

MAT 120 Regression Assignment Using Excel Template in CANVAS

This will count 2% of the semester grade and is due Wednesday, October 31, 2018. Submit your results in CANVAS in Assignments. Email the instructor or drop by the office for help with any aspect of this assignment. You may discuss the assignment and share ideas, but each student must do his own work and type his name on the first sheet of the Regression Assignment, which will count as signing the UNCG Academic Integrity Policy.

Instructions:

1. Download the MAT 120 Regression Assignment Template from CANVAS, Assignments. The file is in Excel format. This software is available in the computer labs on campus. See <https://its.uncg.edu/Software/Available/Microsoft/> for more information.
2. Save the template where you can access it later and find it easily. When you submit your Excel file in CANVAS, name it in the format: Last name, First name, Regression
3. Each problem is to be worked on a separate page in the workbook. The first sheet is the UNCG Academic Integrity Policy: "I HAVE ABIDED BY THE UNCG ACADEMIC INTEGRITY POLICY ON THIS ASSIGNMENT". Type your name in the blank to sign the Academic Integrity Policy. **Assignments that are not signed will not be graded.**
4. The second sheet is the completed Example problem. You may use this to compare to your worked example.
5. Work through the example problem on your own on sheet three of the file.
6. Work through the remaining problems on each of the labelled sheets of the workbook.
7. Save your work and submit it in CANVAS, Assignments by 11:59 PM, October 31, 2018.

Work through the following example before you begin the 4 problems in this assignment. **READ THE INSTRUCTIONS FOR EACH PART CAREFULLY. SOME OF THE PROBLEMS MAY ASK YOU TO MODIFY THE DATA BEFORE CREATING YOUR CHART.** Be sure to round answers **as indicated** for each prediction you are asked to find and include the units with each value of the variables.

In real-world applications we often encounter numerical data in the form of a table. A scatterplot is used to display the data: the explanatory or input variable (**independent variable**) on the horizontal axis and the response or output variable (**dependent variable**) on the vertical axis. The very powerful mathematical tool, Regression Analysis, is the process for finding the function of best fit for the scatterplot. The function can then be used to predict the values of the dependent variable, y , for any given value of the independent variable, x .

From Sullivan, *Precalculus* 9th edition, page 133, 21. The marketing manager at Levi-Strauss wishes to find a function that relates the demand for men's jeans and the price of the jeans. The following data were obtained based on the price history of the jeans.

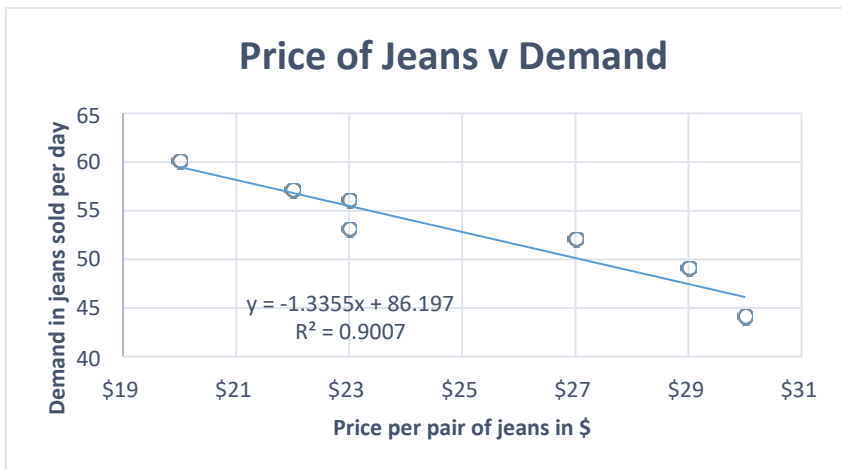
Price per pair (\$), x	Demand (Pairs of Jeans Sold per Day), y
20	60
22	57
23	56
23	53
27	52
29	49
30	44

For the scatterplot to be set up correctly, the independent variable (x) **must** be in column A and the dependent variable (y) in column B. In this example each row represents a point (x, y) in the scatterplot. The first coordinate is always the independent variable and the second coordinate is considered the dependent variable.

Instructions for the Example: We used a computer in the Superlab and Excel 2016 in the Library for these instructions.

- a. Enter the data from the table above in the Excel spreadsheet with price (explanatory) in column A and demand (response) in column B. Be sure to **label each column and include any units**.
- b. Click on **Insert**, then highlight the data you want to use in your scatterplot. Click the scatterplot icon and click on the first scatterplot on the drop down menu. **Be sure you have selected all the data in both columns**.
- c. Click on **CHART TOOLS**. Click on Quick Layout and click on the first option in the drop down menu. This step inserts labels on the horizontal and vertical axes.
- d. Click on **Add Chart Element**, put the pointer on **TRENDLINE**, and then click on **MORE TRENDLINE OPTIONS in the drop-down menu**. On the right side of your screen in a dialog box, click on the three-bar icon to have the list of Trendline options. Select “Linear”, then scroll down and select “Display Equation on Chart” and “Display R-squared value on chart”.
- e. To enter appropriate labels on each axis and a chart title for the scatterplot: Click on the Chart Title and type in a descriptive title for the application. Then doubleclick on each label Axis Title and type in descriptive labels for the axes.
- f. Adjust the scale on the axes if the data points are not spread out over the graph: Double click on the scale on the x-axis, click on the three-bar icon in the dialog box on the right, click on Axis Options. Using Axis Options on the right side of the spreadsheet, you can set the minimum and maximum bounds according to your data in Column A. Click on the scale on the y-axis, click on the three-bar icon in the dialog box on the right, click on Axis Options. Then set the minimum and maximum bounds according to your data in Column B.
- g. Use the model to predict the demand when the price (x) is \$16. **Round to the nearest integer**. Show your work by leaving the formula you used in the spreadsheet or typing it in a separate cell.
- h. Compose a sentence interpreting the statistic R^2 . The value of R^2 , the **coefficient of determination**, is given in decimal. The closer the coefficient of determination is to 1, the better the model fits the data. Change it to percent. Interpret R^2 as follows: “the value of R^2 is the proportion of variation in the output (demand) explained by the regression equation that relates the input (price) and output variable (the demand)”. For this example $R^2 = 90.07\%$, close to 100%. It, along with the scatterplot, indicates that there is a strong association between the price and the demand for jeans. The seven data points appear to follow the line of the model closely. We conclude that approximately 90% of the variation in demand for the jeans can be attributed to its linear relationship to the price. There are other variables, called “lurking variables”, which contribute to the other 10% in variation in demand, such as style and comfort. If the value of R^2 is below 50%, it is not an indication of a strong association between the two variables.

Your scatterplot of the example data should look like the one below. Compare your table and worked example to the scatterplot and table in the assignment template to be sure you are prepared to use Excel and follow the instructions in the problems below.



Rubric for the Regression Assignment:

A more specific rubric is given on the next page

- **4 points:** On the first page, include the honor code and your signature and date.
- **40 points:** Each graph and equation together counts 10 points. Use the independent variable for the horizontal axis and the dependent variable for the vertical axis. If these are backward, the equation will not be correct. Each axis label should name the correct variable and the chart title should further clarify what the scatterplot represents. Lose points for each missing or incorrect label and title. Adjust your axes as appropriate for the data in each problem. The **equation** and **R²** must be shown on the graph.
- **32 points:** Give the correct units for each variable in each problem. The units should be included in the chart labels and in the column headings of the table. Eight points for the correct units in each problem.
- **16 points:** Each prediction calculated will count 4 points, rounded according to directions, and must include the correct units with it. Use complete sentence to relate this information.
- **8 points:** Each problem should include complete statements interpreting the meaning R² for that application. Is the association between the variables good or not very good? (If the value of R² is below 50%, it is not an indication of a strong association between the two variables. If the data points in the scatterplot closely follow the graph of the model equation, we have another indication of a strong

Criteria	Point value
Honor code is signed and dated.	4
Problem 1 – X variable used correctly.	1
Problem 1 – Y variable used correctly.	1
Problem 1 – X-axis label is correct.	1
Problem 1 – Y-axis label is correct.	1
Problem 1 – Scatterplot Title is correct.	1
Problem 1 – Scatterplot graph is correct.	1
Problem 1 – Regression equation is correct and displayed.	2
Problem 1 – Regression line is plotted correctly.	2
Problem 1 – Correct units for x-axis.	2
Problem 1 – Correct units for y-axis.	2
Problem 1 – Correct units for the independent variable in data set.	2
Problem 1 – Correct units for the dependent variable in data set.	2
Problem 1 – Prediction sentence is complete.	1
Problem 1 – Prediction value is correct.	1
Problem 1 – Prediction value is rounded correctly.	1
Problem 1 – Prediction value has correct units.	1
Problem 1 – R-squared estimation is correct.	.75
Problem 1 – R-squared estimation sentence(s) are correct.	.5
Problem 1 – R-squared interpretation is correct.	.75
Problem 2 – X variable used correctly.	1
Problem 2 – Y variable used correctly.	1
Problem 2 – X-axis label is correct.	1
Problem 2 – Y-axis label is correct.	1
Problem 2 – Scatterplot Title is correct.	1
Problem 2 – Scatterplot graph is correct.	1
Problem 2 – Regression equation is correct and displayed.	2
Problem 2 – Regression line is plotted correctly.	2
Problem 2 – Correct units for x-axis.	2
Problem 2 – Correct units for y-axis.	2
Problem 2 – Correct units for the independent variable in data set.	2
Problem 2 – Correct units for the dependent variable in data set.	2
Problem 2 – Prediction sentence is complete.	1
Problem 2 – Prediction value is correct.	1
Problem 2 – Prediction value is rounded correctly.	1
Problem 2 – Prediction value has correct units.	1
Problem 2 – R-squared estimation is correct.	.75
Problem 2 – R-squared estimation sentence(s) are correct.	.5
Problem 2 – R-squared interpretation is correct.	.75
Problem 3 – X variable used correctly.	1
Problem 3 – Y variable used correctly.	1
Problem 3 – X-axis label is correct.	1
Problem 3 – Y-axis label is correct.	1
Problem 3 – Scatterplot Title is correct.	1
Problem 3 – Scatterplot graph is correct.	1
Problem 3 – Regression equation is correct and displayed.	2
Problem 3 – Regression line is plotted correctly.	2
Problem 3 – Correct units for x-axis.	2
Problem 3 – Correct units for y-axis.	2

Problem 3 – Correct units for the independent variable in data set.	2
Problem 3 – Correct units for the dependent variable in data set.	2
Problem 3 – Prediction sentence is complete.	1
Problem 3 – Prediction value is correct.	1
Problem 3 – Prediction value is rounded correctly.	1
Problem 3 – Prediction value has correct units.	1
Problem 3 – R-squared estimation is correct.	.75
Problem 3 – R-squared estimation sentence(s) are correct.	.5
Problem 3 – R-squared interpretation is correct.	.75
Problem 4 – X variable used correctly.	1
Problem 4 – Y variable used correctly.	1
Problem 4 – X-axis label is correct.	1
Problem 4 – Y-axis label is correct.	1
Problem 4 – Scatterplot Title is correct.	1
Problem 4 – Scatterplot graph is correct.	1
Problem 4 – Regression equation is correct and displayed.	2
Problem 4 – Regression line is plotted correctly.	2
Problem 4 – Correct units for x-axis.	2
Problem 4 – Correct units for y-axis.	2
Problem 4 – Correct units for the independent variable in data set.	2
Problem 4 – Correct units for the dependent variable in data set.	2
Problem 4 – Prediction sentence is complete.	1
Problem 4 – Prediction value is correct.	1
Problem 4 – Prediction value is rounded correctly.	1
Problem 4 – Prediction value has correct units.	1
Problem 4 – R-squared estimation is correct.	.75
Problem 4 – R-squared estimation sentence(s) are correct.	.5
Problem 4 – R-squared interpretation is correct.	.75

Mat120 Regression problems

1. The data in the table represent the weight (in grams) of a box of raisins and the number of raisins in each of 10 boxes.

Weight (grams), x	Number of raisins, y
42.3	87
42.7	91
42.8	93
42.4	87
42.6	89
42.4	90
42.3	82
42.5	86
42.7	86
42.5	86

- Enter the data in an Excel spreadsheet. Type the variable name at the top of columns A and B. Put the input data (x) in column A and corresponding output data (y) in column B for each box of raisins. Use Excel to create a scatterplot of the data and produce the line of best fit that models the relation between weight and number of raisins. And find R^2 .
- Use the equation model you just created to predict the number of raisins in a box that weighs 42.5 grams. **Round answer to nearest whole number.** Show your work by leaving the formula you used in the spreadsheet or typing it in a separate cell. Write your results in sentence form. ("If a box weight is 42.5 grams, we expect it to contain about ____ raisins.")
- Change the value of R^2 from decimal to percent. Then write a statement interpreting this statistic in the context of application. Comment on the strength of the association between weight and number of raisins in a box. Does the model give very reliable predictions? Consider the scattering of the data points in the scatterplot, are the points closely following the regression line? (Information on interpretation of R^2 is part h of the example instructions above.)

2. The following data represent nonstop flight time (in minutes) and one-way ticket price (in dollars) for flying from Chicago to various cities on Southwest Airlines.

City	Time(minutes), x	Price (dollars), y
Atlanta	110	140
Los Angeles	260	199
Nashville	80	108
Oklahoma City	125	130
Omaha	85	89
Ft. Myers	170	175
Phoenix	225	191
Houston	155	147
Seattle	265	196
St. Louis	65	93

- The relation between x and y is not linear. It appears to be *logarithmic*. Use Excel to create a scatterplot of the data and produce the graph of logarithmic equation that models the relation between flight time and airfare. In Trendline Options, select “*Logarithmic*” option, also select Display Equation on chart and Display R^2 on chart.
- Use your model to predict airfare for a flight from Chicago to Kansas City, if the flight time is 85 minutes. Show your work by leaving the formula you used in the spreadsheet or typing it in a separate cell. **Round answer to the nearest whole dollar.** Report your results in a complete sentence.
- Change the value of R^2 from decimal to percent. Then write a statement interpreting this statistic in the context of this application. Write another statement about the strength of the association between flight time and the price of a one-way ticket from Chicago. Does the model give very reliable predictions? Consider the scattering of the data points in the scatterplot, are the points closely following the curve?

3. An engineer collects the following data showing the speed of a Toyota Camry and its average mileage per gallon.

Speed (miles per hour), x	Mileage (Miles per Gallon), y
30	18
35	20
40	23
40	25
45	25
50	28
55	30
60	29
65	26
65	25
70	25

- The relation between x and y is not linear. It appears to be *quadratic*, looks like a parabola opening down. Use Excel to create a scatterplot of the data and produce the graph of a polynomial function of degree 2 that models the relation between speed and mileage. In Trendline Options, select "*Polynomial*" option with Order = 2, also select Display Equation on chart and Display R^2 on chart.
- Use your model to predict mileage for a Camry that is traveling 63 miles per hour. Show your work by leaving the formula you used in the spreadsheet or typing it in a separate cell. **Round answer to the nearest whole number.** Report your results in a complete sentence.
- Change the value of R^2 from decimal to percent. Then write a statement interpreting this statistic in the context of this application. Write another statement about the strength of the association between Camry speed and the miles per gallon. Does the model give very reliable predictions? Consider the scattering of the data points in the scatterplot, are the points closely following the curve?

4. Katrina deposited \$20,000 in a well-diversified mutual fund 6 years ago. The data in the table represent the value of the account at the beginning of each year for the last 7 years.

Years, x	Account value \$, y
0	20,000
1	21,516
2	23,355
3	24,885
4	27,484
5	30,053
6	32,622

- The relation between x and y is not linear. It appears to be *exponential*. Use Excel to create a scatterplot of the data and produce the graph of an exponential function that models the relation between year and account value. In Trendline Options, select "*Exponential*" option, also select Display Equation on chart and Display R^2 on chart.
- Use the function found in part (a) to predict the account value after 10 years, $x = 10$. **Round your answer to one the nearest whole dollar.** Show your work by leaving the formula you used in the spreadsheet or typing it in a separate cell. Report your results in a complete sentence.
- Change the value of R^2 from decimal to percent. Then write a statement interpreting R^2 in the context of this data. Write another statement about the strength of the association between year and account value. Comment on the scattering of the data points in the scatterplot too. Information on interpretation of R^2 is in part h of the example instructions above.