Solution *Section* 4.3 – Inferences About Two Means: Independent

Exercise

If the pulse rates of men and women shown in the data below

Women:

76	72	88	60	72	68	80	64	68	68	80	76	68	72	96	72	68	72	64	80
64	80	76	76	76	80	104	88	60	76	72	72	88	80	60	72	88	88	124	64

Men:

68	64	88	72	64	72	60	88	76	60	96	72	56	64	60	64	84	76	84	88
72	56	68	64	60	68	60	60	56	84	72	84	88	56	64	56	56	60	64	72

These data are used to construct 95% confidence interval for the difference between the two population means, the result is $-12.2 < \mu_1 - \mu_2 < -1.6$, where pulse rates of men correspond to population 1 and pulse rates of women correspond to population 2. Express the confidence interval with pulse rates of women being population 1 and pulse rates of men being population 2.

Solution

Reversing the designation of which sample is considered group 1 and which sample is considered group 2 changes the sign of the point estimate and the signs of the endpoints of the interval estimate. The confidence interval using the new designation is $1.6 < \mu_1 - \mu_2 < 12.2$

Exercise

Assume that you want to use a 0.01 significance level to test the claim that the mean pulse rate of men is less than the mean pulse rate of women. What confidence level should be used if you want to test that claim using a confidence interval?

Solution

A one-tailed test of hypothesis at the 0.01 level of significance corresponds to a two-sided confidence interval at the 2(0.01) = 0.02 level of significance –i.e., to an interval with a confidence level of 98%

Exercise

To test the effectiveness of Lipitor, cholesterol levels are measured in 250 subjects before and after Lipitor treatments. Determine whether this sample is independent or dependent.

Solution

Dependent, since cholesterol levels and determined by many factors that the Lipitor treatment cannot change. Treatments to lower cholesterol typically reduce everyone's levels by a certain amount, by persons who were high compared to the others before the treatment, for example, will likely still be high compared to the others after the treatment.

Exercise

On each of 40 different days, you measured the voltage supplied to your home and you also measured the voltage produced by the gasoline-powered generator. One sample consists of the voltages in the house and the second sample consists of the voltages produced by the generator. Determine whether this sample is independent or dependent.

Solution

Independent, since there is no relationship between the voltage supplied to the house by the power company and the voltage generated by a completely separate gasoline-powered generator.

Exercise

In a randomized controlled trial conducted with children suffering from viral croup, 46 children were treated with low humidity while 46 other children were treated with high humidity. Researchers used the Westley Croup Score to assess the results after one hour. The low humidity group had a mean score of 0.98 with standard deviation of 1.22 while the high humidity group had a mean score of 1.09 with standard deviation of 1.11.

a) Use a 0.05 significance level to test the claim that the two groups are from populations with the same mean. What does the result suggest about the common treatment of humidity?
Assume that the two samples are independent simple random samples selected from normally distributed populations.

b) Assume that $\sigma_1 = \sigma_2$, how are the results affected by this additional assumption?

Solution

a) *Original Claim*: $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

$\alpha = 0.05 \quad and \quad df = 45$

Degrees of Freedom	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
45	2.690	2.412	$\boxed{2.014}$	1.679	1.301

Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} = \pm 2.014$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

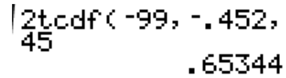
$$= \frac{(.98 - 1.09) - (0)}{\sqrt{\frac{1.22^2}{46} + \frac{1.11^2}{46}}}$$

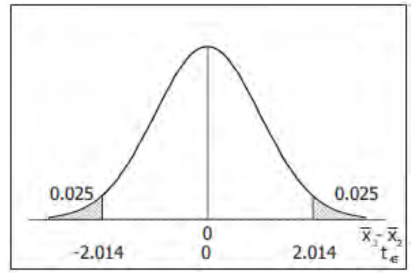
$$= -0.452$$

```
(.98-1.09)/√(1.2
2²/46+1.11²/46
           -.4523
```

$P-value = 2 \cdot tcdf\left(-99, -0.452, 45\right)$

$\underline{= 0.6534}$





Conclusion:

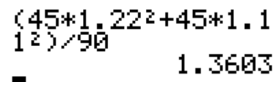
Do not reject H_0; there is not sufficient evidence to reject the claim that $\mu_1 - \mu_2 = 0$. There is not sufficient evidence to reject the claim that the two groups are from populations with the same mean. The results suggest that increasing the humidity does not have a significant effect on the treatment of croup.

b) $df = df_1 + df_2 = 45 + 45 = 90$

$$s_p^2 = \frac{\left(df_1\right)s_1^2 + \left(df_2\right)s_2^2}{df}$$

$$= \frac{(45)(1.22)^2 + (45)(1.11)^2}{90}$$

$\underline{= 1.3603}$

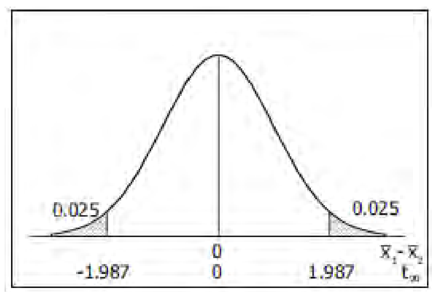


Original Claim: $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$ *and* $df = 90$

Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} \underline{= \pm 1.987}$



$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(.98 - 1.09) - (0)}{\sqrt{\frac{1.3603}{46} + \frac{1.3603}{46}}}$$

$\underline{= -0.452}$

$P-value = 2 \cdot tcdf\left(-99,\ -0.452,\ 90\right)$

$\underline{= 0.6521}$

Conclusion:

Do not reject H_0; there is not sufficient evidence to reject the claim that $\mu_1 - \mu_2 = 0$. There is not sufficient evidence to reject the claim that the two groups are from populations with the same mean. The results suggest that increasing the humidity does not have a significant effect on the treatment of croup.

When $n_1 - n_2 = 0$, the calculated *t* statistic does not change at all. The only difference the assumption of equal standard deviations makes in this instance is to change the *df* from 45 to 90 and the *P*-value from 0.6532 to 0.6521. The conclusion is unaffected.

Exercise

The mean tar content of a simple random sample of 25 unfiltered king size cigarettes is 21.1 mg, with a standard deviation of 3.2 mg. The mean tar content of a simple random sample of 25 filtered 100 mm cigarettes is 13.2 mg, with a standard deviation of 3.7 mg.

Assume that the two samples are independent simple random samples selected from normally distributed populations in part a and b.

a) Construct a 90% confidence interval estimate of the difference between the mean tar content of unfiltered king size cigarettes and the mean tar content of filtered 100 mm cigarettes. Does the result suggest that 100 mm filtered cigarettes have less tar than unfiltered king size cigarettes?

b) Use a 0.05 significance level to test the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. What does the result suggest about the effectiveness of cigarette filters?

c) Assume that $\sigma_1 = \sigma_2$, how are the results affected by this additional assumption?

Solution

a) Let the unfiltered cigarettes be group 1.

$\alpha = 0.1 \quad and \quad df = 24$

Critical value: $t = t_{0.05} = 1.711$

Degrees of Freedom	Area in One Tail 0.005	0.01	0.025	0.05	0.10
24	2.797	2.492	2.064	1.711	1.318

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(21.1 - 13.2) - 1.711\sqrt{\frac{3.2^2}{25} + \frac{3.7^2}{25}} < \mu_1 - \mu_2 < (21.1 - 13.2) + 1.711\sqrt{\frac{3.2^2}{25} + \frac{3.7^2}{25}}$$

$$6.2 < \mu_1 - \mu_2 < 9.6$$

Yes; since the confidence interval includes only positive values, the results suggest that the filtered cigarettes have less tar than the unfiltered ones.

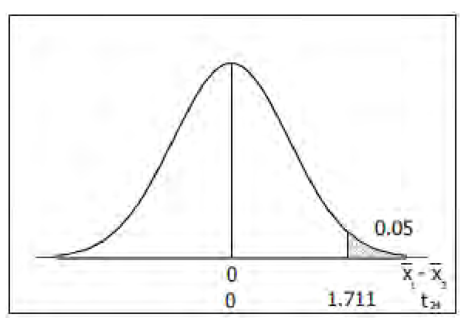
b) *Original Claim*: $\mu_1 - \mu_2 > 0$

$H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 > 0$

$\alpha = 0.1 \quad and \quad df = 24$

Critical value: $t = t_{0.05} = 1.711$



$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(21.1 - 13.2) - 0}{\sqrt{\frac{3.2^2}{25} + \frac{3.7^2}{25}}} = 8.075$$

$$P-value = tcdf\left(8.075,\ 99,\ 24\right)$$

$$\underline{=1.338E-8|} \qquad \approx 0.00000001$$

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $\mu_1 - \mu_2 > 0$. There is sufficient evidence to support the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. The results suggest that filters are effective in reducing the tar content cigarettes.

c) $df = df_1 + df_2 = 24 + 24 \ \underline{= 48|}$

$$s_p^2 = \frac{\left(df_1\right)s_1^2 + \left(df_2\right)s_2^2}{df}$$

$$= \frac{(24)(3.2)^2 + (24)(3.7)^2}{48}$$

$$\underline{= 11.965|}$$

$$\alpha = 0.10 \quad and \quad df = 48$$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(21.1 - 13.2) - 1.676\sqrt{\frac{11.965}{25} + \frac{11.965}{25}} < \mu_1 - \mu_2 < (21.1 - 13.2) + 1.676\sqrt{\frac{11.965}{25} + \frac{11.965}{25}}$$

$$\underline{6.3 < \mu_1 - \mu_2 < 9.5|}$$

Yes; since the confidence interval includes only positive values, the results suggest that the filtered cigarettes have less tar than the unfiltered ones.

When $n_1 = n_2$ the value of   is unchanged. The only difference the assumption of equal standard deviations makes in this instance is to change the *df* from 24 to 48 and the   from 1.711 to 1.676. This makes the interval slightly narrower, but the conclusion is unaffected.

Exercise

The heights are measured for the simple random sample of supermodels Crawford, Bundchen, Pestova, Christenson, Hume, Moss, Campbell, Schiffer, and Taylor. They have a mean of 70.0 in. and a standard deviation of 1.5 in. 40 women who are not supermodels, listed below and they have heights with means of 63.2 in. and a standard deviation of 2.7 in.

64.3	66.4	62.3	62.3	59.6	63.6	59.8	63.3	67.9	61.4	66.7	64.8	63.1	66.7	66.8
64.7	65.1	61.9	64.3	63.4	60.7	63.4	62.6	60.6	63.5	58.6	60.2	67.6	63.4	64.1
62.7	61.3	58.2	63.2	60.5	65.0	61.8	68.0	67.0	57.0					

a) Use a 0.01 significance level to test the claim that the mean height of supermodels is greater than the mean height of women who are not supermodels *b)* Construct a 98% confidence interval level for the difference between the mean height of supermodels and the mean height of women who are not supermodels. What does the result suggest about those two means?

Solution

a) Let the supermodels be group 1. For which $n_1 = 9$

Original Claim: $\mu_1 - \mu_2 > 0$ $(inches)$

$$H_0 : \mu_1 - \mu_2 = 0$$

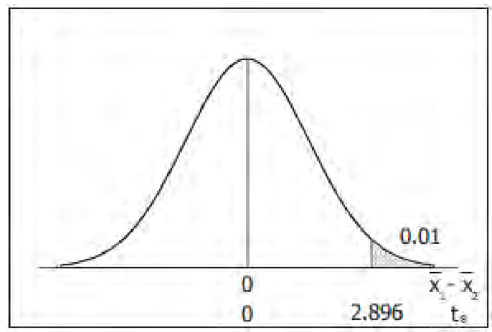
$$H_1 : \mu_1 - \mu_2 > 0$$

$\alpha = 0.01$ *and* $df = 8$

Critical value: $t = t_\alpha = t_{0.01} = 2.896$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{(70.0 - 63.2) - 0}{\sqrt{\frac{1.5^2}{9} + \frac{2.7^2}{40}}}$$

$= 10.343$



$P - value = tcdf\,(10.343,\ 99,\ 8)$

$= 3.29E - 6$ ≈ 0.000003

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $\mu_1 - \mu_2 > 0$. There is sufficient evidence to support the claim that the mean height of supermodels is greater than the mean height of women who are not supermodels.

b) Let the supermodels be group 1. For which $n_1 = 9$.

$\alpha = 1 - .98 = 0.2$ *and* $df = 8$

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(70 - 63.2) - 2.896\sqrt{\frac{1.5^2}{9} + \frac{2.7^2}{40}} < \mu_1 - \mu_2 < (70 - 63.2) + 2.896\sqrt{\frac{1.5^2}{9} + \frac{2.7^2}{40}}$$

$4.9 < \mu_1 - \mu_2 < 8.7$ $(inches)$

Since the confidence interval includes only positive values, the results suggest that the mean height of supermodels is greater than the mean height of women who are not supermodels.

Exercise

Many studies have been conducted to test the effects of marijuana use on mental abilities. In one such study, groups of light and heavy users of marijuana in college were tested for memory recall, with the results given below. Use a 0.01 significance level to test the claim that the population of heavy marijuana users has a lower mean than the light users. Should marijuana use be of concern to college students?

Items sorted correctly by light marijuana users: $n = 64, \quad \overline{x} = 53.3, \quad s = 3.6$

Items sorted correctly by heavy marijuana users: $n = 65, \quad \overline{x} = 51.3, \quad s = 4.5$

Solution

Original Claim: $\mu_1 - \mu_2 > 0$

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 > 0$$

$\alpha = 0.01 \quad and \quad df = 63$

Critical value: $t = t_\alpha = t_{0.01} = 2.390$

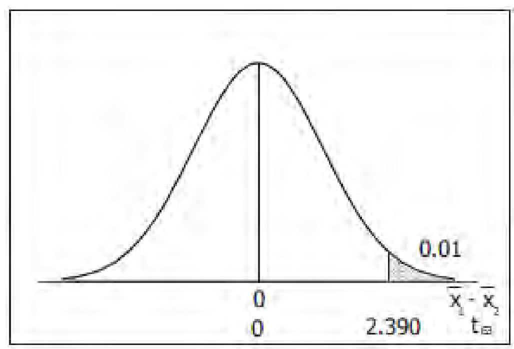
Degrees of Freedom	0.005	0.01	Area in One Tail 0.025	0.05	0.10
60	2.660	2.390	2.000	1.671	1.296

$$t = \frac{\left(\overline{x}_1 - \overline{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(53.3 - 51.3) - (0)}{\sqrt{\frac{3.6^2}{64} + \frac{4.5^2}{65}}}$$

$$= 2.790$$

```
(53.3-51.3)/√(3.
6²/64+4.5²/65
         2.78954
```



$P - value = tcdf\left(2.790,\ 99,\ 63\right)$

$$= 0.0035$$

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $\mu_1 - \mu_2 > 0$. There is sufficient evidence to support the claim that heavy marijuana users have a lower mean number of recalled items than do light users.

Yes; marijuana use should be of concern to college students – and an even more valuable study might one comparing light users to those who do not use marijuana at all.

Exercise

The trend of thinner Miss America winners has generated charges that the contest encourages unhealthy diet habits among young women. Listed below are body mass indexes (BMI) for Miss America winners from two different time periods. Consider the listed values to be simple random samples selected from larger populations.

a) Use a 0.05 significance level to test the claim that recent winners have a lower mean BMI than winners from the 1920s and 1930s.

b) Construct a 90% Confidence interval for the difference between the mean BMI of recent winners and the mean BMI of winners from the 1920s and 1930s.

BMI (from recent winners):	19.5	20.3	19.6	20.2	17.8	17.9	19.1	18.8	17.6	16.8
BMI (from 1920s and 1930s):	20.4	21.9	22.1	22.3	20.3	18.8	18.9	19.4	18.4	19.1

Solution

Group 1: recent ($n = 10$)

$$\bar{x}_1 = \frac{\sum x}{n_1} = \frac{187.6}{10} \underline{=18.76}; \quad s_1 = 1.186$$

```
1-Var Stats
x̄=18.76000
Σx=187.60000
Σx²=3532.04000
Sx=1.18622
σx=1.12534
↓n=10.00000
```

Group 2: 1920,1930 ($n = 10$)

$$\bar{x}_2 = \frac{\sum x}{n_2} = \frac{201.6}{10} \underline{=20.16}; \quad s_2 = 1.479$$

```
1-Var Stats
x̄=20.16000
Σx=201.60000
Σx²=4083.94000
Sx=1.47889
σx=1.40300
↓n=10.00000
```

a) *Original Claim*: $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 < 0$

$\alpha = 0.05 \quad and \quad df = 9$

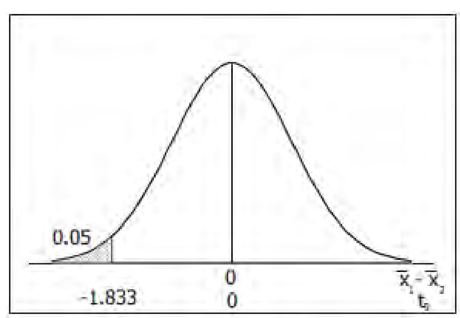
Critical value: $t = -t_\alpha = -t_{0.05} \underline{= -1.833}$

Degrees of Freedom	Area in One Tail 0.005	0.01	0.025	$\boxed{0.05}$	0.10
9	3.250	2.821	2.262	$\boxed{1.833}$	1.383

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(18.76 - 20.16) - (0)}{\sqrt{\frac{1.186^2}{10} + \frac{1.479^2}{10}}}$$

$$\underline{= -2.335}$$



$P-value = tcdf\left(-99, \ -2.335, \ 9\right) \underline{= 0.0222}$

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $\mu_1 - \mu_2 < 0$. There is sufficient evidence to support the claim that recent winners have a lower mean BMI than winners from the 1920s and 1930s.

b) $\alpha = 0.1$ *and* $df = 9$.

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(18.76 - 20.16) - 1.833\sqrt{\frac{1.186^2}{10} + \frac{1.479^2}{10}} < \mu_1 - \mu_2 < (18.76 - 20.16) + 1.833\sqrt{\frac{1.186^2}{10} + \frac{1.479^2}{10}}$$

$$\underline{-2.50 < \mu_1 - \mu_2 < -0.30|}$$

Exercise

Listed below are amounts of strontium-90 (in millibecquerels or mBq per gram of calcium) in a simple random sample of baby teeth obtained from Pennsylvania residents and New York residents born after 1979.

a) Use a 0.05 significance level to test the claim that the mean amount of strontium-90 from Pennsylvania residents is greater than the mean amount from New York residents.

b) Construct a 90% Confidence interval for the difference between the mean amount of strontium-90 from Pennsylvania residents and the mean amount from New York residents.

Pennsylvania:	155	142	149	130	151	163	151	142	156	133	138	161
New York:	133	140	142	131	134	129	128	140	140	140	137	143

Solution

Group 1: PA ($n = 12$)

$$\bar{x}_1 = \frac{\sum x}{n_1} = \frac{1771}{12} \underline{= 147.58|}$$

$s_1 = 10.64$

```
1-Var Stats
 x̄=147.58333
 Σx=1771.00000
 Σx²=262615.000
 Sx=10.63834
 σx=10.18543
↓n=12.00000
```

Group 2: NY ($n = 12$)

$$\bar{x}_2 = \frac{\sum x}{n_2} = \frac{1637}{12} \underline{= 136.42|}$$

$s_2 = 5.21$

```
1-Var Stats
 x̄=136.41667
 Σx=1637.00000
 Σx²=223613.000
 Sx=5.21289
 σx=4.99096
↓n=12.00000
```

a) *Original Claim*: $H_0 : \mu_1 - \mu_2 = 0$

$H_1 : \mu_1 - \mu_2 > 0$

$\alpha = 0.05$ *and* $df = 11$

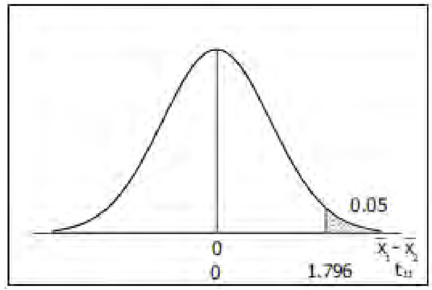
Critical value: $t = t_\alpha = t_{0.05} \underline{= 1.796|}$

$$t = \frac{\left(\bar{x}_1 - \bar{x}_2\right) - \left(\mu_1 - \mu_2\right)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(147.58 - 136.42) - (0)}{\sqrt{\frac{10.64^2}{12} + \frac{5.21^2}{12}}}$$

$$= 3.263$$

```
(147.58-136.42)/
√(10.64²/12+5.21
²/12
             3.26319
```




```
tcdf(3.265,99,11
               .0038
```

$P - value = tcdf\,(3.265,\ 99,\ 11)\ = 0.0038$

Conclusion:

Reject H_0; there is sufficient evidence to conclude that $\mu_1 - \mu_2 > 0$. There is sufficient evidence to support the claim that the mean amount of Strontium-90 from Pennsylvania residents is greater than the mean amount from N.Y. residents.

c) $\alpha = 0.1$ *and* $df = 11$.

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(147.58 - 136.42) - 1.796\sqrt{\frac{10.64^2}{12} + \frac{5.21^2}{12}} < \mu_1 - \mu_2 < (147.58 - 136.42) + 1.796\sqrt{\frac{10.64^2}{12} + \frac{5.21^2}{12}}$$

$$5.0 < \mu_1 - \mu_2 < 17.3 \qquad (mBq)$$

Exercise

Listed below are the word counts for male and female psychology students.

a) Use a 0.05 significance level to test the claim that male and female psychology students speak the same mean number of words in a day.

b) Construct a 95% Confidence interval estimate of the difference between the mean number of words spoken in a day by male and female psychology students. Do the confidence interval limits include 0, and what does that suggest about the two means?

Male	21143	17791	36571	6724	15430	11552	11748	12169	15581	23858	5269
	12384	11576	17707	15229	18160	22482	18626	1118	5319		

Female	6705	21613	11935	15790	17865	13035	24834	7747	3852	11648	25862
	17183	11010	11156	11351	25693	13383	19992	14926	14128	10345	13516
	12831	9671	17011	28575	23557	13656	8231	10601	8124		

Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.

Solution

Group 1: Males ($n = 20$)

$\bar{x}_1 = 15021.9$

$s_1 = 7863.87$

Group 2: Females ($n = 31$)

$\bar{x}_2 = 14704.1$

$s_2 = 6215.35$

a) *Original Claim*: $H_0 : \mu_1 - \mu_2 = 0$ ***words/day***

$H_1 : \mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$ *and* $df = 19$

Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} = \pm 2.093$

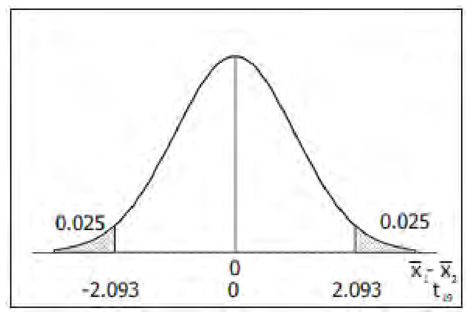
Degrees of Freedom	Area in Two Tails 0.01	0.02	0.05	0.10	0.20
19	2.861	2.539	$\boxed{2.093}$	1.729	1.328

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(15021.9 - 14704.1) - (0)}{\sqrt{\frac{7863.87^2}{20} + \frac{6215.35^2}{31}}}$$

$= 0.153$

```
(15021.9-14704.1
)/√(7863.87²/20+
6215.35²/31
                .1526
```




```
2tcdf(.1526,99,1
9
              .8803
```

$P - value = 2 \cdot tcdf\,(0.1526,\ 99,\ 19) = 0.8803$

Conclusion:

Do not reject H_0; there is not sufficient evidence to reject the claim $\mu_1 - \mu_2 = 0$. There is not sufficient evidence to reject the claim that the male and female students speak the same mean number of words per day.

d) $\alpha = 0.05$ *and* $df = 19$.

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$317.8 - 2.093\sqrt{\frac{7863.87^2}{20} + \frac{6215.35^2}{31}} < \mu_1 - \mu_2 < 317.8 + 2.093\sqrt{\frac{7863.87^2}{20} + \frac{6215.35^2}{31}}$$

$-4041.6 < \mu_1 - \mu_2 < 4677.1$ $(words/day)$

```
317.8-2.093√(786
3.87²/20+6215.35
²/31
        -4041.5578
```

Yes; since the confidence interval includes zero, there does not appear to be significant difference between the mean number of words spoken by the male and female students.

Exercise

Refer to the tables below and test the claim that they contain the same amount of cola, the mean weight of cola cans of regular Coke is the same as the mean weight of cola in cans of Diet Coke. If there is a difference in the mean weights, identify the most likely explanation for that difference.

Coke	0.8192	0.815	0.8163	0.8211	0.8181	0.8247	0.8062	0.8128	0.8172	0.811
	0.8251	0.8264	0.7901	0.8244	0.8073	0.8079	0.8044	0.817	0.8161	0.8194
	0.8189	0.8194	0.8176	0.8284	0.8165	0.8143	0.8229	0.815	0.8152	0.8244
	0.8207	0.8152	0.8126	0.8295	0.8161	0.8192				

Diet	0.7773	0.7758	0.7896	0.7868	0.7844	0.7861	0.7806	0.783	0.7852	0.7879
	0.7881	0.7826	0.7923	0.7852	0.7872	0.7813	0.7885	0.776	0.7822	0.7874
	0.7822	0.7839	0.7802	0.7892	0.7874	0.7907	0.7771	0.787	0.7833	0.7822
	0.7837	0.791	0.7879	0.7923	0.7859	0.7811				

Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal.

Solution

Group 1: Regular Coke ($n = 36$)

$\bar{x}_1 = 0.8168222$

$s_1 = 0.0075074$

Group 2: Diet Coke ($n = 36$)

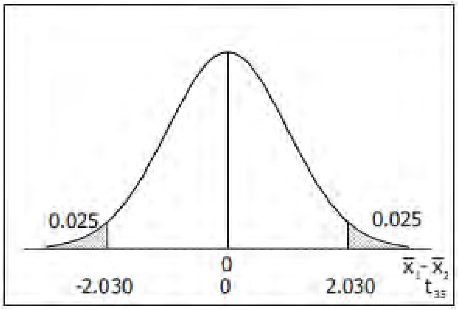
$\bar{x}_2 = 0.7847944$

$s_2 = 0.0043909$

Original Claim: $H_0 : \mu_1 - \mu_2 = 0$ *lbs*

$H_1 : \mu_1 - \mu_2 \neq 0$

$\alpha = 0.05$ *and* $df = 35$

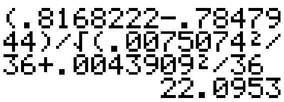


Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} = \pm 2.030$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(0.8168222 - 0.7847944) - (0)}{\sqrt{\frac{.0075074^2}{36} + \frac{.0043909^2}{36}}}$$

$= 22.095$



$P - value = 2 \cdot tcdf\,(22.095,\ 99,\ 35) = 0.8803$

Conclusion:

Reject H_0; there is sufficient evidence to reject the claim that $\mu_1 - \mu_2 = 0$ and conclude that $\mu_1 - \mu_2 \neq 0$ (in fact, that $\mu_1 - \mu_2 > 0$). There is sufficient evidence to reject the claim that the mean weight of cola in cans of regular Coke is the same as the mean weight of cola in cans of Diet Coke. The regular Coke may weigh more because it contains sugar.

Exercise

An Experiment was conducted to test the effects of alcohol. Researchers measured the breath alcohol levels for a treatment group of people who drank ethanol and another group given a placebo. The results are given in the accompanying table. Use a 0.05 significance level to test the claim that the two sample groups come from populations with the same mean.

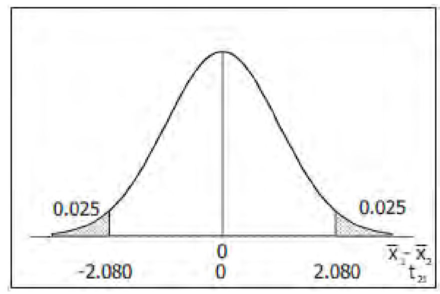
Treatment Group:	$n_1 = 22$	$\bar{x}_1 = 0.049$	$s_1 = 0.015$
Placebo Group:	$n_2 = 22$	$\bar{x}_2 = 0.000$	$s_2 = 0.000$

Solution

Original Claim: $\mu_1 - \mu_2 = 0$ ***lbs***

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 \neq 0$$



$\alpha = 0.05$ *and* $df = 21$

Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} = \pm 2.080$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{(0.049 - 0.0) - (0)}{\sqrt{\frac{.015^2}{22} + \frac{0^2}{22}}}$$

$$= 15.322$$

$P - value = 2 \cdot tcdf\,(15.322,\ 99,\ 21)$

$= 7.14E - 13 \approx 0$

Conclusion:

Reject H_0; there is sufficient evidence to reject the claim that $\mu_1 - \mu_2 = 0$ and conclude that $\mu_1 - \mu_2 \neq 0$ (in fact, that $\mu_1 - \mu_2 > 0$). There is sufficient evidence to reject the claim that the two sample groups come from populations with the same mean.

The fact that there was no variation in the second sample did not affect the calculations or present any special problems. Since there is no variation in x_2, it is really equivalent to the constant value zero – and the test is mathematically equivalent to the one-sample test

$H_0 : \mu_1 = 0$ for which $t = \dfrac{\bar{x}_1 - 0}{s_{\bar{x}_1}}$

Exercise

A researcher was interested in comparing the GPAs of students at two different colleges. Independent simple populations. Do samples of 8 students from college *A* and 13 students from college *B* yielding the following GPAs.

College *A*	3.7	3.2	3.0	2.5	2.7	3.6	2.8	3.4					
College *B*	3.8	3.2	3.0	3.9	3.8	2.5	3.9	2.8	4.0	3.6	2.6	4.0	3.6

Construct a 95% confidence interval for $\mu_1 - \mu_2$. The difference between the mean GPA of college *A* students and the mean GPA of college *B* students.

$$\left(Note:\ \bar{x}_1 = 3.1125,\ \bar{x}_2 = 3.4385,\ s_1 = 0.4357,\ s_2 = 0.5485\right)$$

Solution

$\alpha = 0.05 \quad and \quad df = 21$

Critical value: $t = \pm t_{\alpha/2} = \pm t_{0.025} \underline{= \pm 2.080}$

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(3.1125 - 3.4385) - 2.08\sqrt{\frac{0.4357^2}{8} + \frac{0.5485^2}{13}} \approx -0.78$$

$$(3.1125 - 3.4385) + 2.08\sqrt{\frac{0.4357^2}{8} + \frac{0.5485^2}{13}} \approx 0.13$$

$$\underline{-0.78 < \mu_1 - \mu_2 < 0.13}$$

Exercise

Assume that the two samples are independent simple random samples selected from normal distributed populations. Do not assume that the population standard deviations are equal.
A researcher was interested in comparing the heights of women in two different countries. Independent simple random samples of 9 women from country ***A*** and 9 women from ***B*** yielded to the following heights (in inches).

Country A	64.1	66.4	61.7	62.0	67.3	64.9	64.7	68.0	63.6
Country B	65.3	60.2	61.7	65.8	61.0	64.6	60.0	65.4	59.0

Construct a 90% confidence interval for $\mu_1 - \mu_2$ the difference between the mean height of women in country ***A*** and the mean height of women in country ***B***. Round to two decimal places.

$\left(Note: \ \bar{x}_1 = 64.744 \ in, \ \bar{x}_2 = 62.556 \ in, \ s_1 = 2.192 \ in, \ s_2 = 2.697 \ in\right)$

Solution

$\underline{\bar{x}_1 = 64.744} \quad \underline{\bar{x}_2 = 62.556} \qquad \underline{s_1 = 2.192} \quad \underline{s_2 = 2.697}$

$$A = \frac{s_1^2}{n_1} = 0.53 \quad B = \frac{s_2^2}{n_2} = 0.81$$

$$df = \frac{(A+B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}} \approx 15$$

Degrees of Freedom	0.005	0.01	Area in One Tail 0.025	0.05	0.10
15	**2.947**	2.602	**2.131**	$\boxed{1.753}$	1.341

$$\left(\bar{x}_1 - \bar{x}_2\right) - t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} < \mu_1 - \mu_2 < \left(\bar{x}_1 - \bar{x}_2\right) + t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$(64.744 - 62.556) - 1.753\sqrt{\frac{2.192^2}{9} + \frac{2.697^2}{9}} < \mu_1 - \mu_2 < (64.744 - 62.556) + 1.753\sqrt{\frac{2.192^2}{9} + \frac{2.697^2}{9}}$$

$\underline{0.16 < \mu_1 - \mu_2 < 4.22}$