

One-Way T-Tests in R

The one-sample t -test is not used as frequently as the independent-samples or paired-samples t -test in second language research, but as it could from time to time be useful I will outline briefly here how it can be performed.

When to Use a One-Sample T-Test

To determine whether some obtained value is statistically different from a neutral value, from a previously published population mean, from zero, or from some other externally dictated mean score, a one-sample t -test can be used. The one-sample t -test asks whether the mean score from the sample you have tested is statistically different from the externally determined mean score you are using to compare it to. I use Torres's (2004) study as an example of how the one-sample t -test works (although it is likely that polytomous IRT methods, which are beyond the scope of this book, would be a better way to analyze this data).

Torres gave a 34-item five-point Likert scale questionnaire to 102 adult ESL learners to determine whether the students preferred native or non-native teachers. Torres wanted to know whether the learners would prefer one type of teacher over the other both in general and in specific skill areas such as pronunciation and grammar. In the scale a 5 indicated a preference for native-speaking English teachers (NEST), a 1 indicated a preference for non-native English speaking teachers (non-NEST), and a 3 indicated no particular preference. In order to test whether the mean scores that were recorded were substantially different from a mean of 3, a one-sample t -test was conducted for each of the areas of

investigation.

Calling for a One-Sample T-Test

We will examine the question of whether ESL learners preferred NESTs or non-NESTs in the areas of culture and speaking in this example. I use the Torres.sav file, imported as **torres**. For the one-sample *t*-test, in R Commander choose STATISTICS > MEANS > SINGLE-SAMPLE T-TEST (see Figure 1). For the “Alternative Hypothesis” area, you want to put the value to test against in the “Null hypothesis: mu” box. For Torres’s questionnaire, the number “3” was neither agree nor disagree, so what we want to test is whether values depart from neutral, so I have entered “3” here. However, other numbers are possible for your data. For example, if you wanted to test whether your own students’ scores on an internal test were different from the mean of previous administrations of the test, whose mean score was 456, you could put 456 in the “Test Value” box.

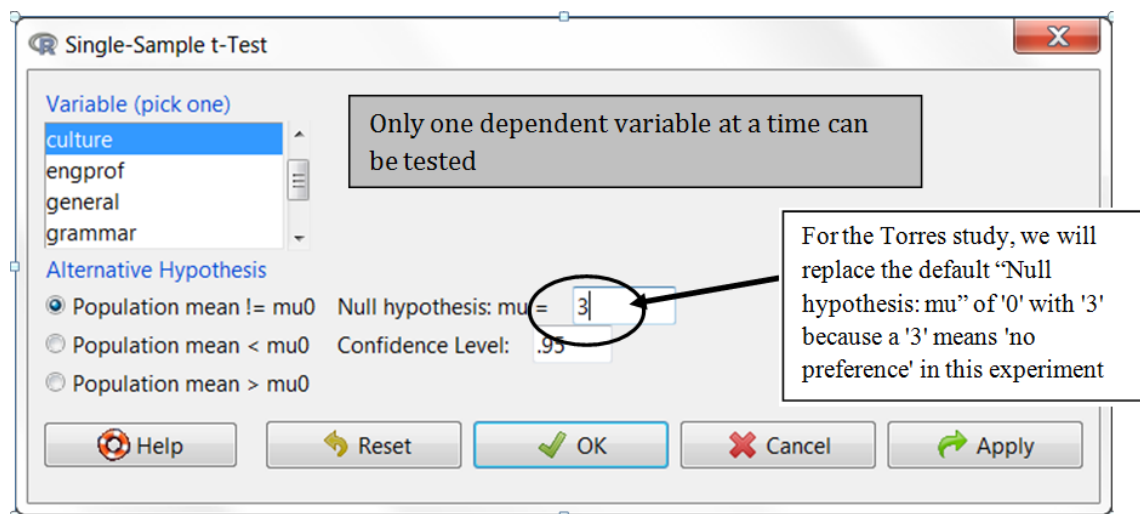


Figure 1 Opening a one-sample *t*-test dialogue box in R Commander.

Notice that if you would like to conduct a one-tailed hypothesis, you can choose that in the dialogue box.

Tip: If you use a one-tailed directional hypothesis, you do not have to adjust the p -value. The one that is returned has already been adjusted for a one-way hypothesis. Remember that using a one-tailed hypothesis will give you more power to find differences.

The output of the one-sample t -test looks like this:

```
One Sample t-test

data:  torres$culture
t = 6.8705, df = 101, p-value = 5.351e-10
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 3.371904 3.673848
sample estimates:
mean of x
 3.522876
```

On the first line you find the variable that you tested for, which in this case was Culture. Make sure to get a feel for your data before looking at the results of the statistical test. Look at the mean score, which can be found on the last line of the output. The mean of the Culture variable is 3.52, which means there is a slight preference above the neutral value for NESTs.

The main result of the t -test that we are interested in is 95% confidence interval, which is [3.37, 3.67]. This means that our questionnaire respondents truly do differ from neutral and have a real preference for NESTs that could be as weak as 3.37 or as strong as 3.67 (at least, we would expect the real mean score to fall in this range 95% of the time!). This

is not substantially larger than 3, so although there does seem to be a real preference, it seems like a fairly mild preference.

The R code for this test is:

```
t.test(torres$culture, alternative='two.sided', mu=3, conf.level=.95)
```

<code>t.test (x, . . .)</code>	Gives the command for all <i>t</i> -tests, not just the one-sample test
<code>torres\$culture</code>	This is the Culture variable in the <code>torres</code> dataset
<code>alternative="two.sided"</code>	This default calls for a two-sided hypothesis test; other alternatives: <code>"less"</code> , <code>"greater"</code>
<code>mu=3</code>	Tells R that you want a one-sample test; it compares your measured mean score to an externally measured mean
<code>conf.level=.95</code>	Sets the confidence level for the mean difference

Performing a One-sample T-Test

- 1 On the R Commander drop-down menu, choose STATISTICS > MEANS > SINGLE-SAMPLE T-TEST. Choose one variable, and then put the value you want to test your data against in the “Alternative Hypothesis” area in the “Null hypothesis: mu” box.
- 2 Basic R code for this command is (N.B. items in red should be replaced with your own data):
`t.test(torres$culture, mu=3)`

Performing a Robust One-Sample T-Test

In this section I will use Wilcox's WRS package to perform robust paired-sample t -tests.

I am assuming that you have read through Section 8.4.4 of the book and have installed all of the packages that are necessary and have loaded the WRS package.

To perform a 20% trimmed mean percentile bootstrap for a one-sample t -test, use Wilcox's command `trimpb()` (this is very similar to the robust command for independent samples t -tests, which was `trimpb2()`). Basically, the only difference in syntax between the two tests is that we add an argument specifying the neutral value: `null.value = 3`. Thus, this test contains means trimming (default is 20%) and uses a non-parametric percentile bootstrap. The syntax to test the Grammar variable is:

```
trimpb(torres$grammar, tr=.2, alpha=.05, nboot=2000, null.value=3)
```

Here is the output from this call:

```
[1] "The p-value returned by the this function is based on the"
[1] "null value specified by the argument null.value, which defaults to 0"
[1] "Taking bootstrap samples. Please wait."
$ci
[1] 2.967742 3.483871

$p.value
[1] 0.102
```

The output shows that the 20% trimmed mean percentile-bootstrap 95% confidence interval for the population mean of the Grammar variable is [2.97, 3.48]. Since the CI contains the neutral value of 3, we cannot reject the null hypothesis that the true population mean might be equal to 3.

Performing a Robust One-sample T-Test

- 1 First, the Wilcox WRS library commands must be loaded or sourced into R (see Section 8.4.4 in the book) and the library opened.
- 2 The basic R code for this command is:
`trimpb(torres$grammar, tr=.2, null.value=3)`

Effect Sizes for One-Sample T-Tests

Effect sizes can be determined quite simply; just take the mean of x listed in the output, subtract your neutral value from it (for the Torres Culture variable the mean is 3.52, and the neutral value is 3, so $3.52 - 3 = .52$) and divide by the standard deviation of the one group you have, just as you would do if the variances were not equal (see Section 8.4.5 of the book). The standard deviation for the Culture variable is .77 so the effect size for Culture is thus $0.52/.77 = .68$, meaning the difference from not caring whether a native English speaker is a teacher for a culture class is 68% of one standard deviation higher. I would say this is a medium-small effect. Previously we looked at the confidence interval and said the preference was statistically different from neutral but not a very strong preference, so the effect size here confirms this.

Application Activities for the One-Sample T-Test

- 1 Torres (2004) data. Use the dataset `Torres.sav`, imported as `torres` for R.

Calculate one-sample t -tests for the variables of listening and reading using a one-sample parametric test. Obtain robust confidence intervals of the mean estimate.

Report on descriptive statistics, 95% CIs and effect sizes. Comment on the size of the effect sizes.

- 2 Using the same dataset as #2, look at the variables of culture and pronunciation using both parametric one-sample tests and robust one-sample tests (means

- trimming and bootstrapping). Do you find any differences? What are the effect sizes?
- 3 Dewaele and Pavlenko Bilingual Emotions Questionnaire (2001–2003) data. Use the BEQ.sav dataset, imported as **beq**. Test the hypothesis that the people who took the online Bilingualism and Emotions Questionnaire will rate themselves as fully fluent in speaking, comprehension, reading, and writing in their first language (ratings on the variable range from 1, least proficient, to 5, fully fluent). Use the variables L1SPEAK, L1COMP, L1READ, and L1WRITE. Calculate effect sizes and comment on their size.

Bibliography

- Chernick, M. (2007). *Bootstrap methods: A guide for practitioners and researchers*. New York: John Wiley & Sons.
- Crawley, M. J. (2007). *The R book*. New York: Wiley.
- Kirby, K. N., & Gerlanc, D. (2013). BootES: An R Package for Bootstrap Confidence Intervals on Effect Sizes. *Behavior Research Methods*, 45(4), 905–927.
- Torres, J. (2004). *Speaking up! Adult ESL students' perceptions of native and non-native English speaking teachers*. Unpublished MA, University of North Texas, Denton.