

Statistics Labs for Psychology

Zakary A. Draper

2021-08-11

Contents

Author's Note	5
1 One Sample t Tests	7
1.1 Learning Objectives	8
1.2 The Study: Is 7 Really Magical?	8
1.3 Data Collection	9
1.4 Lab Report Instructions	9
2 Paired Samples t Tests	13
2.1 Learning Objectives	14
2.2 The Study: Weight-Loss-Related Self-Efficacy	14
2.3 Data Collection	16
2.4 Assignment Instructions	16
3 Independent Samples t Tests	21
3.1 Learning Objectives	21
3.2 The Study: Social Identity	21
3.3 Assignment Instructions	24
4 Correlation	29
4.1 Learning Objectives	29
4.2 The Study:	29
4.3 Data Collection	29
4.4 Assignment Instructions	29
5 Multiple Regression	31
5.1 Learning Objectives	31
5.2 The Study:	31
5.3 Data Collection	31
5.4 Assignment Instructions	31
6 Univariate ANOVA	33
6.1 Learning Objectives	33
6.2 The Study:	33

6.3	Data Collection	33
6.4	Assignment Instructions	33
7	Factorial ANOVA	35
7.1	Learning Objectives	35
7.2	The Study:	35
7.3	Data Collection	35
7.4	Assignment Instructions	35
8	Chi-squared Tests	37
8.1	Learning Objectives	37
8.2	The Study:	37
8.3	Data Collection	37
8.4	Assignment Instructions	37

Author's Note

Hi there. I hope you find this lab manual helpful. If you are an instructor who is interested in using these materials for your course, you can email me and I can provide you with some additional resources.

Chapter 1

One Sample t Tests

One sample t tests are conducted to test questions about the average value of a continuous variable in a population. For example, one-sample t tests might be used to help answer the following research questions:

- Are Canadians aged 65 and older able to recall a 7-digit phone number?
- Can university students type more than 60 words-per-minute?
- Do North Americans consume 2000 calories per day?

Each of these research questions include a definition of a population, a variable measured on a continuous (or at least interval) scale, and a specific value on that scale. Note that there is no mention of how that value might compared with other populations. Research questions that can be informed by one-sample t tests can always be framed in the form:

Is the average value of [continuous variable] among [characteristics of a population] [more than/less than/more or less than] exactly [a number]?

One-sample t tests are rarely used because most research questions cannot be phrased in this way. That's because of the "[a number]" part at the end. We are usually more interested in comparing values between populations (e.g., "is the treatment more effective than a placebo?") or in how a value changes based on some other continuous variable (e.g., "does life satisfaction increase with age?").

Although one-sample t tests are rarely used, it is important to understand how they work. They are the simplest statistical models and the foundation of every other analysis covered in this lab. Mastering the one-sample t test will prepare you for the more complex models that you will be building in future labs.

1.1 Learning Objectives

After completing this lab, you should be able to do the following using R:

- Conduct one-sample t tests.
- Compute Cohen's d_z .
- Produce histograms.
- Conduct Shapiro–Wilk tests.
- Produce descriptive statistics including M , SD , skew, and kurtosis.

Additionally, you should be able to:

- Report results of one-sample t tests in APA style.
- Report results of Shapiro–Wilk tests in APA style.
- Understand the role of the following in evaluating the strength of evidence provided by one-sample t tests to support or refute hypotheses about a population:
 - The correct interpretation of a p value.
 - The assumption of normality in one-sample t tests.
 - Raw and standardized effect sizes.

1.2 The Study: Is 7 Really Magical?

The following is the Introduction and Method sections of a hypothetical study. For this lab, you will conduct the analyses described in the analytic strategy below.

Eminent psychologist George Miller is best known for his seminal paper, “The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information.” “The Magical Number Seven” has been interpreted to claim that humans can hold an average of seven objects in working memory—an idea sometimes referred to as Miller’s Law. Miller’s Law proposes a remarkably simple model of working memory capacity. We tested this model of working memory in a sample of undergraduate psychology students who were in their third or later year of their program and had excelled in their coursework. Because these students had showcased academic excellence over several years of post-secondary education, we hypothesized that their memory spans would be greater than seven.

Participants and Procedure

Participants ($N = 174$) were undergraduate psychology students enrolled in a research methods and statistics course designed to prepare students to conduct research and for graduate school. Entry into this course is competitive based on GPA in psychology courses. As such, these students were above average in terms of academic achievement. Participation in this study was part of the students’ coursework.

Measures

Memory span is the longest sequence of items a person is able to correctly repeat immediately after presentation, in at least 50% of trials. We measured digit span, which is memory span for digits, using the R package `memoryspan`. This package includes a program for measuring digit span by printing digits to the R console and asking participants to type those values into the console. If the correct sequence of values is entered into the console, the user is presented with a new sequence which is one unit longer than the prior sequence. If the incorrect sequence is entered into the console, the trial ends and the user's digit span for that trial is returned. By default, the program begins with a digit span of one, but participants were instructed that they could increase this value to save time. Participants completed 10 trials. Their digit spans were defined as the longest sequence they correctly repeated in at least 5 of the 10 trials.

Analytic Strategy

To test whether Miller's Law applied to digit spans for this population of students, we conducted one-sample t tests against the null hypothesis that our sample was drawn from a population in which average memory span is less than or equal to seven. One-sample t tests assume that variables are drawn from a normally distributed population. We tested this assumption by visually inspecting histograms, examining skew and kurtosis values, and with Shapiro-Wilk tests.

1.3 Data Collection

Use the R package `MemorySpan` to complete 10 trials of a digit-span task. Record your results. Your digit span will be the longest list you correctly recall on at least 5 of the 10 trials. `MemorySpan` is not available on the CRAN. It can be installed from GitHub using the function `install_github()` from the `remotes` package. Install `MemorySpan` by running `remotes::install_github("zakarydraper/MemorySpan")`. As with R packages installed from the CRAN, this only needs to be installed once.

After installing `MemorySpan`, you can load it with `library()` like you would any other R package. Consult the documentation for `measure_memory_span()` for direction on how to use the function to measure your digit span.

1.4 Lab Report Instructions

Read "Is 7 Really Magical? A Simple Test of Miller's Law," which is a manuscript containing an introduction and method section. For this lab, you will be conducting the analyses described in the analytic strategy (part of the method) and reporting the results in APA style. You will submit the following: (a) an R script for conducting the analyses described in the analytic strategy,

(b) a Word document with your results reported in APA style, and (c) short answer responses to the discussion questions.

1.4.1 R Script

Do the following in R:

1. Use `t.test()` to conduct a one-sample t test of the null hypothesis that the average digit span in the population is less than or equal to 7. Use `<-` to assign a name to the output of your call to `t.test()`.
2. Compute Cohen's d_z , either by using the formula $(M - \mu)/SD$ or by converting from the t with `psych::t2d()`.¹
3. Use `hist()` to produce a histogram of the distribution of digit spans.
4. Use `shapiro.test()` to conduct a Shapiro–Wilk test on the distribution of digit spans.
5. Calculate M , SD , skewness, and kurtosis of the digit span variable. Calculate M , SD , and range of participant ages:
 - `psych::describe()` will calculate all of these values in one go.
 - `QuantPsyc::norm()` will calculate skew and kurtosis for one variable and will provide significance tests for each (not provided with `psych::describe()`). It will not provide M , SD , or range.

1.4.2 Results

Your results section should include the following:

- 1 sentence presenting the results of the one-sample t test.
- 1 sentence presenting the results of the Shapiro–Wilk test.
- 1–3 sentences describing the distribution of scores. Reference the results of the Shapiro–Wilk test, the histogram (i.e., Figure 1 in your manuscript), and skew and kurtosis values.
- A histogram showing the distribution of digit spans in the sample. Follow the rules for figures in the *APA Publication Manual*.

1.4.3 Discussion

In future lab reports, you will write complete (albeit brief) discussion sections. For this lab report, answer each of the following questions in complete sentences.

- Do your results support your hypothesis?
- How do your results compare with Miller's work arguing that memory span is equal to seven? Provide an explanation for why your results agree or disagree.
- How does the characteristics of the sample affect your interpretation of these results?

¹The questions are identical to those used by Davies et al. (2008) except “Canadian” replaces “American” in all instances.

- What is an implication of your findings for future research, programs, or policy?

Chapter 2

Paired Samples t Tests

Last lab covered one sample t tests. We learned that one sample t tests compare the mean to a specific value. One common use case for this is when pairs of scores can be represented using a single value. For example, imagine you want to test whether cognitive behavioural therapy (CBT) reduces generalized anxiety. Study participants could complete the GAD-7 (a popular measure of generalized anxiety) prior to receiving treatment. Then, after receiving CBT, they could be take the GAD-7 again. The resulting data would be two GAD-7 scores for each participant. If we want to know whether a participant's anxiety decreased, increased, or stayed the same, we could just subtract their pre-treatment score from their post-treatment score. Doing this for each participant yields a single variable that represents the difference (or change) from pre- to post-treatment. We could then conduct a one sample t test on the change scores.

Another name for this procedure is a paired samples t test. It is so named because the data are two samples that are *paired*. In the example above, samples are paired because they come from the same participant measured on two occasions. This is likely the most common design that leads to paired samples. However, pairs of observations can also come from different individuals. For example, if you wanted to know whether birth influences language acquisition, you could compare the number of words a firstborn child knew at 12 months to the number of words a second born child from the same family knew at 12 months. In this case observations are paired because they come from the same family.

If you are wondering whether a paired samples t test is appropriate for your research design, ask yourself whether it makes sense to produce difference scores for your cases. Note that difference scores are not valid when pairs of observations are measured in different metrics. If a pre-to-post design (such as in the CBT example above) uses a different measurement instrument at each measurement occasion (even if both instruments are intended to measure the same

construct), difference scores are no longer valid. Another situation when difference scores won't work is when groups of observations are dependent, but not paired. For example, if the CBT study above included participants from several different therapists. Participants with the same therapist would be expected to be more similar to one another than to participants from different therapists. But a paired samples t test could not be used to compare them because each participant is not matched to a single other participant.

Because paired samples t tests are so similar (mathematically identical, even!) to one sample t tests, this lab will give you an opportunity to polish the skills you have already learned. It will also introduce to the important considerations that are unique to paired samples.

2.1 Learning Objectives

After completing this lab, you should be able to do the following using R:

- Conduct paired samples t tests.
- Compute Cohen's d_z for paired samples.
- Create appropriate visualizations for comparisons of two paired samples.

Additionally, you should be able to:

- Explain how paired samples t tests are similar (and distinct) from one sample t tests.
- Evaluate the assumptions of paired samples t tests.
- Accurately interpret the results of paired samples t tests.
- Report the results of paired samples t tests in APA style.

2.2 The Study: Weight-Loss-Related Self-Efficacy

2.2.1 Study Information

2.2.1.1 Description

Weight-loss interventions aim to support people in changing their behaviours to be healthier. This includes promoting healthy-eating and exercise. Self-efficacy is thought to be important for adopting and maintaining such behaviours. That is, people will change their behaviours if they believe they can maintain those behaviours and if they believe that changing their behaviour will help them lose weight.

Because self-efficacy is so important to the success of weight-loss interventions, some such interventions now include psychotherapy specifically aimed at improving self-efficacy. Weight-loss interventions provide instrumental support in helping participants develop a plan for eating healthful foods and engaging in

regular physical activity. This has a clear and obvious benefit for self-efficacy. The current study will examine how subsequent psychotherapy further boosts self-efficacy.

2.2.1.2 Hypothesis

The primary hypothesis is that adding psychotherapy to a weight-loss intervention will lead to increased weight-loss self-efficacy.

2.2.2 Design Plan

2.2.2.1 Study Design

This study will use a within-subjects (paired) design. Participants will be enrolled in a weight-loss intervention at UBCO. They will attend a group session where they are provided general information about healthy eating, and physical activity. They will then have one-on-one interviews with a nutritionist to help them build meal plans that fit their budgets and lifestyles. Following the interviews, the nutritionist will ask participants to complete measures of weight-loss self-efficacy (WLSE). Participants will then meet with a trained psychotherapist who will work with the participants to identify potential barriers to their success and develop strategies for overcoming those barriers. After psychotherapy, participants will be asked to complete the same measure of WLSE.

2.2.2.2 Measures

Weight-loss-related behaviour self-efficacy will be measured using the scale proposed by Wilson et al. (2016). This scale has four items, measured on an 11-point scale ranging from 0% (*Not at All Confident*) to 100% (*Completely Confident*). The scale is relatively new; however, it has demonstrated strong psychometric properties in a large sample drawn from a population that is similar to the population used in the current study (Wilson et al., 2016).

2.2.2.3 Analytic Strategy

We will conduct a directional paired samples t test of the null hypothesis that average change in WLSE in the population is equal to zero. Significance will be based on the traditional $\alpha = .05$.

2.2.2.4 Sample Size and Power Analysis

We want to be able to detect a difference of 10 points on the WLSE measure. This measure had a $SD = 21.44$ in a large sample from a similar population (Wilson et al., 2016). Assuming the same SD in our sample, a 10-point change translates to Cohen's $d_z = 0.47$. To achieve 95% power with a population effect of that size requires a sample of $N = 50.4$. We are therefore aiming for a sample of $N = 51$.

2.3 Data Collection

How confident are you that you can **lose weight...**

...even if you need a long time to develop the necessary routines?

(Select an Option) 0% (Not at all Confident) 10% 20% 30% 40% 50% 60% 70%
80% 90% 100% (Completely Confident)

...even if you have to try several times before it works?

(Select an Option) 0% (Not at all Confident) 10% 20% 30% 40% 50% 60% 70%
80% 90% 100% (Completely Confident)

...even if you have to rethink your entire way of losing weight?

(Select an Option) 0% (Not at all Confident) 10% 20% 30% 40% 50% 60% 70%
80% 90% 100% (Completely Confident)

...even if you have to make a detailed plan?

(Select an Option) 0% (Not at all Confident) 10% 20% 30% 40% 50% 60% 70%
80% 90% 100% (Completely Confident)

2.4 Assignment Instructions

2.4.1 R Script

1. Calculate participants' mean scores on the 5-item WLSE measure. Do this separately for the two measurement occasions. See the [additional instructions on scoring][pst-scoring].
2. Calculate M and SD for the pre- and post-psychotherapy means, and M , SD , skew, and kurtosis of the difference scores.
3. Use `t.test()` to conduct a paired t test of the null hypothesis that average pre-to-post change in WLSE scores is zero.
4. Compute Cohen's d_z for the change in WLSE scores.
5. Produce a histogram of the WLSE difference scores.
6. Conduct a Shapiro–Wilk test on the difference scores.
7. Produce a boxplot of pre- and post-psychotherapy WLSE scores. See the [additional instructions][pst-plotting] below.

2.4.2 Additional Instructions

2.4.2.1 Scoring

The “weight-loss.csv” data frame includes item-level data. That is, each column is one item on the WLSE measure. Use `rowMeans()` to calculate each participant's average score on the 5-item measure. Do this separately for the time 1 and time 2 measurement occasions. Add participants' mean scores on times 1 and 2 as columns to your data frame.

Calculate difference/change scores for each participant. This will be the difference between each participant's average WLSE scores at times 1 and 2. When computing difference scores, it is conventional to subtract initial scores from final scores (i.e., time 2 – time 1). Doing this makes the difference scores more interpretable, as a negative difference score means a decrease, and a positive score means an increase.

2.4.2.2 Descriptive Statistics

Produce the following statistics:

- M and SD of average WLSE scores at both measurement occasions.
- M , SD , skew, kurtosis of the difference scores.

2.4.2.3 Paired t Test

Use `t.test()` to conduct a directional paired samples t test comparing WLSE scores before and after psychotherapy.

Use `t.test()` to conduct a one sample t test on the difference scores. What do you notice about the results of this test and the results of the paired t test?

2.4.2.4 Cohen's d_z

Calculate Cohen's d_z either by converting t to d using `psych::t2d()`, or using the M and SD of the difference scores in the formula $d_z = M/SD$.¹

2.4.3 Plots

This section will guide you through the process of producing a figure like the one below.

This is a boxplot with individual data points overlaid. It is a very useful plot because it concisely presents a lot of information.

2.4.3.1 Lengthen the Data

The data are currently arranged such that scores at each measurement occasion are placed in their own columns. That is, there are separate columns for time 1 and time 2 scores.

pid	wlse_m.t1	wlse_m.t2
1	44	50
2	78	62
3	96	84

¹The questions are identical to those used by Davies et al. (2008) except “Canadian” replaces “American” in all instances.

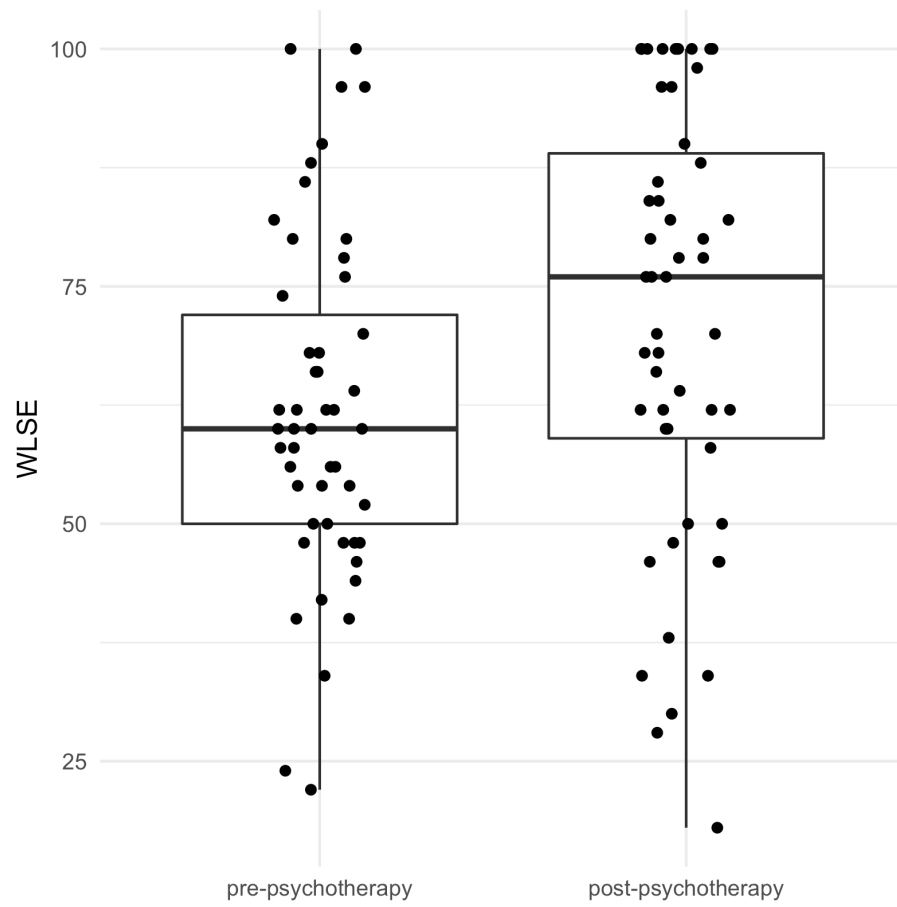


Figure 2.1: boxplot

The same data could be arranged with a single column for each item, and an additional column indicating whether the responses are for time 1 or time 2. That is, instead of time 1 and time 2 responses being differentiated by column, they are differentiated by row.

pid	name	value
1	wlse_m.t1	44
1	wlse_m.t2	50
2	wlse_m.t1	78
2	wlse_m.t2	62
3	wlse_m.t1	96
3	wlse_m.t2	84

Neither of these two options are strictly better than the other. The way the data should be arranged depends on what you are doing with the data. The wider format we have used so far has made sense for what we have been doing. For example, think about how difficult it would be to calculate changer scores if the data were arranged in this longer format.

That said, the plot we want to produce has time on the x axis, and scores on the y axis. To create this plot, time needs its own column in the data frame. So, even though the wider format has made sense so far, we'll need to pivot to the longer format in order to make the plot.

There are many options for reshaping data frames. The easiest (and it isn't close) is to use `tidyr::pivot_longer()` and `tidyr::pivot_wider()`. For a detailed introduction to these functions, see Chapter 12 of R for Data Science.

`tidyr::pivot_longer()` has many arguments, but most are optional. You can lengthen your data frame using just the first two arguments: `data` and `cols`. The first argument, `data`, is the data frame you want to lengthen. The second argument, `cols`, is the names of the variables in that data frame that should be pivoted into the longer format.

Two optional arguments you may find useful are `names_to` and `values_to`. These can be used to specify the names of the new columns in the longer data frame. For example, I've renamed my name column "time," and my value column "wlse."

pid	time	wlse
1	wlse_m.t1	44
1	wlse_m.t2	50
2	wlse_m.t1	78
2	wlse_m.t2	62
3	wlse_m.t1	96
3	wlse_m.t2	84

You'll notice that the pivoted data frame looks different than the data frames you have seen thus far. It prints **#A tibble:** and the number of rows and columns. It also gives the data types of the columns in the data frame. That's because the output of `tidyr::pivot_longer()` (and the related function `tidyr::pivot_wider()`) is not a `data.frame`, but a `tibble`. A `tibble` is a variant of a `data.frame` and functions in much the same way. For more information on tibbles, see Chapter 10 of *R for Data Science*. If you want to turn your `tibble` back into a `data.frame`, use `as.data.frame()`.

Let's turn our attention to the `time` column. This is a character vector. That means that the values in `time` are unordered. But the values represent an ordered variable (i.e., `t1` comes before `t2`). Also, the values are kind of ugly. In the next step, we'll be plotting these values, and we don't want these ugly names showing up as labels on our plots. We can solve both these problems by converting this character vector to a factor using `factor()`.

pid	time	wlse
1	pre-psychotherapy	44
1	post-psychotherapy	50
2	pre-psychotherapy	78
2	post-psychotherapy	62
3	pre-psychotherapy	96
3	post-psychotherapy	84

2.4.3.2 Produce the Plot

With the data frame lengthened, you are now ready to produce the boxplot.

Chapter 3

Independent Samples t Tests

3.1 Learning Objectives

After completing this lab, you should be able to do the following using R:

- Conduct both Student’s and Welch’s independent samples t tests.
- Compute Cohen’s d by conversion from the test statistic using `psych::t2d()` or by computing Cohen’s d directly using `psych::cohen.d()`.
- Create appropriate visualizations for comparisons of two independent samples (as described in the assignment instructions).
- Produce a single figure with plots for each level of a grouping variable using `ggplot2::facet_wrap()`.
- Calculate group-level descriptive statistics using `describeBy()`.

Additionally, you should be able to:

- Evaluate the assumptions of independent samples t tests.
- Explain the difference between Student’s t tests and Welch’s t tests.
- Accurately interpret the results of independent samples t tests.
- Report the results of independent samples t tests in APA style.

3.2 The Study: Social Identity

3.2.1 Study Information

3.2.1.1 Description

People desire positive social identities. A positive social identity is achieved when the social groups to which one belongs (i.e., the “ingroups”) are perceived

as favourable to the social groups to which they do not belong (i.e., the “outgroups”). Such positive social identities are most often fostered by favouring ingroup members, rather than by devaluing outgroups. As such, people tend to feel positively towards those they see as part of their ingroup.

The desire for a positive social identity can lead to discrimination when positive feelings are reserved for one’s ingroup. However it can also promote acceptance of diversity. Preserving positive social identities requires accepting diversity within one’s social groups. As such, broadly defining one’s ingroup can foster a greater acceptance of diversity. Therefore, encouraging the adoption of broad social groups has been proposed as a means of combatting discrimination.

Of course changing how people define their ingroup is difficult. It can occur when one changes social groups, such as when one moves to a new community, makes new friends, or changes political parties. More often however, broadening one’s ingroup is accomplished by strengthening ties with existing groups. That is, rather than changing the social groups to which they belong, people are more likely to change which social groups they see as important to their social identities.

This study will examine one mechanism by which people determine the importance of social groups to their social identities: perceived social group threat. Research has demonstrated that people strengthen ties to social groups when they perceive those groups as threatened in some way. This was evident in the aftermath of the 9/11 attacks on the United States. Many US citizens felt a threat to their American social identity. This led to a dramatic outpouring of displays of patriotism, such as flying the US flag on one’s property.

Research conducted in the time after the 9/11 attacks revealed that one impact of this increased patriotism was a greater tolerance for domestic multiculturalism. That is, people who felt a threat to their American social identity, were more accepting of diversity among American citizens.

One study conducted in that time utilized an experimental manipulation wherein some participants were introduced to a threat American social identity prior to being asked their agreement with statements related to American multiculturalism (Davies et al., 2008). This study will investigate a similar effect using a Canadian sample. Although Canada and United States are similar in many ways, their political climates are distinct. Additionally, the political climate in Canada and the United States has changed in the years since this research was conducted. Political convictions are more rigid than they were. These factors could mean that the effects observed among Americans in 2001 will be different than what can be expected among Canadians today.

3.2.1.2 Hypothesis

It is expected that Canadians who are primed to perceive a threat to Canadian social identity, will be more accepting of multiculturalism in Canada.

3.2.2 Design Plan

3.2.2.1 Study Design

This study will employ an experimental design adapted from experiment 3 of Davies et al. (2008). Participants will read a brief newspaper article describing either a foreign or domestic event that could be perceived as threatening to Canadians. They will then complete a measure of national identity. Participants will be recruited at in-person public events throughout the Okanagan during the summer of 2021. Research assistants will operate tables at events throughout the summer. For completing the study, participants will be offered \$5.00 food vouchers redeemable at participating vendors at the event. Recruitment will continue until sufficient data are obtained or until the summer's end.

Individuals must be at least 18-years-old and Canadian citizens to be eligible to participate.

3.2.2.2 Measures

National identity will be measured with items adapted¹ from Davies et al. (2008). They will be measured on a 7-point Likert-type scale ranging from 1 (*not at all*) to 7 (*completely*).

1. Do you identify with being Canadian?
2. Is being Canadian important to you?
3. Are you proud to be a Canadian?
4. Do you think of yourself as a Canadian?

Participant scores will be averaged across these four items to create a composite measure of national identity.

3.2.2.3 Analytic Strategy

A directional Welch's t test will be conducted with the alternative hypothesis that national identity is higher in the foreign threat condition. Significance will be inferred using the traditional $\alpha = .05$.

3.2.2.4 Sample Size and Power Analysis

We plan to recruit 254 participants divided evenly into the two experimental conditions. This will provide 95% power to detect an effect if the effect size in the population is $d = 0.414$.

Davies et al. (2008) reported mean national identities of 5.57 and 4.99 for participants who read about foreign and domestic threats, respectively. This amounts to a raw difference of 0.58 on the scale of 1 to 7. Because the SD s for national identity were not reported, we cannot know the standardized effect size for this

¹The questions are identical to those used by Davies et al. (2008) except "Canadian" replaces "American" in all instances.

effect. We therefore based our effect size calculation on an assumed pooled *SD* of 1.4. Under this assumption, a raw effect of 0.58 translates to $d = 0.414$.

3.3 Assignment Instructions

Prepare an R script and written report presenting the results of the analysis described in the analytic strategy.

3.3.1 R Script

Your R script should be reproducible and free of errors. It should contain syntax to do the following:

- Import the data and convert the “condition” variable to a factor.
- Calculate participant mean scores on the 4-item national identity measure.
- Compute descriptive statistics, including *M*, *SD*, skew, and kurtosis of participants’ mean national identity. Do this separately for each level of the experimental manipulation. Use `psych::describeBy()` to do this in a single step.
- Conduct Shapiro–Wilk tests of normality for the distributions of national identity in each of the two conditions.
- Compute a Welch’s *t* test as described in the analytic strategy.
- Calculate the corresponding Cohen’s *d* effect size for the *t* test.

3.3.2 R Script (Detailed Instructions)

3.3.2.1 Doing a *t* Test

There are two syntax options for conducting independent samples *t* tests using `t.test()`. The first is similar the syntax used for the paired samples *t* test. This involves supplying a numeric vector to the arguments `x` and `y`, which are the first two arguments of `t.test()`. A paired samples *t* test requires the additional argument `paired = TRUE`. To indicate that the samples are independent, the argument should be `paired = FALSE`. The default value is already `paired = FALSE` so not specifying the argument will treat the samples as independent.

So, if I have a `data.frame` called `dta` that looks like this:

```
##   var1 var2
## 1    7   10
## 2    8    3
## 3    5    2
## 4    7    3
## 5    3    4
```

Then I could conduct a *t* test using the following syntax:


```
t.test(x = dta$var1, y = dta$var2)
```

This would conduct what is called a Welch's t test, which I explain below. When you inspected your data, you probably noticed that it is not arranged as the data above. Your data are arranged in a `data.frame` like this:

```
##      var value
## 1      1      7
## 2      1      8
## 3      1      3
## 4      1      4
## 5      1      9
## 6      2      4
## 7      2      7
## 8      2      1
## 9      2      8
## 10     2      2
```

This is probably the most sensible way to arrange data from independent samples. It places observations from different people on separate rows. The prior arrangement had data from different people in different columns. That arrangement implies some connection between data from the same row, when there is none. It technically works when there are only two variables, but that is never the case with real data. You are going to have other information, such as participant IDs, demographic details, and so on that are unique to each participant. There is no sensible way of arranging the data frame that maintains the information about individual participants while placing scores from different participants side-by-side.

So, how do we make this second layout work with `t.test()`? There are two approaches. The first is to use `subset()`.

```
t.test(
  x = subset(dta, var == 1)$value,
  y = subset(dta, var == 2)$value
)
```

This works, but there is another (in my opinion more elegant) approach as well. This is using the formula notation.

3.3.2.2 Formula Notation

Instead of passing values to `x` and `y`, you can specify a formula and a data frame.

```
t.test(formula = value ~ var, data = dta)
```

This is telling `t.test()` to predict the values of `value` from the levels of `var`. Note that this does not use the `$` to extract the values from the data frame. Instead, the `data` argument tells `t.test()` where the columns in the formula

can be found. Formula notation will continue to show up throughout the course. It always follows the format `outcome ~ predictors`. The `~` (called “tilda”) is on your keyboard above the “tab” key. In the context of formula notation it means “is predicted from.” In the context of a *t* test, the outcome variable must be numeric and the predictor variable must have only two levels. The `t.test()` function will return an error if either of those things are untrue.

You can then pass additional values to arguments of `t.test()`, such as to `alternative`, `paired`, and `conf.level`.

3.3.2.3 Welch’s *t* Test

The above examples have all done a Welch’s *t* test. Welch’s *t* test is the same as a regular *t* test except that it includes an adjustment for the degrees of freedom. One of the assumptions of *t* tests is equal variance in each of the two samples. This assumption is often not met. For example, in the example you are working on for your assignment, it could be that participants who read about a foreign threat responded more similarly to one another than participants who read about a domestic threat. This would violate the assumption of homogeneity of variance, meaning that the regular *t* test would not be valid. Welch’s *t* test adjusts the degrees of freedom based on the degree of heterogeneity of variance between the two samples. This will make the *p* value larger but it means that you don’t need to assume homogeneity of variance for the test to be valid.

If there is no heterogeneity of variance between the samples, Welch’s *t* test will produce the same result as a regular *t* test. As such, it makes sense to just use Welch’s *t* test all the time. It removes an assumption and only affects the *p* value when necessary. The author of `t.test()` agrees with this assessment and has therefore made the default independent samples *t* test a Welch’s *t* test. If you want to conduct a traditional *t* test, you must specify that using `var.equal = TRUE`. This means you are assuming equal variances in the two samples. I don’t recommend ever doing this for your own research. However, you are likely to encounter regular *t* tests in other people’s work. If you want to reproduce their results, you will need to use this argument.

3.3.2.4 describeBy()

3.3.3 Manuscript

Prepare a written document that includes a one-paragraph description of the results and a one-paragraph discussion. The manuscript should follow the APA Journal Article Reporting Standards for Quantitative Research (JARS-Quant). Note that these guidelines are broad and ask for more information than required for the purpose of this lab report. You need not report information about dates of recruitment and data collection or missing data (though this information would be reported if it were a real study!).

The results section should include the following:

- Total participants in each experimental condition (note that the actual number of participants might differ from the planned number of participants in the study preregistration).
- M and SD of national identity for each experimental group.
- Results of the null-hypothesis significance test and corresponding standardized effect size.
- Description of your efforts to diagnose problems with statistical assumptions and data distributions.
- A figure visualizing mean national identity in each of the two groups and error associated with that measurement. The figure should be referenced in the text of the results section.

The discussion section should adhere to the JARS-Quant guidelines. You can use these questions as a guide.

- Was the hypothesis supported? 1 sentence.
- How do your results compare to the results reported by Davies et al. (2008)? Consider both the size of the effect and the results of the significance test. 1–2 sentences.
- Why do you think the results do or do not support the hypothesis? Consider both substantive theoretical reasons as well as design and statistical considerations.
- Canadian adults are the implied population of interest in this study. Given the sampling design, how well does this sample represent that population? How does that affect the interpretation of these results? 2–3 sentences.
- The broad theoretical mechanism under study was one in which the perception of an external threat to a social group leads to increased identification with that social group. In what contexts does this study have implications for this theory? Consider how the study's design might limit its generalizability. 1–3 sentences.
- What is an implication of this research for future research, programs, or policy? 1–2 sentences.

Chapter 4

Correlation

This lab will cover how to test hypotheses about the associations between two continuous variables.

4.1 Learning Objectives

4.2 The Study:

4.3 Data Collection

4.4 Assignment Instructions

Chapter 5

Multiple Regression

5.1 Learning Objectives

5.2 The Study:

5.3 Data Collection

5.4 Assignment Instructions

Chapter 6

Univariate ANOVA

6.1 Learning Objectives

6.2 The Study:

6.3 Data Collection

6.4 Assignment Instructions

Chapter 7

Factorial ANOVA

7.1 Learning Objectives

7.2 The Study:

7.3 Data Collection

7.4 Assignment Instructions

Chapter 8

Chi-squared Tests

8.1 Learning Objectives

8.2 The Study:

8.3 Data Collection

8.4 Assignment Instructions

Bibliography

- Davies, P. G., Steele, C. M., and Rose Markus, H. (2008). A nation challenged: The impact of foreign threat on america's tolerance for diversity. *Journal of Personality and Social Psychology*, 95(2):308–318.
- Wilson, K. E., Harden, S. M., Almeida, F. A., You, W., Hill, J. L., Goessl, C., and Estabrooks, P. A. (2016). Brief self-efficacy scales for use in weight-loss trials: Preliminary evidence of validity. *Psychological assessment*, 28(10):1255–1264.