Lab 7: Two Group Mean Claims

Lab 7: Testing claims with two means

The data you will use for this lab come from some experiments comparing older and younger adults on various tests of cognitive and visual function. To input the data, follow the following steps.

Step 1: Create a new folder for this lab. Download the .csv data which contains this lab's data. It is named "Lab7Data.csv". Make an R Notebook and save it in this folder. Add a code chunk that loads the `tidyverse` package. Add a second code chunk that imports the data to the data frame "Lab7Data.csv" to a dataframe named `d`. Make sure you annotate each code chunk so it is clear what you are doing.

Step 2: Once you do this, inspect your data by clicking it in the environment window. You should have the following variables:

- Sex = 1 if male, 2 if female
- AgeGroup = 1 if younger adult (18-25), 2 if older adult (>65)
- MindEye – score of the Mind in the Eyes test which is an emotion recognition test
- Snellen – vision score. It is the second number of the 20/20 test you might get at the eye doctor. Higher numbers indicate worse vision (e.g. 100 would indicate a person has 20/100 vision)
- MarsForm1 and MarsForm2 – tests of contrast sensitivity which measures a person's up close vision and their ability to read fine lines. Mars Form 1 is a test without reading glasses, Mars Form 2 is a test with reading glasses
- BentonFaceRecog – score on a face recognition test
- Pattern – test which measures processing speed by having people match patterns as fast as they can. The number represents how many patterns they matched in 60 seconds
- Shipley – score on a 40 question vocabulary test.
- Age – the person's actual age.

We are interested in age differences between the tests. We want to see whether there are any differences between older and younger adults. A good way to get summary data is something called the `aggregate` function. We covered this in Lab 3, but I will review it here.

The aggregate function allows us to see summary statistics, like mean, median, and standard deviation of a variable, when dividing that variable into groups based on another grouping variable. It would not be as informative to know the mean Snellen score for all participants versus knowing the mean for older adults and younger adults separately.

As mentioned in Lab 3, the `aggregate()` function has three parts. The first part is a *formula* which describes the dependent variable and the independent or grouping variable. The second part is the data frame itself. The third part is an argument that says what function to apply.

```
aggregate(MindEye~AgeGroup, d, FUN = "mean")
```

```
##   AgeGroup  MindEye
## 1        1 25.69935
## 2        2 24.72222
```

Note the input here divides the data into AgeGroup 1, which is younger adults and AgeGroup 2, which is older adults. It gives the mean for both groups.

Step 3: Use the aggregate function to display the mean and standard deviation for each test separately for older and younger adults. Include this code in your R notebook output and make sure you annotate each code chunk so it's clear what test each output is referring to. In addition, after each test, note if you think that there is a large difference between the two groups. Use the standard deviation as a guide to see whether the difference is large. A difference of one standard deviation between groups is quite substantial.

Conducting T-tests

Now we want to see whether the differences are significant. This involves the t.test function in R. To run a t.test, you are going to use two separate parts. Here is the formula to run a t-test for the MindEye variable.

This formula has two parts, just like the describeBy formula. Instead of a comma, it uses a tilde (~) to connect the two variables.

```
t.test(d$MindEye~d$AgeGroup)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$MindEye by d$AgeGroup
## t = 1.7763, df = 239.67, p-value = 0.07696
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1065317  2.0607801
## sample estimates:
## mean in group 1 mean in group 2
##        25.69935        24.72222
```

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. Note that R will tell you a degrees of freedom that is a decimal value. This is a different way of estimating the degrees of freedom that is more complicated than we do in class, and corrects for the fact that our data may be non-normal. The value that R reports is more accurate, so use that value.

In some cases, the p-value may be in scientifc notation, something like 2.2e-16, especially if the p-value is very, very small. In these cases, I report that the $p < .001$.

The 95% confidence interval is on the 5th line, and the last line has the mean of the variable for each group. Remember that group 1 is younger adults and group 2 is older adults.

Step 4: Do t-tests for each of the variables, except Age (we don't need a t-test to show older adults are older than younger adults). After each t-test, write the output in APA format and then note whether it is a significant difference with an α of .05

Step 5: Find a variable of interest and compare women versus men on one of the tests. Use the Sex variable and set up an appropriate null and alternative hypothesis. Then conduct the t-test to test your hypothesis and report what you do regarding th null and alternative hypothesis.

Once you are finished with these steps, knit your R notebook and upload your knitted R notebook.