Unit 8 Discussion: Is there a correlation?

Unit 8 Discussion Guidance – Main Post

Main Post:

You have 3 options for selecting a dataset:

1. the dataset from your Unit 1 discussion,
2. a dataset available from the dataset link, or
3. a dataset from the additional dataset file.

NOTE: The dataset selected must have at least two quantitative variables.

Complete the following information for the selected dataset.

1. What are the names of the dataset and the two quantitative variables of interest?
2. Create a scatterplot with a simple linear regression. Attach the scatterplot to your post.
3. For the variables selected, run the Regression, available in the Excel Data Analysis Tools or use another linear regression method. Share the output.
4. What is the correlation coefficient (r)?
5. Summarize the results
    - State whether the relation is strong or weak. What evidence supports your assertation?
    - State whether the relation is positive or negative. What evidence supports your assertation?
    - Is this result what you expected?
6. Use the regression analysis output to determine the linear regression (best fit prediction line) equation and state the equation.

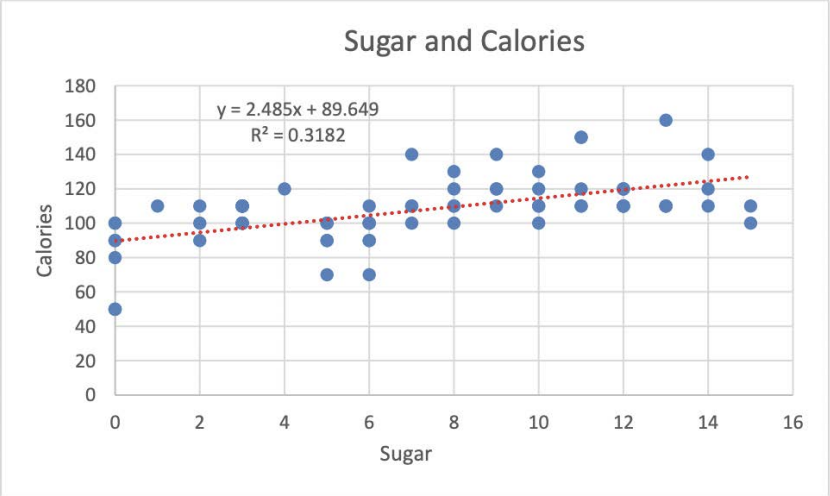
You can also view a Discussion Board starter video to assist you with the Unit 8 Discussion Board in the Unit 8 LiveBinder.

These responses are meant to be a guide on how to address the initial post and do not include all possible responses.

1. I choose to download and analyze the Cereal_Data.xlsx

The two quantitative variables are: Calories and Sugar.

2. The scatter plot for sugar vs calories is:



3. The regression analysis output is below.

The coefficient of determination, R Squared = 0.318200033. The square root of that is 0.56409222, which is the multiple R value shown below in the regression output. Now as the multiple R value is the absolute value of r, I look to the slope of the regression line, which is positive, and determine the r value is 0.56409222.

SUMMARY OUTPUT									
Regression Statistics									
Multiple R	0.56409222								
R Square	0.318200033								
Adjusted R Square	0.309109367								
Standard Error	16.19516686								
Observations	77								
ANOVA									
	df	*SS*	*MS*	*F*	*Significance F*				
Regression	1	9180.690823	9180.691	35.00293874	9.1704E-08				
Residual	75	19671.25723	262.2834						
Total	76	28851.94805							
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*	
Intercept	89.64935269	3.44838444	25.99749	2.92627E-39	82.77981862	96.51888676	82.77981862	96.51888676	
X Variable 1	2.485018432	0.420027147	5.916328	9.1704E-08	1.648281448	3.321755416	1.648281448	3.321755416	

4. The correlation coefficient is r = 0.56409222.

5. This gives a positive but relatively weak correlation. This [is/is not] what I expected when I was doing the analysis because … .

6. y = 2.485x + 89.649 or Predicted Calories = 2.485 (Sugars) + 89.649