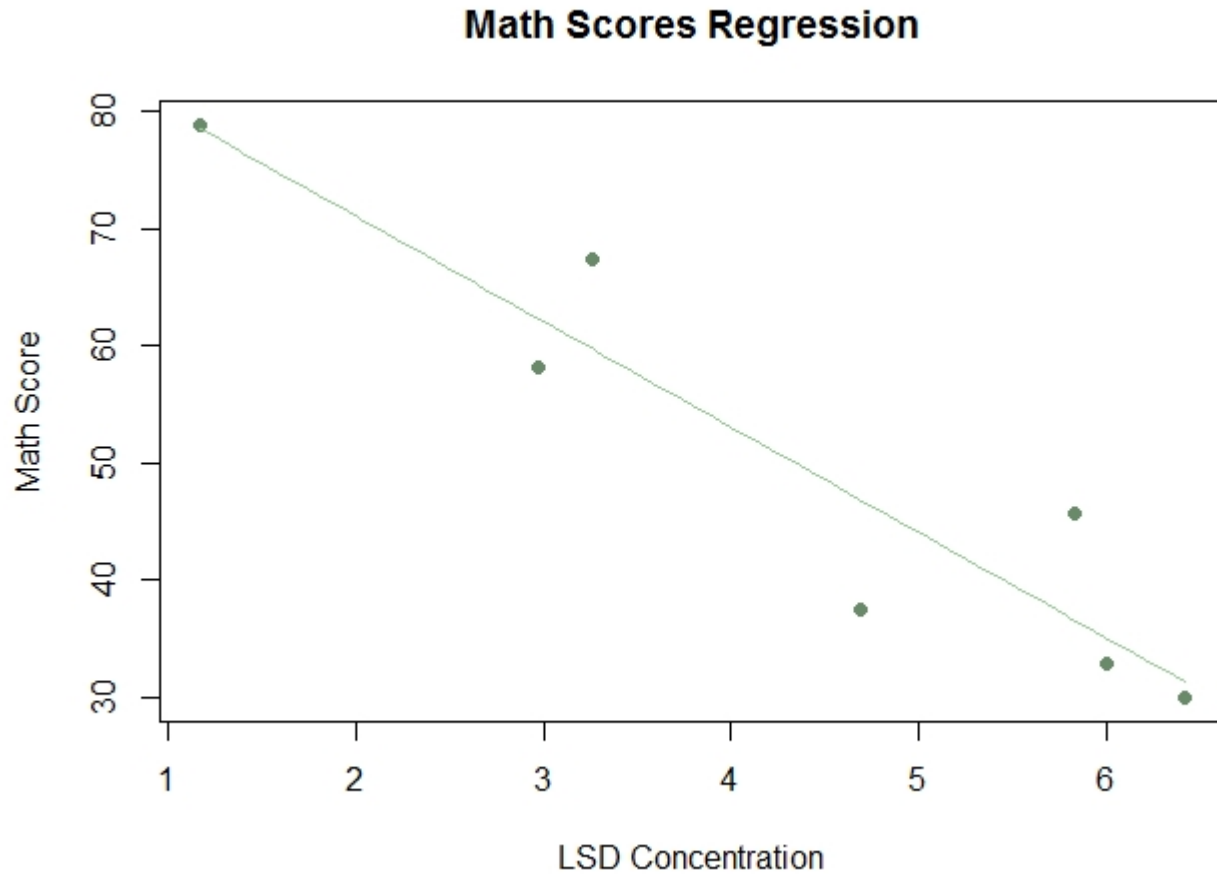


Figure 1: Math Scores Regression Plot

(b) Parameter estimates: slope = -9.00946 intercept = 89.123  
95 CI for slope and intercept parameter:

	2.5	97.5
Intercept	71.0	107.24
LSD Concentration	-12.873	-5.125

(c) RMSE = 6.02235

- A. A concentration of less than 0.457 is needed to score above an 85 percent.
- B. Overall, LSD concentration is a good predictor for test scores below 89 percent because that is the baseline and someone cannot have a negative LSD concentration. Based on the RMSE of 6.022, the model is an okay predictor, however it should be noted there is a very small sample size of only 7 data points and there is no "control" or person who took the test with no LSD.
- C. A normal distribution might be inappropriate to model this data because there is not equal data across the LSD concentrations. There are more data points with higher concentrations and fewer points with less concentrations meaning the data is not centered around some mean.

## Question 2

- (a) a scatter plot of the dependent and independent variable with a line added (using curve) representing best fit least squares model (See Figure 2)
- (b) Parameter estimates: slope = -0.5251 intercept = -0.179  
95 CI for slope and intercept parameter:

	2.5	97.5
Intercept	-1.4089937	1.05097
Number of Pomegranates	-0.8864	-0.16379

- (c) RMSE = 9.96

- A. I disagree with the farmer's association claim because the RMSE is high, and there is not a strong linear relationship within the data. There is also very high variance in the data.

## Question 3

- A. MAE= function(n,y,yhat)  
return ((1/n)\*sum(abs(y-yhat)))  
(see R code for MAE function working)
- B.

Parameter	Math Scored	Miracle Food
RMSE	6.022	9.961
MAE	4.890	7.981
R2	0.877	0.0080

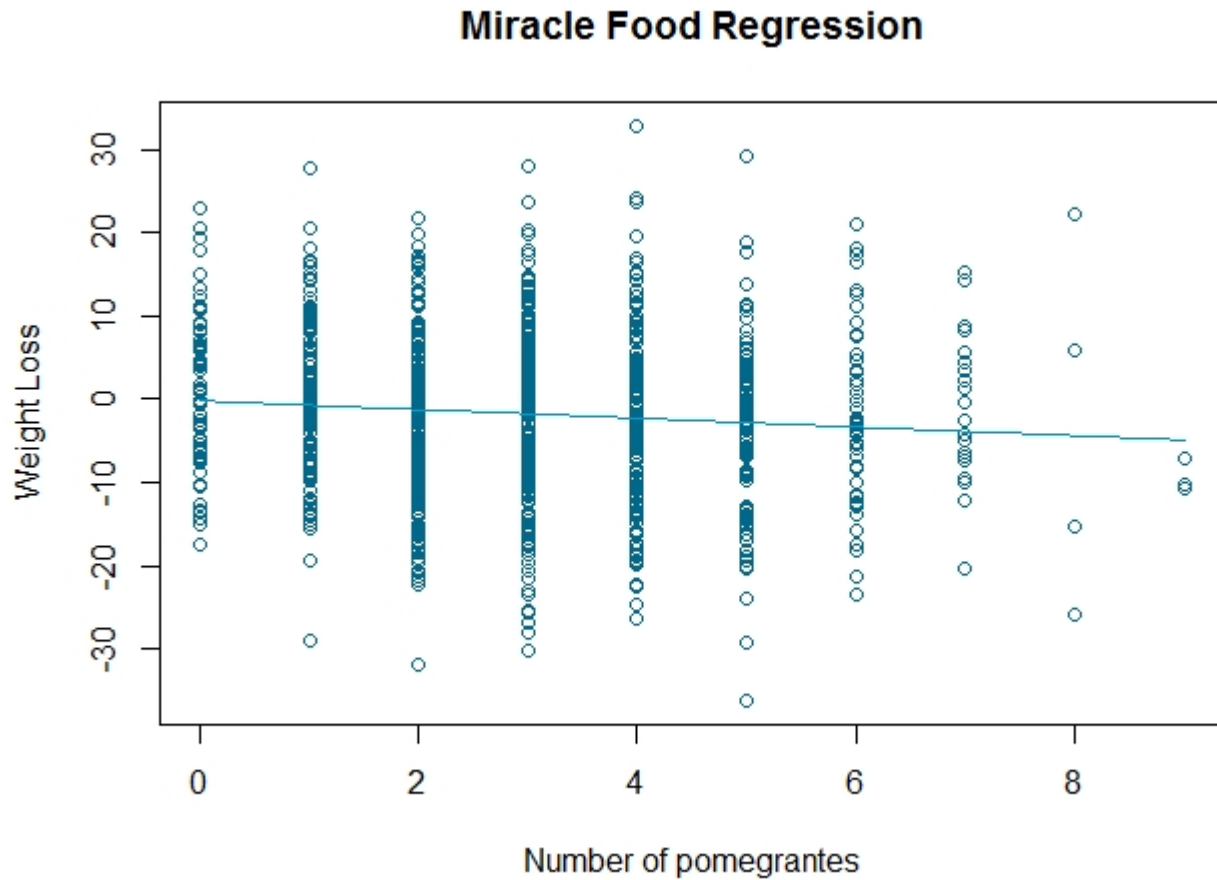


Figure 2: Miracle Regression Plot

**Question 4**

A. (see Figure 3 below)

B.

Coef	Assgined	Predicted
Slope	23.2	23.38
Intercept	2.5	2.31

C. My estimates, while very closer were off from my actual parameters by the order of 0.10 , the slope was actually higher than the actual slope while the intercept was actually less than that of the slope.

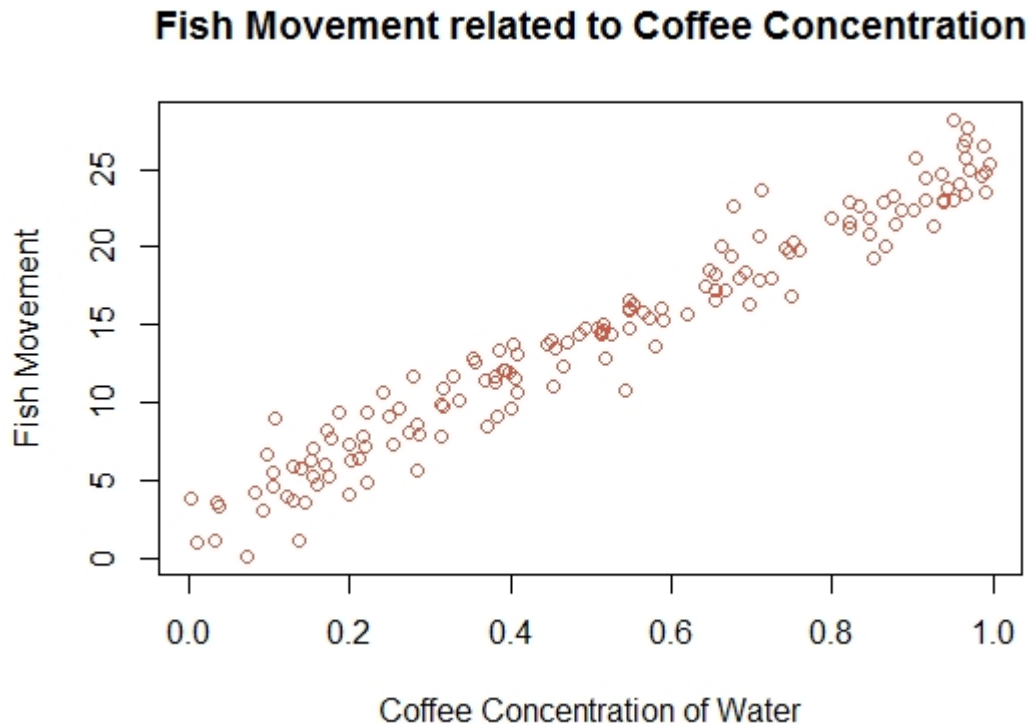


Figure 3: Generated Data Regression Plot

## Question 5

A. (See Figure 4)

B. A biological explanation for the unequal variance is past a concentration of 0.6 some fish start to react differently to the coffee. It may be they are now tired from the excessive movement and the excess coffee isn't having as much of an effect where other fish continue to be hyperactive with the increase in coffee concentration.

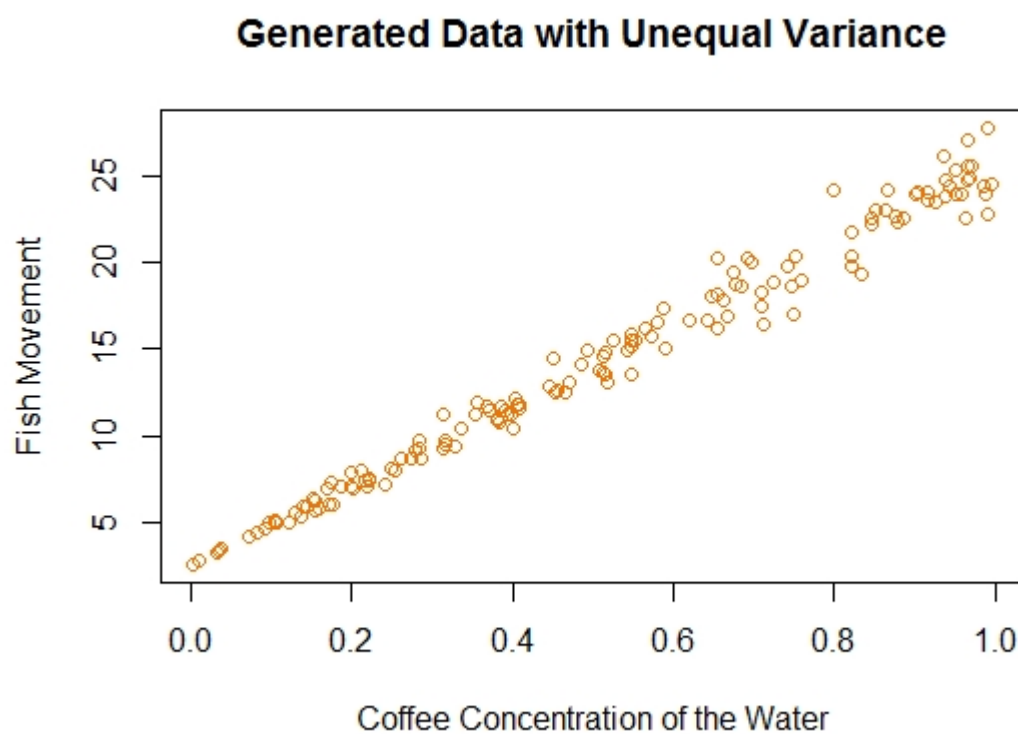


Figure 4: Plot with Unequal Variance