Independent-samples t-test

Relevant research questions and data requirements

Research question: Independent-sample t-tests are *mean difference tests*. Specifically, we are seeking to compare if there is a difference between the mean of population 1 (e.g., untreated) and the mean of population 2 (e.g., treated). Because we cannot usually get data from everyone in our population of interest, we take a sample from each population and compare the means. We make inferences from our samples to our populations – hence, this is an example of Inferential Statistics.
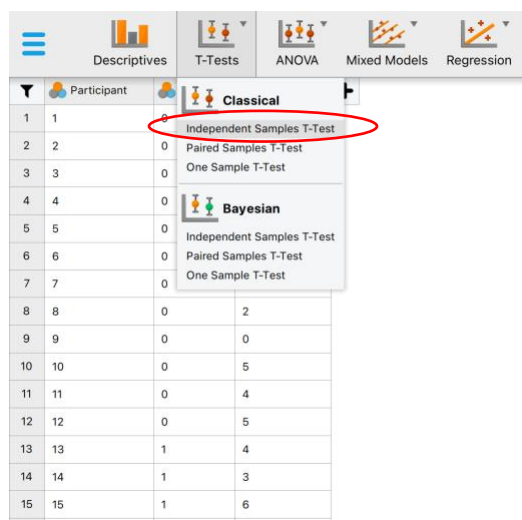
Let's say we want to know if students at Hogwarts wearing an invisibility cloak commit more mischief than students not wearing an invisibility cloak. These are two different populations in theory (students with cloak and students without cloak), but we can't measure everyone, so we sample 24 students and randomly assign them to wear a cloak vs. no cloak; we measure acts of mischief. This is a *between-subjects design* because the participants are either in one group or the other.

How many groups? Two (e.g., experimental and control group). Fun fact: While a t-test cannot do what ANOVA can do (namely, compare 3+ groups at once), an ANOVA can do whatever a t-test can do (namely, compare two groups). So, you could also analyze these data using an ANOVA!

Data requirements? Interval/ratio outcome data (here: mischievous acts), ideally the data are (roughly) normally distributed and independent (scores are not related). You also assume that the variance in the two groups is *roughly* the same. This is called homogeneity/ equality of variance. Check for outliers.
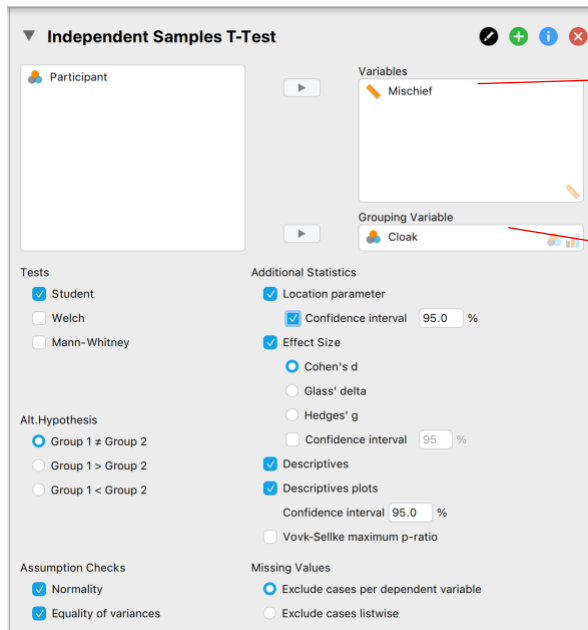
Running the test in JASP

Once you have opened your data in JASP, click on "t-tests" in the main menu and select "independent-samples t" under "classical" (we will ignore all Bayesian analyses for now).

Move your outcome variable of interest into the "variable" window. Notice the little ruler in the bottom right reminds you this variable must be identified as a scale variable in your data set.

Move your independent variable into "grouping variable". Notice the little symbols on the right reminding you this must be an ordinal or nominal variable.

Tests: Select "Student" (Fun fact: This is not a special test for students, but named after a Statistician who published about the t-distribution by the synonym "Student"). Welch is what you want to run if your assumption of homogeneity/equality is violated and Mann-Whitney is the nonparametric equivalent of the independent samples t, i.e., the test to use when your data are not normally distributed (see the output below to tell whether data are normally distributed and if homogeneity of variance is met).

Alt. hypothesis: Here, the two-tailed option is chosen automatically as the alternative hypothesis, i.e. "the acts of mischief in group 1 are different from the acts of mischief in group 2". (Recall a one-tailed test would specify a direction, e.g. we might hypothesize "students with invisibility cloaks conduct *more* acts of mischief than students without cloaks." Unless you have good reason to run a one-tailed test, however, you should always run a two-tailed test because it requires slightly more extreme cut-offs for the critical region, i.e. it is more stringent.)

Assumption Checks:
- Check "Normality" to test the hypothesis that mischief in your sample is, in fact, (roughly) normally distributed. Note: You can also get a visual impression by running the distribution plot under Descriptives as outlined in the handout on "Descriptive Statistics"
- Check equality of variances to test that the variability in both samples is roughly the same

Additional Statistics:
- Check "Location Parameter" – this is the difference between the sample means; then click "confidence interval"
- Check "effect size" and "Cohen's d" – this will tell you the distance between the sample means in SD units
- Descriptives will calculate M and SD as well as standard error (SE) for your conditions
- Descriptive Plots will present a visual overview with 95% CI around each mean

Invisibility Cloak: Independent samples t-test ▼

Independent Samples T-Test ▼

Independent Samples T-Test ▼

	t	df	p	Mean Difference	SE Difference	95% CI for Mean Difference Lower	95% CI for Mean Difference Upper	Cohen's d
Mischief	−1.713	22	0.101	−1.250	0.730	−2.763	0.263	−0.700

Note. Student's t-test.

This is our hypothesis test. Shown in order are the t-value, the df (for an independent samples t this is n-2), the p-value, the mean difference between sample and test value (sample mean-test value), the 95% CI around the mean difference, and the effect size Cohen's d

Here we learn that the difference between our sample means is -1.25 with a standard error of 0.73. Although the cloaked students conducted an average of 1.25 more acts of mischief compared to the students not wearing an invisibility cloak (also see Descriptives below), this difference is not statistically significant at an alpha level of .05, because $p = .101$, which is larger than .05. We therefore do not have enough evidence to reject our null hypothesis that $\mu_{cloak} = \mu_{no\ cloak}$ (or $\mu_{cloak} - \mu_{no\ cloak} = 0$).

We also see that 0 is contained in the 95% CI [-2.76; 0.26] for the mean difference, so 0 is a likely value of the true difference between the acts of mischief in the population of cloaked students and the population of uncloaked students. If a 0 difference is a distinct possibility, this means we are left to conclude that invisibility cloaks have no effect on acts of mischief based on our experiment.

BUT: Notice that we had a sample size of n=24, which is very small. Cohen's d is -.70, which tells us our sample means are 0.70 SDs apart. This makes it a medium-large effect (d=0.20 is small, d=0.50 is medium, d=0.80 is a large effect). In other words, we may have a decent effect here that we cannot detect because our sample size is too small (we don't have enough "power" to detect a possible effect). The verdict? Replicate, increase sample size.

Assumption Checks ▼

Test of Normality (Shapiro–Wilk) ▼

		W	p
Mischief	no cloak	0.913	0.231
	cloak	0.973	0.936

Note. Significant results suggest a deviation from normality.

Test of Equality of Variances (Levene's)

	F	df	p
Mischief	0.545	1	0.468

These are our assumption checks. The top tests our assumption of normally distributed data, the bottom tests our assumption that variances in both groups are roughly equal. For both tests the null hypothesis is that assumptions are met.

Looking at the p-values, you can see that all of them are >.05. Hence, we have no reason to reject the null for either test, so we conclude assumptions are met.

*Note: If you detect serious departures from normality (one of the p-values for the Shapiro-Wilk is <.05), you might need to re-run your analysis using the Mann-Whitney option.
If Levene's test p-value is <.05 (i.e., the assumption of homogeneity of variance is violated), you should re-run your analysis using the Welch option.
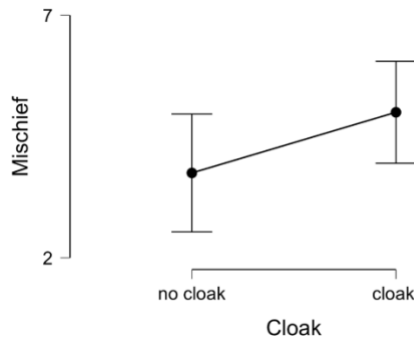
Descriptives ▼

Group Descriptives

	Group	N	Mean	SD	SE
Mischief	no cloak	12	3.750	1.913	0.552
	cloak	12	5.000	1.651	0.477

Descriptives Plots ▼

Mischief

Lastly, our descriptive statistics and descriptive plots. Condition labels will be reflected here, so always label your conditions in a way that makes the output easy to read (here: "cloak" vs "no cloak" rather than "0" vs "1").

Our plots also show that the cloaked students conducted more mischief than the uncloaked students, but notice that the 95% CIs overlap, which means that the true population means likely overlap, i.e. there is no difference.

Also notice that the y-axis is scaled from 2-7, not 0-10, so differences between the two groups appear slightly more pronounced.

Writing up results in APA style

An independent samples t-test indicated that students wearing an invisibility cloak (n = 12, M = 5.00, SD = 1.65) did not conduct more acts of mischief than students not wearing an invisibility cloak (n = 12, M = 3.75, SD = 1.91) $t(22)$ = -1.71, p = .101, 95% CI [-2.76; 0.26], Cohen's d = -0.70.

*Note that there are spaces before and after equal signs; M, SD, t, p are italicized; everything is rounded to two digits *except* for p-values, which should be reported exactly as given in the output. Only report leading zeros for values that can exceed 1 (hence, p-values should *not* be reported with leading 0s)