Two-Sample Hypothesis Test Examples

(Chapter 11)

1. In a test of the reliability of products produced by two machines, machine A produced 15 defective parts in a run of 280, while machine B produced 10 defective parts in a run of 200. Do these results imply a difference in the reliability of these two machines? (Use $\alpha = 0.01$.)

Step 0: **Check Assumptions**

$n_A p_A = y_A = 15 \geq 10$ and $n_A(1 - p_A) = n_A - y_A = 265 \geq 10$

$n_B p_B = y_B = 10 \geq 10$ and $n_B(1 - p_B) = n_B - y_B = 190 \geq 10$

Step 1: **Hypotheses**

H_0: $\pi_A - \pi_B = 0$

H_a: $\pi_A - \pi_B \neq 0$

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value(s) and Rejection Region(s)**

Critical Value: $\pm z_{\alpha} = \pm z_{0.005} = \pm 2.58$

Reject the null hypothesis if $Z \leq -2.58$ or if $Z \geq 2.58$.

Step 4: **Test Statistic**

$$p_c = \frac{y_A + y_B}{n_A + n_B} = \frac{15 + 10}{280 + 200} = \frac{25}{480}$$

$$Z = \frac{(p_A - p_B) - \delta_0}{\sqrt{p_c(1 - p_c)\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} = \frac{\left(\frac{15}{280} - \frac{10}{200}\right) - 0}{\sqrt{\left(\frac{25}{480}\right)\left(\frac{455}{480}\right)\left(\frac{1}{280} + \frac{1}{200}\right)}} = 0.1736$$

$$p\text{-value} = 2 * P(z \geq 0.1736) \approx 2 * P(z \geq 0.17) = 2 * 0.4325 = 0.8650$$

Step 5: **Conclusion**

Since $-2.58 \leq 0.1736 \leq 2.58$ (p-value $\approx 0.8650 > 0.01 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.01$ level of significance, there is not enough evidence to conclude that there is a difference in the reliability of the two machines.

2. Two sections of a class in statistics were taught by two different methods. Students' scores on a standardized test are shown below. Do the results present evidence of a difference in the effectiveness of the two methods? (Use $\alpha = 0.01$.)

Class A		Class B	
74	76	78	79
97	75	92	76
79	82	94	93
88	86	78	82
78	100	71	69
93	94	85	84
		70	

Step 1: **Hypotheses**

$H_0:\ \mu_A - \mu_B = 0$

$H_a:\ \mu_A - \mu_B \neq 0$

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we don't know the population variances (σ_A^2 and σ_B^2) and don't think that they are equal, we'll use the non-pooled t-test.

$$s_A^2 = \frac{\left(\sum y_A^2 - \frac{\left(\sum y_A\right)^2}{n_A}\right)}{n_A - 1} = \frac{\left(87960 - \frac{(1022)^2}{12}\right)}{12-1} = 83.6061$$

$$s_B^2 = \frac{\left(\sum y_B^2 - \frac{\left(\sum y_B\right)^2}{n_B}\right)}{n_B - 1} = \frac{\left(85841 - \frac{(1051)^2}{13}\right)}{13-1} = 72.6410$$

$$df = \upsilon = \frac{\left[\left(s_A^2/n_A\right) + \left(s_B^2/n_B\right)\right]^2}{\frac{\left(s_A^2/n_A\right)^2}{n_A - 1} + \frac{\left(s_B^2/n_B\right)^2}{n_B - 1}} = \frac{\left[(83.6061/12) + (72.6410/13)\right]^2}{\frac{(83.6061/12)^2}{12-1} + \frac{(72.6410/13)^2}{13-1}} = 22.0147 \rightarrow 22$$

Critical Values: $\pm t_{\alpha/2, df=\upsilon} = \pm t_{0.005, df=22} = \pm 2.82$

Reject the null hypothesis if $T \le -2.82$ or if $T \ge 2.82$.

Step 4: **Test Statistic**

$$\bar{y}_A = \frac{\sum y_A}{n_A} = \frac{1022}{12} = 85.1667 \qquad \bar{y}_B = \frac{\sum y_B}{n_B} = \frac{1051}{13} = 80.8462$$

$$T = \frac{(\bar{y}_A - \bar{y}_B) - \delta_0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{(85.1667 - 80.8462) - 0}{\sqrt{\frac{83.6061}{12} + \frac{72.6410}{13}}} = 1.2193$$

$$p\text{-value} = 2 * P(t \ge 1.2193) \approx 2 * P(t \ge 1.2) = 2 * 0.121 = 0.242$$

Step 5: **Conclusion**

Since $-2.82 \le 1.2193 \le 2.82$ (p-value $\approx 0.242 > 0.01 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.01$ level of significance, there is not enough evidence to conclude that there is a difference in the effectiveness of the two methods.

3. The table below shows the observed pollution indexes of air samples in two areas of a city. Test the hypothesis that the mean pollution indexes are the same for the two areas. (Use $\alpha = 0.05$.)

Area A		Area B	
2.92	4.69	1.84	3.44
1.88	4.86	0.95	3.69
5.35	5.81	4.26	4.95
3.81	5.55	3.18	4.47

Step 1: **Hypotheses**

$H_0: \mu_A - \mu_B = 0$

$H_a: \mu_A - \mu_B \neq 0$

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we don't know the population variances (σ_A^2 and σ_B^2) but think that they are not equal (air varies across different areas of the same city due to industrialization, vegetation, etc.), we'll use the non-pooled *t*-test.

$$s_A^2 = \frac{\left(\sum y_A^2 - \frac{\left(\sum y_A\right)^2}{n_A}\right)}{n_A - 1} = \frac{\left(165.3737 - \frac{(34.87)^2}{8}\right)}{8-1} = 1.9120$$

$$s_B^2 = \frac{\left(\sum y_B^2 - \frac{\left(\sum y_B\right)^2}{n_B}\right)}{n_B - 1} = \frac{\left(102.4812 - \frac{(26.78)^2}{8}\right)}{8-1} = 1.8336$$

$$df = \upsilon = \frac{\left[\left(s_A^2/n_A\right) + \left(s_B^2/n_B\right)\right]^2}{\frac{\left(s_A^2/n_A\right)^2}{n_A - 1} + \frac{\left(s_B^2/n_B\right)^2}{n_B - 1}} = \frac{\left[(1.9120/8) + (1.8336/8)\right]^2}{\frac{(1.9120/8)^2}{8-1} + \frac{(1.8336/8)^2}{8-1}} = 13.9939 \to 13$$

Critical Values: $\pm t_{\alpha/2, df=\upsilon} = \pm t_{0.025, df=13} = \pm 2.16$

Reject the null hypothesis if $T \le -2.16$ or if $T \ge 2.16$.

Step 4: **Test Statistic**

$$\bar{y}_A = \frac{\sum y_A}{n_A} = \frac{34.87}{8} = 4.3588 \qquad \bar{y}_B = \frac{\sum y_B}{n_B} = \frac{26.78}{8} = 3.3475$$

$$T = \frac{(\bar{y}_A - \bar{y}_B) - \delta_0}{\sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}} = \frac{(4.3588 - 3.3475) - 0}{\sqrt{\frac{1.9120}{8} + \frac{1.8336}{8}}} = 1.4780$$

$$p\text{-value} = 2 * P(t \ge 1.4780) \approx 2 * P(t \ge 1.5) = 2 * 0.079 = 0.158$$

Step 5: **Conclusion**

Since $-2.16 \le 1.4780 \le 2.16$ (*p*-value $\approx 0.158 > 0.05 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there is not enough evidence to conclude that there is a difference in the mean pollution indexes for the two areas.

4. A closer examination of the records of the air samples in Example 3 reveals that each line of the data actually represents readings on the same day: 2.92 and 1.84 are from day 1, and so forth. Since this affects the validity of the results obtained in Example 10, reanalyze. (Use $\alpha = 0.05$.)

Area A	Area B	$y_d = A - B$
2.92	1.84	1.08
1.88	0.95	0.93
5.35	4.26	1.09
3.81	3.18	0.63
4.69	3.44	1.25
4.86	3.69	1.17
5.81	4.95	0.86
5.55	4.47	1.08

Step 1: **Hypotheses**

H_0: $\mu_d = 0$

H_a: $\mu_d \neq 0$

Step 2: **Significance Level**

$\alpha = 0.05$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we have paired data and don't know the population variance of the differences (σ_d^2), we'll use the paired *t*-test.

Critical Values: $\pm t_{\alpha/2, df = n_d - 1} = \pm t_{0.025, df = 7} = \pm 2.37$

Reject the null hypothesis if $T \leq -2.37$ or if $T \geq 2.37$.

Step 4: **Test Statistic**

$$\bar{y}_d = \frac{\sum y_d}{n_d} = \frac{8.09}{8} = 1.0113 \qquad s_d = \sqrt{\frac{\sum y_d^2 - \frac{\left(\sum y_d\right)^2}{n_d}}{n_d - 1}} = \sqrt{\frac{8.45 - \frac{(8.09)^2}{8}}{7}} = 0.1960$$

$$T = \frac{\bar{y}_d - \delta_0}{s_d / \sqrt{n_d}} = \frac{1.0113 - 0}{0.1960 / \sqrt{8}} = 14.5938$$

$$p\text{-value} = 2 * P(t \geq 14.5938) \approx 2 * P(t \geq 4.0) = 2 * 0.003 = 0.006$$

Step 5: **Conclusion**

Since $14.5938 \geq 2.37$ (p-value $\approx 0.006 \leq 0.05 = \alpha$), we shall reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.05$ level of significance, there exists enough evidence to conclude that there is a difference in the mean pollution indexes for the two areas.

5. Eight quantities of effluent from a pulp mill were each divided into ten batches. From each quantity, five randomly selected batches were subjected to a treatment process intended to remove toxic substances. Five fish of the same species were placed in each batch, and the mean number surviving in the five treated and untreated portions of each effluent quantity after five days were recorded and are given below. Test to see if the treatment increased the mean number of surviving fish. (Use $\alpha = 0.01$.)

Quantity No.	1	2	3	4	5	6	7	8
				Mean Number Surviving				
Untreated	5	1	1.8	1	3.6	5	2.6	1
Treated	5	5	1.2	4.8	5	5	4.4	2
$y_d = U - T$	0	-4	0.6	-3.8	-1.4	0	-1.8	-1

Step 1: **Hypotheses**

H_0: $\mu_d = 0$

H_a: $\mu_d < 0$

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we have paired data and don't know the population variance of the differences (σ_d^2), we'll use the paired *t*-test.

Critical Value: $-t_{\alpha, df = n_d - 1} = -t_{0.01, df = 7} = -3.00$

Reject the null hypothesis if $T \le -3.00$.

Step 4: **Test Statistic**

$$\bar{y}_d = \frac{\sum y_d}{n_d} = \frac{-11.4}{8} = -1.425$$

$$s_d = \sqrt{\frac{\sum y_d^2 - \frac{\left(\sum y_d\right)^2}{n_d}}{n_d - 1}} = \sqrt{\frac{37 - \frac{(-11.4)^2}{8}}{7}} = 1.7219$$

$$T = \frac{\bar{y}_d - \delta_0}{s_d / \sqrt{n_d}} = \frac{-1.425 - 0}{1.7219/\sqrt{8}} = -2.3407$$

$$p\text{-value} = P(t \le -2.3407) = P(t \ge 2.3407) \approx P(t \ge 2.3) = 0.027$$

Step 5: **Conclusion**

Since $-2.3407 > -3.00$ (*p*-value $\approx 0.027 > 0.01 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.01$ level of significance, there is not enough evidence to conclude that the treatment increased the mean number of surviving fish.

6. In a test of the effectiveness of a device that is supposed to increase gasoline mileage in automobiles, 12 cars were run, in random order, over a prescribed course both with and without the device in random order. The mileages (mpg) are given below. Is there evidence that the device is effective? (Use $\alpha = 0.01$.)

Car	Without Device	With Device	y_d = With − Without
1	21.0	20.6	−0.4
2	30.0	29.9	−0.1
3	29.8	30.7	0.9
4	27.3	26.5	−0.8
5	27.7	26.7	−1.0
6	33.1	32.8	−0.3
7	18.8	21.7	2.9
8	26.2	28.2	2.0
9	28.0	28.9	0.9
10	18.9	19.9	1.0
11	29.3	32.4	3.1
12	21.0	22.0	1.0

Step 1: **Hypotheses**

H_0: $\mu_d = 0$

H_a: $\mu_d > 0$

Step 2: **Significance Level**

$\alpha = 0.01$

Step 3: **Critical Value(s) and Rejection Region(s)**

Since we have paired data and don't know the population variance of the differences (σ_d^2), we'll use the paired *t*-test.

Critical Value: $t_{\alpha, df = n_d - 1} = t_{0.01, df = 11} = 2.72$

Reject the null hypothesis if $T \geq 2.72$.

Step 4: **Test Statistic**

$$\bar{y}_d = \frac{\sum y_d}{n_d} = \frac{9.2}{12} = 0.7667 \quad s_d = \sqrt{\frac{\sum y_d^2 - \frac{\left(\sum y_d\right)^2}{n_d}}{n_d - 1}} = \sqrt{\frac{27.54 - \frac{(9.2)^2}{12}}{11}} = 1.3647$$

$$T = \frac{\bar{y}_d - \delta_0}{s_d / \sqrt{n_d}} = \frac{0.7667 - 0}{1.3647 / \sqrt{12}} = 1.9462$$

$$p\text{-value} = P(t \geq 1.9462) \approx P(t \geq 1.9) = 0.042$$

Step 5: **Conclusion**

Since $1.9462 < 2.72$ (p-value $\approx 0.042 > 0.01 = \alpha$), we fail to reject the null hypothesis.

Step 6: **State conclusion in words**

At the $\alpha = 0.01$ level of significance, there is not enough evidence to conclude that the device is effective for increasing gasoline mileage.