Math 140 & Math 140X Project#1

Directions & Grading Rubric: Collecting Your Own Data

Use with Teachout Textbook Sections 1A,1B & 1C

Updated Fall 2024

Project#1 Directions *(You will be collecting your data, typing the data into two columns in Excel, Creating an "Other" category if needed, Custom sorting your data, then answering 15 questions below identifying your population of interest, explaining how your data was collected, and identifying sources of bias.)*

I. **Choose one categorical question and one quantitative question from the lists provided at the "projects" page at www.matt-teachout.org or on your instructor's home page in Canvas.**

- This is called paired data since you will be asking each person you talk to both questions. Choose your two questions from one of the various lists provided on your home page in Canvas. Each list has a different topic emphasis *(business/marketing, food/drink, health, sports, politics, other miscellaneous topics)*.
- Email your instructor and let them know the categorical and quantitative question (paired data) you chose, and be sure to include the "pair data number" corresponding to your two questions. Each student must use different paired data. It is ok if one of the questions is the same, but both questions cannot be the same as another student. So, if you see on the list that the paired data says "taken", that means another student has chosen that pair and you will have to chose another. First come first serve.

II. **Collect your categorical and quantitative data.**

- Identify your large population of interest. The population must be large and cannot be your friends and family or just people in a class. Examples of the population of interest might be all people in Santa Clarita, or all students at COC, or all cars in Southern California.
- Get data from 50 or more people or objects in your population. This is called a sample. You cannot get data from everyone in your population. You can get your sample data by asking people in your COC classes your two questions (as long as it is ok with the professor), or asking people on social media, or asking people you know, or using any other venue to ask people questions.
- You will need to ask each person you talk to BOTH of the questions listed in your paired data choice. Or you can choose objects to get the two bits of data from.
- Only keep data from people that answered both questions**.** Keep track of how many people refused to answer one or both of the questions. This will be asked in your report and is part of non-response bias. However, do not keep data from people that only answered one of the questions. The only data you will type into your spreadsheet, are those that answered both questions.
- You need to get data from at least 50 people or objects. Keep track of the number of people that refused to answer one or both of the questions. You may need to ask 75 people to get at least 50 that answered both questions. Remember, more data is better so it is better if you can get data from significantly more than 50 people.

III. **Type your data into an Excel spreadsheet (Windows) or a Numbers spreadsheet (Apple).**

- Your spreadsheet should have two columns, one for the categorical data (words) on the left column and one for the quantitative data (count or numerical measurement) on the right column.
- Make sure both of your columns have a title describing what was asked.

- When typing in the categorical data words, make sure the words for each answer are typed exactly the same. For example, if you ask people their favorite sports team, you should not have "dodgers", "Dodgers", "LA Dodgers", "Los Angeles Dodgers.", etc. in the same column. They should all be spelled and capitalized the same. In this case, people that answered Los Angeles Dodgers should all say "Dodgers". The reason for this is that the computer will see "Dodgers" and "dodgers" as separate different answers. They need to be spelled exactly the same for the computer to recognize it.

IV. **If you have more than five different categorical answers, create an "Other" category.**

- Note: If your categorical question only has five or less answers like "Yes or NO" or "Hamburgers, Tacos or Pizza", or "small, medium, large and extra-large", then DO NOT make an "Other" category. Just keep the answers you have.
- If you have many categorical answer options, then you will want to create an "Other" category. Identify the top four answers given to your categorical question. Change all answers that were not in your top four to say "Other". That will cut down on how many categories you have. It is not good to have 18 different answers for your categorical data with most of them coming from 1 person. Creating an "Other" category will limit the total number of categories to five (four categorical answer options and "other" is your fifth category).
- For example, suppose my categorical question was favorite coffee shop. Suppose your top answers were "Starbucks", "Honu", "Peets", and "Coffee Bean". A few people answered other answers like "McDonalds" or "Home Coffee". These other answers can now become the "Other" category.
- Notice now we have a total of five categories (Starbucks, Coffee Bean, Peets, Honu and Other).

Favorite Coffee Shop	*$ spent on gas per week*
Starbucks	*50*
Coffee Bean	*40*
Starbucks	*0*
Honu	*25*
Starbucks	*60*
Peets	*75*
Coffee Bean	*20*
Home Coffee	*35*
Honu	*80*
Peets	*100*
McDonalds	*65*

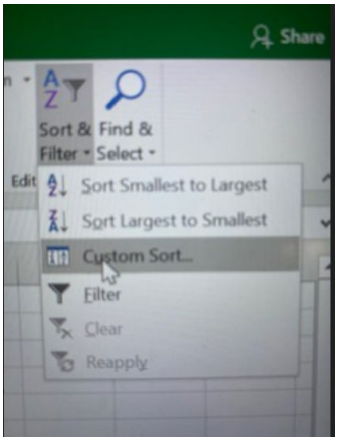
Favorite Coffee Shop	*$ spent on gas per week*
Starbucks	*50*
Coffee Bean	*40*
Starbucks	*0*
Honu	*25*
Starbucks	*60*
Peets	*75*
Coffee Bean	*20*
Other	*35*
Honu	*80*
Peets	*100*
Other	*65*

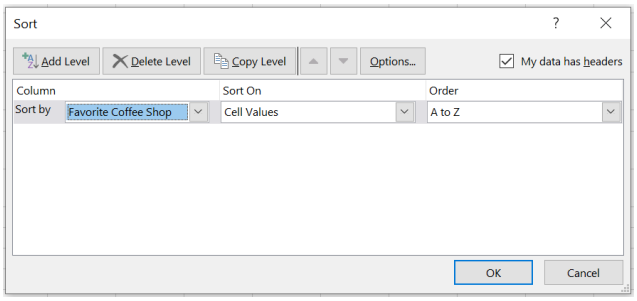
V. Custom Sort the paired categorical and quantitative data you collected by category.

- Highlight both of the two columns of data. Here is an example using the favorite coffee shop and $ spent on gas per week questions.

Favorite Coffee Shop	*$ spent on gas per week*
Coffee Bean	*40*
Coffee Bean	*20*
Honu	*25*
Honu	*80*
Other	*35*
Other	*65*
Peets	*75*
Peets	*100*
Starbucks	*50*
Starbucks	*0*
Starbucks	*60*

- **Custom Sorting Directions in Excel:** After highlighting both of the columns, click on "sort and filter" and then click on "custom sort". The sort window will open. Make sure you check the box that says "My data has headers". Under "Sort by" click on your categorical column. In this example, that is "Favorite Coffee Shop". Then push "OK". Your categorical data should now be in alphabetical order without losing the pairing to your quantitative data. Remember you have to highlight both columns before sorting. Do NOT just sort the left column alone. If you do that you lose the pairing and mess up your data!





Favorite Coffee Shop	*$ spent on gas per week*
Coffee Bean	*40*
Coffee Bean	*20*
Honu	*25*
Honu	*80*
Other	*35*
Other	*65*
Peets	*75*
Peets	*100*
Starbucks	*50*
Starbucks	*0*
Starbucks	*60*

- You should now see your left column in alphabetical order, but you have not lost the pairing to your quantitative right column of data. Notice the numbers next to Coffee Bean still say 40 and 20. The numbers next to Starbucks still say 50, 0 and 60. The numbers next to Honu still say 25 and 80. The number next to Peets still say 75 and 100. The numbers next to "Other" still say 65 and 35.

VI. Save and take a picture of the two columns of sorted data

- Save your two columns of sorted data in Excel or Tables. You will need this data for future projects in the class. If you used Excel, you can turn in the Excel file. If you used a Numbers spreadsheet with a MAC, you will need to take a picture of it to prove you have done it. You will attach either the Excel file or the picture or the Tables file with your report to prove you have typed the data into a spreadsheet. Remember, you should have at least 51 rows including the title. The data should be in alphabetical order without losing the pairing. The spreadsheet should a maximum of three categories including the "Other" category. Save the spreadsheet. You will need it to complete other projects in the class.

Favorite Coffee Shop	*$ spent on gas per week*
Coffee Bean	*40*
Coffee Bean	*20*
Honu	*25*
Honu	*80*
Other	*35*
Other	*65*
Peets	*75*
Peets	*100*
Starbucks	*50*
Starbucks	*0*
Starbucks	*60*

Project#1 Report Questions: Turn in the Excel Spreadsheet with two columns of sorted data, name and project title, anti-cheating statement, and answers to these 15 questions. There is 100 points possible.

A. **Attach a picture of your Excel or Tables spread sheet to prove you have typed the two columns of data into Excel or Tables.** Notes: The data must be custom sorted by category buy not lose the connection to the quantitative numerical question answer. The data should have a maximum of four categories by using the "Other" category. I prefer your data to be in Excel. If your two columns are in Excel, you can attach the Excel file with your report. If you have a MAC computer, you can type your data into "Tables". However, Mr. Teachout will **NOT** be able to open a Tables file. You will need to submit a picture of your two columns of data instead.

B. **Put the following title at the top of your report.**
First and Last Name
Math 140 Project#1 Collecting Data
Instructor's Name
Section# for your class
Semester and Year

C. **Confirm that you did not cheat. If true, write the following:**
I did not cheat on this project. I collected the data and did the work myself.
No one else did this work for me.

D. **Answer the following fifteen questions.**

1. **What was your categorical question?** *(Answer was words)*

2. **What was your quantitative question?** *(Answer was number)*

3. **What was your pair data number corresponding to these two questions on the lists provided on your instructor's home page?**

4. **How many people answered both questions?**

5. **How many people did not answer one or both of the questions for any reason?**

6. **Describe how you collected the data? Be specific and give details.** Here are examples: "I walked around the Valencia mall and asked people I met both questions." "I posted my questions on Instagram and kept data from only those people that answered both of my questions". "I sat in the COC cafeteria and asked COC students I met both of my questions". "I asked all of the students in the COC classes I am enrolled in during the semester."

7. **What is your population of interest?** *(i.e. What large group of people could you apply your sample data to? For example: If you collect data from 47 COC students you attend class with, your population of interest might be all students at College of the Canyons. If you collect data from 33 people on Instagram, your population of interest may be all people on Instagram. If you collected data from your friends and family, your population of interest can NOT be your friends and family. If your friends and family live in Santa Clarita, your population of interest could be all people in Santa Clarita. If your friends and family live all round Southern CA, your population of interest could be all people in Southern CA.)*

8. **What method of collecting data did you use: convenience, voluntary response, simple random sample, random cluster, non-random cluster?** (Note: You cannot have a census. The data you collect must be a sample from a small group that you will attempt to apply to a larger population.)

9. Data has sampling bias if the sample data was not collected randomly or if the sample size is too small. **Do you think that sampling bias may be present in your data?** (Has sampling bias or does not have sampling bias.) **Explain why you think your data may or may not have sampling bias.**

10. Data has question bias if the person collecting the data asked the questions in such a way as to to sway the person answering to give a certain answer. **Do you think that question bias may be present in your data?** (Has question bias or does not have question bias.) **Explain why you think your data may or may not have question bias.**

11. Data has deliberate bias if there are groups in the population that were left out and not represented in the data. Data can also have deliberate bias if the person collecting the data deletes or falsifies the data after collecting it. **Do you think that deliberate bias may be present in your data?** (Has deliberate bias or does not have deliberate bias.) **Explain why you think your data may or may not have deliberate bias.**

12. Data has response bias if the people answering the questions lie or cannot give an accurate answer. **Do you think that response bias may be present in your data?** (Has response bias or does not have response bias.) **Explain why you think your data may or may not have response bias.**

13. Data has non-response bias if there are people that refused to answer the questions or for any reason were not able to answer the questions. **Do you think that non-response bias may be present in your data?** (Has non-response bias or does not have non-response bias.) **Explain why you think your data may or may not have non-response bias.**

14. **What industries, companies, colleges, hospitals, public health, sports teams, governments, etc. would be interested in this data?**

15. **List at least three specific questions that could be answered by analyzing your data?**

Project #1 Grading Rubric *(100 points total)*

- **Collect data by asking many people both questions.** *Must have at least 30 or more people that answered both questions. Also keep track of how many people refused to answer one or both of the questions. (23 points total)*
- **Typing two columns of data into Excel or Tables spreadsheet.** *(20 points)*
- **Create an "Other" category if needed. Your data should have a maximum of five categories.** *(5 points)*
- **Custom sort your paired data so that the categorical data is in alphabetical order but you have not lost the pairing.** *(5 points)*
- **Name and Title, Cheating Statement, Answers #1-5,7,8** *(2 points each, 18 points total)*
- **Answers #6,14,15** *(3 points each, 9 total points)*
- **Answers #9-13** *(4 points each, 20 total points)*