

T-test

A **t-test** is a hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis. There are several different test statistics that fall into the category of a t-test.

One-Sample t-test

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
$$df = n - 1$$

Independent Two-Sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2} \sqrt{\frac{2}{n}}}$$
$$s_{x_1x_2} = \sqrt{\frac{1}{2}(s_{x_1}^2 + s_{x_2}^2)}$$
$$df = 2n - 2$$

Unequal Sample Size Two-Sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$s_{x_1x_2} = \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}}$$
$$df = n_1 + n_2 - 2$$

Unequal Sample Size and Unequal Variance Two-Sample t-test

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2}}$$

$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}\right]}$$

Dependent t-test for two samples

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

$$df = n - 1$$

Some assumptions are made in order to use the t-test.

Assumptions for t-test

1. The data comes from normal distribution or the sample size is greater than 30.
2. The sample is a simple random sample

The t-test is a hypothesis test. There are seven steps for a hypothesis test.

1. State the null hypothesis
2. State the alternative hypothesis
3. State the level of significance
4. State the test statistic
5. Calculate
6. Statistical Conclusion
7. Experimental Conclusion

Example One-Sample t-test

The age of the head of household residents of Phoenix, Arizona has increased in the recent years. Ten years ago the average age of head of household in Phoenix, Arizona was 33. A random sample of 40 head of household residents in Phoenix, Arizona was taken and the ages were recorded as follows.

25
46
63
37
72
20
47
41
59
64
32
37
80
59
63
31
38
35
45
36
45
22
36
44
48
69
31
36
34
37
70
33
65
29
31
76
51
25
27
21
27

We wish to use this data to test if the hypothesis that the average age of head of household residents in Phoenix, Arizona has increased.

This is a hypothesis test so we will need to go through the seven steps of a hypothesis testing.

Step 1: Null Hypothesis

Since in the past the mean Head of Household age was 33 that is the statement of no effect which is what the null hypothesis gives.

$$H_0 : \mu = 33$$

Step 2: Alternative Hypothesis

We want to know if the mean head of household age has increased so the alternative is

$$H_A : \mu > 33$$

This is a one-tailed test as we are only looking at a one-sided alternative.

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test Statistic

The test statistic needed is for a one-sample t-test. It is one-sample because we are only looking at one set of data values. The other requirements for a t-test have been met as the sample size is $40 > 30$ and the sample is a simple random sample.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$
$$df = n - 1$$

Step 5: Calculations

First the sample mean of the data should be calculated.

$$\bar{x} = \frac{\sum x}{n} = \frac{1787}{40} = 43.58537$$

Now find the sample standard deviation

$$s = \sqrt{\frac{\sum(x - \bar{x})^2}{(n-1)}} = 16.53931$$

(For explanations of how to calculate the mean and the standard deviation see Measure of Center and Variation respectively)

Plug this information into the formula for the test statistic.

$$t_{obs} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{43.58537 - 33}{\frac{16.53931}{\sqrt{40}}} = \frac{10.58537}{2.615095}$$

$$t_{obs} = 4.047795$$

$$df = n - 1 = 40 - 1 = 39$$

Find the critical value for a level of significance of 0.05 and 39 degrees of freedom. Use the student t-distribution table or a t-score calculator. (i.e. <http://www.stattek.com/online-calculator/t-distribution.aspx>)

$$t_{\alpha=0.05, df=39} = 1.685$$

Step 6: Statistical Conclusion

The null hypothesis is rejected if the observed test statistic is more extreme than the critical value. Since the alternative hypothesis is right tailed (greater than) then we reject if the observed value is greater than the critical value.

Since $t_{obs} = 4.047795 > 1.685 = t_{\alpha=0.05, df=39}$ then we reject the null hypothesis at a level of significance of 0.05.

Step 7: Experimental Conclusion

There is sufficient evidence to indicate that the mean age of the Head of Household in Phoenix, Arizona has increased at a level of significance of 0.05.

Example Dependent t-test for two samples

A research wishes to prove that a diet results in significant weight loss. A random sample of 40 people are weighed and then follow the diet for 6 weeks and weighed. The 40 participants were all in need of weight loss and of approximately the same health standings. The before and after weights are in the following table.

Before	After
257	240
244	236

151	145
229	200
151	146
296	277
151	147
219	201
265	215
177	160
202	190
194	172
198	180
175	170
194	183
281	233
148	240
260	200
159	250
176	267
255	202
212	195
287	243
242	204
248	205
224	199
185	164
210	173
157	150
170	152
293	232
152	140
263	202
154	142
287	260
249	195
161	153
244	204
267	207
300	245
300	234

Step 1: Null Hypothesis

The null hypothesis is the statement of no effect, which in this case, means no weight loss, so the mean weight loss would be 0.

$$H_0 : \mu_D = 0$$

Step 2: Alternative Hypothesis

The researcher believes that the diet causes weight loss, so the difference between the weight before and after the diet should be greater than 0.

$$H_A : \mu_D > 0$$

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test statistic

We are comparing before and after results. This means we have paired data so the test statistic needed is the dependent test statistic for two samples.

$$t = \frac{\bar{x}_D - \mu_D}{\frac{s_D}{\sqrt{n}}}$$
$$df = n - 1$$

Step 5: Calculations

To start the calculation, we need to find the difference between the before and after weights for the 40 participants.

Before	After	Difference
257	240	17
244	236	8
151	145	6
229	200	29
151	146	5
296	277	19
151	147	4
219	201	18
265	215	50
177	160	17
202	190	12

194	172	22
198	180	18
175	170	5
194	183	11
281	233	48
148	140	8
260	200	60
159	150	9
176	167	9
255	202	53
212	195	17
287	243	44
242	204	38
248	205	43
224	199	25
185	164	21
210	173	37
157	150	7
170	152	18
293	232	61
152	140	12
263	202	61
154	142	12
287	260	27
249	195	54
161	153	8
244	204	40
267	207	60
300	245	55
300	234	66

Find the sample mean of the differences and the sample standard deviation. (For explanations of how to calculate the mean and the standard deviation see Measure of Center and Variation Respectively)

$$\bar{x} = \frac{\sum x}{n} = \frac{1134}{40} = 27.65854$$

$$s = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = 19.85146$$

This information can be plugged into the test statistic.

$$t_{obs} = \frac{\bar{x}_D - \mu_D}{s_D / \sqrt{n}} = \frac{27.65854 - 0}{19.85146 / \sqrt{40}} = \frac{27.65854}{3.13879} = 8.81184$$
$$df = n - 1 = 40 - 1 = 39$$

The critical value can be found using the t-table or a t-value calculator.

(<http://www.statrek.com/online-calculator/t-distribution.aspx>)

$$t_{\alpha=0.05, df=39} = 1.685$$

Step 6: Statistical Conclusion

Reject the null hypothesis if the observed test statistic is more extreme than the critical value. Since the alternative hypothesis is right tailed (greater than) then we reject if the observed value is greater than the critical value.

Since $t_{obs} = 8.81184 > 1.685 = t_{\alpha=0.05, df=39}$ then we reject the null hypothesis.

Step 7: Experimental Conclusion

There is significant evidence to indicate that at a level of significance of 0.05 the diet leads to statistically significant weight loss.

Example Unequal Sample Size Two-Sample t-test

An advertising company wishes to place billboards for their new client in cities where they will do the most good. They believe this means cities with the youngest population for the new product. They wish to compare Phoenix and Los Angeles. They took a random sample of adult residents from each city. The ages of 45 LA residents and 40 Phoenix residents are shown below. The researchers wish to prove that the mean age of the adult residents of Phoenix is less than the mean age of the adult residents of LA.

Phoenix	LA
85	88
50	37
64	54
46	85
24	84
32	73
87	25
24	37
18	23

85	59
31	56
21	78
35	26
29	72
58	22
69	57
31	31
27	68
89	88
34	77
26	90
57	90
50	89
26	82
28	68
23	90
47	30
50	19
86	22
65	35
55	42
30	46
46	64
19	78
28	58
19	83
80	56
81	30
21	63
31	54
	70
	78
	79
	82
	50

Step 1: Null Hypothesis

The null hypothesis is the statement of no effect so that would be that the mean age of the adult residents of the two cities is the same.

$$H_0 : \mu_1 = \mu_2$$

Where population 1 is Phoenix and population 2 is LA.

Step 2: Alternative Hypothesis

The researchers' theory is that the mean age of the adult residents in Phoenix (population 1) is less than the mean age of the adult residents in LA (population 2)

$$H_A : \mu_1 < \mu_2$$

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test statistic

The test is comparing two data sets that have unequal sample size. $n_1 = 40$ and $n_2 = 45$ so we use the independent two sample t-test statistic for unequal sample size.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1, x_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
$$s_{x_1, x_2} = \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}}$$
$$df = n_1 + n_2 - 2$$

Step 5: Calculations

The calculations start with the sample mean for Phoenix and LA.

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{1807}{40} = 45.175$$
$$\bar{x}_2 = \frac{\sum x}{n} = \frac{2688}{45} = 59.733$$

Now find the standard deviation for each sample.

$$s_{x_1} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = 22.9904$$
$$s_{x_2} = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = 23.0478$$

Plug this information into the test statistic.

$$\begin{aligned}
 s_{x_1x_2} &= \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}} \\
 s_{x_1x_2} &= \sqrt{\frac{(40 - 1)(22.9904)^2 + (45 - 1)(23.0478)^2}{40 + 45 - 2}} \\
 s_{x_1x_2} &= \sqrt{\frac{(39)(528.5585) + (44)(531.2011)}{83}} \\
 s_{x_1x_2} &= \sqrt{\frac{20613.7815 + 23372.8477}{83}} \\
 s_{x_1x_2} &= \sqrt{\frac{43986.6292}{83}} \\
 s_{x_1x_2} &= \sqrt{529.9594} \\
 s_{x_1x_2} &= 23.0208 \\
 t_{obs} &= \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{45.175 - 59.733}{23.0208 \sqrt{\frac{1}{40} + \frac{1}{45}}} \\
 t_{obs} &= \frac{-14.558}{23.0208 \sqrt{0.04722}} = \frac{-14.558}{5.0026} \\
 t_{obs} &= -2.9101 \\
 df &= n_1 + n_2 - 2 = 40 + 45 - 2 = 83
 \end{aligned}$$

Find the critical value by using either the t-distribution table or t-distribution calculator.

$$t_{\alpha=0.05, df=83} = -1.663$$

Step 6: Statistical Conclusion

Reject the null hypothesis if the observed test statistic is more extreme than the critical value. Since we are interested in an alternative hypothesis that the mean of population 1 is less than the mean of population 2. The Rejection will occur if the observed test statistic is less than the critical value.

Since $t_{obs} = -2.9101 < -1.663 = t_{\alpha=0.05, df=83}$ then we reject the null hypothesis.

Step 7: Experimental Conclusion

There is significant evidence to indicate that the mean age in Phoenix residents is less than the mean age of the adult residents in LA.

Example Independent Two-Sample t-test

An advertising company wishes to place billboards for their new client in cities where they will do the most good. They believe this means cities with the youngest population for the new product. They wish to compare Phoenix and Los Angeles residents. They took a random sample of adult residents from each city. The ages of 40 LA residents and 40 Phoenix residents are shown below. The researcher wishes to prove that the mean age of the adult residents of Phoenix is less than the mean age of the adult residents of LA.

Phoenix	LA
85	88
50	37
64	54
46	85
24	84
32	73
87	25
24	37
18	23
85	59
31	56
21	78
35	26
29	72
58	22
69	57
31	31
27	68
89	88
34	77
26	90
57	90
50	89
26	82
28	68
23	90
47	30
50	19

86	22
65	35
55	42
30	46
46	64
19	78
28	58
19	83
80	56
81	30
21	63
31	54

Step 1: Null Hypothesis

The null hypothesis is the statement of no effect which is that the mean age of population 1 (Phoenix) is the same as the mean age of population 2 (LA).

$$H_0 : \mu_1 = \mu_2$$

Step 2: Alternative Hypothesis

The alternative hypothesis of interest here is that the mean age of the population 1 (Phoenix) is less than the mean age of the population 2 (LA).

$$H_A : \mu_1 < \mu_2$$

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test statistic

Independent two sample t-test statistic.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2} \sqrt{\frac{2}{n}}}$$

$$s_{x_1x_2} = \sqrt{\frac{1}{2}(s_{x_1}^2 + s_{x_2}^2)}$$

$$df = 2n - 2$$

Step 5: Calculations

To begin the calculations find the sample mean and variance of each sample.

$$\bar{x}_1 = \frac{\sum x}{n} = \frac{1807}{40} = 45.175$$

$$\bar{x}_2 = \frac{\sum x}{n} = \frac{2329}{40} = 58.225$$

$$s_{x_1}^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = 528.5583$$

$$s_{x_2}^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = 561.0506$$

Use this to begin calculating the test statistic.

$$s_{x_1x_2} = \sqrt{\frac{1}{2}(s_{x_1}^2 + s_{x_2}^2)}$$

$$s_{x_1x_2} = \sqrt{\frac{1}{2}(528.5583 + 561.0506)}$$

$$s_{x_1x_2} = \sqrt{\frac{1}{2}(1089.6089)}$$

$$s_{x_1x_2} = \sqrt{544.80445}$$

$$s_{x_1x_2} = 23.341$$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{s_{x_1x_2} \sqrt{\frac{2}{n}}} = \frac{45.175 - 58.225}{(23.341) \sqrt{\frac{2}{40}}} = \frac{-13.05}{(23.341)(0.2236)}$$

$$t_{obs} = \frac{-13.05}{5.2192} = -2.5004$$

$$df = 2n - 2 = 2(40) - 2 = 78$$

Now find the critical value for $\alpha = 0.05$ and degrees of freedom = 78 using either a t-distribution table or calculator.

$$t_{\alpha=0.05,df=78} = -1.665$$

Step 6: Statistical Conclusion

We reject the null hypothesis if the observed test statistic is more extreme than the critical value. We reject the null hypothesis if the observed test statistic is less than the critical value.

Since $t_{obs} = -2.5004 < -1.665 = t_{\alpha=0.05,df=78}$ then reject the null hypothesis.

Step 7: Experimental Conclusion

There is significant evidence to indicate that the mean age of Phoenix residents is less than the mean age of LA residents at a level of significance of 0.05.

Example Unequal Sample Size and Unequal Variance Two-Sample t-test

The publishers of a health magazine are going to publish an article on the healthiest cities. They believe LA and Phoenix are equal on all accounts except they need to compare the average weights of the adult residents. The variance of the two populations is known to be different. A random sample of 40 adult residents from each city was taken and their weights measured.

Phoenix	LA
280	195
146	186
121	258
256	125
279	190
167	232
186	142
264	281
290	257
279	158
153	204
246	162
134	118
151	142
126	134
246	215
270	132
228	221
274	174

159	145
134	256
146	163
295	168
212	226
135	183
127	151
158	235
255	255
164	267
231	156
179	231
251	233
220	255
174	199
146	212
121	163
296	155
120	166
130	127
237	269

Step 1: Null Hypothesis

The null hypothesis is the statement of no effect. For this problem that means that the average weights of the adult residents of Phoenix and LA are the same. Let Phoenix be population 1 and LA be population 2.

$$H_0 : \mu_1 = \mu_2$$

Step 2: Alternative Hypothesis

The alternative hypothesis is would be that one city has an average weight less than the other. We will test the claim that the average weight of the adult residents of Phoenix (population 1) is less than the average weight of the adult residents of LA (population 2)

$$H_A : \mu_1 < \mu_2$$

Step 3: Level of Significance

$$\alpha = 0.05$$

Step 4: Test statistic

Since we are comparing two different populations with different variances the following test statistic is needed. This test statistic can be used even though we have equal sample size for each population.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1\bar{x}_2}}$$
$$s_{\bar{x}_1\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1 - 1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2 - 1)}\right]}$$

Step 5: Calculations

Find the mean and variance of each set of sample data.

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = 199.65$$

$$\bar{x}_2 = \frac{\sum \bar{x}_2}{n_2} = 193.525$$

$$s_{x_1}^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = 3701.721$$

$$s_{x_2}^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = 2254.358$$

Use this information to calculate the test statistic

$$s_{\bar{x}_1, \bar{x}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{3701.721}{40} + \frac{2254.358}{40}}$$

$$s_{\bar{x}_1, \bar{x}_2} = \sqrt{92.5401 + 56.3589} = \sqrt{148.8991}$$

$$s_{\bar{x}_1, \bar{x}_2} = 12.2024$$

$$t_{obs} = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1, \bar{x}_2}} = \frac{199.65 - 193.525}{12.2024} = \frac{6.125}{12.2024}$$

$$t_{obs} = 0.5020$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{(n_1-1)} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{(n_2-1)}\right]} = \frac{\left(\frac{3701.721}{40} + \frac{2254.358}{40}\right)^2}{\left[\frac{\left(\frac{3701.721}{40}\right)^2}{(40-1)} + \frac{\left(\frac{2254.358}{40}\right)^2}{(40-1)}\right]}$$

$$df = \frac{(148.9019)^2}{[219.5952 + 81.4444]} = \frac{22171.7982}{301.0396} = 73.6508 \approx 74$$

Find the critical value for the t-distribution with alpha=0.05 and degrees of freedom of 74. Use either a table or a t-distribution calculator available on the web.

$$t_{\alpha=0.05, df=74} = -1.666$$

Step 6: Statistical Conclusion

Reject the null hypothesis if the observed test statistic is more extreme than the critical value. Since we are looking at a left tailed alternative hypothesis that we reject the null hypothesis if the observed test statistic is less than the critical value.

Since $t_{obs} = 0.5020 > -1.666 = t_{\alpha=0.05, df=74}$ then we fail to reject the null hypothesis.

Step 7: Experimental Conclusion

There is not sufficient evidence to indicate that the mean weight of Phoenix is different than the mean weight of LA at a level of significance of 0.05.