Interactive Assignment 2: Reviewing Inferential Statistics

Introduction

In this lab, we will review how to do t-tests and correlations in R. This lab assumes that you have covered the theory behind these tests in a previous class, including what is meant by t-values, r-values, and p-values.

After completing this lab, you should be able to:

1. Import data into R from a csv file
2. Know how to conduct a t-test in R and interpret the output
3. Know how to conduct a correlation in R and interpret the output.

In this lab, we will use some data that some students collected for a research project previously. In this project, the researchers gave students a series of questionnaires looking how personality traits affected a participant's likelihood to conform.

Whenever you share data with other people, you should always include a guide that tells people what each variable is and what the variables represent. I will give you a guide for each of your analyses in this class.

In this case, the variables are named as follows:

- Participant: a participant number
- E: extraversion score
- A: agreeableness score
- C: conscientiousness score
- N: Neuroticism score
- O: Openness score
- Gender: gender coded F and M
- Age: participant age
- Conformity: their score on a survey about how much they would conform
- RATING: a participant's rating to the question "how much do I think I conform to others' behaviors"
- SelfEsteem: a score on a self-esteem inventory with higher scores indicating more self-esteem

One last note about data sharing. In this case, the individuals who shared this data with me did not use a consistent capitalization scheme. I left the data as is in order to illustrate why this is not a good idea. You should develop a consistent way you name your variables when you enter data. Here are my suggestions:

- Never use capital letters, unless it's a normally capitalized abbreviation, like "IQ". Or you could just capitalize the first letter all the time.
- Never, ever use spaces. I use underscores to divide spaces. Some others will bunch all the words together, capitalizing the first letter of each word. Either way is fine, but be consistent.

In Interactive Assignment 1, we covered how to import data from a csv file. We will do that here.

Step 1: Go ahead and make an R notebook for this lab. Download the IA2.csv data and save it in the R notebook.

Step 2: Create a code chunk to import the IA2.csv into your R notebook as the data frame `d`.

Summary Statistics in R

We are going to use the `summary()` and `aggregate()` functions in the `psych` package to first look at some summary statistics. You can see more about these functions in the last assignment. As mentioned in the last

assignment, you can use the `summary()` function to get a summary about a variable, with several summary statistics. For instance, to look at a summary of the extraversion variable, we would type:

```
summary(d$E)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   19.00   26.00   25.67   31.25   40.00
```

Notice that this tells us the mean for extraversion is 25.67. To get the standard deviation, we have to use the sd() command.

```
sd(d$E)
```

```
## [1] 7.158417
```

Step 3: Use the `summary()` and `sd()` function to find the mean and standard deviation of the other 4 personality tests (agreeableness, conscientiousness, neuroticism, and openness, along with the Conformity variable). Write a table below with the means and standard deviations of each of these variables.

When looking at a dataset, it is a good idea to look at patterns in the summary data. In this case, we are not that interested in the means for each variable. It's meaningless if the mean for extraversion is higher than the mean for agreeableness, because those are not two variables that can be compared, since they are on different scales. However, it may be useful to look at the range and standard deviation, to see if there are any outliers or variables with a very low standard deviation.

In another lab, we'll discuss graphical ways to do this, but here we will just look at the raw statistics to see if there are any major differences. Make a note in your annotation about whether you see any differences in the variables

Examining Gender Differences Using T-tests

Now we are going to see if there are any differences between male and female participants in their scores on the personality inventories and in conformity. To conduct t-tests, we can use the `t-test()` function in R.

There are two ways to do t-tests in R. The first way is to give R two different variables as two different arguments. For instance, if we wanted to compare to see whether extraversion is higher than neuroticism, we would type the below code.

```
t.test(d$N, d$E)
```

However, like I mentioned above, comparing how high extraversion is to neuroticism doesn't make sense. We want to compare a single personality trait with another variable that tells us which gender a person is. To do that, we have to enter data different way.

In R, we use what is called a *formula notation* to describe when we want to input one variable **by** another variable. What this means is that we have a dependent variable or an outcome variable, and then we have one or more variables that tell us which groups to use to divide the dependent variable.

This function has one input, but it has two variables divided with a tilde. The first variable is the dependent variable and the other is the the grouping variable that tells what category an observation is in. The independent variable must have exactly 2 levels, because a t-test only compares 2 groups. This makes the `Gender` variable a good candidate.

To do a t-test for the Extraversion variable with Gender as an independent variable we would type:

```
t.test(d$E~d$Gender)
```

```
##
##  Welch Two Sample t-test
```

```
##
## data:  d$E by d$Gender
## t = 0.20057, df = 25.11, p-value = 0.8426
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.740950  4.548403
## sample estimates:
## mean in group F mean in group M
##        25.76087        25.35714
```

Go ahead and add this as a code chunk in your notebook.

This is the output you should get with this test. The second line of the output is the one you are most interested in. This tells us the t value, the degrees of freedom (abbreviated df), and the p value. The degrees of freedom is a corrected value, so it is often a decimal.

The 95% confidence interval (CI) is on the 5th line. In this case, it gives us a 95% confidence interval about the difference between the two groups. Generally, if a 95% confidence interval of a difference does not contain 0, it is equivalent to being significant at a $p < .05$ level.

The last line has the mean of the variable for each group. Notice in this case the mean for group F, for female, is very close to the value for group M, for male, which is why this t-test is not significant and why the 95% CI contains 0.

Step 4: Using this function, do a t-test examining whether men and women vary in the other 4 personality variables and whether they vary in conformity. Report your t-value and p-value for each of the tests below, using APA formatting: $t(\text{df}) = \text{tvalue}$, $p = \text{pvalue}$, or $t(25) = .20$, $p = .84$. Remember that gender should be the independent variable here.

Correlation with R

The researchers were interested in whether there was a correlation between the personality variables and conformity. To test whether a correlation is present, we can use the `cor.test()` function in R. This function requires two input variables, which will be correlated. To see the correlation between Extraversion and Conformity, we would type:

```
cor.test(d$E, d$Conformity)
```

```
##
##  Pearson's product-moment correlation
##
## data:  d$E and d$Conformity
## t = 1.2445, df = 58, p-value = 0.2183
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.09660536  0.39886592
## sample estimates:
##       cor
## 0.1612755
```

The last number gives us the correlation. The significance of the correlation is listed as a p-value in the second row, along with the degrees of freedom. In this case, we have a small and insignificant correlation between Extraversion and Conformity, $r(58) = .16$, $p = .21$.

Step 5: Using the correlation function, examine if there are any correlations between the other 4 personality variables and conformity. Report your r-value and p-value for each of the tests after each code chunk, using APA formatting.

Step 6: Using the `cor.test()` function, examine if there are any correlations between SelfEsteem and conformity. Make sure you annotate your code.

Step 7: Is the Conformity variable correlated with the RATING variable? How so?

Step 9: Now you're all done! Knit the notebook and submit.