

MATH 336: Project Description
Correlation and Linear Regression Statistical Study
Due: Friday, December 11, 2020.

In this project you will gather data, analyze it using MS Excel, and write your conclusions. The project takes several weeks to complete. The steps are as follows:

1. Gather Data
2. Excel Spreadsheet
3. Introduction
4. Data analysis using MS Excel
5. Analysis
6. Written description of data analysis and conclusions

1. GATHER DATA You must gather real-life data for which you can perform a statistical study of a relationship between two variables. You must use at least 30 data points (x,y) and for most projects no more than 50. The data must be quantitative. Try to choose data that you think may be correlated, and where you think values of one variable are 'explaining' values of the other. For example, you could research how the wealth of countries impact infant mortality rates.

2. EXCEL SPREADSHEET: Enter the data for the independent variable x in one column and the corresponding data for the dependent variable y in another column. Be sure to label your columns and include units.

3. INTRODUCTION: (Approximately 2-3 pages) Describe what the data is, where you found it, why you think it is of interest, and any other background information that you think is important. Then identify what the independent and dependent variables are *likely* to be, whether you think the

response variable (this is y) will have a normal or skewed distribution, whether you expect to see any correlation between x and y , and if you think it will be positive or negative, weak or strong. Do not study your data before writing the introduction, or at least try not to let it affect what you write. In fact, it may be better if you do not even look very carefully at the raw data. The idea is to say what you think will happen in your own words. It does not matter if your predictions turn out to be wrong after you do the data analysis as this is part of the process.

4. DATA ANALYSIS: Please use MS EXCEL to analyze the data. As you do your analysis keep in mind that it must be clear and understandable to the reader. That means that data, numerical descriptions, calculations and diagrams must be clearly labeled. You must also make sure that information is not split (especially diagrams) across pages. If necessary, use additional sheets within the workbook. Your EXCEL analysis must include the following:

- **1-variable analysis of the dependent (response) variable y .** This should include the five-number summary of the dependent variable, a frequency distribution table, and a histogram. Do not calculate the mean and standard deviation here, since that is done in the next part.
- **Correlation and Linear Regression.** Start this on a new sheet, but include the data again (cut and paste it.) Include a scatterplot with the linear regression line (trend-line) labeled. Also compute the means, standard deviations, correlation coefficient r and the coefficient of determination r^2 . Compute the slope

and y-intercept for the linear regression line: then give the equation of the regression line.

5. ANALYSIS: In your analysis describe your findings from your Excel spreadsheets. Try to incorporate your knowledge from class in order to not only state what you observe, but also to describe the mathematical reasons and consequences. Try to phrase it in terms of what the data is about rather than just as a number (For example, say 'the mean travel time is...' rather than 'the mean of y is...'). Write your analysis as a report, not just a list of statistical facts and numbers. Some things you should include are:

- Skewness/symmetry of histograms
- Mean and median, and why they are similar or different.
- Outliers
- Spread
- Five number summary and mean/standard deviation
- Correlation (positive, negative, weak, strong, none-non-linear-exponential?)
- Correlation vs. causation
- Regression line (Interpret the slope and y-intercept for your data.)
- The coefficient of determination (r -squared)
- Your predictions (How accurate do you think they are and why)
- Cite the references that you used.

6. CONCLUSION: Your conclusion should summarize what you have discovered. For example, it should include what kind of correlation you found, what you think this means in 'real life', whether it was stronger or weaker than you expected, and if you think there are any lurking variables or outliers affecting

the results. You may think of things that not mentioned above but which you think are important about your experience.

Your project should be about 5 to 8 pages, double-spaced and in 16pt font. Please note the page lengths are guidelines. Quality is more important than quantity. Please submit your project by uploading it to Blackboard using a link provided. Don't forget that the Writing Center (L-118) is available to help.

Your project is graded by the following criteria:

1. Presentation: did you follow the instructions and include everything required?
2. Clarity: how easy is it for the reader to understand your report?
3. Mathematics: did you understand, use and interpret the mathematics (formulas) correctly?
4. EXCEL: did you do all the calculations/diagrams in EXCEL correctly?