

Lab 7 (90 pts. + 15 pts. BONUS): Two Sample t Independent and 2-Sample t Pairs Procedure

Objectives: Confidence interval and significance tests for two samples.

A (45 points + 5 points BONUS) Is the departure delay at Chicago's O'Haire Airport (ORD) different from the departure delay at Chicago's Midway Airport (MDW)? (Data Set: cleaned airline data set)

We are interested if the Departure Delay ("DepDelay") is different at the two Chicago Airports. Remember in this study, you need to restrict the input to only these two airports. See the appropriate tutorial for details. This data again needs to be transformed. However, since DepDelay has negative values and you can not take the log of a negative value, you need to translate the DepDelay so that all the values are positive. In this case, the minimum value for all airports is -25, therefore, please add 25.1 to the DepDelay before you take the log of the variable. This will not change the answer to the hypothesis test, however, be careful if you transform the confidence interval or bound back to the original values. See the tutorial for Lab 6 for details on how to do a log transformation.

1. (5 pts.) Code. BONUS (5 pts.) Determine the minimum value of the departure delay for all of the airports via your software package.
2. (5 pts.) Should you use two sample independent or 2-sample pairs procedure to analyze the data? Please explain your answer without referring to the format of the data. If this is a pairs situation, please state the common characteristic that makes these data paired.
3. (5 pts.) Would you consider using a one-sided or two-sided alternative for this analysis? Explain your decision.
4. (5 pts.) Do you think these data are Normally distributed using the transformed data? Use graphical methods to examine the appropriate distribution. Write a short summary of your findings.
5. (10 pts) No matter how you answered in part 3, determine and interpret the 95% confidence interval of the population mean transformed time difference of the Departure Delay of the two Chicago airports.
6. (10 pts) No matter how you answered in part 3, test the alternative hypothesis that the two airports have different population mean departure delay types. Please use the transformed data. You may assume a 0.05 significance level for the test. Remember to use the full 4-step process.
7. (5 pts) Compare the answers of 5 and 6. Are they saying the same thing? Please explain your answer. In addition, answer the original question in one English sentence, that is, in practical terms are the departure delays of the two airports different? You may need additional output to justify your answer.

B (45 points + 10 points BONUS) Is Taxiing in to the gate less than Taxiing out before takeoff? (Data Set: AirlineTaxi_In_Out.txt)

Because planes often have to wait in line before they can take off, it is expected that the time between landing and arriving at the gate (aveTaxiIn) will be less than the time that it takes to go from the gate until the plane has taken off (avgTaxiOut). In fact, previous studies have indicated that the difference is more than 10 minutes. Is that statement true in these data? To save time, we have generated the data with the average taxi in times and the average taxi out times for each airport. As a bonus (8), you can create code that will generate the dataset used in this part from the original cleaned data set.

1. (5 pts.) Code.
2. (5 pts.) Should you use two sample independent or 2-sample pairs procedure to analyze the data? Please explain your answer without referring to the format of the data. If this is a pairs situation, please state the common characteristic that makes these data paired.
3. (5 pts.) Would you consider using a one-sided or two-sided alternative for this analysis? Explain your decision.
4. (5 pts) Do you think these data are Normally distributed? Use graphical methods to examine the appropriate distributions. Write a short summary of your findings.
5. (10 pts.) No matter how you answered in part 3, determine and interpret the appropriate 99% confidence bound of the population time difference between the taxi times.
6. (10 pts) No matter how you answered in part 3, test the alternative hypothesis that the difference between the population means of TaxiIn and TaxiOut is more than 10 minutes. You may assume a 0.01 significance level in your test. Remember to use the full 4-step procedure.
7. (5 pts) Compare the answers of 5 and 6. Are they saying the same thing? Please explain your answer. In addition, answer the original question in one English sentence, that is, in practical terms is the difference between the population means of TaxiIn and TaxiOut more than 10 minutes. You may need additional output to justify your answer.
8. (10 pts BONUS) Provide the code (in any language, we will only provide the code for R and SAS in the answer key) that will convert the cleaned airport dataset into the dataset used in this part.