Math 140 Project#3 Directions
Estimating Population Parameters with Confidence Intervals
Use with Teachout Textbook Chapter 2 / Updated Spring 2025

GRADING RUBRIC PROJECT#3 REPORT *(100 points total)*

- Name, project title, & section number *(1 point)*
- Anti-cheating statement *(2 points)*
- **StatKey Graphs** *(3 points each, 12 points total)*
  Bar Chart of Categorical Data
  Categorical Data Statistics (counts and proportions)
  Histogram of Quantitative Data (Should have 5 bars to determine shape of quantitative data.)
  Quantitative Data Statistics (Showing sample size, mean, median, standard deviation, etc.),
- **StatKey Bootstrap Confidence Interval Distributions** *(4 points each, 12 points total)*
  One-population Proportion Bootstrap Confidence Interval graph (0.95 in middle top and the upper and lower confidence interval limits in the bottom of the distribution.)
  Two-population Proportions Bootstrap Confidence Interval graph (0.95 in middle top and the upper and lower confidence interval limits in the bottom of the distribution.)
  One-population Mean **OR** One-population Median Bootstrap Confidence Interval graph
  (0.95 in middle top and the upper and lower confidence interval limits in the bottom of the distribution. Do not do both. If your quantitative data is normal bell shaped, make a bootstrap for the population mean. If your data is skewed or not normal, make a bootstrap for the population median.)
- **Confidence Interval and Margin of Error Interpretation Sentences** *(4 points each, 16 points total)*
- **#6,#7 Checking Conditions Random & Independence** *(3 points each, 6 points total)*
- **#20 Counts for each group and total** *(3 points)*
- **All Other Questions** *(2 points each, 48 total points)*

IMPORTANT NOTE: You must have the three bootstrap confidence interval distributions with upper and lower limits and statistics included. You must have the bar chart and summary statistics (counts and proportions) for your categorical data. You must have the 5-bar histogram and summary statistics (sample size, mean, standard deviation, etc) for your quantitative (numerical measurement) data. 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#3 REPORT: Answer the following questions. This is what you will turn in to your instructor to be graded. There is 100 points possible.

Part I: Put your name, title, and cheating statement at the top of your report.

First and Last Name
Math 140 Project#3 Confidence Intervals
Instructor's Name
Section# for your class
Semester and Year

Confirm that you did not cheat. If true, write the following on your report:
I did not cheat on this project. I did the work myself. No one else did this work for me. I did not copy anyone else's work.

PART II: Put the answers to the following data questions on your report.

1. **What was the categorical question you asked when you collected the data?** *(Answers were words)*
2. **What was the quantitative question you asked when you collected the data?** *(Answers were numbers)*
3. **What were the units?** *(inches, $ spent on shoes, number of units taken at college per semester, number of times you get gas per month, etc.)*
4. **What method did you collect your sample data?** *(Convenience, Voluntary Response, Simple Random Sample, etc. NOTE: You cannot say "Census". The data was a sample and has a larger population to apply to.)*
5. **What is your population of interest?** *(For example, if you collected data from a few math 140 students, your population of interest might be ALL math 140 students at COC. If you collected data from your followers on Instagram, your population of interest might be ALL people on Instagram.)*
6. **Does your data pass or fail the random condition? Explain why.** *(Your sample data will not pass the random condition if your data was collected with convenience or voluntary response method.*
7. **Does your data pass or fail the independent observations condition? Explain why.** *(If the individuals are not related, it passes the independence condition. If the individuals are related in some way, it fails the independence condition. For example, if your population is all adults in Santa Clarita, and you got your data only from friends and family, the individuals in your data would be related and fail the independence condition. If your population is all COC students, and you collected data from students in the same stat classes, the individuals in your data would be related and fail the independence condition.)*

Part III: Create your bar chart and statistics for your categorical data.

Open the data Excel spreadsheet (Windows) or Numbers spreadsheet (Apple) that has the sorted 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. It should be sorted in alphabetical order and have an "Other" category if needed. This should be the same data that you collected in project#1 unless Mr. Teachout had you correct something. If that is the case, use the corrected data.

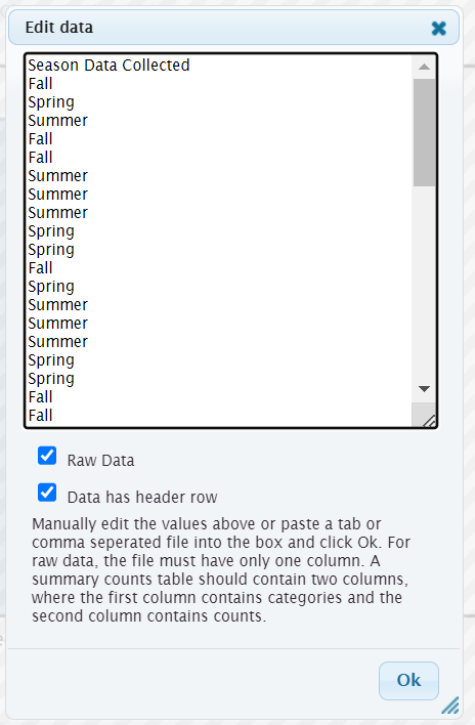
Put the categorical column of data (left column of words) into StatKey and create a bar chart and calculate proportion and counts for your categories.

Go to www.lock5stat.com, click on "StatKey", and then click on "One Categorical Variable" under the descriptive statistics and graphs menu.

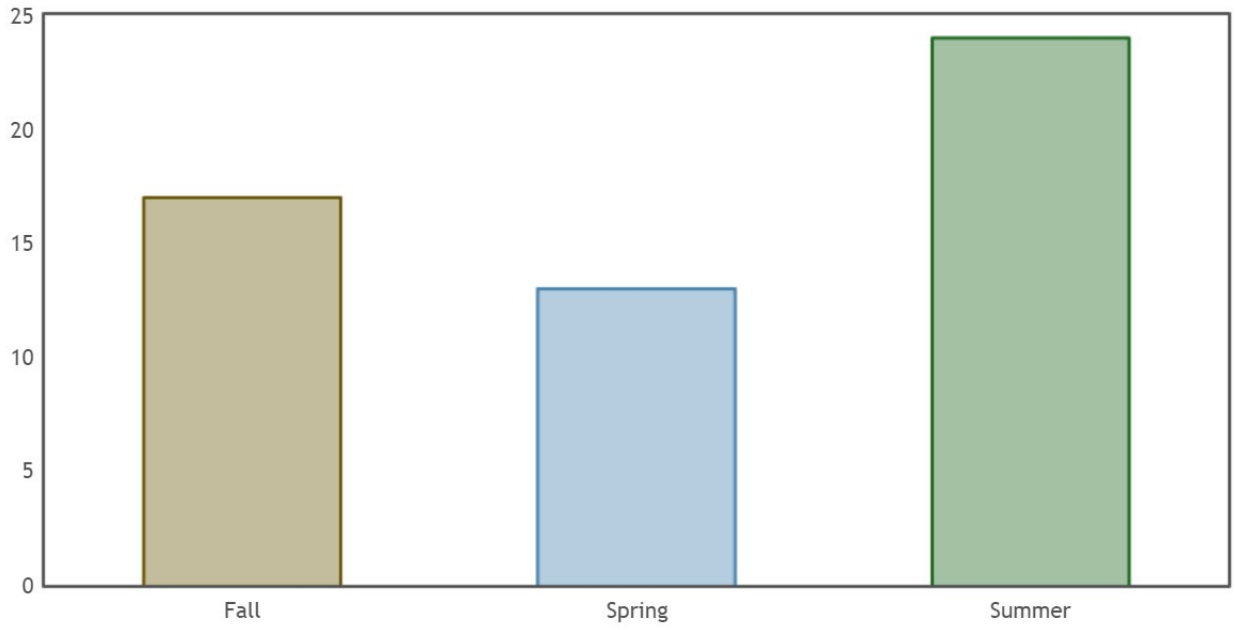
Click the "Edit Data" button. Push Control A and then delete to delete out any data listed.

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 a picture of your categorical summary statistics. Do not make the pictures too small. Make sure the statistics and categories are readable.



Summary Statistics

	Count	Proportion
Fall	17	0.315
Spring	13	0.241
Summer	24	0.444
Total	54	1.000

Put the following answers on your report:

8. **Put a picture of your bar chart showing your sample categorical data on your report.**

9. **Put a picture of the categorical summary statistics on your report showing the counts and proportions for your sample categorical data.**

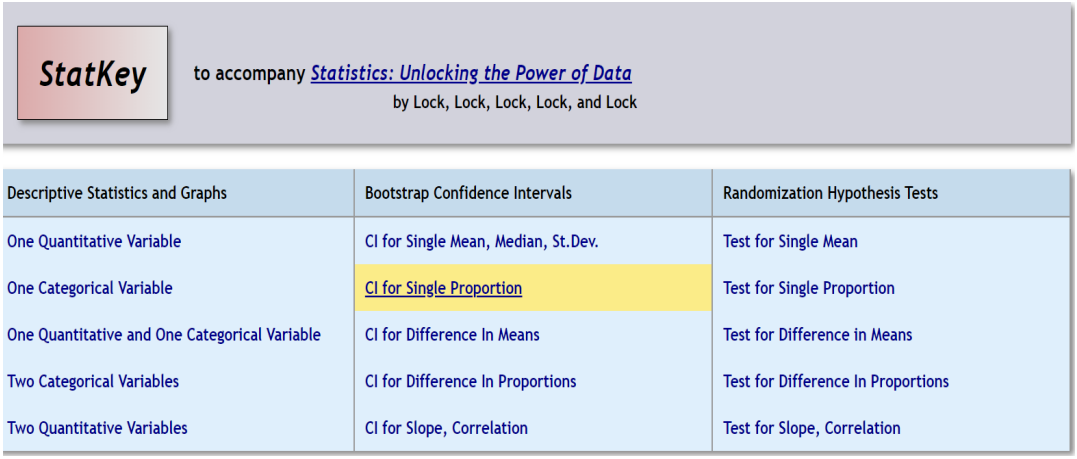
10. **Pick one of your categories that you are interested in. Which category did you pick? Give the count for that category AND the total count.** (Bear data example: I picked the fall bears as my category. There were 17 bears measured in fall. The total count (total number of bears) is 54.)

Part IV: Create a One Population Proportion (%) Confidence Interval with Bootstrapping

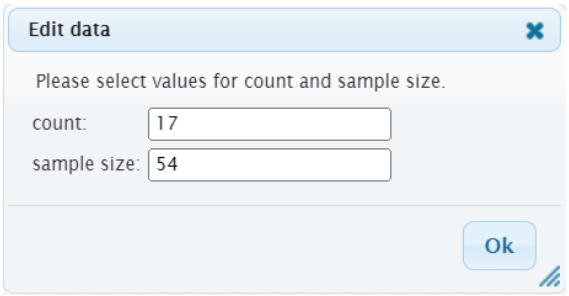
Look at the bar chart and categorical summary statistics you made for your categorical data in #8 and #9. You should see the sample proportions and counts in the summary statistics. In this example, 31.5% (0.315) of the bears were weighed in Fall, 0.241 (24.1%) of the bears were weighed in Spring, and 0.444 (44.4%) of the bears were weighed in Summer.

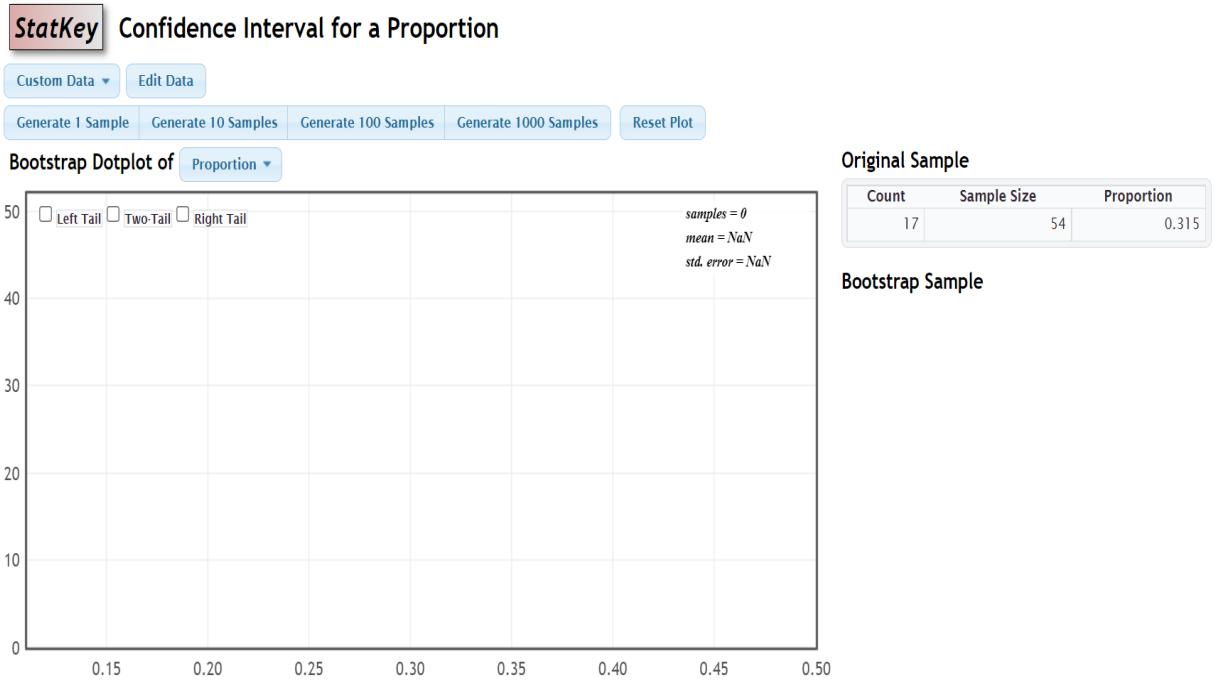
Pick one sample proportion that you want to see what the population proportion may be. We will need the counts for the category you picked and the total. In the bear example, I picked the Bears measured in Fall. There were 17 bears measured in Fall out of 54 total.

Go to www.lock5stat.com. Click on "StatKey", then look under the "Bootstrap Confidence Intervals" menu and click on "CI for Single Proportion".

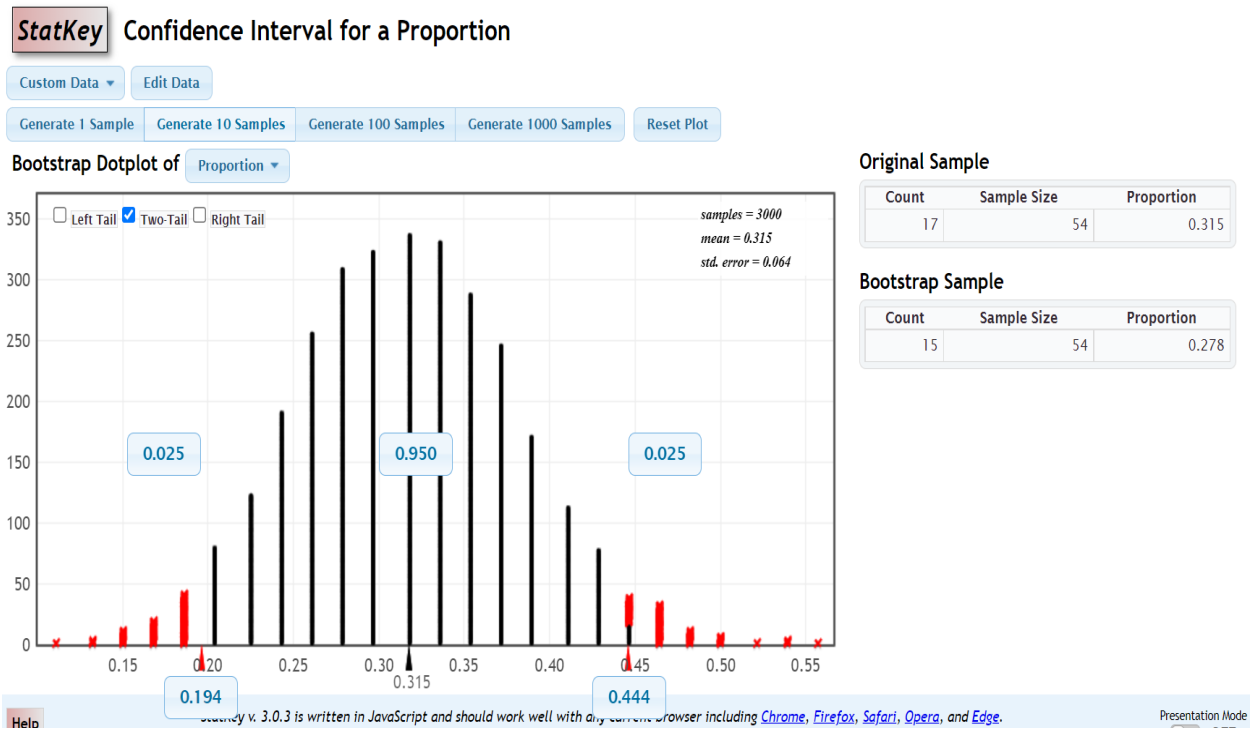


Click on the "Edit Data" button and enter the count and the total for your first category. For my example I will enter the sample counts for bears measured in Fall. The total goes in the box that says "Sample Size". Then push ok.





Now we will create a one population proportion bootstrap confidence interval to estimate the proportion of bears that are measured in Fall. Click the button that says "Generate 1000 samples" at least 3 times. You should have a total of 3000 samples or more. Then check the box that says "Two-Tail". The default is 95% confidence and that is the one we want. Take a picture of the bootstrap distribution showing your 95% confidence interval.



Put the following answers on your report:

11. **Show StatKey bootstrap distribution for the one-population proportion confidence interval. Make sure you can also see the "original sample" part one the right side of the StatKey bootstrap distributions.** *Note: Make sure the graph and statistics are large enough to read! Your instructor will not be able to grade the report if the numbers are too small to read! (Here is an example.)*

12. **What is the lower limit proportion for the bootstrap confidence interval picture in #11?** (In the bottom left box of your bootstrap confidence interval distribution. For the Fall bears example, the lower limit confidence interval proportion was 0.194)

13. **Convert the lower limit in #12 into a percentage.** (For the Fall bears example, the lower limit confidence interval proportion was 0.194, so the percentage is 19.4%.)

14. **What is the upper limit proportion for the bootstrap confidence interval picture in #11?** (In the bottom right box of your bootstrap confidence interval distribution. For the Fall bears example, the upper limit confidence interval proportion was 0.444)

15. **Convert the upper limit in #14 into a percentage.** (For the Fall bears example, the upper limit confidence interval proportion was 0.444, so the percentage is 44.4%.)

16. **Write the 95% confidence interval sentence for your one-population proportion bootstrap confidence interval. Make sure the sentence includes what the percentages represent in your categorical data.** (Example Sentence for Fall Bears: "I am 95% confident that the POPULATION PERCENTAGE of bears measured in Fall is between 19.4% and 44.4%".

17. **Calculate the Margin of Error proportion for your confidence interval using the answers in #12 and #14 and the formula below.** (For bear example upper and lower limit, the margin of error would be (0.444 – 0.194) ÷ 2 = 0.125)

    (Upper Limit Proportion – Lower Limit Proportion) ÷ 2 = ???

18. **Convert the Margin of Error in #17 into a percentage.** (For the Fall bears example, the margin of error proportion was 0.125, so the percentage is 12.5%.)

19. **Write a sentence explaining the margin of error percentage in #18.** (For Bear Example: "The sample percentage of bears measured in fall could be off as much as 12.5% from the population percentage.")

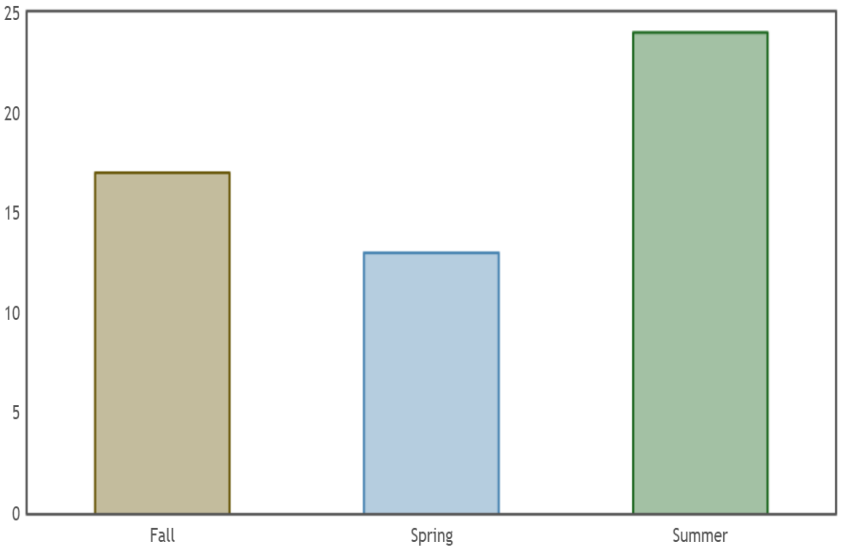
Part V: Create a Two-Population Proportion (%) Confidence Interval with Bootstrapping

20. **Look again at your categorical summary counts and proportions. We will now be comparing two categories. Pick TWO of your categories that you are interested in comparing. Which TWO categories did you pick? For the first category, give the counts for that category AND the <u>total count</u>. For the second category, give the counts for that category AND the <u>total count</u>.** (In the bear example: I want to compare the bears measured in Fall to the bears measured in Spring.

    Fall bears sample count = 17 bears measured in Fall.
    Total number of bears in sample = 54 total bears.
    Spring bears sample count = 13 bears measured in Spring.
    Total number of bears in sample = 54 total bears.

StatKey Descriptive Statistics for One Categorical Variable

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



Summary Statistics

	Count	Proportion
Fall	17	0.315
Spring	13	0.241
Summer	24	0.444
Total	54	1.000

Go to www.lock5stat.com. Click on "StatKey", then look under the "Bootstrap Confidence Intervals" menu and click on "CI for Difference in Proportions".

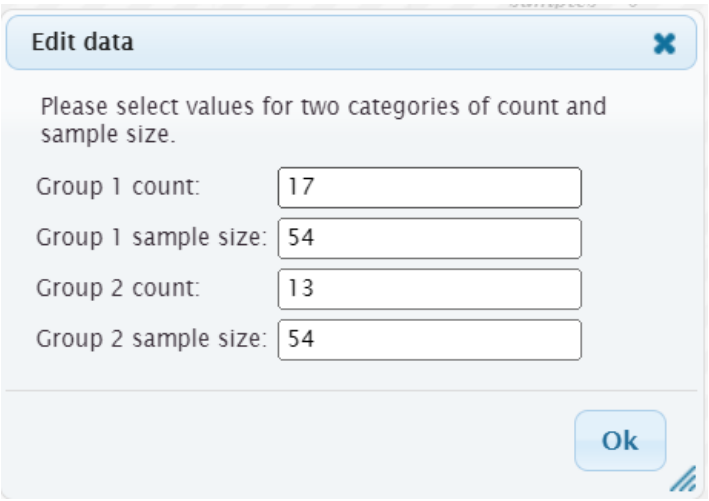
StatKey

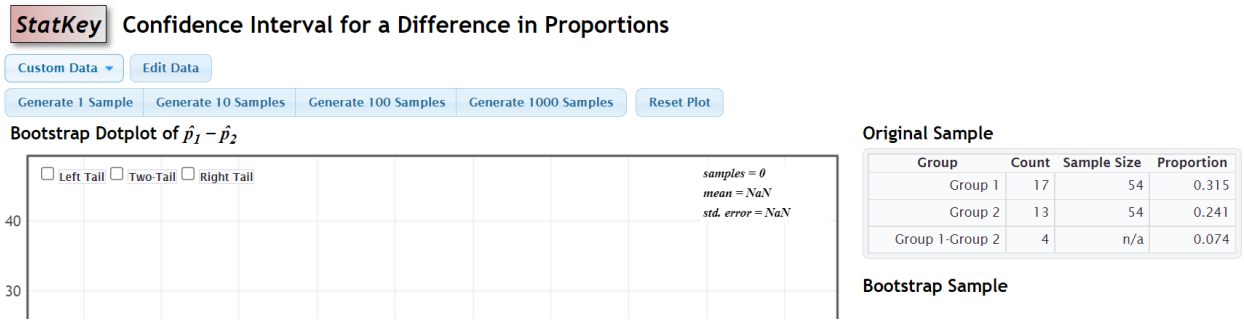
to accompany *Statistics: Unlocking the Power of Data*
by Lock, Lock, Lock, Lock, and Lock

Descriptive Statistics and Graphs	Bootstrap Confidence Intervals	Randomization Hypothesis Tests
One Quantitative Variable	CI for Single Mean, Median, St.Dev.	Test for Single Mean
One Categorical Variable	CI for Single Proportion	Test for Single Proportion
One Quantitative and One Categorical Variable	CI for Difference In Means	Test for Difference in Means
Two Categorical Variables	CI for Difference In Proportions	Test for Difference In Proportions
Two Quantitative Variables	CI for Slope, Correlation	Test for Slope, Correlation

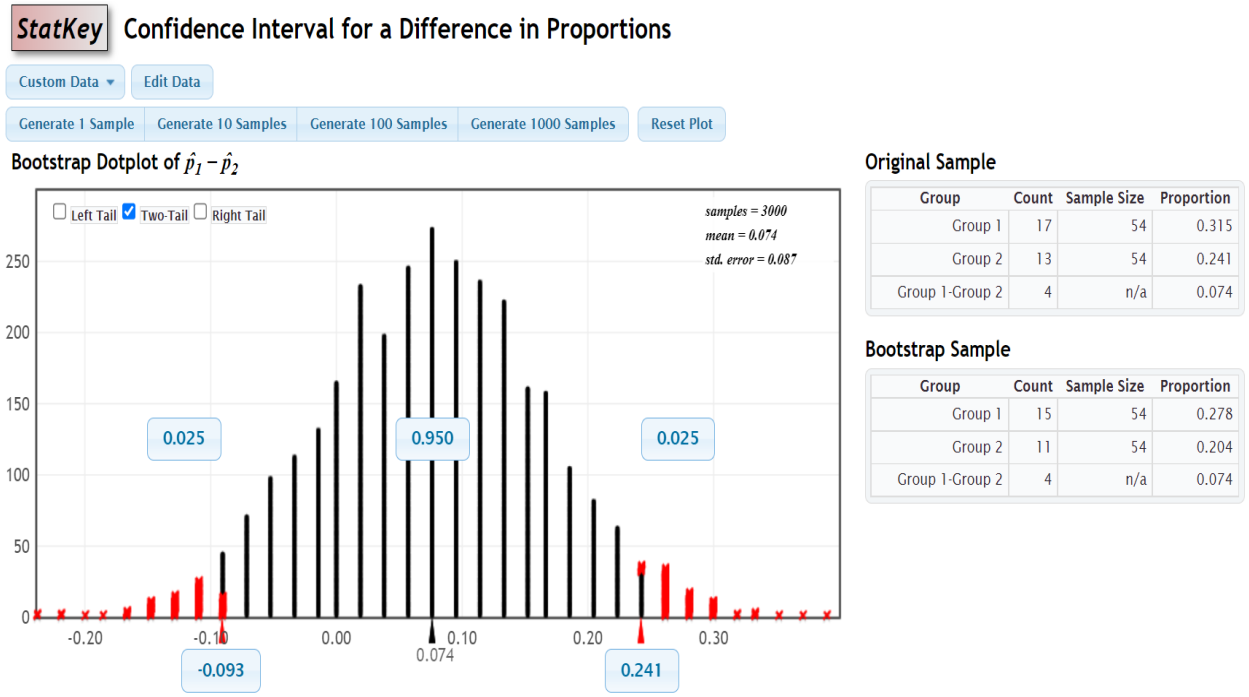
Click on the “Edit Data” button and enter the count for your first category you picked in #17 and the total count where it says “sample size”. Enter the count for your second category you picked in #17 and the total count where it says “sample size”. For the bear example, I enter the sample count for bears measured in Fall (17) and the total count sample size (54). I then enter the sample count for bears measured in Spring (13) and the total count sample size (54). Then push “OK”.

Under the “Original Sample” We see that the proportion for fall bears was 0.315 (31.5%) and the proportion for spring bears was 0.241 (24.15%) and there was a difference of 0.074 (7.4%). But is 7.4% difference significant in this case? We will need to make the bootstrap confidence interval to find out.





Now we will create a two-population proportion bootstrap confidence interval to estimate the difference between the two population proportions. Click the button that says "Generate 1000 samples" at least 3 times. You should have a total of 3000 samples or more. Then check the box that says "Two-Tail". The default is 95% confidence and that is the one we want. You will see the upper and lower confidence interval limits in the bottom right and bottom left boxes. Remember these are measuring the <u>population difference</u>. Note: Take careful note of the signs of the upper and lower limits. Are they both negative, both positive or is the lower limit negative and the upper limit positive? Take a picture of the bootstrap distribution showing your 95% confidence interval for the population difference between your categories.



Put the answers to the following on your report:

21. **Show StatKey bootstrap distribution two-population proportion confidence interval. Make sure you can also see the "original sample" part one the right side of the StatKey bootstrap distribution.** *Note: <u>Make sure the graph and statistics are large enough to read! Your instructor will not be able to grade the report if the numbers are too small to read!</u> (There is an example below.)*

22. **What group did you pick for population 1?** (Bear Example: All bears measured in Fall.)

23. **What group did you pick for population 2?** (Bear Example: All bears measured in Spring.)

24. **What is the lower limit proportion for the bootstrap confidence interval picture in #21?** (Include the negative or positive sign. In the bear example, the two-population confidence interval lower limit proportion was −0.093)

25. **Convert the lower limit proportion in #24 into a percentage.** (Include the negative or positive sign. In the bear example, the two-population confidence interval lower limit proportion was −0.093, so the percentage is −9.3%.)

26. **What is the upper limit proportion for the bootstrap confidence interval picture in #21?** (Include the negative or positive sign. In the bear example, the two-population confidence interval upper limit proportion was +0.241)

27. **Convert the upper limit proportion in #26 into a percentage.** (Include the negative or positive sign. In the bear example, the two-population confidence interval upper limit proportion was +0.241, so the percentage is +24.1%.)
28. **Is the population 1 percentage significantly higher** ((+,+) rule)**, significantly lower** ((−,−) rule)**, or not significantly different** ((−,+) rule) **than the population 2 percentage? Explain why.**

29. **Write the two-population proportion 95% confidence interval sentence in context comparing the percentages from the two populations.** (In the bear example, the signs of the upper and lower limits were (neg, pos) (−0.093, +0.241). This means there is no significant difference. Sentence: "I am 95% confident that there is no significant difference between the population percentage of bears measured in fall and population percentage of bears measured in spring.")

    (Example 2: If both upper and lower limits were positive (+0.093, +0.241), then the bear sentence would be: "I am 95% confident that the population percentage of bears measured in fall is between 9.3% and 24.1% higher than the population percentage of bears measured in spring.")

    (Example 3: If both upper and lower limits were negative (−0.093, −0.241), then the bear sentence would be: "I am 95% confident that the population percentage of bears measured in fall is between 9.3% and 24.1% lower than the population percentage of bears measured in spring.")

Part VI: Use the 5-bar Histogram and summary statistics (sample size, mean, standard deviation, etc.) **from project#2 or create a new one with your quantitative (numerical measurement) data.**

Creating the Histogram and Summary Statistics

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

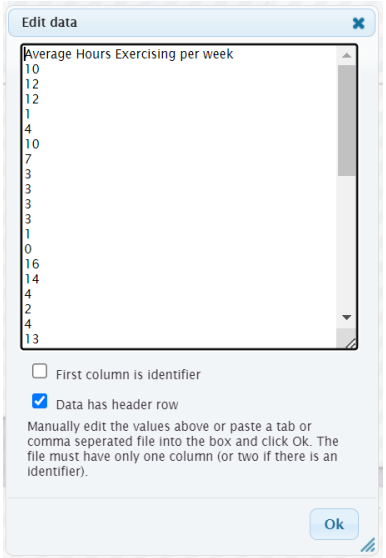
Go to www.lock5stat.com, click on "StatKey", and then click on "One Quantitative Variable" under the descriptive statistics and graphs menu.

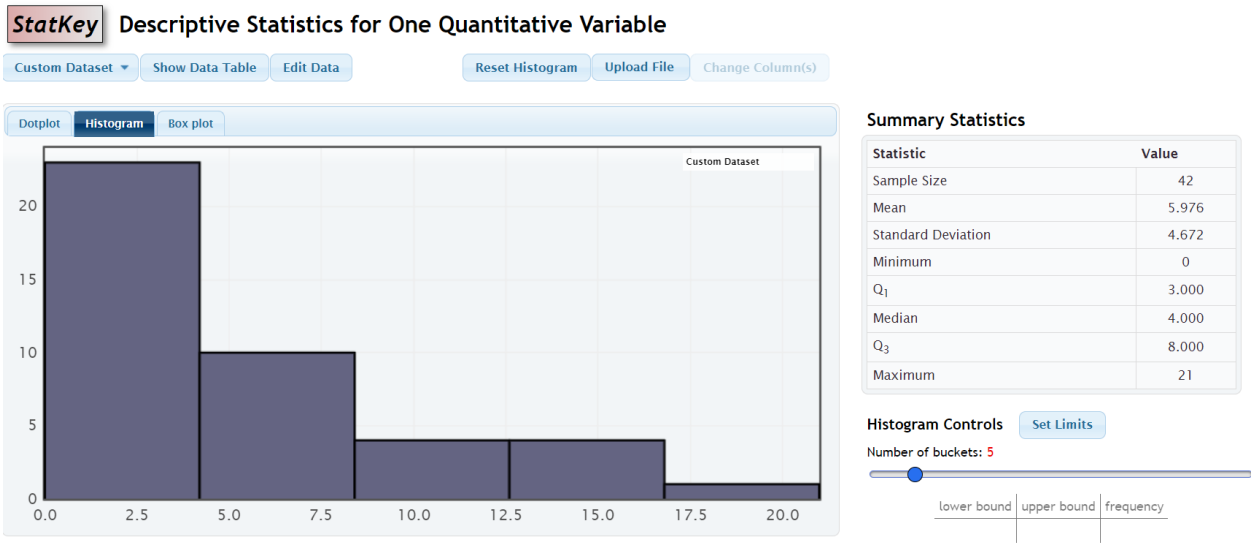
Click the "Edit Data" button. Push Control A and then delete to delete out any data listed.

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





You should see the sample mean and sample median in the summary statistics. In this example, the sample mean average was 5.976 hours of exercise per week and the sample median average was 4 hours of exercise per week. You will be making a confidence interval for either the population mean or the population median, but not both. If your data was nearly normal (bell shaped), you should make a confidence interval for the population mean. If your data was skewed, then you should make a confidence interval for the population median.

Put the following answers on your report.

30. **Copy and paste a picture of the 5-bar histogram.**

31. **What was the shape of your quantitative data?** (Nearly Normal if highest bar is in center. Skewed Right if highest bar is on left and long tail to right. Skewed Left if highest bar is on the right and long tail is on the left. Uniform if all the bars are about the same height and the graph looks rectangular.)

32. **Copy and paste the summary statistics for the quantitative data you collected.** *Note: Make sure the graph and statistics are large enough to read! Your instructor will not be able to grade the report if the numbers are too small to read!*

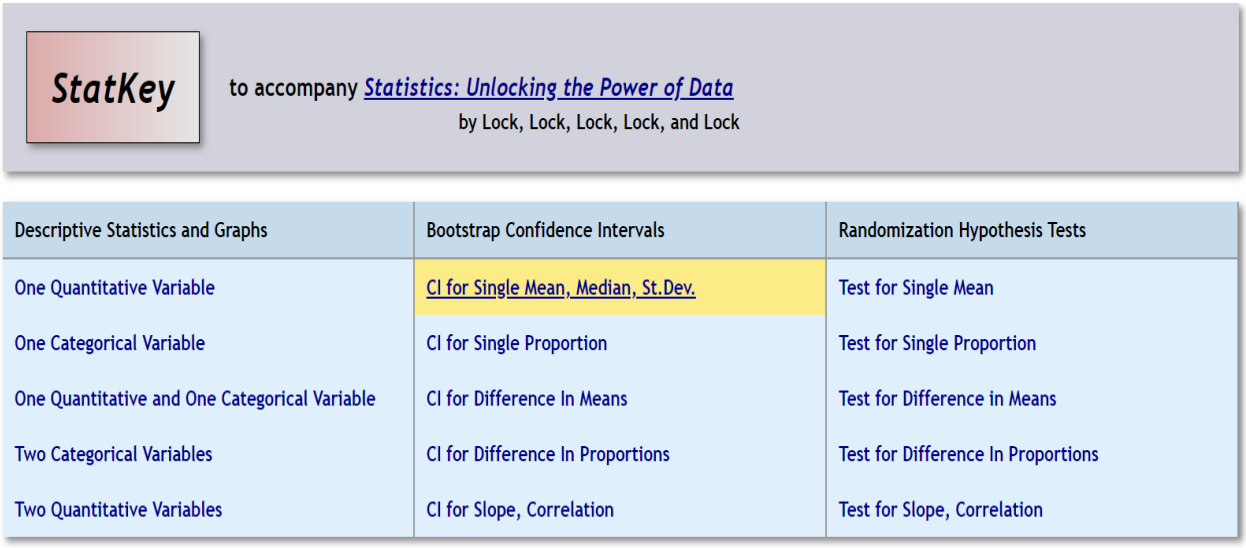
33. **Should you pick the mean or the median as your most accurate average?** (If your data was nearly normal bells shaped, we should use the mean average. If your data was skewed, Uniform, or not normal, we should use the median average.

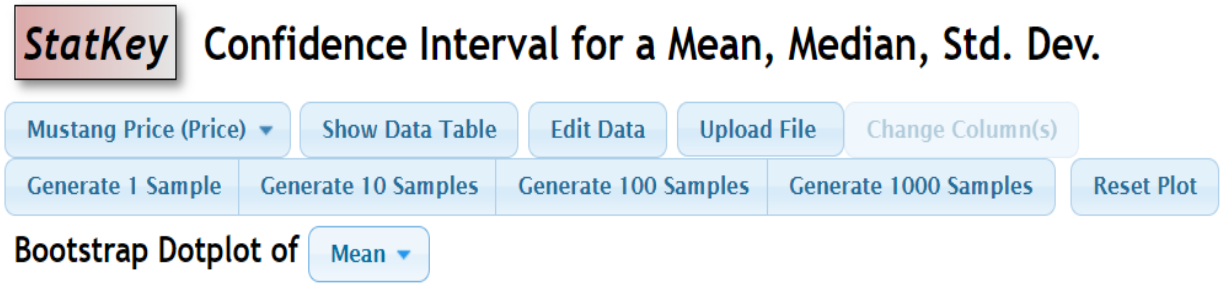
34. **What is the most accurate sample average including units?** (Give the number with units for either the mean or median depending on what you picked in #33.)

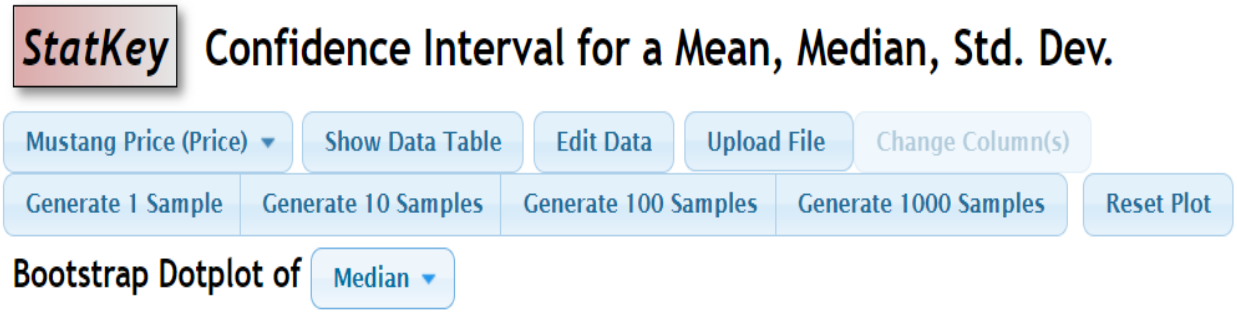
Part VII: Create a One-Population Mean or Median Confidence Interval with Bootstrapping

Open the data spreadsheet you collected in project#1. Highlight the quantitative column of data and copy it. Go to www.lock5stat.com. Click on "StatKey", then look under the "Bootstrap Confidence Intervals" menu and click on "CI for Single Mean, Median, St.Dev.". You will see "Bootstrap Dotplot of". Click the

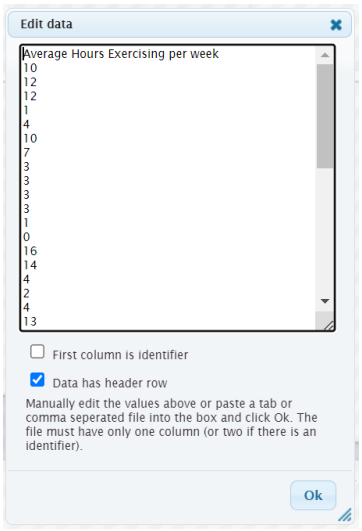
arrow and change it to “mean” or “median” depending on the shape of your quantitative sample data listed in #27.





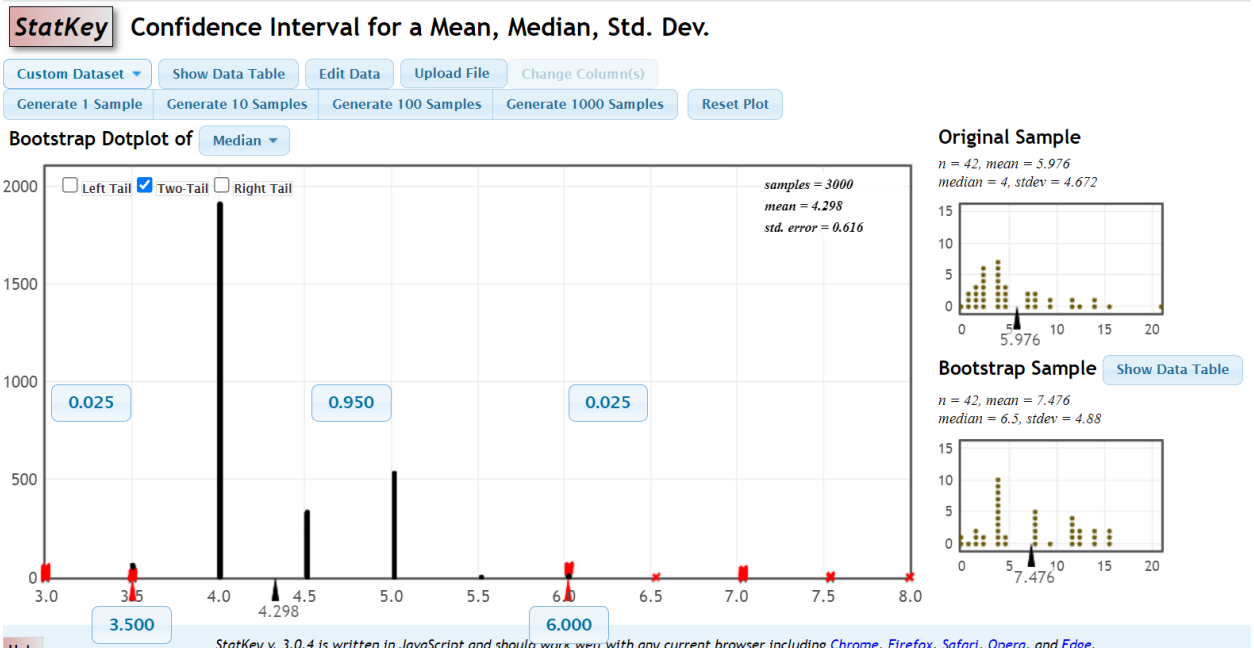


Click on the “Edit Data” button, delete any old data present, and then paste your quantitative data with title. Do NOT check the box that says “First column is identifier”. If you have your title, then check the box that says “Data has header row”. Then push OK.



Now we will create a one population mean or one population median bootstrap confidence interval to estimate the population mean average **OR** the population median average. Remember, do NOT do both. Just the one that applies to your data. Click the button that says "Generate 1000 samples" at least 3 times. You should have a total of 3000 samples or more. Then check the box that says "Two-Tail". The default is 95% confidence and that is the one we want. Take a picture of the bootstrap distribution showing your 95% confidence interval. The lower limit for your confidence interval is located in the bottom left box. The upper limit for your confidence interval is located in the bottom right box.

NOTE: Bootstrap distributions for the median often look very weird. That is OK. The computer calculated the middle 95% directly.



Put the answers to the following questions on your report.

35. **Show the StatKey bootstrap distribution for the one-population mean confidence interval OR the StatKey bootstrap distribution for the one population median**

confidence interval. Only show the picture for the mean OR the median but <u>not both</u>. Just the correct one for the shape of your data. *Note: <u>Make sure the graph and statistics are large enough to read! Your instructor will not be able to grade the report if the numbers are too small to read!</u>*

36. **Give the lower limit of the confidence interval with units.** (This is the number in the box on the lower left of the bootstrap distribution. This is a quantitative statistic so do <u>not</u> convert into a percentage. In the number of hours exercise per week example above the lower limit was 3.5 hours per week.)

37. **Give the upper limit of the confidence interval with units.** (This is the number in the box on the lower right of the bootstrap distribution. This is a quantitative statistic so do <u>not</u> convert into a percentage. In the number of hours exercise per week example above the upper limit was 6 hours per week.)

38. **Write the confidence interval sentence explaining the population mean average or the population median average in context with units.**

    Example 1 (Skewed Exercise Hours Data Median Ave Bootstrap): “I am 95% confident that the POPULATION MEDIAN average amount of exercise for all COC students is between 3.5 hours per week and 6 hours per week.”)

    Example#2: If the exercise data had been normal (bell) shaped, we would have created a bootstrap confidence interval for the mean average. The sentence would then have been “I am 95% confident that the POPULATION MEAN average amount of exercise for all COC students is between # hours per week and # hours per week.”
