

Mat 170-770 Online Elementary Statistics

Semester Project: Disney World Wait Times in StatCrunch

Spring 2024 – Due Friday, May 3rd (by 11:59pm)

200 Points - 10 point per day penalty if late.

This project is broken up into two parts. Part 1 will focus on descriptive statistics, which will come from your knowledge primarily in chapters 2 and 3. You will compute various descriptive statistics and construct graphs for each of the two sets of data and answer questions using your descriptive statistics. Part 2 will focus on inferential statistics, which will come primarily from chapters 7 and 8. You will be constructing confidence intervals, sample sizes and doing hypothesis testing. You will be using the StatCrunch dataset called **'Disney World Wait Times'** (dataset #33 in StatCrunch) to analyze the wait times of various theme park rides at Walt Disney World in Orlando Florida.

*****Before starting the assignment:** Please view my screencast video about the project. I will show you how to use the 'bins' option for frequency tables and histograms and how to copy and paste tables, graphs etc. from StatCrunch. This is posted in MML under 'StatCrunch Project' link – this is an important part of your instructions.

Statistics are collected on theme park wait times in order to better improve the efficiency of the wait times, as well as to inform current theme park patrons of how long certain wait times are in order for them to better organize their visit. Disney World posts their wait times in real time within the theme park and on their theme park app, but people inside the parks also report live wait times to social media websites, travel blogs, etc. Use this data to complete the following:

Part 1: Descriptive Statistics: Tables, graphs, statistics, and answer questions. Chapters 2 and 3 topics are used here.

For Part 1, you will be looking at two specific rides and times: **Space Mountain at 5:00pm** and **Flight of Passage at 5:00pm**. Each one of these rides has a sample of 50 data values that represent the amount of time waiting in line before getting on the ride (wait times). In part 1, you will work with these two rides only: nothing else from the data set.

1. Using the 50 wait times from each of the two rides, put each one into a frequency table. So, you will have two frequency tables. You must use the option 'bin numeric values' in the frequency table selection when using StatCrunch. This is where you enter the first lower class limit in the 'start at' box and the class width in the 'width' box. The frequency table for Space Mountain at 5:00pm should have six categories with a class width of 20 minutes and start with a lower class limit of 0. The frequency table for Flight

of Passage at 5:00pm should have seven categories with a class width of 20 minutes and start with a lower class limit of 40.

2. Create a histogram for each dataset. Use StatCrunch to generate the histograms. Again, you must use the 'bins' option in the histogram selection the same way did when creating the frequency tables. Enter the same starting lower class limit in the 'start at' box and the class width in the 'width' box that you used in number 1. This way, your histograms and frequency tables will match perfectly. Use the original data sets in the two columns as given to make each histogram.
3. Run descriptive statistics on BOTH datasets. Find the following statistics for each of the two rides: mean, median, mode, range, standard deviation, variance, minimum, maximum, Q1 and Q3. Use all the original 50 data values for each ride – **DO NOT use your frequency tables to compute the descriptive statistics!**
4. List the 5-number summaries and construct box plots for both rides. Construct both a regular boxplot **AND** a modified boxplot for each ride. The modified boxplot will be used to identify any outliers in the wait times.

Answer the following questions based on your descriptive statistics. These are not simply 'yes' or 'no' answers; parts a, b, e, and f require a couple of sentences. Parts c and d can each be answered in a few short words.

- a. Comment on the 'shape' of the data from each ride. Are the wait times bell-shaped, skewed, close to bell-shaped, etc.
- b. Does there appear to be a significant difference in the wait times between these two rides? Can you draw any conclusions about the two rides from this data?
- c. Calculate a z-score for 30 minutes, 60 minutes and 120 minutes for each ride (six z-score calculations total). **Show your calculation for each z-score.** Are any of the z-scores for these times considered unusual?
- d. Are there any outlier wait times for each ride in the datasets? If so, what are they and for which ride? Use the modified boxplot you created in #4 to locate any outliers. (**Hint:** one of the datasets has outliers, one does not!)
- e. Are there any conclusions that you think can be drawn from the data? Do any of the results surprise you or not?
- f. These wait times come from a travel website where people in the theme park actually post the wait times to the website in real time. These wait times often differ from the actual wait times posted by Disney in their theme parks (Disney World posted wait times tend to be longer than the actual wait times posted by the people in line waiting for the rides. Do you think one source of the wait times (the theme park guests, or the 'official' Disney World posted wait times) is more accurate than the other? Would you rely on one more than the other?

All your calculations should be applied to both data sets separately. Never combine any of the two data sets together!

To help you check that you are on the right track with your descriptive statistics calculations, here are two of the answers from #3 in part 1:

Mean wait time for Space Mountain at 5:00pm is **55.28 minutes**.

Mean wait time for Flight of Passage at 5:00pm is **93.2 minutes**.

If you do not get those mean times from the data set, then you have done something incorrect. Check over your steps in using the StatCrunch program.

Part 2: Inferential Statistics: Confidence intervals, sample sizes, hypothesis testing, and answer questions: Chapter 7 and 8 topics are used here.

For this part, you will continue using the Disney World Wait times data set. Now, you will focus on five of the rides. Space Mountain, Small World, Rock N Roller Coaster, Tower of Terror, and Flight of Passage. You will use the 5PM time for all five rides.

5. Construct 95% confidence intervals for the **mean** wait time for the following five rides: Space Mountain at 5PM, Small World at 5PM, Rock N Roller Coaster at 5PM, Tower of Terror at 5PM, and Flight of Passage at 5PM. Using the confidence intervals, answer the following questions:
 - a. Which ride has the longest mean wait time? Which has the shortest?
 - b. Calculate the margin of error used in the Space Mountain confidence interval. Round the answer to one decimal place. Show your calculation.
 - c. Would it be reasonable to expect a mean wait time of 40 minutes for the Tower of Terror at 5PM? Why? Would it be reasonable to expect a wait time of 40 minutes for Space Mountain at 5PM? Why?
 - d. Compare the confidence interval of Small World to Tower of Terror. Is it reasonable to state that Small World has a shorter wait time than Tower of Terror at 5PM time? Why?
 - e. In one sentence write the formal interpretation of the confidence interval for Flight of Passage at 5PM.
 - f. What are some reasons for the differences in mean wait times among rides when you compare rides such as Small World (an original ride in Magic Kingdom since 1971) and Flight of Passage (a ride that opened in Animal Kingdom in 2017)? (2-3 sentences is sufficient)

6. Now focus on the Tower of Terror at 5PM sample data. Use a .05 significance level to test the claim that the **mean** wait time for Tower of Terror at 5PM is equal to 34 minutes. Complete the following:

- a. Write out the hypothesis statements.
 - b. Compute the test statistic and the p-value.
 - c. Make the initial decision using the P-Value method and determine if you should reject the null hypothesis or fail to reject the null hypothesis.
 - d. State the formal conclusion addressing the original claim.
 - e. Compare the sample mean (you will get this in your output when you do the hypothesis test) to the claimed mean time of 34 minutes. Is the difference between those two values statistically significant? Does the difference between those values seem to have any practical significance to someone who is waiting in line for this ride?
 - f. Use the 95% confidence interval for the **mean** wait time for Tower of Terror at 5PM that you constructed in #5 to test the claim that the mean wait time for Tower of Terror is equal to 34 minutes. What do you conclude about the claim based on the confidence interval?
7. Now focus on the Space Mountain at 5PM sample data. A theme park administrator claims that the **majority** of wait times for this ride at 5PM is longer than forty minutes. Test this claim using a .05 level of significance and use the following summary statistics: There are 29 wait times in the Space Mountain at 5PM sample that are longer than forty minutes out of the total sample of 50 observations (**use the ‘with summary’ option in StatCrunch, do not use the with data option**).
- a. Write out the hypothesis statements.
 - b. Compute the test statistic and the p-value.
 - c. Make the initial decision using the P-Value method and determine if you should reject the null hypothesis or fail to reject the null hypothesis.
 - d. State the formal conclusion addressing the original claim.
 - e. Construct a 90% confidence interval for the **proportion** of wait times for Space Mountain at 5PM that are greater than forty minutes in order to test the claim that the majority of wait times for Space Mountain at 5PM are longer than forty minutes. What do you conclude about the claim based on the confidence interval?
8. Compute Sample sizes. Compute the following sample sizes using the given information and the sample size program in StatCrunch.
- a. Compute the minimum sample size required to get a 95% confidence interval estimate for the **mean** wait times for Flight of Passage at 5PM. Assume a standard deviation of 30 minutes and a margin of error of 9 minutes. The dataset for Flight of Passage used a sample size of 50. Compare this to your answer. Was 50 an appropriate sample size to use? Why or why not? Show what you entered into the StatCrunch program and what you got for an answer.
 - b. Compute the minimum sample size required to get a 95% confidence interval estimate for the **mean** wait times for Small World at 5PM. Assume a standard deviation of 17 minutes and a margin of error of 4 minutes. The dataset for Small World at 5PM used a sample size of 50. Compare this to your answer. Was 50 an

appropriate sample size to use? Why or why not? Show what you entered into the StatCrunch program and what you got for an answer.

Grading:

Part 1:

Descriptive statistics:

Question 1: 20 total points

Question 2: 10 total points

Question 3: 20 total points

Question 4: 15 total points

Answer questions:

a: 5 points

b: 5 points

c: 10 points

d: 5 points

e: 5 points

f: 5 points

Part 2:

Question 5:

Confidence intervals (5) – 10 points

Part a: 5 total points

Part b: 5 total points

Part c: 5 total points

Part d: 5 total points

Part e: 5 total points

Part f: 5 total points

Question 6:

Part a: 4 total points

Part b: 5 total points

Part c: 3 total points

Part d: 3 total points

Part e: 5 total points

Part f: 5 total points

Question 7:

Part a: 4 total points

Part b: 5 total points

Part c: 3 total points

Part d: 3 total points

Part e: 5 total points

Question 8:

Part a: 7.5 total points

Part b: 7.5 total points

Points are based on completeness, following instructions, accuracy of calculations and graphs/tables, answers to all questions.

Most of the work should be completed using the functions in the StatCrunch video (watch my StatCrunch Videos under Mr. J's StatCrunch Videos in MML if you need a refresher on how to use these StatCrunch functions)

Your final submission should include the following:

For Part 1:

Two frequency tables, two histograms, two sets of descriptive statistics, two five number summaries, four box plots (a regular boxplot and a modified boxplot for each ride), six z-score calculations, and the answers to ALL questions. All tables, graphs and statistical calculations should be done on StatCrunch (except the z-score calculations – those you must calculate by hand and include in your final submission). You can easily copy all of your graphs, tables, and output tables from StatCrunch into WORD, a google document, PDF, etc. Written answers to questions should be typed out neatly. If I cannot read something, you may lose points!!

For Part 2:

All confidence intervals, all parts of the hypothesis tests, and computation of the sample sizes, and answers to all the questions. You may copy and paste in your output directly from StatCrunch into your final document. Written answers to questions should be typed out neatly. If I cannot read something, you may lose points!!

Completed projects are to be emailed to me (glennjablonski@triton.edu) by the end of day of the deadline (by 11:59pm). Email it directly to my email, do NOT use Blackboard.