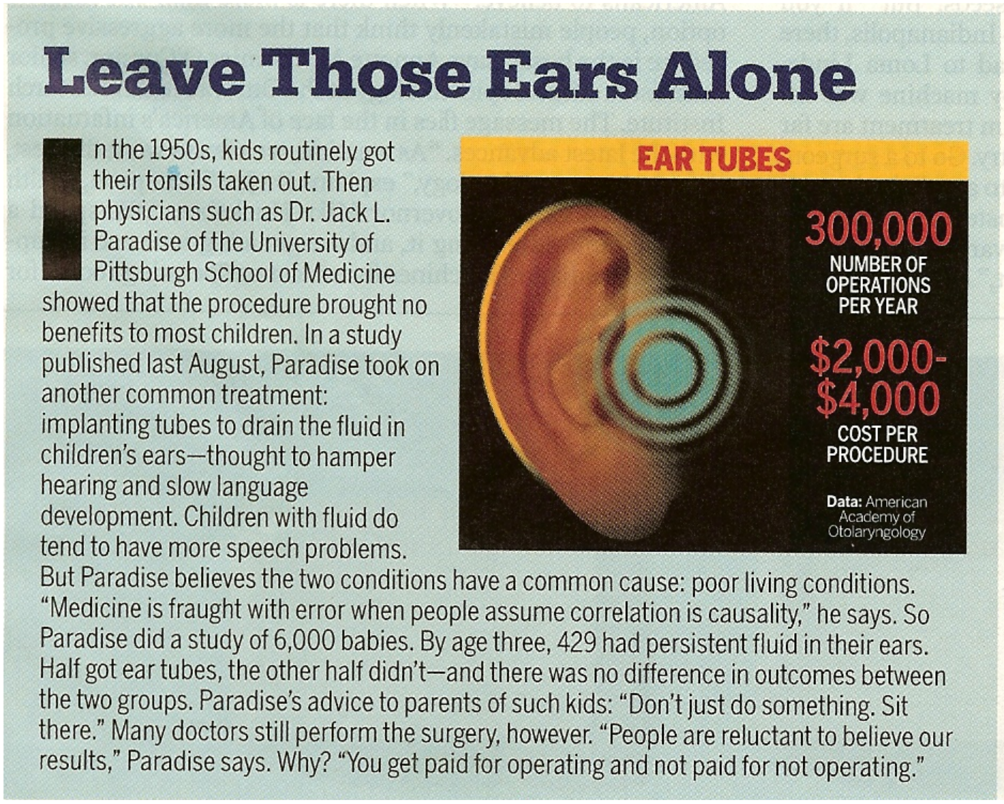
Two Sample (Two Group) Hypothesis Tests



Leave Those Ears Alone

In the 1950s, kids routinely got their tonsils taken out. Then physicians such as Dr. Jack L. Paradise of the University of Pittsburgh School of Medicine showed that the procedure brought no benefits to most children. In a study published last August, Paradise took on another common treatment: implanting tubes to drain the fluid in children's ears—thought to hamper hearing and slow language development. Children with fluid do tend to have more speech problems. But Paradise believes the two conditions have a common cause: poor living conditions. "Medicine is fraught with error when people assume correlation is causality," he says. So Paradise did a study of 6,000 babies. By age three, 429 had persistent fluid in their ears. Half got ear tubes, the other half didn't—and there was no difference in outcomes between the two groups. Paradise's advice to parents of such kids: "Don't just do something. Sit there." Many doctors still perform the surgery, however. "People are reluctant to believe our results," Paradise says. Why? "You get paid for operating and not paid for not operating."

Source: Business Week May 29, 2006

The above is a good example of how medical research is conducted. Note that there are two groups: half received ear tubes and the other half did not. Subjects are assigned to the two groups randomly. The researcher was interested in determining whether there was a difference between the two groups in hearing and language development.

EXAMPLE:
Compare the following 2 groups

<u>Drug Group</u>
$\overline{X}_1$ = 4.4 colds per year
S_1 = 0.7 colds per year
n_1 = 81

<u>Placebo Group</u>
$\overline{X}_2$ = 4.8 colds per year
S_2 = 0.8 colds per year
n_2 = 64

Test at α = .05 (these tests are usually two-tailed)

How to do this?

Two-Sample Z-Tests

If the samples are large, random, and independent, then $\left(\overline{X}_1 - \overline{X}_2\right)$ is a random variable and has approximately a normal distribution, with

$$E\left(\overline{X}_1 - \overline{X}_2\right) = \mu_1 - \mu_2$$

and $\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}$

So, $$Z = \frac{\left(\overline{X}_1 - \overline{X}_2\right) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}}$$

However, since H_0 is usually $\mu_1=\mu_2$, then $\mu_1-\mu_2=0$ and:

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{\sigma_1^{\ 2}}{n_1} + \frac{\sigma_2^{\ 2}}{n_2}}}$$ for σ_1, σ_2 known

for σ_1, σ_2 unknown

$$Z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad \text{if } n_1 + n_2 \geq 32$$

For a 1-α % Confidence Interval for the difference between two means:

$$\left(\overline{X}_1 - \overline{X}_2\right) \pm Z_\alpha \sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}$$

When we construct the confidence interval for the mean difference in the population **($\mu_1 - \mu_2$),** we check to see whether 0 is in the interval. If 0 is in the interval, then we are basically saying that there may be no difference between the two groups and the observed difference between the two sample means is simply sampling error.

The Null hypothesis that $\mu_1 = \mu_2$, is the same as saying that the hypothesized mean difference is 0, i.e., ($\mu_1 - \mu_2$) = 0.

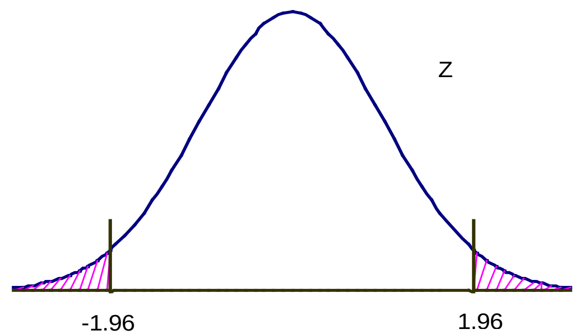
EXAMPLE:
Compare the following 2 groups

<u>Drug Group</u>
$\bar{X}_1$ = 4.4 colds per year
S_1 = 0.7 colds per year
n_1 = 81

<u>Placebo Group</u>
$\bar{X}_2$ = 4.8 colds per year
S_2 = 0.8 colds per year
n_2 = 64

(a) Test at α = .05 (these tests are usually two-tailed)

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{4.4 - 4.8}{\sqrt{\frac{(.7)^2}{81} + \frac{(.8)^2}{64}}} = \frac{-.4}{\sqrt{.01605}} = \frac{-.4}{.127} = -3.15$$

Therefore, REJECT H_0 $p < .05$. The two groups are indeed statistically different.

(b) Construct a 95% Confidence Interval Estimate for the difference between the two population means

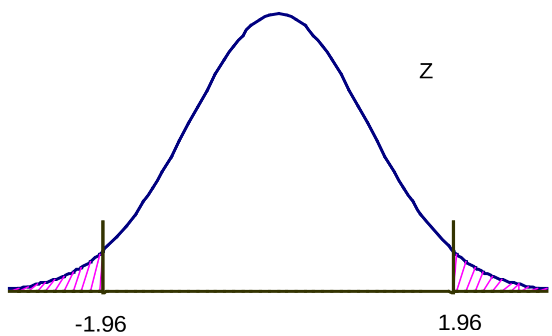
$$-.4 \pm \underbrace{1.96(.127)}_{.25} \Rightarrow \text{-.15} \longleftarrow\!\!\longrightarrow \text{-.65}$$

The Drug group has fewer colds per year. Since 0 is not in the above interval, we see that there is a difference between the two groups. If 0 were in the interval, then we could not reject a 0 mean difference. The drug group has fewer colds. We are 95% sure that the true population mean difference between the drug group and the placebo (control) group is between .15 and .65 fewer colds per year.

EXAMPLE
Scores on a standardized science test. Is there a difference between men and women? Test at α = .05.

	Men	Women
$\bar{X}$	80.0	76.5
s	10	16
n	100	64

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{80.0 - 76.5}{\sqrt{\frac{(10)^2}{100} + \frac{(16)^2}{64}}} = \frac{3.5}{\sqrt{5}} = 1.56$$

Do not reject H_0 $p > .05$
There is no statistically difference between men and women on test scores in science.

(b) 95% CIE

$$3.5 \pm 1.96\left(\sqrt{5}\right) \Rightarrow 3.5 \pm 4.4$$

$$-0.9 \longleftarrow\!\!\longrightarrow +7.9$$

$$\uparrow$$

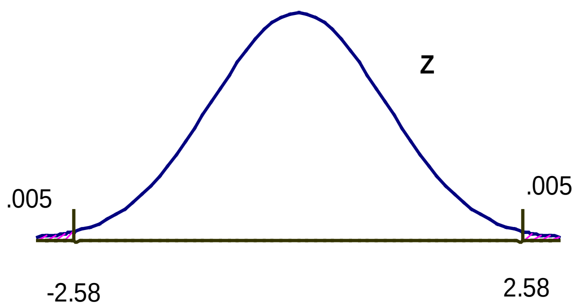
0 difference is possible. When difference goes from negative to positive or positive to negative, 0 is contained in the interval. There may indeed be no difference between men and women on the science test.

EXAMPLE
Typing Speed using MS Word. Who types faster?

	Men	Women
$\bar{X}$	65 wpm	68 wpm
s	10 wpm	14 wpm
n	50	60

Test at $\alpha = .01$.

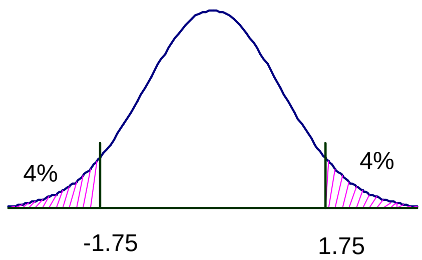
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{65 - 68}{\sqrt{\frac{(10)^2}{50} + \frac{(14)^2}{60}}} = \frac{-3}{2.29} = -1.30$$

Do not reject H_o. $p > .01$.
There is no statistically difference between men and women in typing speed.

(b) Construct a 92% C.I. for different between the means



$$-3 \pm 1.75(2.29)$$

$$-3 \pm 4$$

-7 ←——→ +1

↑

Note that 0 is in the range.

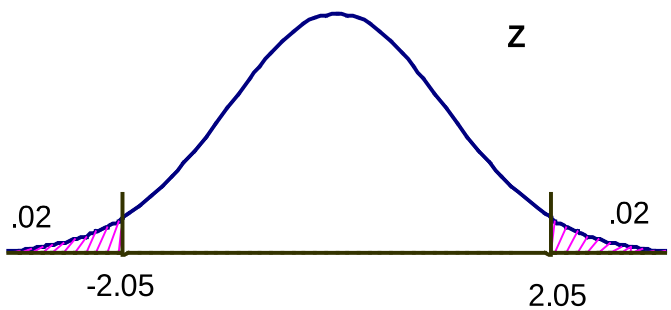
EXAMPLE

Take-Home Pay. Who earns more: Married or unmarried people?

	Married	Not Married
$\overline{X}$	$639.60	$658.20
S	$60	$90
n	40	60

(a) Test at $\alpha = .04$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{-18.60}{\sqrt{\frac{(60)^2}{40} + \frac{(90)^2}{60}}} = \frac{-18.60}{\sqrt{225}} = \frac{-18.60}{15} = -1.24$$

DO NOT REJECT H_0. $p > .04$

There is no statistically difference between married and single individuals with regard to take-home pay.

(b) Construct a 95% CIE for the difference between two means

$$-18.60 \pm \underbrace{1.96(15)}_{29.40}$$

-\$48 ←——→ +\$10.80

↑

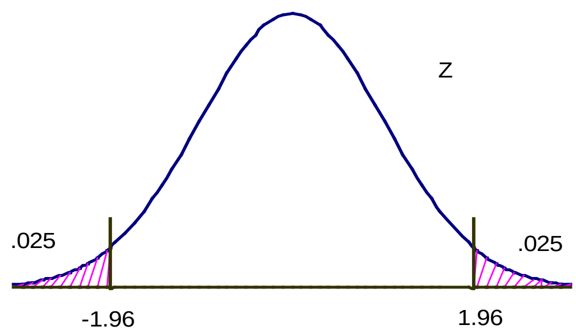
Note that 0 is in the interval.

EXAMPLE
Life Span of Marijuana and non-Marijuana Users

	Non users	Marijuana users
$\overline{X}$	75.2 years	73.2 years
s	8.0 years	7.0 years
n	200	100

(a) Test at $\alpha = .05$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{75.2 - 73.2}{\sqrt{\frac{(8)^2}{200} + \frac{(7)^2}{100}}} = \frac{2}{\sqrt{.81}} = \frac{2}{.9} = 2.22$$

Reject H_0 $p < .05$. . The two groups are indeed statistically different. Non-users live longer.

(b) Construct a 95% CIE for the difference between the two means

$$2 \pm 1.96(.9) \quad \Rightarrow \quad 2 \pm 1.8$$

$$+.2 years \longleftrightarrow +3.8 years$$

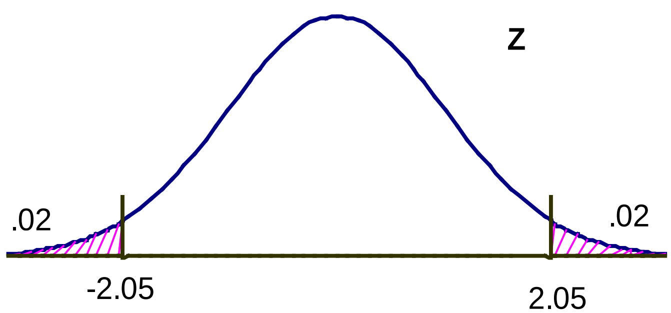
EXAMPLE

Are the machine tools manufactured by Company X and Y different with regard to how long they last?

	Company X	Company Y
$\bar{X}$	16.2 weeks	15.9 weeks
s	.2 weeks	.2 weeks
n	40	40

(a) Test at $\alpha = .04$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$Z = \frac{16.2 - 15.9}{\sqrt{\frac{(.2)^2}{40} + \frac{(.2)^2}{40}}} = \frac{.3}{\sqrt{.002}} = \frac{.3}{.045} = 6.67$$

Reject H_0 $p<.04$. . The two groups are indeed statistically different. Machine tools manufactured by Company X last longer.

(b) Construct a 92% Confidence Interval for difference

$.3 \pm 1.75(.045) \quad \Rightarrow \quad .3 \pm .08$

$+.22$ weeks $\longleftrightarrow$ $+.38$ weeks

Two-Sample t-Tests

Use Z if σ_1 , σ_2 are known OR samples are large
Use ***t*** when σ_1 , σ_2 are unknown AND samples are small ($n_1 + n_2 < 32$)

$$t_{n_1+n_2-2} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S^2_{pooled}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where $$S^2_{pooled} = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

NOTE: S^2_{pooled} is a weighted average of S_1^2 and S_2^2, weighted by degrees of freedom.

Technically, to use this formula one must know (or prove statistically) that the two variances are "equal" – this property is called *homoskedasticity*. (Incidentally, some spell "homos*k*edasticity" with a k and others spell it with a c, "homos*c*edasticity.") If the variances are not statistically equivalent, one may not pool the variances. An F-test may be performed to test for whether S_1^2 and S_1^2 are statistically equivalent, i.e., homoskedasticity:

$$F_{d.f._L, d.f._S} = \frac{S^2_{LARGER}}{S^2_{SMALLER}}$$

We will learn about this F-test in other courses.

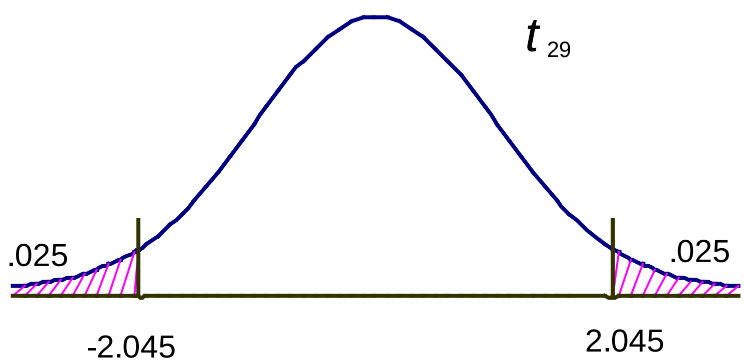
If you do not have homoskedasticity, i.e. the variances are not equal, then you have to make adjustment to the degrees of freedom.

EXAMPLE
Score on a standardized reading examination.

	Men	Women
$\overline{X}$	$80 = \overline{X}_1$	$84 = \overline{X}_2$
S	$16 = S_1$	$20 = S_2$
n	$16 = n_1$	$15 = n_2$

Test at $\alpha = .05$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$S^2_{pooled} = \frac{15(256) + 14(400)}{29} = \frac{9440}{29} = 325.5$$

Note that S^2_{pooled} is between S_1^2 (= 256) and S_2^2 (= 400).

$$t_{29} = \frac{80 - 84}{\sqrt{325.5(\frac{1}{16} + \frac{1}{15})}} = \frac{-4}{\sqrt{42.04}} = \frac{-4}{6.48} = -.62$$

Do not reject H_0. $P > .05$
There is no statistically difference between men and women on test scores.

EXAMPLE

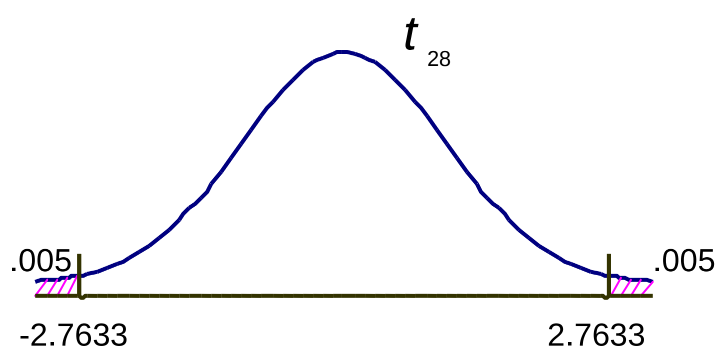
Which company, if any, has a better daily wage

Company A	Company B
$\overline{X}_1 = \$210$	$\overline{X}_2 = \$175$
$S_1 = \$25$	$S_2 = \$20$
$n_1 = 10$	$n_2 = 20$

Test at α = 0.01

$$H_0 : \mu_1 = \mu_2$$

$$H_1 : \mu_1 \neq \mu_2$$



$$S^2_{pooled} = \frac{9(625) + 19(400)}{28} = 472.3$$

$$t_{28} = \frac{210 - 175}{\sqrt{472.3(\frac{1}{10} + \frac{1}{20})}} = \frac{35}{\sqrt{70.845}} = \frac{35}{8.42} = 4.16$$

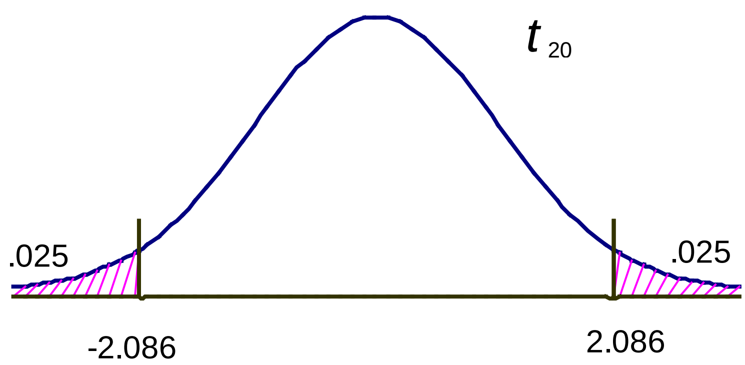
Reject H_0 at $p < .01$. The two groups are indeed statistically different.

EXAMPLE

Two types of precast concrete beams are being considered for sale. The only difference between the two beams is in the type of material used. Strength is measured in terms of pounds per square inch (psi) of pressure. Which beams are stronger, those from Supplier A or those from Supplier B? Test at $\alpha = 0.05$

<u>Supplier A</u>	<u>Supplier B</u>
$\bar{X}_1 = 5000$ *psi*	$\bar{X}_2 = 4975$ *psi*
$S_1 = 50$ *psi*	$S_2 = 60$ *psi*
$n_1 = 12$ *batches*	$n_2 = 10$ *batches*

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$S^2_{pooled} = \frac{11(2500) + 7(3600)}{20} = 2995$$

$$t_{20} = \frac{5000 - 4975}{\sqrt{2995(\frac{1}{12} + \frac{1}{10})}} = \frac{25}{\sqrt{549}} = \frac{25}{23.4} = 1.07$$

Do not reject H_0. P>.05

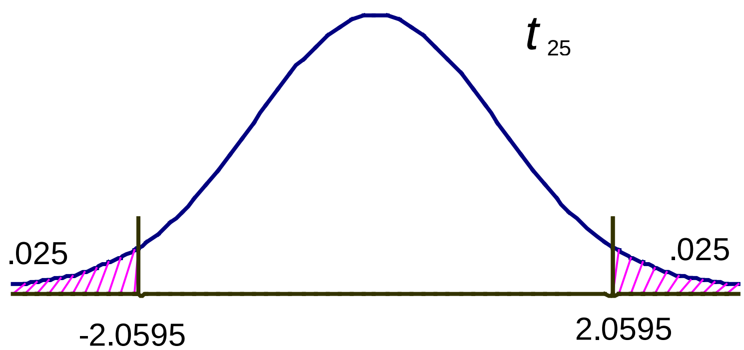
There is no statistically difference between the two concrete suppliers.

EXAMPLE

Two approaches to dealing with liver cancer (additional years of life) Are they different? Test at $a = .05$

Approach A	Approach B
$\overline{X}_1 = 6.2\,years$	$\overline{X}_2 = 5.6\,years$
$S_1 = .69\,years$	$S_2 = .60\,years$
$n_1 = 10$	$n_2 = 17$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$S^2_{pooled} = \frac{9(.476) + 16(.36)}{25} = .402$$

$$t_{25} = \frac{6.2 - 5.6}{\sqrt{.402(\frac{1}{10} + \frac{1}{17})}} = \frac{.6}{\sqrt{.064}} = \frac{.6}{.253} = 2.37$$

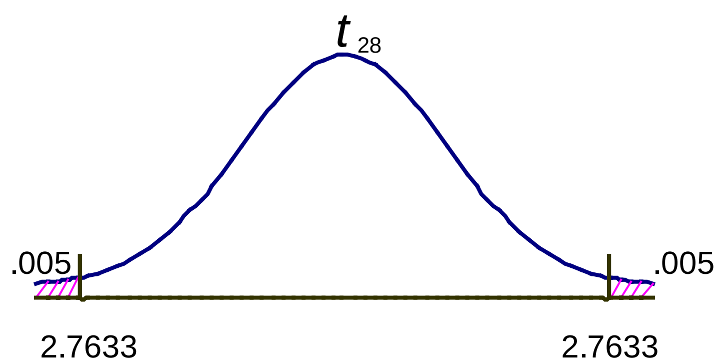
REJECT H_0

Reject H_0 at $p < .05$. There is a statistically significant difference between the two treatment approaches.

EXAMPLE
Who lives longer? Test at α =.01

Single Men	Married Men
$\overline{X}_1 = 72.5 years$	$\overline{X}_2 = 74.5 years$
$S_1 = 7.0 years$	$S_2 = 8.0 years$
$n_1 = 14$	$n_2 = 16$

$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$



$$S^2_{pooled} \frac{13(49)+15(64)}{28} = 57$$

$$t_{28} = \frac{-2}{\sqrt{57(\frac{1}{14}+\frac{1}{16})}} = \frac{-2}{\sqrt{7.6}} = \frac{-2}{2.76} = -.72$$

Do not reject H_0.   P >.01. There is no significant difference in lifespan.

EXAMPLE
Absenteeism of Executives

Test at $\alpha = .05$

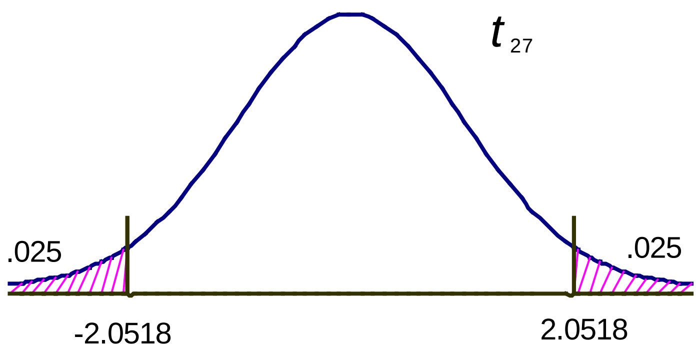
<u>Female Executives</u>

$\overline{X}_1 = 9.2\ days$
$S_1 = 1.1\ days$
$n_1 = 9$

<u>Male Executives</u>

$\overline{X}_2 = 10.4\ days$
$S_2 = .9\ days$
$n_2 = 20$

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$



$$S^2_{pooled} = \frac{8(1.21) + 19(.81)}{27} = .929$$

$$t_{27} = \frac{9.2 - 10.4}{\sqrt{.929(\frac{1}{9} + \frac{1}{20})}} = \frac{-1.2}{\sqrt{.14967}} = \frac{-1.2}{.387} = -3.1$$

Reject H_0. at $p < .05$. There is a statistically significant difference. Male executives miss work more often than females.

EXAMPLE USING EXCEL:
A marketer wants to determine whether men and women spend different amounts on wine. (It is well known that men spend considerably more on beer.) A researcher decides to test this. She randomly samples 34 people (17 women and 17 men) and finds that the average amount spent on wine (in a year) by women is $437.47. The average amount spent by men is $552.94**.** **Given the Excel printout below, is the difference statistically significant?**

t-Test: Two-Sample Assuming Equal Variances

	Variable 1	*Variable 2*
Mean	437.4705882	552.9412
Variance	97784.13971	104221.6
Observations	17	17
Pooled Variance	101002.8493	
Hypothesized Mean Difference	0	
df	32	
t Stat	**-1.05928794**	
P(T<=t) one-tail	0.148699644	
t Critical one-tail	1.693888407	
P(T<=t) two-tail	0.297399288	
t Critical two-tail	2.036931619	

Answer: If a two-tail test was done, the probability of getting the sample evidence (or sample evidence showing an even larger difference) given that there is no difference in the population means of job satisfaction scores for men and women is .30 (rounded from .297399288). In another words, if men and women spend the same on wine, there is a 30% chance of getting the sample evidence (or sample evidence indicating a larger difference between men and women) we obtained. Statisticians usually test at an alpha of .05 so we do not have evidence to reject the null hypothesis. Conclusion: There is no statistically significant difference between men and women on how much they spend on wine consumption.

The calculated t-statistic is -1.059287941. Why is it negative?
Answer: The amount spent on wine by women is less than that spent by men (although the difference is not statistically significant). If you make men the first variable the t-value will be positive but the results will be

exactly the same (the t-distribution is symmetric).

What would the calculated t-value have to be for us to reject it?
Answer: If a two-tail test is being done, the critical value of t is 2.036931619. To reject the null hypothesis, we would need a calculated t-value of more than 2.036931619 or less than -2.036931619.

See more Excel examples in:
TBA

Matched T-Test (Optional Topic)
(non-independent samples, not two groups)

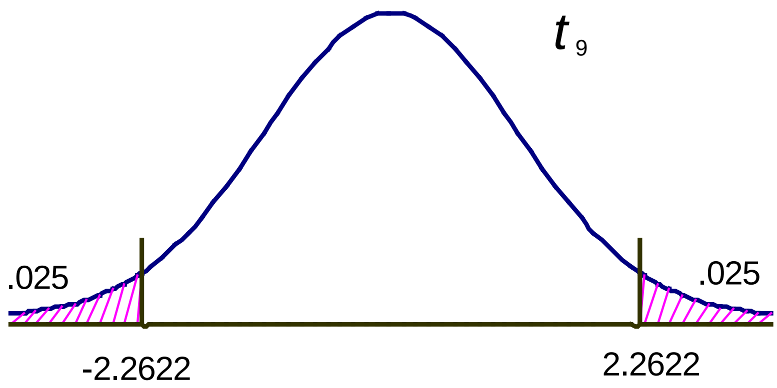
If you do this in MS Excel, this test is known as the
t-Test: Paired Two Sample for Means

(Example 1)

X_1(Before)	X_2(After)	D =(X_1-X_2)	D^2
150	130	20	400
160	145	15	225
155	150	5	25
170	162	8	64
150	140	10	100
145	125	20	400
170	155	15	225
160	140	20	400
165	163	2	4
160	160	0	0
		115	1,843

$H_0 : \mu_D = 0$ (mean difference is 0)

$H_1 : \mu_D \neq 0$



$$t_9 = \frac{\overline{D}}{S_D / \sqrt{n}} = \frac{\sum D}{\sqrt{\frac{n\sum D^2 - (\sum D)^2}{n-1}}}$$

$$t_9 = \frac{115}{\sqrt{\frac{10(1843) - (115)^2}{9}}} = \frac{115}{24.05} = 4.78$$

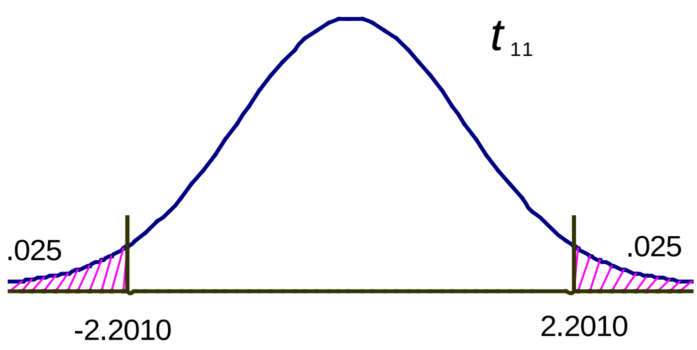
Reject H_0 at $P<.05$. There is a significant difference between the before and after measures.

(Example 2)

X_1(Before)	X_2(After)	$D=(X_1-X_2)$	D^2
12	15	-3	9
4	9	-5	25
17	16	1	1
3	9	-6	36
5	5	0	0
18	14	4	16
15	7	8	64
14	8	6	36
8	1	7	49
6	17	-11	121
16	25	-9	81
15	12	3	9
		-5	447

$$H_0 : \mu_D = 0$$

$$H_1 : \mu_D \neq 0$$



$$t_{11} = \frac{-5}{\sqrt{\dfrac{5(447)-(-5)^2}{11}}} = \frac{-5}{22} = -.23$$

Do Not Reject H_0.   P >.05. Difference is not significant.