z-tests and t-tests

z-test (One-sample location test)

The goal of the one-sample z-test is to determine whether a sample mean ($\bar{X}$) shows a statistically significant difference from a known/hypothesized constant (μ) when the standard deviation of the population (σ) is *known*. In this case, the test statistic $z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ follows a standard normal distribution, and we can quantify effect size using Cohen's $d = \frac{\bar{X}-\mu}{\sigma}$.

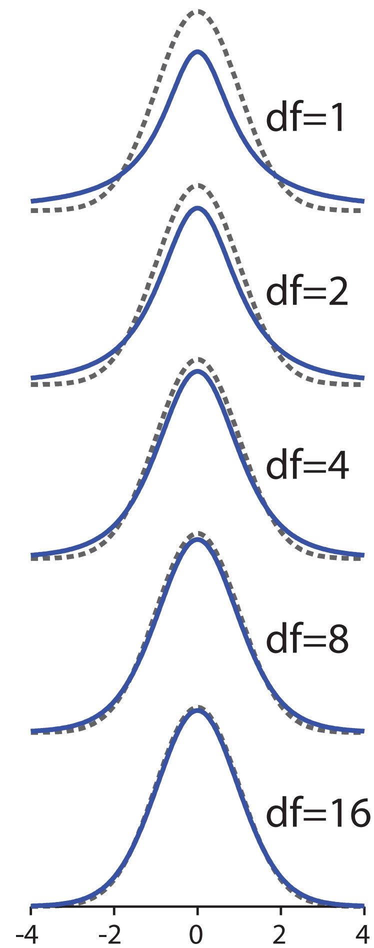

Figure 1: Comparison of the t-distribution (solid line) and standard normal distribution (dashed) as the degrees of freedom (df) increase.

One-sample t-test

The goal of the one-sample t-test is to determine whether a sample mean ($\bar{X}$) shows a statistically significant difference from a known/hypothesized constant (μ) when the standard deviation of the population (σ) is *unknown*. Instead, we estimate σ with the sample standard deviation (s).

The main difference between the one-sample z-test and one-sample t-test is that estimating the variance has the effect of spreading out the distribution a little beyond what it would be if σ were known. The effect is pronounced at small sample sizes, but with larger n, the difference between the normal and the t-distribution is really not very important.

Here the test statistic $t = \frac{\bar{X}-\mu}{s/\sqrt{n}}$ follows a t-distribution with $df = n-1$. To quantify effect size we can use Cohen's $d = \frac{\bar{X}-\mu}{s}$, and confidence intervals can be found using $\mu = \bar{X} \pm t_c s/\sqrt{n}$ where t_c is a critical value of the t-distribution. If we want 95% confidence intervals, for instance, we can use the t-table and find the t_c corresponding to a two-tailed test with $df = n-1$ and $\alpha = 0.05$.

Paired-sample t-test

In a within-subjects or matched experimental design, rather than comparing a group mean to a constant, we instead want to determine whether the sample means for two measurements (from the same subjects) show a statistically significant difference. Here use a test statistic very similar to the one-sample t-test, but the trick is to first compute difference scores $X_D = X_2 - X_1$ for each score pair (e.g. the difference between pre- and post-test for each participant).

Here the test statistic $t = \frac{\bar{X}_D-\mu_D}{s_D/\sqrt{n}}$ follows a t-distribution with $df = n-1$ where $\bar{X}_D$ is the mean of the difference scores and s_D is the standard deviation of the difference scores. Note that n here is the number of differences *not* the total number of scores ($n_1 + n_2$). In almost all cases, we will set $\mu_D = 0$, meaning that the "expected difference under the null hypothesis" is zero. Similar to the one-sample t-test, Cohen's d is $d = \frac{\bar{X}_D - \mu_D}{s_D}$ and confidence intervals are given by $\mu_D = \bar{X}_D \pm t_c s_D / \sqrt{n}$.

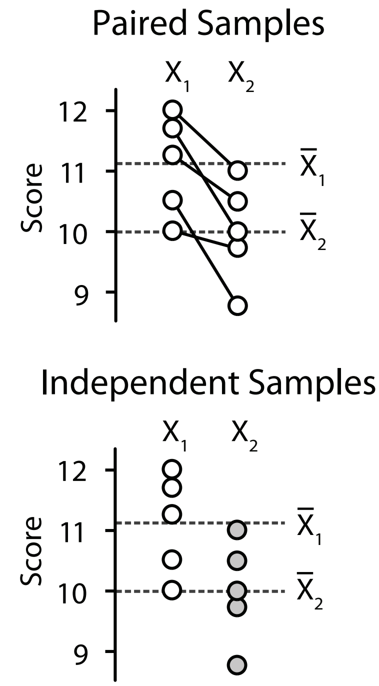

Figure 2: Paired vs Independent samples. Within-subjects designs are often preffered because they allow us to control for individual differences. Even though the differences between the means are identical, only the paired data shows a significant difference. When the data is paired, it's clear that there is a consistent decrease from condition 1 to condition 2 ($s_D < s_p$).

Independent-sample t-test

Finally, in a between-subjects (independent measures) experimental design we want to determine whether the sample means for two independent groups show a statistically significant difference. Here we cannot use differences between pairs of scores, since these are completely different participants. Instead, we need to first compute the pooled variance:

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

where $SS_1 = \sum(X - \bar{X}_1)^2$ is the sum of squares for group 1, which has a mean $\bar{X}_1$. n_1 is the sample size for group 1, and $df_1 = n_1 - 1$ is the degrees of freedom for group 1. Similarly, $SS_2 = \sum(X - \bar{X}_2)^2$ is the sum of squares for group 2, which has mean $\bar{X}_2$, sample size n_2, and $df_2 = n_2 - 1$.

Then, the test statistic

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

follows a t-distribution with $df = n_1 + n_2 - 2$. As with the paired t-test, in almost all cases we set $\mu_1 - \mu_2 = 0$, meaning that the "expected difference under the null hypothesis" is zero. Cohen's d, in this case, is $d = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p}$, and the confidence intervals are given by $(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_c \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$.

Note: this version of the unpaired t-test assumes that the two groups have equal variances. This is called the "homogeneity of variance assumption" and is something we will need to watch out for (SPSS returns results for tests with and without). Additionally, keep in mind that all z-tests and t-tests make the assumption that the population is normally distributed. For sufficiently large n this is not an issue, since the distribution of sample means becomes normal, but for small sample sizes it can lead to problems.

Paired Sample t-test Example

Subject	No Imagery	Imagery
1	5	11
2	6	6
3	5	4
4	9	13
5	8	11
6	8	7
7	6	11
8	7	14
9	9	13

Dr. Smith is a cognitive psychologist interested in the role of imagery on memory. He decides to conduct a counter-balanced repeated-measures experiment where, in one condition, participants are asked to memorize a list of words without explicit instructions, and, in a second condition, using a different list of words, participants are instructed to think of a picture of each individual word in the list. He records how many words subjects can recall under the two conditions. Data from 9 subjects is shown in the table at right.

1. State the null and alternate hypotheses,
2. Calculate the appropriate test statistic,
3. Find the critical value
4. State your conclusion about the null hypothesis
5. How large is the effect size in the study?

Note: all assumptions are satisfied, use alpha = .05, two-tailed test.

Independent Sample t-test Example

After analyzing the data from his first experiment, Dr. Smith decides that there may have been practice effects in his within-subjects design. He decides to conduct a second experiment using completely separate groups. Participants are now randomly assigned to one of two groups: in group one, participants are asked to memorize a list of words without explicit instructions, and, in group two, using the same list of words, participants are instructed to think of a picture of each individual word in the list. He records how many words subjects can recall under the two conditions. Data from 16 subjects is shown in the table at right.

No Imagery	Imagery
5	7
5	6
4	4
10	7
6	10
5	7
4	8
9	7

1. State the null and alternate hypotheses,
2. Calculate the appropriate test statistic,
3. Find the critical value
4. State your conclusion about the null hypothesis
5. How large is the effect size in the study?

Note: all assumptions are satisfied, use alpha = .05, two-tailed test.

Paired Sample t-test Example - Solution

1. H_0: $\mu_D = 0$ (there is no difference between the imagery and no imagery condition)
   H_A: $\mu_D \neq 0$ (there is a difference between the conditions)
2. We're using a paired t-test. In this case, the test statistic

$$t = \frac{\bar{X}_D - \mu_D}{s_D/\sqrt{n}}$$

   follows a t-distribution with $df = n - 1$ where $\bar{X}_D$ is the mean of the difference scores and s_D is the standard deviation of the difference scores. After calculating the individual differences

Subject	No Imagery	Imagery	X_D
1	5	11	6
2	6	6	0
3	5	4	-1
4	9	13	4
5	8	11	3
6	8	7	-1
7	6	11	5
8	7	14	7
9	9	13	4
			$\bar{X}_D = 3$
			$s_D = 3$

   we find $\bar{X}_D = 3$ and $s_D = 3$. Our null hypothesis is that $\mu_D = 0$ so $t = \frac{3-0}{3/\sqrt{9}} = 3$.
3. Using the t-table we look up the critical value. Look for the row for $df = n - 1 = 9 - 1 = 8$ and the column for the $\alpha = 0.05$ two-tailed, where we find $t_c = 2.306$.
4. Since $t > t_c$, we reject the null hypothesis, and conclude that the groups have a statistically significant difference.
5. For the paired t-test the effect size is given by $d = \frac{\bar{X}_D - \mu_D}{s_D}$. Plugging in our previous values we get $d = 3/3 = 1.0$.

Independent Sample t-test Example - Solution

1. H_0: $\mu_1 = \mu_2$ (there is no difference between the imagery and no imagery condition)
   H_A: $\mu_1 \neq \mu_2$ (there is a difference between the conditions)
2. We're using a independent sample t-test. In this case, the test statistic

   $$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$

   follows a t-distribution with $df = n_1 + n_2 - 2$. First, let's find the means and SS values for the two groups

No Imagery	Imagery
5	7
5	6
4	4
10	7
6	10
5	7
4	8
9	7
$\bar{X}_1 = 6$	$\bar{X}_2 = 7$
$SS_1 = 36$	$SS_2 = 20$
$df_1 = n_1 - 1 = 7$	$df_2 = n_2 - 1 = 7$

   Given SS values we can now calculate the pooled variance

   $$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{36 + 20}{7 + 7} = 4$$

   Since our null hypothesis assumes that $\mu_1 = \mu_2$ we have $t = \frac{(7-6)-0}{\sqrt{4/8+4/8}} = 1$
3. Using the t-table we look up the critical value. Look for the row for $df = n_1 + n_2 - 2 = 8 + 8 - 2 = 14$ and the column for the $\alpha = 0.05$ two-tailed, where we find $t_c = 2.145$.
4. Since $-2.145 < t < 2.145$, we fail to reject the null hypothesis, and cannot conclude that the groups are different.
5. For the independent samples t-test the effect size is given by

   $$d = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p}.$$

   Plugging in our previous values we get $d = \frac{(7-6)-0}{\sqrt{4}} = 0.5$.