

Unit 8 Discussion: Is there a correlation?

Unit 8 Discussion Example – Main Post

Main Post: Consider the dataset that you analyzed in Unit 1. If your dataset did not have two quantitative variables or if you would prefer using a different dataset, visit the [dataset link](#) to select a data set of interest to you for this unit. See the DB Starter video in the [Unit 8 LiveBinder](#).

Determine the following information on your selected data set. Be sure to answer all questions using complete sentences.

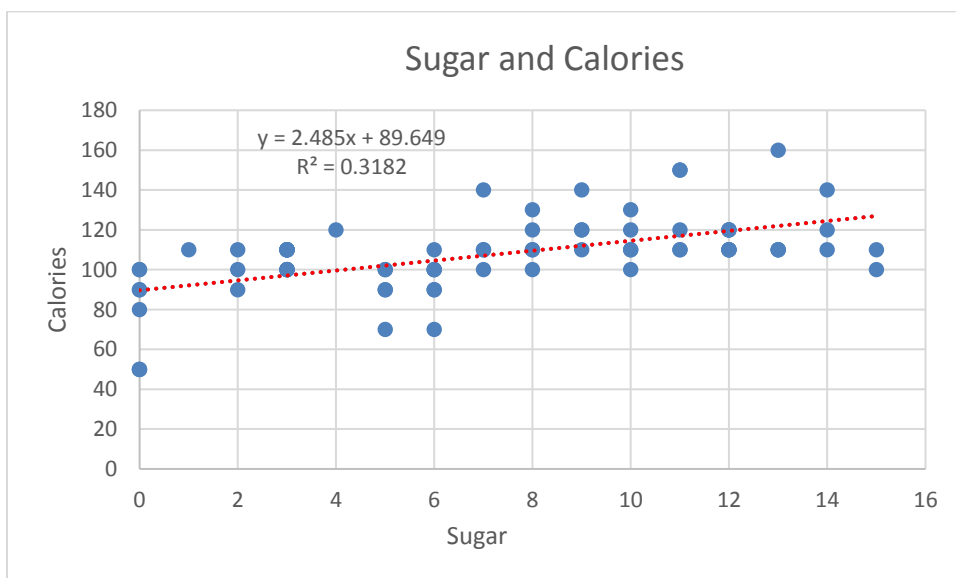
1. State the dataset and the two quantitative variables of interest. Do you think there might be a correlation between the two variables (before you analyze the data)? Why or why not.
2. Create a scatterplot with a simple linear regression (see video in [Unit 8 LiveBinder](#)). Attach the scatterplot to your post.
3. Run the regression analysis in Excel Data Analysis. Share the output.
4. What is the coefficient of determination r^2 ? Summarize the results by stating whether this is a strong or weak relationship. State whether it is a positive or negative relationship. Is this result what you expected?
5. Use the regression analysis output to determine the linear regression (best fit prediction line) equation.

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

Two quantitative variables are: Calories and Sugars

Yes! I think there is a positive correlation between sugar and calories. The more grams of sugar there will be the more calories there will be!

2. The scatter plot for sugar vs calories is:



3. The regression analysis output is below. The linear regression equation (prediction line) is :

$$y = 2.485x + 89.649 \text{ or Predicted Calories} = 2.485 (\text{Sugars}) + 89.649$$

The coefficient of determination, R Square = 0.318200033.

SUMMARY OUTPUT								
<i>Regression Statistics</i>								
Multiple R	0.56409222							
R Square	0.318200033							
Adjusted R Square	0.309109367							
Standard Error	16.19516686							
Observations	77							
<i>ANOVA</i>								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	1	9180.690823	9180.691	35.00293874	9.1704E-08			
Residual	75	19671.25723	262.2834					
Total	76	28851.94805						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
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. This gives a positive but relatively weak correlation.

5. See above scatter plot.