Inference for Paired Data

Matt Allen

06/19/2018

Create a Word docx from this R Markdown file for the following exercise. Submit the R markdown file and resulting Word docx file via D2L Dropbox.

## Exercise 1

To reduce ankle injuries, restrictive appliances such as taping and spatting (applying tape over the shoe and sock) have been employed. As part of a study at UWL, subjects also completed a 5-point Likert-type scale survey regarding their perceptions of the movement of each ankle appliance during exercise.

Researchers would like to compare the central values for perceptions of the movement of taped ankles compared to spatted ankles using and to estimate the difference with 90% confidence.

### Part 1a

Load the data set AnkleMovement.rda from the DS705 package.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 1b

Create a new variable of the differences, with the perceptions of the spatted ankle (spat) subtracted from the perceptions of the taped ankle (tape).

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1b -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 1c

Create a boxplot and histogram for the sample of differences.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1c -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 1d

Comment on the suitability of this data for the paired t-test, the Wilcoxon signed rank test, and the sign test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 1d -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 1e

Because the choice of test is somewhat unclear, as happens often in real life, try all three tests to compare the central values for subject’s perceptions of the movement of taped ankles compared to spatted ankles using .

Do the t-test first:

#### Step 1

Define the parameter in words in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step1 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 2

State the null and alternative hypotheses for the test using the symbol you defined previously.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step2 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 3

Use R to generate the output for the test you selected.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step3 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

#### Step 4

State a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step4 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 5

Write an interpretation in the context of the problem for the 90% CI for the population mean difference.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step5 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 6

Perform the Wilcoxon Signed Rank Test.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step6 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

#### Step 7

Perform the sign test.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step7 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

#### Step 8

Construct a bootstrap confidence interval at a 90% level of confidence for the mean difference in population mean perception of movement for taped and spatted ankles. Use a bootstrap sample size of 5000. Compare this interval with the results of the 90% *t*-interval.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step8 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

#### Step 9

Compare the results of the three hypothesis tests and also whether or not the 90% bootstrap interval agrees with the result of each test. Which procedure should be reported and why?

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 1e.step9 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

## Exercise 2

In a nationwide study of insurance claims (in dollars) filed in the previous year, a random sample of 125 claims was selected from all claims for vehicles classified as small, meaning the gross vehicle weight rating (GVWR) is less than 4500 pounds.

For each claim, the insurance company’s estimate for the claim was also provided.

The data frame SmallAuto.rda contains the claims and estimates for each vehicle class.

### Part 2a

Load the data SmallAuto from the DS705 package.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 2b

Construct histograms and boxplots for both the estimated claims and actual for the small class of vehicle. Describe the shapes of these distributions.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2b -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

### Part 2c

Create a new variable of the differences for small vehicles, with the difference being the estimated claim amount minus the actual claim amount. The estimated claim amounts in the first half of the vector are paired with the actual claim amounts in the second half of the vector so that row 1 and row 126 form a pair, rows 2 and 127, etc.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 2c

Create a boxplot, histogram, and normal probability plot for the sample of differences. Also, obtain the P-value for a Shapiro-Wilk normality test.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2c -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 2d

Comment on the shape of the distribution of differences and the suitability of this data for the paired *t*-test, the Wilcoxon signed rank test, and the sign test. Which test will you use?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2d -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 2e

Conduct an appropriate test to see if the population central values for the estimated claim amount is less than for the actual claim amounts for vehicles in the small class using .

#### Step 1

Define the parameter in words in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e.step1 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 2

State the null and alternative hypotheses for the test using the symbol you defined previously.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e.step2 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

#### Step 3

Use R to generate the output for the test you selected. Pay close attention to the order of subtraction done behind the scenes in R.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e.step3 -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

#### Step 4

State a statistical conclusion at and interpret it in the context of the problem.

#### -|-|-|-|-|-|-|-|-|-|-|- Answer 2e.step4 -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 2f

Write an interpretation in the context of the problem for a 95% two-sided confidence interval.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 2f -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

Replace this text with your answer here.

## Exercise 3

The data frame AutoIns is very similar to Small Auto.

In a nationwide study of insurance claims filed in the previous year, a random sample of 125 claims was selected from all claims for vehicles classified as small, meaning the gross vehicle weight rating (GVWR) is less than 4500 pounds A separate sample of 125 claims for vehicles classified as standard, meaning the GVWR is between 4500 and 8500 pounds.

For each claim, the insurance company’s estimate for the claim was also provided.

The data frame AutoIns.rda contains the claims and estimates for each vehicle class. The variables in the data frame are defined as follows:

claim.small = the actual claim amount in dollars for a vehicle in the small class

est.small = the estimated claim amount in dollars for a vehicle in the small class

claim.standard = the actual claim amount in dollars for a vehicle in the standard class

est.standard = the estimated claim amount in dollars for a vehicle in the standard class

### Part 3a

Load the data AutoIns from the DS705 package and look at the structure of the data in the file.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3a -|-|-|-|-|-|-|-|-|-|-|-

# Insert your R code here.

### Part 3b

Is the data “stacked” or “side-by-side” (“tall” or “wide”)?

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3b -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 3c

Which pairs of variables in the data frame are independent of each other? You can use the variable names to identify them.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3c -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.

### Part 3d

Which pairs of variables in the data frame are paired (matched pairs)? You can use the variable names to identify them.

### -|-|-|-|-|-|-|-|-|-|-|- Answer 3d -|-|-|-|-|-|-|-|-|-|-|-

Replace this text with your answer here.