Homework 07 - t-tests and effect sizes (part 2)

partner names

type date

Part A

General Questions

1. **QUESTION:** Describe how the independent-samples *t*-test is different from the paired-samples *t*-test.

ANSWER: Samples are independent of one another; different experimental units; assigned randomly to groups

2. **QUESTION:** What is an assumption of the independent-samples *t*-test that is different from the other *t*-tests.

ANSWER: Homogeneity of variance; variances of the populations should be equal

Part B

1. Installing and using libraries in RStudio

1.1. Use the RStudio interface to install packages/libraries. Go to the Tools option and select Install Packages. Type the package name(s) correctly using the proper letter casing. Also, make sure that you *check the box to Install Dependencies*. Do not install with code.

- Mac users, most of you should already have XQuartz installed from the first R class, but please make sure that you have it downloaded and installed in your Applications folder. See the original download instruction file if you need help.

1.2. Key functions used for this assignment:

- t.test() for calculating t-tests; built-in stats library

Change from scientific notation if you want.

```
options(scipen=999) #options(scipen=0)
```

2. Overview of the independent-samples *t*-test

2.1. The independent-samples *t*-test is use to compare two independent groups on a single dependent variable; experimental units provide only one dependent variable. Independent groups simply refers to the groups as indendent of each other; experimental units that are in one group are not the same as those in another group. This is very different from the dependent aspect of the paired-samples *t*-test. The independent-samples *t*-test has the same assumptions of normality and measurement scale as the other *t*-tests. These assumptions should be tested as usual, but are not always done for this assignment. In addition, there is a new assumption related to the groups.

2.2. Under H0, the assumption for this test is that the population values (e.g., means) of two independent groups is the same (e.g., mu1 = mu2). Because samples estimate those populations, the independent-samples *t*-test test whether a difference between the two population means is different enough from 0 to suggest that H0 is not a reasonable account of the data.

2.3. Because the assumption is that the two populations are equal to one another, the samples take from them should yield the same means and variances. Having equivalent variances is an assumption of this test called the *homogeneity-of-variance* assumption. To the extent that the variances are not equivalent, there is a violaiton of the assumption known as *heterogeneity-of-variance*. Because the sample size can influence variance, having roughly equivalent ns for the two groups is advised. There are two ways to address this concern of equivalent variance:

a. compare the variances statistically using a test like the Levene’s test using the levene.test() function from the lawstat library.

b. do not assume the variances are equal by specifying this by setting your t.test() function to adjust the results of *t*-test for you using a Welch correction.

3. Doing the independent-samples *t*-test

3.1. We can test some data that have met the conditions for doing an independent-samples *t*-test by reading in the class Survey data (“SurveyNames.csv”). Because men and women are independent samples, we can compare them t-test examining Men vs. Women on how many hours a day they spend on social media.

The variables that we are interested in are:

- Gender (Men = 0, Women = 1)
- Vector/variable SocialMedia, which asked “How many hours a day do you spend on social media.”

We can read in the SurveyNames data file.

```
#setwd("c:/users/gcook/desktop/Psyc109")
SURVEY <- read.csv("SurveyNames.csv")
str(SURVEY)
```

```
## 'data.frame':    54 obs. of  48 variables:
##  $ X          : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ ID         : int  4 5 6 7 8 9 10 11 12 13 ...
##  $ Gender     : int  0 1 1 1 1 1 0 0 0 1 ...
##  $ PolParty   : int  0 0 1 0 2 2 3 3 4 1 ...
##  $ NonFiction : int  1 1 1 0 1 0 1 1 1 0 ...
##  $ Religion   : int  7 1 7 1 1 6 1 5 1 1 ...
##  $ Dress      : int  3 2 1 1 1 2 2 2 2 1 ...
##  $ Superpower : int  5 2 2 3 4 5 3 3 3 3 ...
##  $ Fight      : int  2 7 1 1 3 6 1 4 4 5 ...
##  $ GlobalIssue: int  7 6 3 2 7 1 2 2 2 2 ...
##  $ MovGenre   : int  6 3 1 2 4 1 1 1 5 2 ...
##  $ Music      : int  3 3 3 3 4 3 6 3 3 3 ...
##  $ Electronic : int  1 1 2 1 1 2 1 1 2 1 ...
##  $ Sport      : int  2 7 5 2 1 4 1 6 6 6 ...
##  $ FavTime    : int  2 1 4 4 2 4 4 1 4 3 ...
##  $ Hand       : int  3 4 4 4 4 4 4 4 4 4 ...
##  $ Quad       : int  3 3 3 3 2 3 3 2 3 3 ...
##  $ MusGenre   : int  4 7 7 2 4 4 6 5 5 4 ...
##  $ Voting     : int  3 1 2 2 1 1 3 2 2 2 ...
##  $ Partners   : int  2 2 2 1 2 2 2 3 2 1 ...
##  $ YearBorn   : int  5 4 6 6 5 5 5 6 6 7 ...
##  $ Tip        : int  15 10 10 10 15 15 20 15 15 20 ...
##  $ SocialMedia: num  3 10 2 3 2 1 2 0.5 3 1 ...
##  $ Siblings   : int  1 1 3 3 3 0 3 1 2 1 ...
##  $ MovMonth   : int  1 5 1 2 7 2 0 2 2 4 ...
##  $ Potter     : int  0 2 0 0 0 3 0 3 3 1 ...
##  $ MilesHome  : Factor w/ 41 levels "0","1","1,000",..: 12 13 20 17 40 15 14 5 20 38
...
##  $ Excercise  : num  0 12 4 6 20 10 6 20 20 20 ...
##  $ HourSleep  : num  6 7 5 5 6 7 6 7 5 7 ...
##  $ HourPhone  : num  2 100 24 5 21 0 4 1 2 1 ...
##  $ Commuting  : num  0 2 1 0.5 0 0 0 0 0 1 ...
##  $ Parents    : num  3 20 3 7 3 0.5 5 4 0 1 ...
##  $ SportsPlay : int  6 5 1 6 5 4 6 7 8 2 ...
##  $ GELeft     : int  2 6 2 1 3 3 2 4 2 0 ...
##  $ Alone      : int  2 10 4 3 2 0 0 10 0 50 ...
##  $ Countries  : int  0 4 1 10 2 2 0 7 8 9 ...
##  $ Dogs       : int  2 0 5 0 0 0 6 1 1 0 ...
##  $ Snack      : int  3 1 2 1 2 2 3 20 2 5 ...
##  $ OffCampus  : int  2 3 2 2 2 2 2 3 2 3 ...
##  $ Sleep7     : int  1 5 3 2 4 4 5 5 1 5 ...
##  $ Reusable   : int  5 5 4 3 2 5 3 5 4 5 ...
##  $ Recycle    : int  4 3 2 3 2 3 1 4 1 5 ...
##  $ JobSec     : int  3 3 2 2 2 4 1 1 5 1 ...
##  $ Enjoyment  : int  3 4 2 4 4 3 5 4 4 4 ...
##  $ Present    : int  5 4 5 4 4 4 5 5 3 4 ...
##  $ FamPhone   : int  4 5 4 5 4 3 4 4 3 4 ...
##  $ GoodNight  : int  3 5 3 3 4 4 4 4 1 5 ...
##  $ Q44        : int  1 1 1 1 1 1 1 1 1 1 ...
```

Gender: Males = 0 Females = 1

Hours a day is a continuous variable.

3.2. We should compare whether the data for men and women are distributed normally using the shapiro.test() function. However, obtaining a subset of the data from a signle variable depending on whether respondents are coded as 0 (men) or 1 (women)

Without subseting we basically have:

```
shapiro.test(SURVEY$SocialMedia)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  SURVEY$SocialMedia
## W = 0.785, p-value = 0.0000001821
```

However, such a test includes both men and women. Also, the test suggests that the data are not normal in shape. In order to create a subset, specify the data frame and the variable as you normally would, but then use brackets to select cases based on responses to the Gender variable:

```
shapiro.test(SURVEY$SocialMedia[SURVEY$Gender == 0])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  SURVEY$SocialMedia[SURVEY$Gender == 0]
## W = 0.88421, p-value = 0.008451
```

```
shapiro.test(SURVEY$SocialMedia[SURVEY$Gender == 1])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  SURVEY$SocialMedia[SURVEY$Gender == 1]
## W = 0.66902, p-value = 0.0000007789
```

3.3. Notice that a normal distribution is not a reasonable assumption of the data. Your data are skewed and the mean is likely not a good measure of center. Thus, you should not use an inferential statistic (e.g., the *t*-test) that uses the mean. Rather, use the Wilcoxon test. However, if we pretend that the normality assumption is met, we can forge ahead and compare if there is a difference between males and females with social media usage.

As with the one-sample and paired-samples *t*-tests, the t.test() function will be used to compare independent, or between-subjects groups. With a paired-samples *t*-test, the function was:

t.test(X1, X2, paired = TRUE)

However, now that the samples are not paired, the “parired”" argument needs to be change to FALSE. The nice thing is that paired = FALSE is the default setting (see help section for t.test()), so the function can be:

t.test(X1, X2)

is the same as,

t.test(X1, X2, paired = FALSE)

In order to compare subgroups for the t.test() function, we need to subset the variables. The t.test() arguments are displayed on separate lines to make this easier to read. You can see that you are selecting two variables from the SURVEY data frame and the SocialMedia variable, plus a subset Gender that is equal to 0 or 1. If we also add the var.equal argument and set it to FALSE, this will matematcially adjust the t.test() to provide a result that corrects for any issue with heterogeneity of variance. This is my recommendation.

```
MediaGender.t <- t.test(SURVEY$SocialMedia[SURVEY$Gender == 0],
                        SURVEY$SocialMedia[SURVEY$Gender == 1],
                        var.equal = FALSE)
```

If you examine the output of the t-test, you will notice that the df are now not whole numbers, but decimals or fractions. This is what the Welch correction does. If you use this correction, you should report your test using the fractional df.

QUESTION What is the calculated *t*-test value?

Answer: 0.636625

QUESTION What is the *p*-value?

Answer: 0.5271867

At the bottom of the output, you can see the mean values for the two variables you entered, men and women. You should have noticed that that men report using Social Media more than do women, however the result is not statistically different from 0. Perhaps the differences are due to sampling error.

3.4. Also, because the normality assumption was not met, the *t*-test is potentially inappropriate. The alternative test for independent groups with two levels of an independent variable that can be used when normality is not met is a Wilcoxon Rank-Sum Test. Because this test is comparing ranks (ordinal data) and not means based on ratio data, we do not have the normality assumption. Because groups are independent, you will need to set the paired argument to FALSE or omit it. Including the paired argument might be a useful reminder of what you are doing.

```
wilcox.test(SURVEY$SocialMedia[SURVEY$Gender == 0],
            SURVEY$SocialMedia[SURVEY$Gender == 1],
            paired = FALSE)
```

```
## Warning in wilcox.test.default(SURVEY$SocialMedia[SURVEY$Gender == 0],
## SURVEY$SocialMedia[SURVEY$Gender == : cannot compute exact p-value with
## ties
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  SURVEY$SocialMedia[SURVEY$Gender == 0] and SURVEY$SocialMedia[SURVEY$Gender ==
1]
## W = 404, p-value = 0.4672
## alternative hypothesis: true location shift is not equal to 0
```

QUESTION: Does this test reveal a significant difference between men and women?

Answers No

Part C.

Try for yourself!

QUESTION: Another variable asked was whether people liked to read non-fiction books (NonFiction variable. 0 = No, 1 = Yes). Another variable the number of hours slept per night. Compare whether non-fiction readers and non-readers obtain the same amount of sleep. Use the Welch correction.

CODED ANSWER:

```
#hide
ReadingSleep.t <- t.test(SURVEY$HourSleep[SURVEY$NonFiction == 0],
                    SURVEY$HourSleep[SURVEY$NonFiction == 1],
                    var.equal = FALSE)
ReadingSleep.t
```

```
##
##  Welch Two Sample t-test
##
## data:  SURVEY$HourSleep[SURVEY$NonFiction == 0] and SURVEY$HourSleep[SURVEY$NonFiction
== 1]
## t = -0.12534, df = 10.426, p-value = 0.9026
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.8121436  0.7251871
## sample estimates:
## mean of x mean of y
##  6.750000  6.793478
```

QUESTION: What was the *t*-test value?

ANSWER: -0.1253366

QUESTION: Calculate the r-squared effect size based on your results.

CODED ANSWER

```
#hide
ReadingSleep.r2 <- ReadingSleep.t$statistic^2 / (ReadingSleep.t$statistic^2 + ReadingSlee
p.t$parameter)
ReadingSleep.r2
```

```
##           t
## 0.001504488
```

COMMENTS

If you have comments to include for improving the assignment, please type them here.

DONE!

E-mail your knit HTML file that you completed with your partner. Make sure to include your names.