Inference about a Population Mean

Diana Mindrila, Ph.D.
Phoebe Balentyne, M.Ed.

Based on Chapter 18 of The Basic Practice of Statistics (6th ed.)

Concepts:
- Conditions for Inference about a Mean
- The t Distributions
- The One-Sample t Confidence Interval
- The One-Sample t Test
- Using Technology
- Matched-Pairs t Procedures
- Robustness of t Procedures

Objectives:
- Describe the conditions necessary for inference.
- Describe the t distributions.
- Check the conditions necessary for inference.
- Construct and interpret a one-sample t confidence interval.
- Perform a one-sample t test.
- Perform a matched-pairs t test.
- Describe the robustness of the t procedures.

References:
Moore, D. S., Notz, W. I, & Flinger, M. A. (2013). *The basic practice of statistics* (6th ed.). New York, NY: W. H. Freeman and Company.

Conditions for Inference about a Mean

Making inferences about a population mean requires several assumptions:

1) **Random:** The data come from a random sample of size n from the population of interest or a randomized experiment.
2) **Normal:** The population has a Normal distribution. In practice, it is enough that the distribution be symmetric and single-peaked unless the sample is very small.
3) **Independent:** The population must be much larger than the sample (at least 20 times as large).
4) **Standard Deviation:** The population standard deviation is known.

- When all of these assumptions are met, z scores can be used in the computation process.
- However, many times these assumptions are not met and even more often the population standard deviation is not known for the variable of interest.
- In this case, t procedures are used instead, which are based on a distribution of standardized scores called t scores.

Standard Error

- *T* procedures are very similar to z procedures, and they are used when the data are not perfectly Normal and when the population standard deviation is unknown.
- *T* procedures use the standard deviation of the sample instead of the standard deviation of the population. The notation changes from σ to *s* when *t* procedures are used.

Formula for standard error:

z Procedures: **t Procedures:**

Population standard deviation: **Sample standard deviation (also called standard error):**

$$\frac{\sigma}{\sqrt{n}}$$ $$\frac{s}{\sqrt{n}}$$
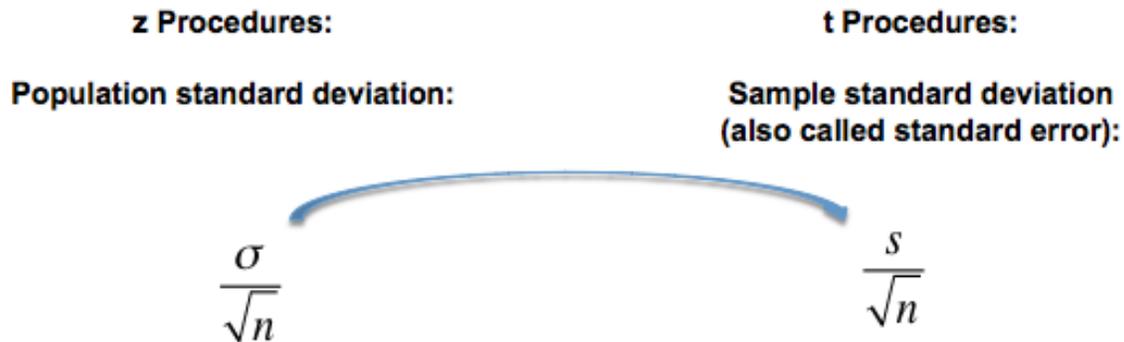
> The **standard error** of the sample mean (x) is the sample standard deviation, and shows how far the sample mean will be from the population mean (μ), on average, in repeated random samples of size *n*.

- The formula for standard error stays the same for *z* procedures and *t* procedures. In both cases, the standard deviation is divided by the square root of the sample size. The only difference is that instead of using the population standard deviation, as is done in *z* procedures, the standard deviation of the sample is used for *t* procedures. The result is called the **standard error**.
- The interpretation of this statistic remains the same: if there were many random samples and the mean was computed for each one, the standard error shows the average distance of these samples from the actual population mean.

The *t* Distributions

- After estimating the standard error, researchers can compute confidence intervals and conduct tests of significance.
- Again, the same formula is used as with the z procedures, except the sample standard deviation is used instead of the population standard deviation.
- The other difference is that the notation *t* is now used instead of *z** in the confidence interval formula, and a *t* test statistic instead of the *z* test statistic for the tests of significance.

	z Procedures:	t Procedures:
Confidence Intervals:	$\bar{x} \pm z * \dfrac{\sigma}{\sqrt{n}}$	$\bar{x} \pm t * \dfrac{s}{\sqrt{n}}$
Tests of Significance:	$z = \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$	$t = \dfrac{\bar{x} - \mu_0}{s / \sqrt{n}}$

The *t* Distributions

- The interpretation of *t* scores is the same as the interpretation of *z* scores: they are a standardized measure of how far the sample mean or a certain given value is from the population mean.
- However, the distribution of *t* scores has a slightly different shape than the distribution of *z* scores, which is Normal.

It has a *different shape* than the standard Normal curve:
- ✓ It is symmetric with a single peak at 0, however
- ✓ it has much more area in the tails.

Z distribution
(standard normal)

t-distribution
(*n* close to 30)

t-distribution
(*n* smaller than 30)

$\mu = 0$
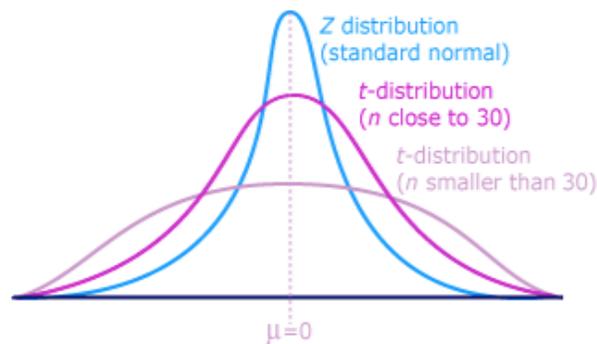
There is a different *t* distribution for each sample size, specified by its **degrees of freedom (df):**
df = n-1.

- The distribution of *t* scores is symmetric, but it is not Normal.
- In the figure above, the *z*, or the Normal distribution (represented in blue) and the *t* distributions (represented in different shades of pink) are overlapped for comparison.
- Like the Normal distribution, the *t* distribution has a single peak and a mean of zero, which is the center of the distribution.
- However, the tails of the *t* distribution are higher and wider than for the Normal distribution.
- The *t* distribution also has more spread because the standard error is larger than the population standard deviation.
- Therefore, there is more error when *t* scores are used to make inferences. There is less precision with *t* scores than with *z* scores. This is the tradeoff for being able to use these scores without meeting some of the assumptions required for *z* statistics.
- Another important difference is that the curve looks different for different sample sizes. The figure above shows that as the sample size increases, the *t* curve gets closer to the Normal curve. Therefore, whenever *t* scores are used, the sample size must be taken into account.

The *t* Distributions – Summary of Facts

When comparing the density curves of the standard Normal distribution and *t* distributions, several facts are apparent:
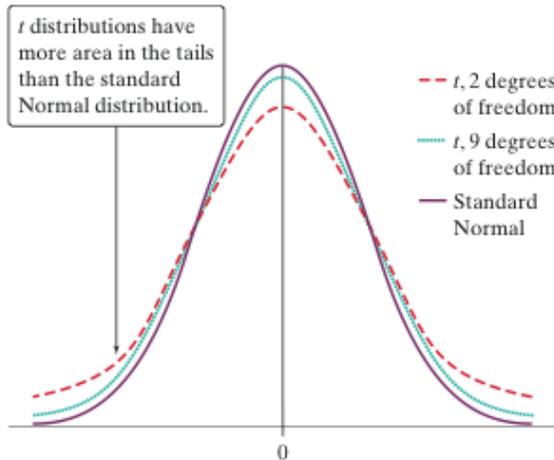
- ✓ The density curves of the *t* distributions are similar in shape to the standard Normal curve.
- ✓ The spread of the *t* distributions is slightly greater than that of the standard Normal distribution.
- ✓ The *t* distributions have more probability in the tails and less in the center than does the standard Normal distribution.
- ✓ As the degrees of freedom increase, the *t* density curve approaches the standard Normal curve more closely.

One-Sample *t* Confidence Interval

The **one-sample *t* interval for a population mean** is similar in both reasoning and computational detail to the one-sample *z* interval for a population proportion.

The One-Sample *t* Interval for a Population Mean

The confidence interval for the population mean (μ) is:

$$\bar{x} \pm t^* \frac{s_x}{\sqrt{n}}$$

Margin of error

where t^* is the critical value.

- *T* scores can be used just like *z* scores to compute a confidence interval for a population mean.
- To obtain the upper and the lower limit of this confidence interval, add and subtract the margin of error from the sample mean.
- To find the sample mean, the sample standard deviation is used instead of the population standard deviation and t^* is used as the critical value instead of z^*.
- T^* is also called the critical value (just like z^*) and can be obtained from a table in a statistics textbook.
- This type of confidence interval can also be computed using statistical software.

Using a Table

- The shape of the t distribution differs based on the sample size.
- Therefore, for the same confidence level, the value of t^* will be different for different sample sizes.
- In order to find the value of t^* using a table in a statistics textbook, the degrees of freedom must be known.
- Degrees of freedom = n – 1 $\Rightarrow$ Take the number of individuals in the sample and subtract one.
- Use the rows to find the degrees of freedom for the specific sample.
- Use the columns to find the specific confidence level that has been chosen.

Example:

To construct a 95% confidence interval for the mean μ for a sample size: $n = 12$, what critical t^* should be used?

Move across columns to the entry that is directly above the desired confidence level (e.g. 95%) .

Confidence level C

df	90%	95%	96%	98%
10	1.812	2.228	2.359	2.764
11	1.796	2.201	2.328	2.718
12	1.782	2.179	2.303	2.681
		...		
z*	1.645	1.960	2.054	2.326

In Table C, consult the row corresponding to $df = n – 1 = 11$.

The desired critical value is $t^* = 2.201$.

- In the above example, the number of individuals in the sample is 12. To find the degrees of freedom, subtract 1 and obtain 11. Then find the 11th row.
- The confidence level desired is 95%. Find the 95% confidence column.
- The 11th row and the 95% confidence column meet at the value: $t^* = 2.201$
- Notice that for the same confidence level the value of z^* is 1.96, which is slightly smaller than t^*. The margin of error with the z procedures is slightly smaller than with the t procedures.
- ❖ T procedures are not as precise as z procedures, so they are only used when the assumptions for z procedures cannot be met.

Example

- The following example shows how to compute a t confidence interval by hand.
- In this example, researchers would like to estimate the population mean on a student achievement motivation scale.
- The sample contains 61 students. The mean score on the scale for the sample is 100 and the standard deviation is 15.62.
- Researchers want to estimate the population mean on this scale with a 95% confidence level.
- The population standard deviation for this scale is unknown, so a t procedure must be used instead of a z procedure.

Motivation Scale: $\bar{x} = 100$ and $s_x = 15.62$ n=61

Using the t Distribution to find the 95% confidence interval for the population mean:

Confidence level C

df	90%	95%	96%
50	1.676	2.009	2.109
60	1.671	2.000	2.099
80	1.664	1.990	2.088

n = 61 => df =61-1= 60 ⟶ 60

From Table C, we find $t* = 2.000$.

Therefore, the 90% confidence interval for μ is:

$$CI = \bar{x} \pm t * \frac{s_x}{\sqrt{n}} = 100 \pm 2 * \frac{15.62}{\sqrt{61}} = 100 \pm 4$$
$$= (96,104)$$

CONCLUDE: *Researchers can be 95% confident that the interval from 96 to 104 captures the true achievement motivation mean for the entire population of students.*

- First, use a table to obtain the value of $t*$ that corresponds to the correct confidence level and sample size. (The value of $t*$ is 2.000).
- Next, place the value of $t*$ into the formula for the margin of error. In this case, the margin of error is 4.
- Finally, add and subtract the margin of error to the sample mean of 100 to obtain the upper and lower limits of the confidence interval. The lower limit is 96 and the upper limit is 104.
- ❖ Conclusion: Researchers can say with 95% confidence that the population mean on this motivation scale is between 96 and 104.
- Generally, this type of result is obtained using statistical software, but it can also be done by hand. Knowing how to compute statistics by hand helps researchers catch potential errors in statistical software.

The One-Sample t Test

- Just like z scores, t scores can be used to conduct tests of significance.

Choose a sample of size n from a large population that contains an unknown mean μ. To test the hypothesis $H_0: \mu = \mu_0$, compute the one-sample t statistic:

$$t = \frac{\bar{x} - \mu_0}{s_x / \sqrt{n}}$$

Find the P-value by calculating the probability of getting a t statistic this large or larger in the direction specified by the alternative hypothesis H_a in a t-distribution with $df = n - 1$.
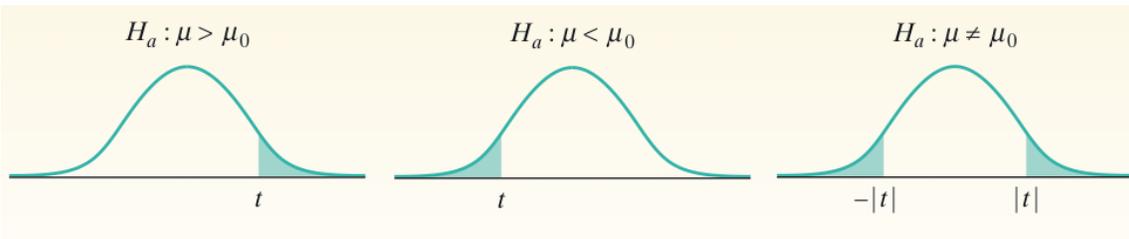
- H_0 is the same when conducting tests of significance using t scores as it is when using z scores. It states that the sample mean is equal to the population mean, or that there is no significant difference between the two values.
- The alternative hypothesis is also formulated in the same way. It can be one-sided, stating that the sample mean is either higher or lower than the population mean, or it can be two-sided, stating that the sample mean is different (not necessarily higher or lower) that the population, which means that the test is looking for very high or very low values.
- Similarly, the formula for the t test statistic stays the same as for the z statistic. The only difference is that the sample standard deviation, rather than the population standard deviation, is used. Also, the value of the t statistic differs based on sample size, or degrees of freedom.
- After the test statistic is computed, this information can be used to determine whether the null hypothesis should be rejected or accepted, just like is done with a z test.
- The p-value for the test statistic can be obtained using statistical software.
- A table can also be used to decide whether to accept or reject the null hypothesis.

The One-Sample *t* Test Example

The following example illustrates how to conduct a test of significance by hand using the *t* statistic. Statistical software can also be used to obtain the same results.

- Suppose that prior studies have found that the population mean for the motivation scale being studied is 94, but researchers want to challenge this statement to see if it holds true.
- The null hypothesis is that the population mean is equal to 94.
- The alternative hypothesis is that the mean is different. (It is a two-sided alternative hypothesis).

- The first step is to collect a sample. In the example, the sample includes 61 individuals.
- Next, the mean is computed for the sample, in this case, the sample mean is 100.
- The population standard deviation is unknown, so a *t* test must be conducted instead of a *z* test. The sample standard deviation, which is 15.62, must be used for the *t* test.
- The value of alpha, the significance level, must be specified before any tests are run. In this case, the alpha level is .05.

STUDENT T DISTRIBUTION : f(x)

Motivation Scale:
Ho: μ=94
Ha: μ≠94

Population Mean: μ=94
Sample mean: x̄=100
Sample size: n=61
Sample st dev: s$_x$=15.62
Significance level: α=0.05

t=3
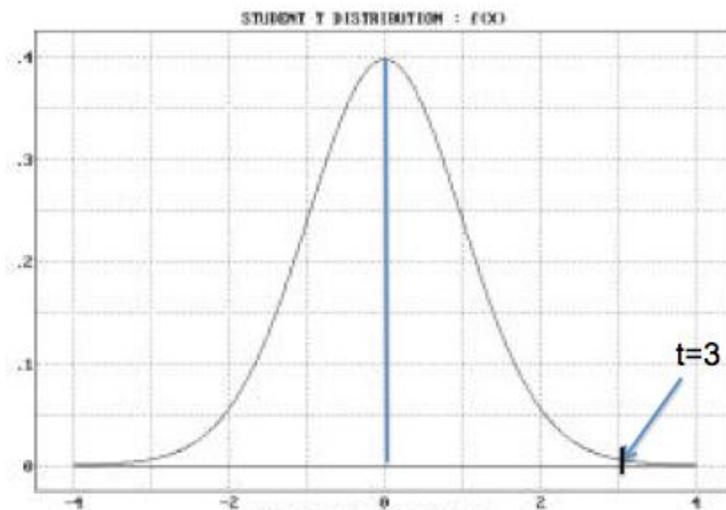
Test statistic $t = \dfrac{\overline{x} - \mu_0}{s_x / \sqrt{n}} = \dfrac{100 - 94}{15.62 / \sqrt{61}} = \dfrac{6}{2} = 3$

- Now that the *t* value has been found (*t* = 3), the next step is to determine whether this test statistic is far enough from the mean to reject the null hypothesis.

The One-Sample *t* Test

- After the test statistic has been computed, researchers must determine whether this standardized scores is far enough from the mean to reject the null hypothesis.
- If statistical software is being used, the p-value is computed by the software for any value of *t*. If the p-value is lower than alpha, then the null hypothesis should be rejected.
- This process can also be done by hand by using a table in a statistics textbook. However, the table does not provide p-values, rather, it gives the value of the *t* statistic at different probability levels and for different degrees of freedom.
- Therefore, after the test statistic *t* has been computed for the test of significance, the *t* score that corresponds to the alpha level must be found. This value is called the critical value of *t*, and represents the cut-off value for *t*:
  - ❖ If the test statistic *t* is larger than the critical value, the null hypothesis can be rejected.
  - ❖ If the test statistic *t* is smaller than the critical value, the null hypothesis should be accepted.

The One-Sample *t* Test Example

Example:

Statistical software:
-provides the p_value for t

p_value <= α -> reject H_0
p_value > α -> accept H_0

Table C:

α=0.05 -> $t_{Critical Value}$: t_{CV}=2.00

df		Table C	
50	1.676	2.009	2.109
60	1.671	2.000	2.099
80	1.664	1.990	2.088
One-sided p_value	.05	.025	.02
Two sided p_value	.10	.05	.04

t_{CV}=2

t=3

-8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8
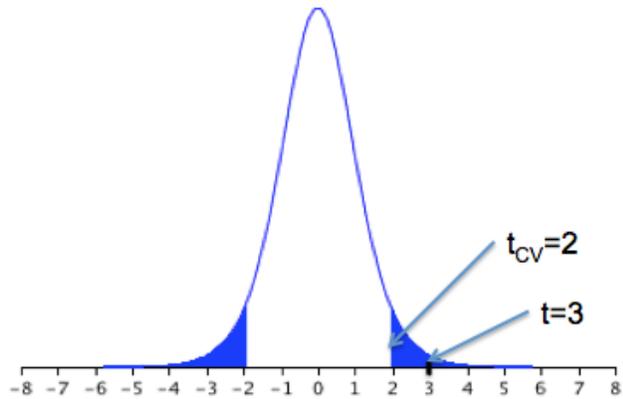
t=3
t_{CV}=2.00 } t>≥t_{CV}=> reject H_0

(When t < t_{CV}=> accept H_0)

- In the above example, the alternative hypothesis is two-sided, so the two-sided p-value must be used. For the alpha level of .05, the critical value for degrees of freedom equal to 60 is 2.000.
- The test statistic *t* was 3, which is larger than the critical value of 2. In this case, the null hypothesis should be rejected because *t* is further from the mean than the critical value or beyond the cut-off level.

Matched-Pairs *t* Procedures

Comparative studies are more convincing that single-sample investigations. For that reason, one-sample inference is less common than comparative inference. Study designs that involve making two observations on the same individual, or one observation on each of two similar individuals, result in **paired data**.

Many times in social science research comparisons need to be made between two equivalent groups to assess the effectiveness of a certain intervention or to compare the results of the same group before and after a certain intervention. When paired data result from measuring the same quantitative variable twice, comparisons can be made by analyzing the difference in each pair. If the conditions for inference are met, one-sample *t* procedures can be used to perform inference about the mean difference μ_d.

> **Matched-Pairs *t* Procedures**
> To compare the responses to the two treatments in a matched-pairs design, find the difference between the responses within each pair. Then apply the one-sample *t* procedures to these differences.

Matched-Pairs *t* Procedures

Example:

Population:
Ho: $\mu_d=0$
Ha: $\mu_d>0$

Sample:
x̄=10.05
s=7.56
α=0.05

Test statistic: $t = \dfrac{\bar{x} - \mu_d}{s_x / \sqrt{n}}$

$= \dfrac{10.05 - 0}{7.56 / \sqrt{20}} = \dfrac{10.05}{7.56/4.47} = \dfrac{10.05}{1.69} = 5.94$

Name	Pre-test	Post-test	Posttest - Pretest Difference
Dan	100	100	0
Marian	98	100	2
Sam	67	78	11
Gene	56	68	12
Lena	89	96	7
Richard	88	95	7
Dorian	84	98	14
Ernest	57	79	22
John	90	98	8
Linda	78	89	11
Martha	69	75	6
Geta	89	95	6
Delia	95	98	3
Damian	79	86	7
Sylvia	87	98	11
Wynona	98	100	2
Steve	67	78	11
Gregory	56	89	33
Julian	78	89	11
Steve	77	94	17
		Average Difference	**10.05**
		Standard Deviation of the Difference	**7.56**

- The example above involves paired data. The sample contains 20 students with a score recorded for a pretest and a posttest.
- The goal is to determine if there is a significant improvement from the pretest score to the posttest score.
- The third column in the table above shows the difference in the posttest and pretest scores for each individual.
- The average difference in scores for the entire sample is 10.05, but simply looking at the scores cannot determine if the differences are statistically significance. A test of significance must be conducted.
- The null hypothesis for the test of significance says that the average difference from the pretest to the posttest across all individuals is equal to zero, meaning there is no significant difference between the scores on the tests.
- The alternative hypothesis states that the difference is larger than zero because researchers are expecting an improvement from the pretest to the posttest. In this case, the alternative hypothesis is one-sided because it only includes an increase in scores, not simply a difference.
- In this case, the mean difference is 10.05 and the standard deviation of the differences is 7.56, but the population standard deviation is unknown, so a *t* test of significance must be used. In this case, the test statistic equals 5.94.

Matched-Pairs *t* Procedures

Statistical Software:

t=5.94
Two-sided p_value=0.000

p_value <= alpha => reject Ho

t=1.729   t=5.95

Table C:
df=20-1=19
α=0.05

t=5.94
t$_{cv}$=1.729

t>t$_{cv}$=> reject Ho
Conclusion: Evidence from the sample shows that the difference
between posttest and pretest scores is statistically significant.
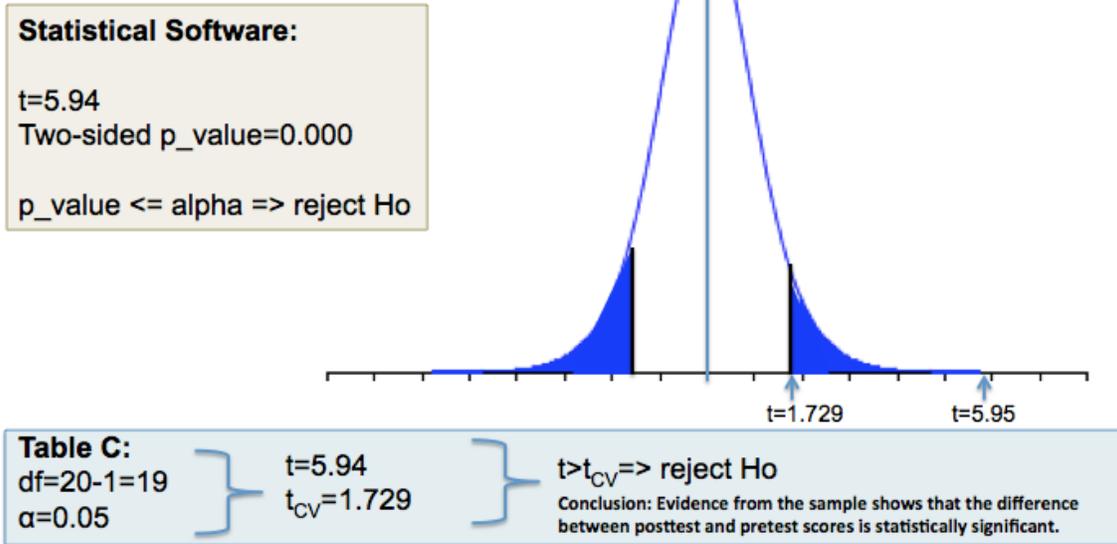
- If statistical software is used, the test statistic and p-value are automatically obtained.
- When *t* = 5.94, calculators or statistical software show that the p-value is equal to zero. This value is obviously smaller than alpha, which is .05, so the null hypothesis can be rejected.
- Statistical software provides the two-sided p-value, so when the alternative hypothesis is one-sided this value must be divided by 2 to obtain the correct p-value.
- If a statistical table, instead of statistical software, is used, the critical value is 1.729. The test statistic (*t* = 5.94) is larger than the critical value, so the null hypothesis can be rejected.
- ❖ In conclusion, based on the evidence from the sample, the difference between the posttest scores and the pretest scores is statistically significant.

Robustness of _t_ Procedures

A confidence interval or significance test is called **robust** if the confidence level or _p_-value does not change very much when the conditions for use of the procedure are violated.

> **Using the _t_ Procedures**
> - Except in the case of small samples, the condition that the data are a random sample from the population of interest is more important than the condition that the population distribution is Normal.
> - _Sample size less than 15_: Use _t_ procedures if the data appear close to Normal. If the data are clearly skewed or if outliers are present, do not use _t_.
> - _Sample size at least 15_: The _t_ procedures can be used except in the presence of outliers or strong skewness.
> - _Large samples_: The _t_ procedures can be used even for clearly skewed distributions when the sample is large, roughly $n \geq 40$.