STAT 303 Lecture 24 Outline: p-Values and the Tests of Hypotheses Lab

H.J. Newton

March 18, 1996

1 p-Values

We have seen that the basic idea of a test of hypothesis is to compare the observed value of a test statistic to 'critical values' of the Z, t, χ^2, or F curve to see if the observed value is 'too extreme' to be consistent with the hypothesized value of the parameter.

An alternative to this procedure is to measure 'how extreme' the observed value of the test statistic actually is. How we do this depends on whether the test is right sided (the measure is the area to the *right* of the observed value), left sided (area to the *left*, or two sided (twice the smaller of the area to the right and the area to the left). This measure is called the p-value of the test. Note the following about a p-value:

1. Since it is a probability, it must be between 0 and 1.
2. We reject H_0 if the p-value is less than α.
3. The smaller the p-value, the more extreme is the observed value of the test statistic and thus the more evidence there is against the null hypothesis. This is the usual way of interpreting a p-value.
4. The p-value is the largest value of α for which we would not reject the null hypothesis.

The formulas for finding the p-values for all of our one and two sample situations are in the new table at the end of this handout. See Section 3 below for examples of finding and interpreting p-values.

2 The Testing Hypotheses Lab

In the last week, I have written a new lab which will do the tests of hypotheses and p-values for all the one and two sample situations we cover in STAT 303. The lab is very similar to the one that calculates confidence intervals. The dialog box for the lab is given below. Again you choose which case you want to use and then supply the information asked for by that case. The dialog box below has the information required for the first example.

Note that the lab displays information both in the results window and in the graphics window.

3 Some Examples

In this section we consider three examples, the last of which has two hypothesis tests.

3.1 Example 1

In a study of the keyboard height that typists prefer, a sample of 15 trained typists was selected and their prefered heights determined. The resulting average height was 80 cm. Assuming that the population of all typists prefered heights is normally distributed with standard deviation 2 cm, is this enough evidence to conclude that the average prefered height is more than 79 cm?

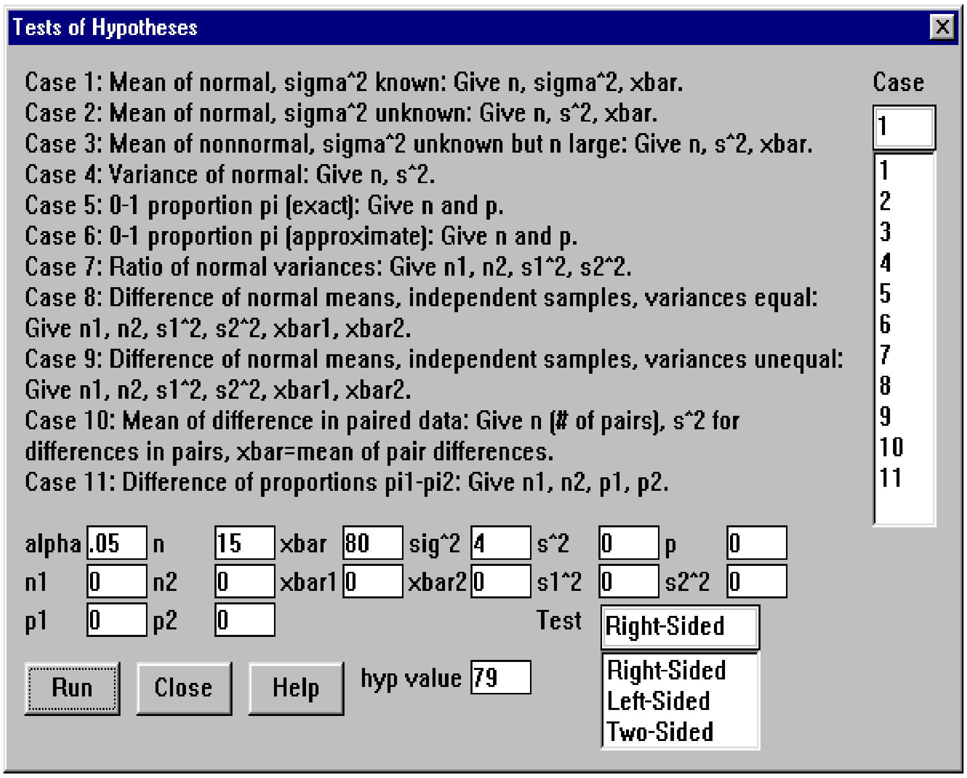


Figure 1: The Test of Hypotheses Dialog Box

The Analysis of Example 1

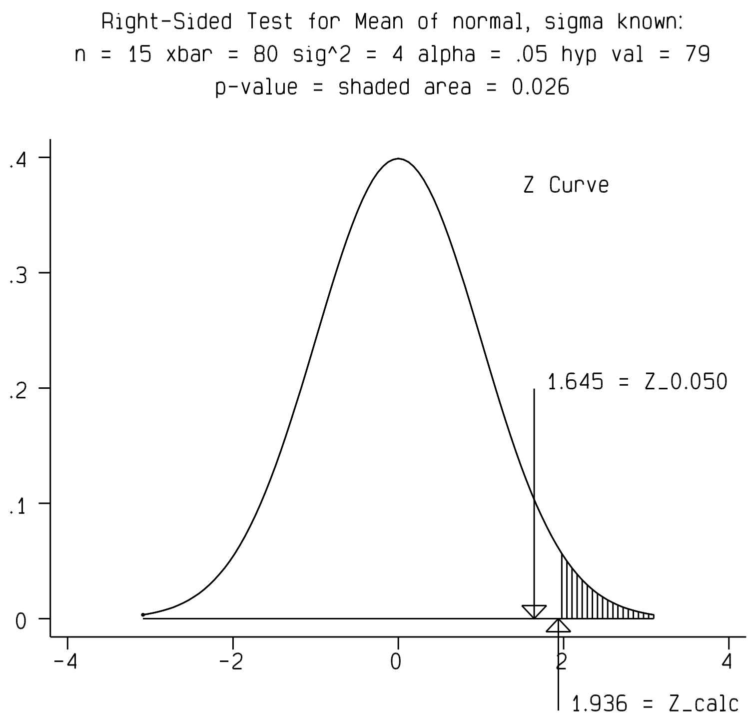
Since $Z_C = 1.936 > 1.645 = Z_{.05}$, we reject $H_0 : \mu \leq 79$ at the $\alpha = .05$ level of significance and conclude that there is enough evidence to say the average prefered height is more than 79. The graph makes this very clear since the calculated Z is to the right of the critical value. Note that the p-value is .026 (the shaded area to the right of the calculated value of Z) so this is strong but not overwhelming evidence against H_0 (for example, we would not reject H_0 for $\alpha = .025$).

The Results Window for Example 1

```
Right-Sided Test for Mean of normal, sigma known:
n = 15, sigma^2 = 4, xbar = 80
alpha = .05
Hypothesized value = 79

Z_calc = 1.9364917
Critical value: 1.6448536
Reject H_0
p-value = .02640376
```

The Graphics Window for Example 1



3.2 Example 2

A ball bearing manufacturer claims he has developed a new process that leads to a reduction in the variation in the diameters in the ball bearings his plant produces. The current variance is 10. A random sample of 50 ball bearings is obtained using the new process and for these it is found that $s = 3$. Does this justify his claim?

The Analysis of Example 2

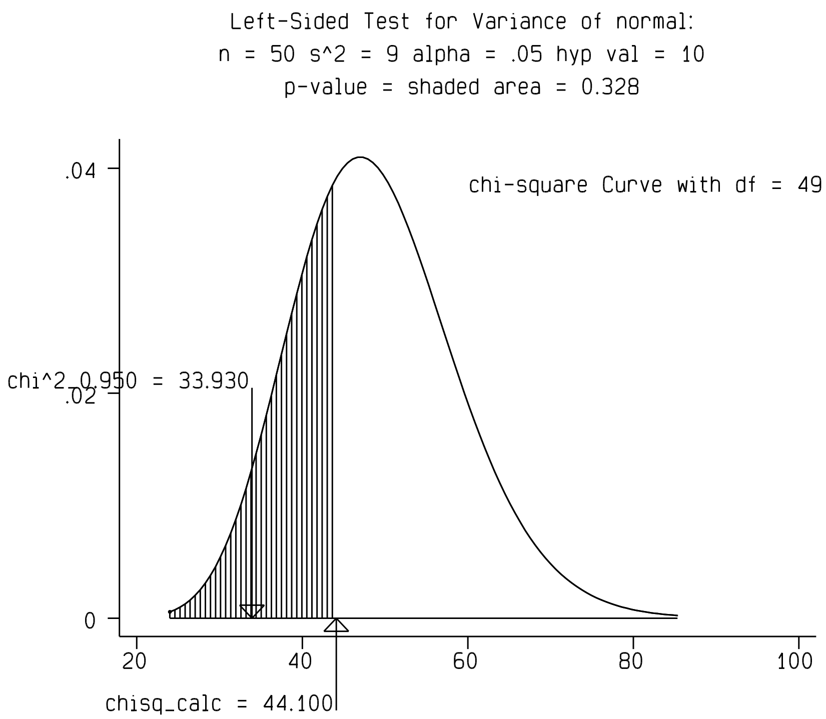
Now we illustrate a left sided test since the claim is that $\sigma^2 < 10$. Since the parameter of interest is σ^2, the χ^2 statistic is used. The calculated value is $\chi^2_C = 44.1$ which is not less than the critical value $\chi^2_{.95,49} = 33.930$ so we can't reject H_0. Note that now the p-value is the area to the *left* of χ^2_C (since we're doing a left sided test) and is equal to .328, which shows there is in fact little evidence against H_0.

The Results Window for Example 2

```
Left-Sided Test for Variance of normal:
n = 50, s^2 = 9
alpha = .05
Hypothesized value = 10

chisquare_calc= 44.1
Critical value: 33.929901
Fail to reject H_0
p-value = .32833193
```

The Graphics Window for Example 2



3.3 Example 3

To compare two teaching methods, a school found 45 students who were as similar in backgrounds as they could and randomly selected 25 students to receive the first method and 20 students for the second. At the end of the school year, the same test was administered to both classes and the results recorded as follows: $\bar{X}_1 = 75$, $\bar{X}_2 = 79$, $s_1^2 = 9$, and $s_2^2 = 16$.

a. Assuming that the populations of test scores are normally distributed, is it reasonable to conclude that the variability in the two populations is the same?

b. Based on your result in part (a), test the hypothesis that the two testing methods lead to the same average test score.

The Analysis for Example 3 (a)

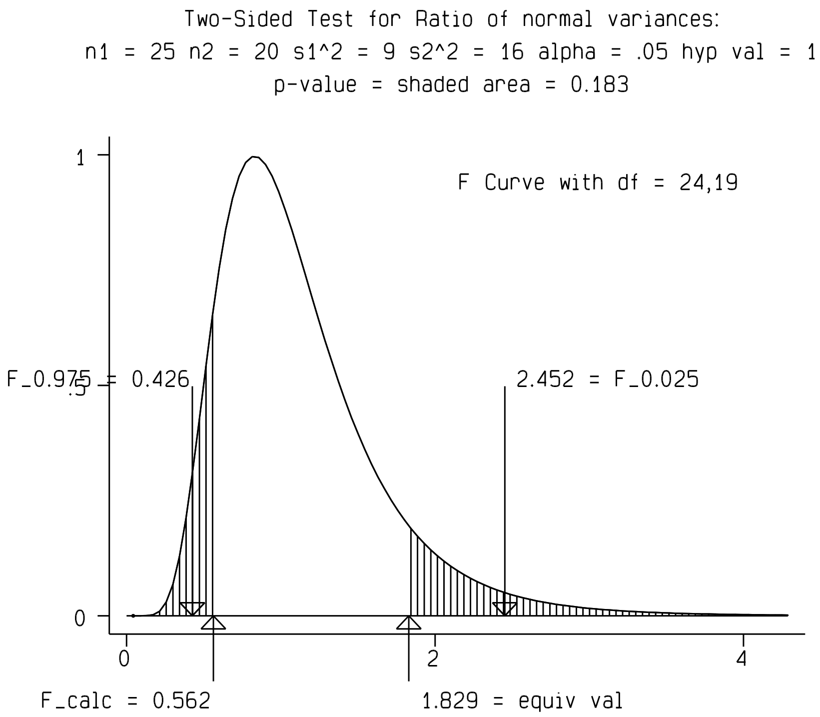
Now we have the two sided version of Case 7, that is, if we let σ_1^2 and σ_2^2 be the variances of the two populations, then we have $H_0 : \sigma_1^2 = \sigma_2^2$, with the alternative being that the two variances are unequal. There are now two critical values, namely $F_{.025,24,19} = 2.452$ and $F_{.975,24,19} = 0.426$ and the calculated value is $F_C = 0.562$ which is not more extreme (further from 1) than either critical value. Thus we don't reject H_0 and so we conclude it is reasonable to assume the two variances are the same. Note that the p-value is 0.183 and is made up of the sum of the two shaded areas.

The Results Window for Example 3 (a)

```
Two-Sided Test for Ratio of normal variances:
n1 = 25, n2 = 20, s1^2 = 9, s2^2 = 16
alpha = .05
Hypothesized value = 1
```

```
F_calc = .5625
Critical values: .42641163 , 2.4523139
Fail to reject H_0
p-value = .18282865
```

The Graphics Window for Example 3 (a)



The Analysis for Example 3 (b)

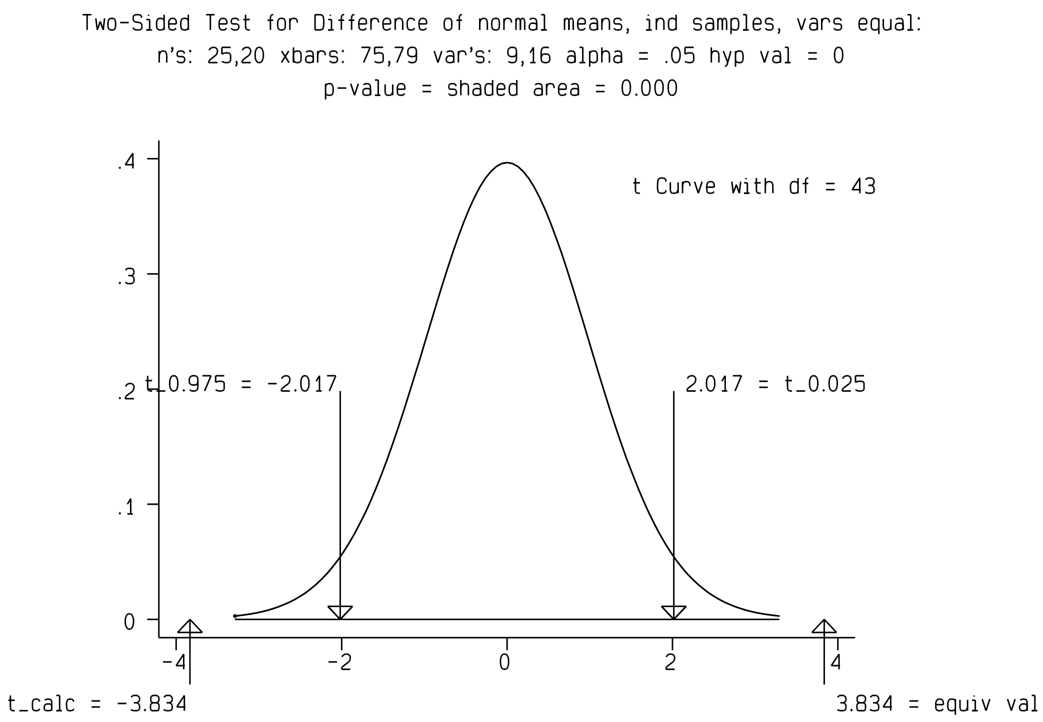
Because we concluded that the variances were not different in the previous example, we now have Case 8, and it is two sided because the question only asks if the two average test scores are the same. The graph shows that for these sample sizes and variances, the means $\bar{X}_1 = 75$ and $\bar{X}_2 = 79$ are very significantly different. The test statistic is $t_C = -3.834$ while the critical values are ± 2.017. The location of the arrows on the graph emphasize how extreme the difference in means is as does the fact that to three decimal places, the p-value is zero (in the results window you can see it is actually 0.0004).

The Results Window for Example 3 (b)

```
Two-Sided Test for Difference of normal means, ind samples, vars equal:
n1 = 25, n2 = 20, xbar1 = 75, xbar2 = 79
s1^2 = 9, s2^2 = 16, sp^2 = 12.093023, df = 43
alpha = .05
Hypothesized value = 0

t_calc = -3.8341694
Critical values: -2.0166922 , 2.0166922
Reject H_0
p-value = .0004064
```

The Graphics Window for Example 3 (b)



Confidence Intervals and Tests of Hypotheses for One and Two Samples

Situation				
Test Statistic	H_0	H_a	Rej H_0 if	p-value
Confidence Interval				

Case 1: Normal mean μ (σ^2 known) (`1 sample normal`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$Z_C = \dfrac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$	$\mu \le \mu_0$	$\mu > \mu_0$	$Z_C > Z_\alpha$	$\Pr(Z > Z_C)$
	$\mu \ge \mu_0$	$\mu < \mu_0$	$Z_C < -Z_\alpha$	$\Pr(Z < Z_C)$
	$\mu = \mu_0$	$\mu \ne \mu_0$	$\lvert Z_C\rvert > Z_{\alpha/2}$	$2\Pr(Z > \lvert Z_C\rvert)$

$$\bar{X} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Case 2: Normal mean μ (σ^2 unknown) (`1 sample t`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$t_C = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}}$	$\mu \le \mu_0$	$\mu > \mu_0$	$t_C > t_{\alpha,\nu}$	$\Pr(t > t_C)$
	$\mu \ge \mu_0$	$\mu < \mu_0$	$t_C < -t_{\alpha,\nu}$	$\Pr(t < t_C)$
	$\mu = \mu_0$	$\mu \ne \mu_0$	$\lvert t_C\rvert > t_{\alpha/2,\nu}$	$2\Pr(t > \lvert t_C\rvert)$

$$\bar{X} \pm t_{\alpha/2,\nu}\frac{s}{\sqrt{n}}, \qquad \nu = n-1$$

Case 3: Nonnormal mean μ (σ^2 unknown, n 'large') (`1 sample t`): Formulas are the same as Case 2

Case 4: Normal variance σ^2 (`1 sample variance`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$\chi^2_C = \dfrac{(n-1)s^2}{\sigma_0^2}$	$\sigma^2 \le \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\chi^2_C > \chi^2_{\alpha,\nu}$	$\Pr(\chi^2 > \chi^2_C)$
	$\sigma^2 \ge \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\chi^2_C < \chi^2_{1-\alpha,\nu}$	$\Pr(\chi^2 < \chi^2_C)$
	$\sigma^2 = \sigma_0^2$	$\sigma^2 \ne \sigma_0^2$	$\chi^2_C > \chi^2_{\alpha/2,\nu}$ or $\chi^2_C < \chi^2_{1-\alpha/2,\nu}$	$2\min(\Pr(\chi^2 > \chi^2_C),$ $1 - \Pr(\chi^2 > \chi^2_C))$

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,\nu}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,\nu}}\right), \qquad \nu = n-1$$

Case 5: Proportion π (exact CI for small n)

$$\left(\frac{XF_{1-\alpha/2,\nu_1,\nu_2}}{(n-X+1) + XF_{1-\alpha/2,\nu_1,\nu_2}}, \frac{(X+1)F_{\alpha/2,\nu_3,\nu_4}}{(n-X) + (X+1)F_{\alpha/2,\nu_3,\nu_4}}\right), \qquad X = np, \qquad \nu_1 = 2X,$$

$$\nu_2 = 2(n-X+1), \quad \nu_3 = 2(X+1), \quad \nu_4 = 2(n-X)$$

Case 6: Proportion π ($\min(n\pi, n(1-\pi)) \ge 5$) (`1 sample proportion`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$Z_C = \dfrac{p - \pi_0}{\sqrt{\pi_0(1-\pi_0)/n}}$	$\pi \le \pi_0$	$\pi > \pi_0$	$Z_C > Z_\alpha$	$\Pr(Z > Z_C)$
	$\pi \ge \pi_0$	$\pi < \pi_0$	$Z_C < -Z_\alpha$	$\Pr(Z < Z_C)$
	$\pi = \pi_0$	$\pi \ne \pi_0$	$\lvert Z_C\rvert > Z_{\alpha/2}$	$2\Pr(Z > \lvert Z_C\rvert)$

$$p \pm Z_{\alpha/2}\sqrt{p(1-p)/n}$$

Confidence Intervals and Tests of Hypotheses for One and Two Samples (cont)

Situation

Test Statistic	H_0	H_a	Rej H_0 if	p-value

Confidence Interval

Case 7: Ratio of normal variances (independent samples) (`2 sample variances`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$F_C = \dfrac{s_1^2/s_2^2}{\delta}$	$\sigma_1^2/\sigma_2^2 \leq \delta$	$\sigma_1^2/\sigma_2^2 > \delta$	$F_C > F_{\alpha,\nu_1,\nu_2}$	$\Pr(F > F_C)$
	$\sigma_1^2/\sigma_2^2 \geq \delta$	$\sigma_1^2/\sigma_2^2 < \delta$	$F_C < F_{1-\alpha,\nu_1,\nu_2}$	$\Pr(F < F_C)$
	$\sigma_1^2/\sigma_2^2 = \delta$	$\sigma_1^2/\sigma_2^2 \neq \delta$	$F_C > F_{\alpha/2,\nu_1,\nu_2}$ or $F_C < F_{1-\alpha/2,\nu_1,\nu_2}$	$2\min(\Pr(F > F_C), 1 - \Pr(F > F_C))$

$$\left(\frac{s_1^2/s_2^2}{F_{\alpha/2,\nu_1,\nu_2}}, \frac{s_1^2/s_2^2}{F_{1-\alpha/2,\nu_1,\nu_2}}\right), \qquad \nu_1 = n_1 - 1, \qquad \nu_2 = n_2 - 1$$

Case 8: Difference of two normal means (σ_1^2, σ_2^2 unknown but equal, independent samples) (`2 sample t`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$t_C = \dfrac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{s_P^2(1/n_1 + 1/n_2)}}$	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$t_C > t_{\alpha,\nu}$	$\Pr(t > t_C)$
	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$t_C < -t_{\alpha,\nu}$	$\Pr(t < t_C)$
	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 \neq \delta$	$\lvert t_C \rvert > t_{\alpha/2,\nu}$	$2\Pr(t > \lvert t_C \rvert)$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\nu}\sqrt{s_P^2(1/n_1 + 1/n_2)}, \qquad s_P^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}, \qquad \nu = n_1 + n_2 - 2$$

Case 9: Difference of two normal means (σ_1^2, σ_2^2 unknown, unequal, independent samples) (`2 sample t`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$t_C = \dfrac{(\bar{X}_1 - \bar{X}_2) - \delta}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$	$\mu_1 - \mu_2 \leq \delta$	$\mu_1 - \mu_2 > \delta$	$t_C > t_{\alpha,\nu}$	$\Pr(t > t_C)$
	$\mu_1 - \mu_2 \geq \delta$	$\mu_1 - \mu_2 < \delta$	$t_C < -t_{\alpha,\nu}$	$\Pr(t < t_C)$
	$\mu_1 - \mu_2 = \delta$	$\mu_1 - \mu_2 \neq \delta$	$t_C > \lvert t_{\alpha/2,\nu} \rvert$	$2\Pr(t > \lvert t_C \rvert)$

$$(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2,\nu}\sqrt{s_1^2/n_1 + s_2^2/n_2}, \qquad \nu = \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{\left(s_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(s_2^2/n_2\right)^2}{n_2 - 1}}$$

Case 10: Mean of differences in normal paired data (σ^2 unknown) (`paired t`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$t_C = \dfrac{\bar{d} - \mu_0}{s_d/\sqrt{n}}$	$\mu_d \leq \mu_0$	$\mu_d > \mu_0$	$t_C > t_{\alpha,\nu}$	$\Pr(t > t_C)$
	$\mu_d \geq \mu_0$	$\mu_d < \mu_0$	$t_C < -t_{\alpha,\nu}$	$\Pr(t < t_C)$
	$\mu_d = \mu_0$	$\mu_d \neq \mu_0$	$\lvert t_C >> t_{\alpha/2,\nu}$	$2\Pr(t > \lvert t_C \rvert)$

$\bar{d} \pm t_{\alpha/2,\nu}\dfrac{s_d}{\sqrt{n}}$, $\qquad \nu = n - 1$, $\bar{d}$ and s_d are mean and standard deviation of differences in pairs.

Case 11: Difference $\pi_1 - \pi_2$ of proportions (independent samples, $\min(n_1\pi_1, n_1(1-\pi_1)) \geq 5$, $\min(n_2\pi_2, n_2(1-\pi_2)) \geq 5$) (`2 sample proportions`)

Test Statistic	H_0	H_a	Rej H_0 if	p-value
$Z_C = \dfrac{p_1 - p_2}{\sqrt{p(1-p)(1/n_1 + 1/n_2)}}$	$\pi_1 - \pi_2 \leq 0$	$\pi_1 - \pi_2 > 0$	$Z_C > Z_\alpha$	$\Pr(Z > Z_C)$
	$\pi_1 - \pi_2 \geq 0$	$\pi_1 - \pi_2 < 0$	$Z_C < -Z_\alpha$	$\Pr(Z < Z_C)$
	$\pi_1 - \pi_2 = 0$	$\pi_1 - \pi_2 \neq 0$	$\lvert Z_C \rvert > Z_{\alpha/2}$	$2\Pr(Z > \lvert Z_C \rvert)$

$$(p_1 - p_2) \pm Z_{\alpha/2}\sqrt{\frac{p_1(1 - p_1)}{n_1} + \frac{p_2(1 - p_2)}{n_2}}, \qquad p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$