Exercise 2: Descriptive Statistics and *t*-Tests

A recent paper by Ed O'Brien and Ellen Roney examines people's tendency to order work activities before leisure activities. O'Brien and Roney hypothesize that people *believe* that they will enjoy leisure activities to a greater degree if they do these activities after rather than before work activities. However, the authors also predict that the *actual* enjoyment of leisure will be similar, regardless of whether work comes before or after. This can lead people to order their experiences in a way that does not necessarily maximize enjoyment.

Because the authors provided their raw data, this assignment will rely on their actual data to reproduce a subset of their results and to test additional hypotheses.

First, download the O'Brien and Roney paper. If you are on campus using the university network, you can download the paper by searching the following citation on Google Scholar:

> O'Brien, E., & Roney E. (2017). Worth the wait? Leisure can be just as enjoyable with work left undone. *Psychological Science*. Advanced online publication.

Since we are only looking at the data from the first experiment in this exercise, you need to read only the introduction and Experiment 1 methods and results (the first three pages of the paper).

The authors posted their materials and data to an online repository. For your convenience, the data are reposted in an Excel sheet on the course website (where each tab is one study). For this assignment, you will need only the first four tabs (Studies 1A–D).

You will need to turn in two documents: (1) a copy of the online Excel data sheet, with your calculations at the bottom of the first four tabs to implement the computations requested in the following questions, and (B) a .pdf file with your brief verbal answers to the following questions.

A. Describe the experimental designs for Studies 1A–D. (Refer to Exercise 1 if you need a refresher on this.) Calculate the descriptive statistics (*M*, *SD*, *N*) for each condition of each of these studies. For Studies 1C and 1D, compute these means both for *each* measure of positivity and for the *composite* of these measures (i.e., the mean for each participant of these different measures).

B. In analyzing Studies 1A and 1B, the authors used one-sample *t*-tests to compare responses to the midpoint of the scale. There is no option using the TTEST function to conduct a one-sample *t*-test directly. However, recall that a paired (one-group) *t*-test is the same as calculating the *difference* between each participant's scores and conducting a one-sample *t*-test on that difference. Given that the TTEST function can execute a paired *t*-test, how can you use this function to perform a one-sample *t*-test? (Hint: You will need to create a new column of numbers to enter as the second argument of the *t*-test function. What numbers would give you the right answer?) Using your method, verify that you find the same *t*-test results for Studies 1A and 1B as the authors did.

C. What test did the authors use to analyze the results of Studies 1C and 1D? Conduct these tests yourself on the *composite* scores you computed in Part A.

D. In Session 1 of the workshop, we briefly discussed confidence intervals. Intuitively, a 95% confidence interval on some variable can be taken to mean that you are "95% confident" that the variable's value falls within that interval. For instance, if participants answered a mean of 6.0 on some scale, the 95% confidence interval might be the range [5.0, 7.0], indicating that any number between 5 and 7 would be plausible as the population mean (at a 95% confidence level). If we wanted to be 99% confident, the range would be wider and if we only wanted to be 90% confident, the range would be narrower.

   But what exactly do we mean by "plausible" or "confident"? The way that confidence intervals are constructed is closely related to p-values. Essentially, if a value Y is *outside* the confidence interval, then a one-sample t-test comparing the sample mean to Y would be statistically significant (at $p < .05$ for a 95% confidence interval, at $p < .01$ for a 99% confidence interval, etc.), whereas if Y is *inside* the confidence interval, such a t-test would not be significant. This means that a 95% confidence interval is centered at the sample mean and ranges approximately two standard errors in each direction. Thus, a confidence interval includes the range of possible population means that cannot be ruled out as implausible through significance testing for a given significance level.

   In all of the studies, the authors reported 95% confidence intervals on the *difference* between participants' scores and the midpoint (for Studies 1A and 1B) or between participants' scores across conditions (for Studies 1C and 1D). Intuitively, what do these confidence intervals mean? What does it mean if 0 is or is not included in these intervals?

E. Calculate this interval yourself for Studies 1A and 1B, using the CONFIDENCE.T function to compute the length of each side of the interval. The endpoints of the confidence interval are the mean difference plus or minus the length given by the above function. (You do not need to do this for Studies 1C and 1D, as this function will not give the correct answer.)

F. The authors report a statistic called Cohen's d for each significance test. This is a standardized measure of *effect size*. (Recall that p-values tell you whether a difference is *reliable* or likely to be *real*. Effect size statistics tells you whether the difference is *big* or *meaningful*.) Cohen's d represents the size of a difference (either between the sample mean and a number, or between the means of two conditions) in terms of standard *deviations* (not standard errors). For example, a Cohen's d of 0.5 would mean that two conditions differ by about half a standard deviation. Thus, d is not sensitive to the unit of measurement or to sample size. A rough guideline is that a Cohen's d of 0.2 represents a "small" effect, while d of 0.5 or 0.8 would be "medium" and "large" effects, respectively.

   Cohen's d can be calculated by taking the mean difference and dividing it by the standard deviation of the difference. This is the same procedure as computing a t-statistic (discussed in class), except the denominator is the standard *deviation* instead of standard *error*. Calculate Cohen's *d* for Studies 1A and 1B. (Hint: You may not get the same answer as the authors of the paper.)

G. Choose either Study 1C or 1D, and perform one or more *t*-tests that go beyond the results reported in the published paper. These should be organized around some additional research question that builds on the issues discussed in the paper. For example, is there a gender difference in either study? (1 = male, 2 = female in the "sex" column.) Do intuitions differ for participants under the age of 40 compared to participants age 40 and over? Do responses systematically differ across the different measures of positivity used? (Feel free to try out multiple different options, but please focus on just one issue in your final analysis.)

H. Write up a short results section for *either* Study 1C or 1D, including all of the results computed in Parts A–F pertaining to that study as well as the analysis you conducted in Part G. Be sure to write the results in a way that underscores what research question each statistical test is answering. Include a brief discussion that ties the results back to the authors' original questions and that explains why you might have found the results you did for the Part G comparison(s). This should be no longer than 1 short paragraph for the results and 1 short paragraph for the discussion (and possibly shorter still).