

One-Way T-Tests in SPSS

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. For a one-sample t -test choose ANALYZE > COMPARE MEANS > ONE-SAMPLE T-TEST, and the dialogue box in Figure 1 will appear. Move as many variables you like over to the “Test Variable(s)” box. Consider the “Test Value” box. You want to put the neutral value in here. For Torres’s questionnaire, the number “3” was neither agree nor disagree, so what we want to test is whether values depart from neutral. 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. Last of all, you can open the BOOTSTRAP button and check the “Perform bootstrapping” box. Change the number of samples to 10,000 instead of 1000 (Chernick, 2007). Change the confidence intervals to “BCa” (Crawley, 2007, p. 322 says statisticians prefer this interval).

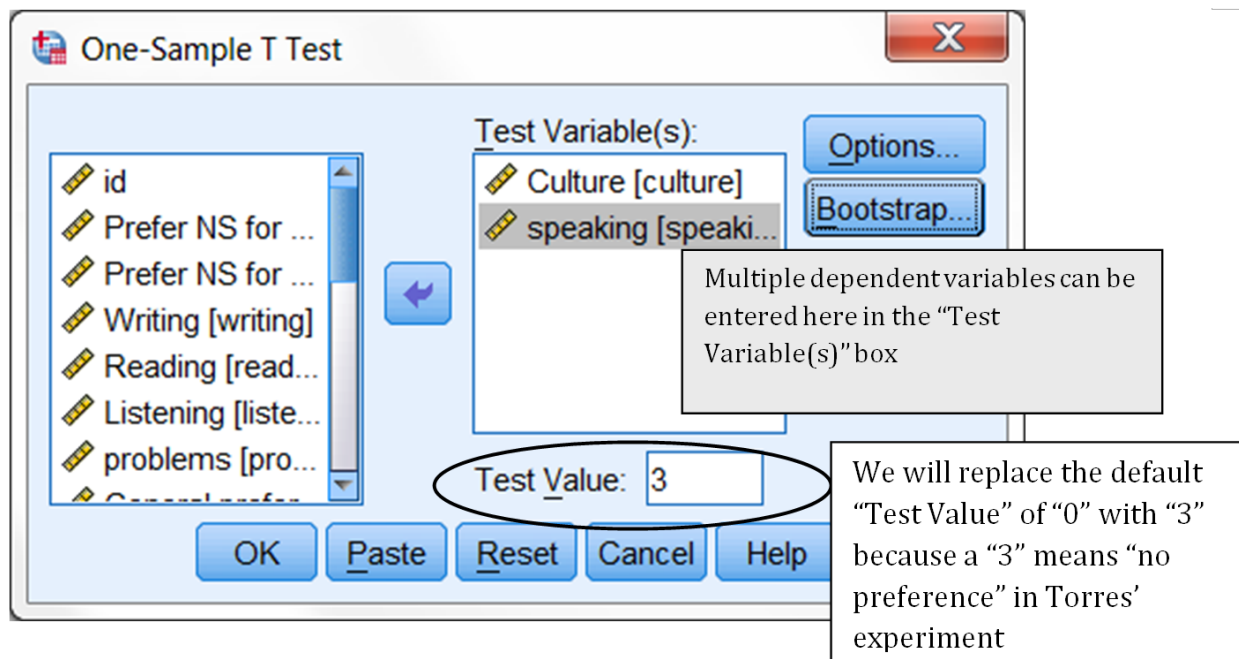


Figure 1 Opening a one-sample *t*-test dialogue box in SPSS.

When you run the one-sample *t*-test on the Torres data, the output shown in Table 1 will appear in the SPSS Viewer. I am omitting the descriptive statistics, and will just tell you that for the Culture variable the descriptive statistics are $M = 3.53$, $sd = .77$, $N = 102$, and for the Speaking variable the descriptive statistics are $M = 3.30$, $sd = .95$, $N = 102$. Since the mean score for Culture is 3.5, this means there is some preference for NESTs, but the standard deviation is fairly large as well. For Speaking, the mean score is even more towards the neutral level, and the standard deviation is even higher.

One-Sample Test						
	Test Value = 3					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Culture	6.870	101	.000	.52288	.3719	.6738
speaking	3.241	101	.002	.30392	.1179	.4899

Bootstrap for One-Sample Test						
	Mean Difference	Bootstrap ^a				
		Bias	Std. Error	Sig. (2-tailed)	BCa 95% Confidence Interval	
					Lower	Upper
Culture	.52288	.00081	.07574	.001	.36601	.67320
speaking	.30392	.00225	.09517		.10784	.48056

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Table 1 Output from a one-way *t*-test.

The results of the parametric *t*-test are found in the “One-Sample Test” table (the first in the output shown in Table 1). The main result of the *t*-test that we are interested in is 95% confidence interval, which for Culture 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. For the Speaking variable, the 95% CI is even closer to the neutral value of 3, at [3.12, 3.48]. This means that although there is a clear preference for NESTs for Speaking, the magnitude of the effect is smaller than for the Culture variable since the lower limit of the CI for Speaking (3.12) is closer to neutral than the lower limit for Culture (3.37). We can thus conclude that the respondents for this questionnaire had a preference for NESTs for both Culture and Speaking classes, although the preference for a native speaker teacher for Culture was stronger than that of Speaking.

The results of the non-parametric BCa bootstrapping show the same general results as the parametric test, although the details of the confidence intervals are different, and just very slightly show a smaller interval than the parametric results.

If you would like to run a one-tailed hypothesis, SPSS does not provide any way to do this directly but you could use the method of running a 90% CI and using the upper or lower limit that you need, as explained in Section 8.5.2 of the book.

Tip: If we had had a directional alternative hypothesis (also called a one-tailed test), such as "The true mean is greater than 3," then we would have had to divide the p -value that SPSS returns by 2 to obtain the correct p -value result.

Performing a One-sample T-Test in SPSS

Performing an Independent-Samples T-Test in SPSS

- 1 On the drop-down menu, choose ANALYZE > COMPARE MEANS > ONE-SAMPLE T-TEST. When a dialogue box comes up, move the variables you want to test to the "Test Variable(s)" box.
- 2 Decide what your "Test Value" is. This is the value you want to test against your data and see whether your data are different from this value.
- 3 Open the BOOTSTRAP button and check the "Perform bootstrapping" box. Change the number of samples to 10,000 and the confidence intervals to "BCa." Press CONTINUE. Run the test.

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.

We can use the R package **bootES** (Kirby & Gerlanc, 2013) to calculate confidence intervals for Cohen's *d* effect sizes. The following code gets the package installed and the package open, and then specifies that we want to test the contrast between the NP and PTP group using Cohen's *d* results, with BCa bootstrapping.

```
install.packages("bootES")
```

```
library(bootES)
```

```
bootES(leow, R=2000, data.col="recprescore", group.col="group",
```

```
contrast=c("Non Think Aloud"=1, "Think Aloud"=-1), effect.type=c("cohens.d"),
```

```
ci.type=c("bca"))
```

Application Activities for the One-Sample T-Test

- 1 Torres (2004) data. Use the dataset *Torres.sav*. Calculate one-sample *t*-tests for the variables of listening and reading using a one-sample parametric test. Obtain bootstrapped estimates of the mean difference between the neutral value of 3. 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 (use 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. 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.