Math 140 & Math 140X

Project#2 Directions: Analyzing Your Data

Use with Teachout Textbook Sections 1E, 1F & 1G

Updated Spring 2024

GRADING RUBRIC FOR PROJECT#2 REPORT *(100 points total)*

- **First and Last Name, Title, Section Number, Instructors Name, Semester and Year:** *1 point*
- **Non-cheating statement:** *1 point*
- **StatKey Pictures #2,#15,#16:** *5 points each (15 points total)*
- **Answers to Questions #1,4,5,6,7,8,9,10,11,13,14,17,18,20,21,23,29,31:** *2 points each (36 total points)*
- **Answer to Question#3:** *3 points*
- **Answers to Questions #12,19,22,24,25,26,27,28,30,32,33:** *4 points each (44 total points)*

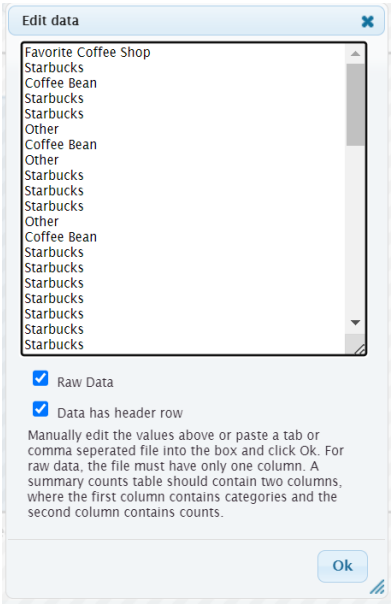
IMPORTANT NOTE: You must have the bar chart from your categorical column with summary statistics (counts and proportions) included. You must have a 5-bar histogram from your quantitative column with summary statistics (sample size, mean standard deviation, etc.) included. You must have a box plot from your quantitative column with summary statistics (sample size, mean standard deviation, etc.) included. All of the graphs and statistics must NOT be too small. They must be readable. Without readable StatKey graphs and statistics, your instructor will NOT be able to grade your report!

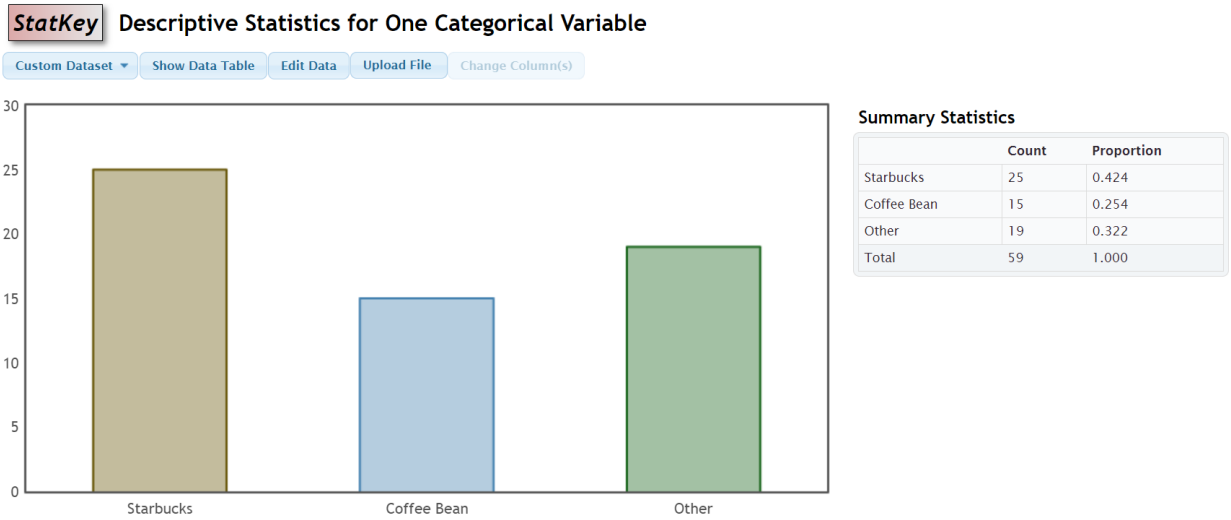
PROJECT#2 DIRECTIONS

I. **Open the data Excel spreadsheet (Windows) or Numbers spreadsheet (Apple) that has the paired data you collected in project#1.**
   - 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. The data should be sorted with the categories in alphabetical order but have not last the pairing. It should be the same data that you collected in project#1 unless Mr. Teachout had you make corrections. In that case, use the corrected data.

II. **Put the categorical column of data** (left column of words) **into StatKey.**
   - **Go to www.lock5stat.com, click on "StatKey", and then click on "One Categorical Variable" under the descriptive statistics and graphs menu.**
   - **Click the "Edit Data" button. Push Control A and then delete to delete out any data listed.**
   - **Go back to your Excel or Tables spreadsheet. Highlight your left column of categorical** *(words)* **data. Then push "control C" to copy. Or you can right click and copy.**
   - **Go back to the edit data screen in StatKey, and paste the column of categorical data into StatKey. This is raw categorical data, so you will need to check the box that says "raw data". If you have the title, check the box that says "Data has a header row". If you do not have the column title, do not check that box that says "Data has a header row". Then push the "OK" button.**
   - **You should now see a bar chart and the summary statistics with the counts and proportions for each category. Take a picture of your bar chart and summary**

statistics. Do not make the picture too small. Make sure the statistics and categories are readable.



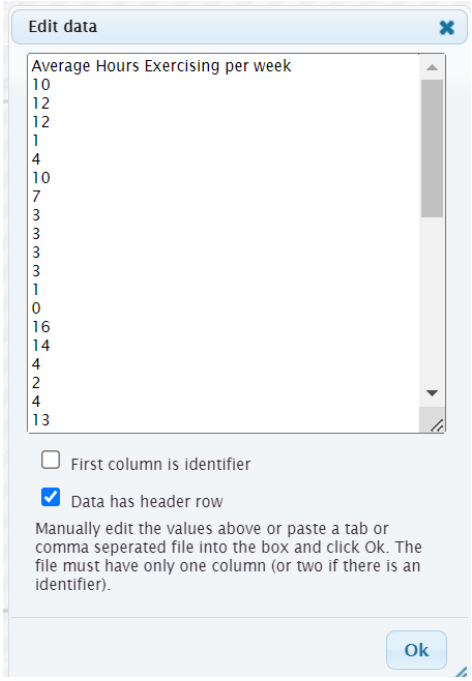


III. Put the quantitative column of data (right column of numbers) **into StatKey.**

- **Go to www.lock5stat.com, click on "StatKey", and then click on "One Quantitative Variable" under the descriptive statistics and graphs menu.**
- **Click the "Edit Data" button. Push Control A and then delete to delete out any data listed.**
- **Go back to your Excel or Tables spreadsheet. Highlight your right column of numerical measurement data. Then push "control C" to copy. Or you can right click and copy.**
- **Go back to the edit data screen in StatKey, and paste the column of quantitative data into StatKey. This is one column of numerical data, so you do not have an "identifier" column. Make sure the box that says "First Column is Identifier" is <u>not</u> checked. If you have the title, check the box that says "Data has a header row". If**

you do not have the column title, do <u>not</u> check that box that says "Data has a header row". Then push the "OK" button.

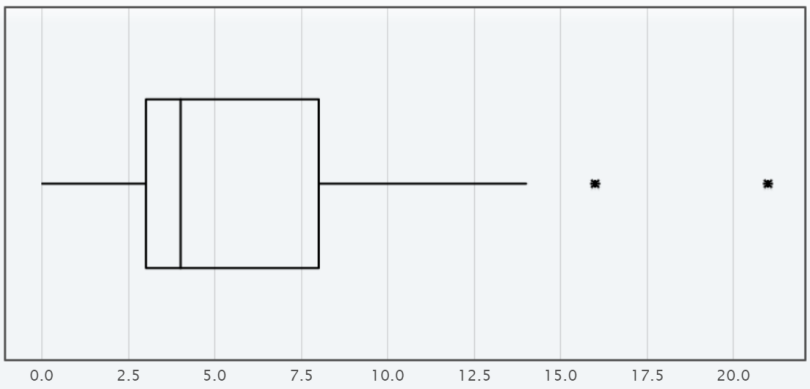
- **Click on the "Box plot" button. You should also see the summary statistics on right with the sample size, mean, standard deviation, min, Q1, median, Q3, and max. Take a picture of your box plot and summary statistics. Do not make the picture too small. Make sure the statistics are readable.**
- **Click on the "Histogram" button. Use the slider on the bottom right to change the number of buckets (bars) to 5 bars. You may need to keep sliding the buckets slider back and forth. Make sure your histogram has 5 bars. Take a picture of your histogram and summary statistics. If your highest bar is in the middle, we can classify the data as "nearly normal" and use the mean as our best average and standard deviation as our best measure of spread. If the highest bar is on the far right, we can classify the data as skewed left. If the highest bar is on the far left, we can classify the data as skewed right. For skewed data, we will use the median as our best average and IQR as our best measure of spread. Do not make the picture too small. Make sure the statistics are readable.**



StatKey Descriptive Statistics for One Quantitative Variable

Custom Dataset | Show Data Table | Edit Data | Upload File | Change Column(s)

Dotplot | Histogram | Box plot



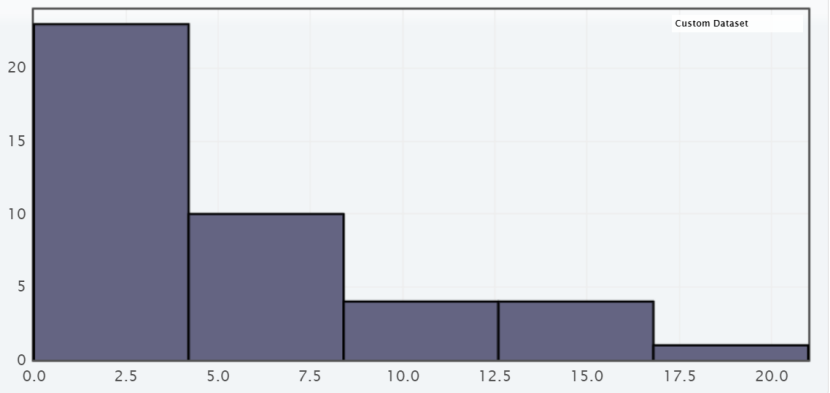
Summary Statistics

Statistic	Value
Sample Size	42
Mean	5.976
Standard Deviation	4.672
Minimum	0
Q_1	3.000
Median	4.000
Q_3	8.000
Maximum	21

StatKey Descriptive Statistics for One Quantitative Variable

Custom Dataset | Show Data Table | Edit Data | Reset Histogram | Upload File | Change Column(s)

Dotplot | Histogram | Box plot



Summary Statistics

Statistic	Value
Sample Size	42
Mean	5.976
Standard Deviation	4.672
Minimum	0
Q_1	3.000
Median	4.000
Q_3	8.000
Maximum	21

Histogram Controls Set Limits

Number of buckets: 5

lower bound | upper bound | frequency

PROJECT#2 REPORT QUESTIONS: Answer the following questions. This is what you will turn in to your instructor to be graded. There is 100 points possible.

Put the following title at the top of your report.
First and Last Name
Math 140 Project#2 Analyzing Data
Instructor's Name
Section# for your class
Semester and Year

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.

PART 1: CATEGORICAL DATA ANALYSIS

1. **What was the categorical question you asked when you collected the data?** *(Answers were words)*
2. **Show the StatKey picture showing your summary statistics and bar chart for your categorical data.** *(Look at your categorical data bar chart and summary statistics counts and proportions to answer the following.)*
3. **Convert all of your proportions listed in your summary statistics into percentages. Do <u>not</u> round.**
4. **Which category answer besides "Other" was answered the most** (i.e. had the highest bar besides the "Other" category)**?**
5. **How many people gave this most common non-other answer?** *(This is the highest count in summary statistics that was not "other".)*
6. **What was the proportion for the most common non-other answer?** *(This is the highest proportion listed in summary statistics that was not "other".)*
7. **What was the percentage for the most common non-other answer?** *(This is the percentage corresponding to the highest proportion listed in summary statistics that is not "Other".)*
8. **Which category answer besides "Other" was answered the least?** (i.e. had the lowest bar besides the "Other" category)
9. **How many people gave this least common non-other answer?** *(This is the lowest count in summary statistics that was not "other".)*
10. **What was the proportion for the least common non-other answer?** *(This is the lowest proportion listed in summary statistics that was not "other".)*
11. **What was the percentage for the least common non-other answer?** *(This is the percentage corresponding to the lowest proportion listed in summary statistics that is not "Other".)*
12. **Look at your categorical data bar chart and summary statistics counts, proportions, and percentages. List something you found surprising or interesting? Explain why you found them surprising or interesting. If you did not find anything surprising or interesting, explain why you did not find anything surprising or interesting.**

PART 2: QUANTITATIVE DATA ANALYSIS

13. **What was the quantitative question you asked when you collected the data?** *(Answers were numbers)*
14. **What were the units of your quantitative data?** *(Dollars spent on shoes, number of units taken at college, number of times you get gas per month, etc.)*
15. **Show the StatKey picture showing your summary statistics and 5-bar histogram for your quantitative data.**
16. **Show the StatKey picture showing your summary statistics and box plot for your quantitative data.**
17. **Look closely at the histogram and box plot. Was your data normal** (close to bell shaped)**, skewed right** (long right tail)**, or skewed left** (long left tail)**?**
18. **Based on the shape, should we use the <u>mean average</u> or the <u>median average</u> as our most accurate average?**
19. **Look at mean and median calculated in your summary statistics for the quantitative data. Write a sentence using the most accurate average in context with units.** *(For example: People in my data exercise an average of 8 hours per week.)*

20. **Look at the first quartile (Q1) and the third quartile (Q3) calculated in your summary statistics for the quantitative data. Calculate the interquartile range (IQR) using the formula** IQR = Q3 – Q1**.**
21. **Based on the shape, should we use the <u>standard deviation</u> or the <u>interquartile range (IQR)</u> as our most accurate measure of spread?**
22. **Look at the mean and the standard deviation listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas.**
    mean – standard deviation = ?
    mean + standard deviation = ?
23. **Based on the shape, typical values for this data will be between which two values?**
    <u>Between Q1 & Q3</u> **OR**
    <u>Between mean – standard deviation & mean + standard deviation</u>
24. **Write a sentence in context with units explaining the two cutoffs where typical values will be in between.** *(This should correspond to your answer in #23. Either use Q1 and Q3 listed in your summary statistics or use the two answers listed in #22. For example: Typical people in my data exercise between 5 and 12 hours per week.)*

25. **Look at your box plot for your quantitative data. High outliers are stars on the far right. Were there any high outliers** (unusually high values) **shown in the box plot? If so, hold your curser over each of the stars and write down the numbers. Write a sentence with units listing the unusually high values.** *(For example: In my data, there were nine unusually high values that people exercise per week: 24 hours, 24 hours, 25 hours, 25 hours, 25 hours, 27 hours, 27 hours, 30 hours and 40 hours.)* **Note: If there are no high outliers, just say "There were no unusually high values in my data."**
26. **Look at your box plot for your quantitative data. Low outliers are stars on the far left. Were there any low outliers** (unusually low values) **shown in the box plot? If so, hold your curser over each of the stars and write down the numbers. Write a sentence with units listing the unusually low values.** *(For example: In my data, there were no unusually low hours of exercise per week. Zero hours of exercise per week was not unusual.)*
27. **Look at the mean and the standard deviation listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas. Make sure to do the multiplication before doing the addition or subtraction.** (Note: It is common for the lower value to be negative. It does not mean you did something wrong.)
    mean – (2 x standard deviation) =
    mean + (2 x standard deviation) =
28. **Look at the answer for #20 and the Q1 and Q3 listed in your summary statistics for the quantitative data. Calculate the following two values using the provided formulas. Make sure to do the multiplication before doing the addition or subtraction.** (Note: It is common for the lower value to be negative. It does not mean you did something wrong.)
    Q1 – (1.5 x IQR) =
    Q3 + (1.5 x IQR) =
29. **Based on the shape of the data, should you use** <u>mean – (2 x standard deviation)</u> **OR** <u>Q1 – (1.5 x IQR)</u> **as your low outlier cutoff?**

30. **Write a sentence in context with units about the low outlier cutoff.** *(For example: It is unusual for a person in my data to work out less than -5.5 hours per week. This means zero hours was not unusual.)*
31. **Based on the shape of the data, should you use** mean + (2 x standard deviation) **OR** Q3 + (1.5 x IQR) **as your high outlier cutoff?**
32. **Write a sentence in context with units about the high outlier cutoff.** *(For example: It is unusual for a person in my data to work out more than 22.5 hours per week.)*
33. **Look at your quantitative data graphs, statistics and calculations. List something you found surprising or interesting? Explain why you found them surprising or interesting. If you did not find anything surprising or interesting, explain why you did not find anything surprising or interesting.**