

## 11.11: Summary

- A one sample t-test is used to compare a single sample mean against a hypothesised value for the population mean. (Section 13.2)
- An independent samples t-test is used to compare the means of two groups, and tests the null hypothesis that they have the same mean. It comes in two forms: the Student test (Section 13.3 assumes that the groups have the same standard deviation, the Welch test (Section 13.4) does not.
- A paired samples t-test is used when you have two scores from each person, and you want to test the null hypothesis that the two scores have the same mean. It is equivalent to taking the difference between the two scores for each person, and then running a one sample t-test on the difference scores. (Section 13.5)
- Effect size calculations for the difference between means can be calculated via the Cohen's d statistic. (Section 13.8).
- You can check the normality of a sample using QQ plots and the Shapiro-Wilk test. (Section 13.9)
- If your data are non-normal, you can use Wilcoxon tests instead of t-tests. (Section 13.10)

### References

- Student, A. 1908. "The Probable Error of a Mean." *Biometrika* 6: 1–2.
- Box, J. F. 1987. "Guinness, Gosset, Fisher, and Small Samples." *Statistical Science* 2: 45–52.
- Welch, B. L. 1947. "The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved." *Biometrika* 34: 28–35.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Lawrence Erlbaum.
- McGrath, R. E., and G. J. Meyer. 2006. "When Effect Sizes Disagree: The Case of r and d." *Psychological Methods* 11: 386–401.
- Hedges, L. V. 1981. "Distribution Theory for Glass's Estimator of Effect Size and Related Estimators." *Journal of Educational Statistics* 6: 107–28.
- Hedges, L. V., and I. Olkin. 1985. *Statistical Methods for Meta-Analysis*. New York: Academic Press.
- Shapiro, S. S., and M. B. Wilk. 1965. "An Analysis of Variance Test for Normality (Complete Samples)." *Biometrika* 52: 591–611.

184. We won't cover multiple predictors until Chapter 15

185. Informal experimentation in my garden suggests that yes, it does. Australian natives are adapted to low phosphorus levels relative to everywhere else on Earth, apparently, so if you've bought a house with a bunch of exotics and you want to plant natives, don't follow my example: keep them separate. Nutrients to European plants are poison to Australian ones. There's probably a joke in that, but I can't figure out what it is.

186. Actually this is too strong. Strictly speaking the z test only requires that the sampling distribution of the mean be normally distributed; if the population is normal then it necessarily follows that the sampling distribution of the mean is also normal. However, as we saw when talking about the central limit theorem, it's quite possible (even commonplace) for the sampling distribution to be normal even if the population distribution itself is non-normal. However, in light of the sheer ridiculousness of the assumption that the true standard deviation is known, there really isn't much point in going into details on this front!

187. Well, sort of. As I understand the history, Gosset only provided a partial solution: the general solution to the problem was provided by Sir Ronald Fisher.

188. More seriously, I tend to think the reverse is true: I get very suspicious of technical reports that fill their results sections with nothing except the numbers. It might just be that I'm an arrogant jerk, but I often feel like an author that makes no attempt to explain and interpret their analysis to the reader either doesn't understand it themselves, or is being a bit lazy. Your readers are smart, but not infinitely patient. Don't annoy them if you can help it.

189. Although it is the simplest, which is why I started with it.

190. A funny question almost always pops up at this point: what the heck is the population being referred to in this case? Is it the set of students actually taking Dr Harpo's class (all 33 of them)? The set of people who might take the class (an unknown number) of them? Or something else? Does it matter which of these we pick? It's traditional in an introductory behavioural stats class to mumble a lot at this point, but since I get asked this question every year by my students, I'll give a brief answer. Technically yes, it does matter: if you change your definition of what the "real world" population actually is, then the sampling distribution of your observed mean  $\bar{X}$  changes too. The t-test relies on an assumption that the observations are sampled at random from an infinitely large population; and to the extent that real life isn't like that, then the t-test can be wrong. In practice, however, this

isn't usually a big deal: even though the assumption is almost always wrong, it doesn't lead to a lot of pathological behaviour from the test, so we tend to just ignore it.

191. Yes, I have a “favourite” way of thinking about pooled standard deviation estimates. So what?
192. A more correct notation will be introduced in Chapter 14.
193. Well, I guess you can average apples and oranges, and what you end up with is a delicious fruit smoothie. But no one really thinks that a fruit smoothie is a very good way to describe the original fruits, do they?
194. This design is very similar to the one in Section 12.8 that motivated the McNemar test. This should be no surprise. Both are standard repeated measures designs involving two measurements. The only difference is that this time our outcome variable is interval scale (working memory capacity) rather than a binary, nominal scale variable (a yes-or-no question).
195. At this point we have Drs Harpo, Chico and Zeppo. No prizes for guessing who Dr Groucho is.
196. This is obviously a class being taught at a very small or very expensive university, or else is a postgraduate class. *I've never taught an intro stats class with less than 350 students.*
197. The `sortFrame()` function sorts factor variables like `id` in alphabetical order, which is why it jumps from “student1” to “student10”
198. This is a massive oversimplification.
199. Either that, or the Kolmogorov-Smirnov test, which is probably more traditional than the Shapiro-Wilk, though most things I've read seem to suggest Shapiro-Wilk is the better test of normality; although Kolomogorov-Smirnov is a general purpose test of distributional equivalence, so it can be adapted to handle other kinds of distribution tests; in R it's implemented via the `ks.test()` function.
200. Actually, there are two different versions of the test statistic; they differ from each other by a constant value. The version that I've described is the one that R calculates.

---

This page titled [11.11: Summary](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.

- [13.11: Summary](#) by [Danielle Navarro](#) is licensed [CC BY-SA 4.0](#). Original source: <https://bookdown.org/ekothe/navarro26/>.