CTIS 365

APPLIED DATA ANALYSIS

Dr. Seyid Amjad Ali
syedali@bilkent.edu.tr
C202, 290 5335

T-STATISTIC

Limitations of using z-score

$$z = \frac{X - \mu}{\sigma_M}$$

Need to know σ, which is often not known.

When σ is not known, we use the **t-statistic** rather than a z-score

T-STATISTIC

Instead of using variability for population, we use the sample variability as an unbiased estimate.

Let's recall the formulas for sample variability and standard deviation:

$$s^2 = \frac{SS}{n-1} \qquad s = \sqrt{\frac{SS}{n-1}}$$

Using the sample statistics, we can estimate standard error.

T-STATISTIC

Standard error = $\sigma_M = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{\sigma^2}{n}}$

Estimated standard error = $s_M = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$

s_M provides an estimate of the standard distance between a sample mean and the population mean.

T-STATISTIC

THE *t* STATISTIC

The t statistic is used to test hypotheses about an unknown population mean when the population standard deviation is unknown.

$$t = \frac{M - \mu}{s_M} = \frac{M - \mu}{\sqrt{s^2/n}}$$

T-STATISTIC

Degrees of freedom and t statistic

df describes the number of scores in a sample that are free to vary.

$$df = n - 1$$

Because t-statistic is calculated by using sample variance, each t-statistic is related to a *df*.

T-STATISTIC

THE *t* DISTRIBUTION

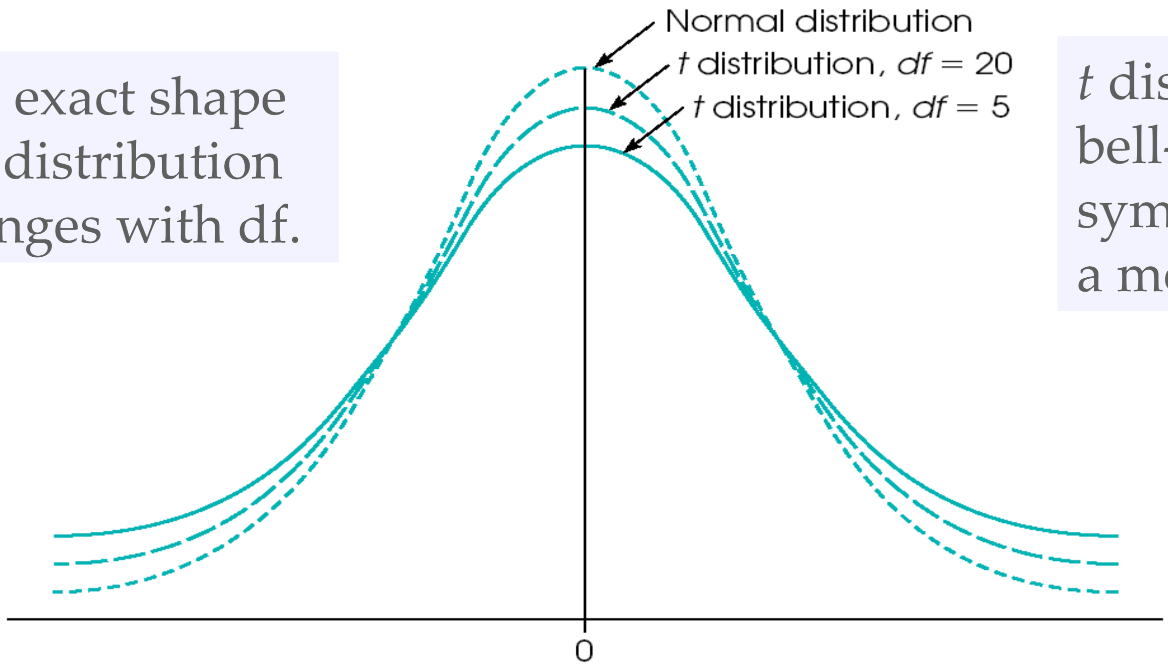
When the population is normally distributed or when sample size (n) is large, the distribution of z-scores computed from sample means is normal.

When the population is normally distributed or when sample size (n) is large, the distribution of t-statistics is t-distribution with the df (n – 1).

T-STATISTIC

THE *t* DISTRIBUTIONS

The exact shape of t distribution changes with df.



t distributions are also bell-shaped and symmetrical and have a mean of zero.

As df gets larger, t distribution gets closer to normal distribution.

T-STATISTIC

THE *t* DISTRIBUTION TABLE

	PROPORTION IN ONE TAIL					
	0.25	0.10	0.05	0.025	0.01	0.005
	PROPORTION IN TWO TAILS COMBINED					
df	0.50	0.20	0.10	0.05	0.02	0.01
1	1.000	3.078	6.314	12.706	31.821	63.657
2	0.816	1.886	2.920	4.303	6.965	9.925
3	0.765	1.638	2.353	3.182	4.541	5.841
4	0.741	1.533	2.132	2.776	3.747	4.604
5	0.727	1.476	2.015	2.571	3.365	4.032
6	0.718	1.440	1.943	2.447	3.143	3.707

See Appendix B.2 (page 651)

LEARNING CHECK

1. Under what circumstances is a t statistic used instead of a z-score for a hypothesis test?

2. A sample of $n = 9$ scores has $SS = 288$.
   a. Compute the variance for the sample.
   b. Compute the estimated standard error for the sample mean.

3. In general, a distribution of t statistics is flatter and more spread out than the standard normal distribution. (True or false?)

4. A researcher reports a t statistic with $df = 20$. How many individuals participated in the study?

5. For $df = 15$, find the value(s) of t associated with each of the following:
   a. The top 5% of the distribution.
   b. The middle 95% of the distribution.
   c. The middle 99% of the distribution.

ANSWERS

1. A t statistic is used instead of a z-score when the population standard deviation and variance are not known.

2. a. $s^2 = 36$
   b. $s_M = 2$

3. True.

4. $n = 21$

5. a. $t = +1.753$
   b. $t = \pm 2.131$
   c. $t = \pm 2.947$

Example

Web page users prefer to look at attractive GUI compared to less attractive faces (Albayrak, 2000). In the study, participants from 18 to 25 years old were shown two different GUI of web pages. Previously, a group of adults had rated one of the GUI as significantly more attractive than the other. The participants were positioned in front of a screen. The pair of GUI remained on the screen until the participant accumulated a total of 20 seconds of looking at one or the other. The number of seconds looking at the attractive GUI was recorded for each participant. Suppose that the study used a sample of n = 9 participant and the data produced an average of M =13 seconds for the attractive GUI with SS = 72. Note that all of the available information comes from the sample. Specifically, we do not know the population mean or the population standard deviation.

T-STATISTIC

HYPOTHESIS TESTING WITH THE *t* STATISTIC

STEP 1. State the hypotheses

STEP 2. Set the criteria for a decision

STEP 3. Collect data and compute test statistic

STEP 4. Make a decision

T-STATISTIC

STEP 1: State the hypotheses

The *null hypothesis* H_0: $\mu_{\text{attractive GUI}} = 10$

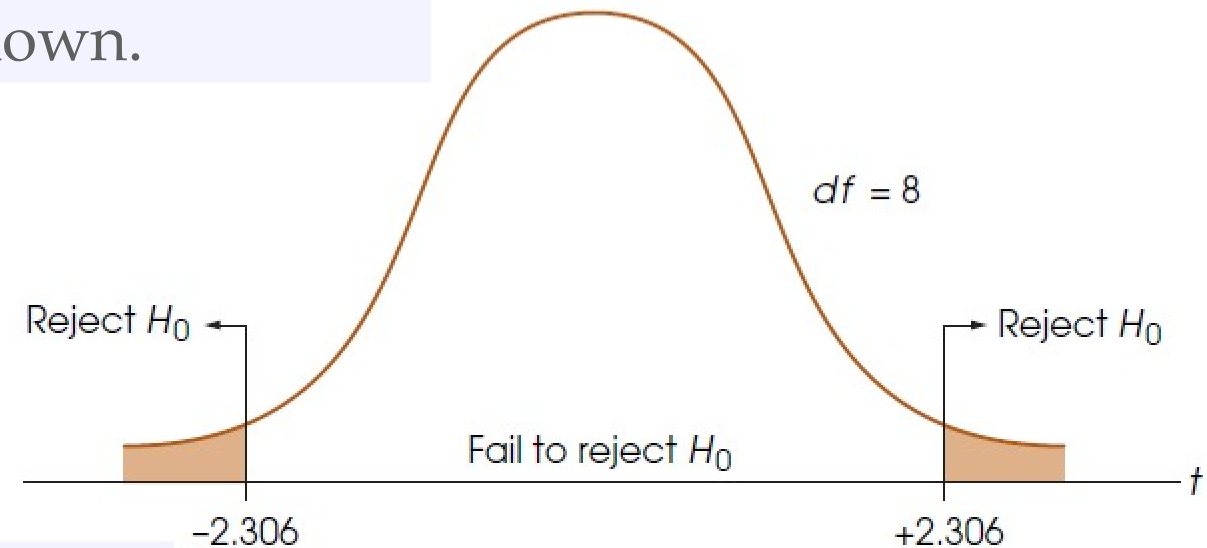
The *alternative hypothesis* H_1: $\mu_{\text{attractive GUI}} \neq 10$

$\alpha = 0.05$ for two tails

T-STATISTIC

STEP 2: Set the criteria for the decision

We will use t-test rather than z, because σ is unknown.



$df = n – 1 = 9 – 1 = 8$

T-STATISTIC

STEP 3: Collect data and compute test statistic

Sample data: M = 13 and SS = 72

$$s^2 = \frac{SS}{n-1} \qquad s_M = \sqrt{\frac{s^2}{n}} \qquad t = \frac{M-\mu}{s_M}$$

T-STATISTIC

STEP 3: Collect data and compute test statistic

Sample data: M = 13 and SS = 72

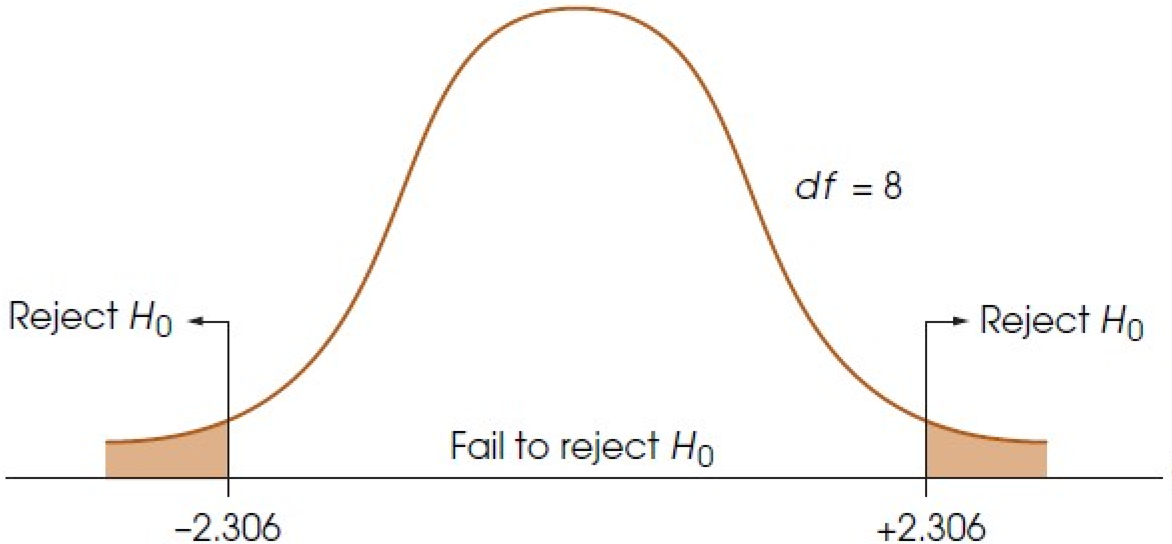
$$s^2 = \frac{SS}{n-1} = \frac{72}{8} = 9$$

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{9}{9}} = 1$$

$$t = \frac{M - \mu}{s_M} = \frac{13 - 10}{1} = 3$$

T-STATISTIC

STEP 4: Make a decision



The calculated t statistic (3.0) is greater than the critical t value (2.306). Thus, we reject the null hypothesis and conclude that there is a tendency for the web users to spend more time looking at the attractive GUI.

T-STATISTIC

HYPOTHESIS TESTING WITH THE *t* STATISTIC

Example 1. The average test score of 5th graders is 86. A researcher claims that the average test score of children from low SES families is not different from that of the population. To test whether this claim is true, a sample of 16 children from low SES families was selected and their test scores were secured.

T-STATISTIC

HYPOTHESIS TESTING WITH THE *t* STATISTIC

STEP 1. State the hypotheses

STEP 2. Set the criteria for a decision

STEP 3. Collect data and compute test statistic

STEP 4. Make a decision

T-STATISTIC

STEP 1: State the hypotheses

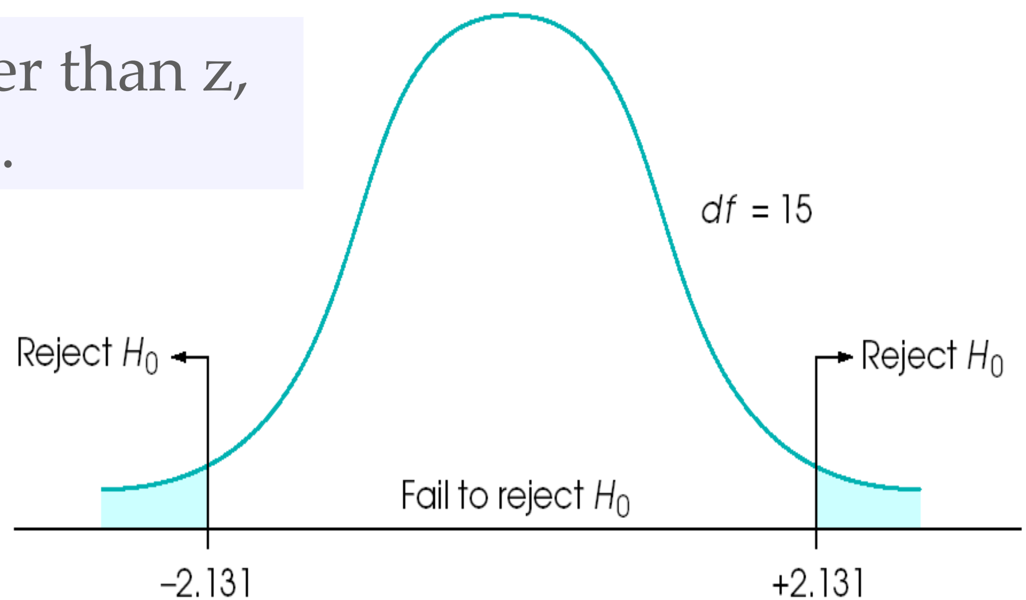
The *null hypothesis* H_0: $\mu_{\text{low SES}} = 86$

The *alternative hypothesis* H_1: $\mu_{\text{low SES}} \neq 86$

T-STATISTIC

STEP 2: Set the criteria for the decision

We will use t-test rather than z, because σ is unknown.



df = n – 1 = 16 – 1 = 15

T-STATISTIC

STEP 3: Collect data and compute test statistic

Sample data: M = 84 and SS = 240

$$s^2 = \frac{SS}{n-1} \qquad s_M = \sqrt{\frac{s^2}{n}} \qquad t = \frac{M-\mu}{s_M}$$

T-STATISTIC

STEP 3: Collect data and compute test statistic

Sample data: M = 84 and SS = 240

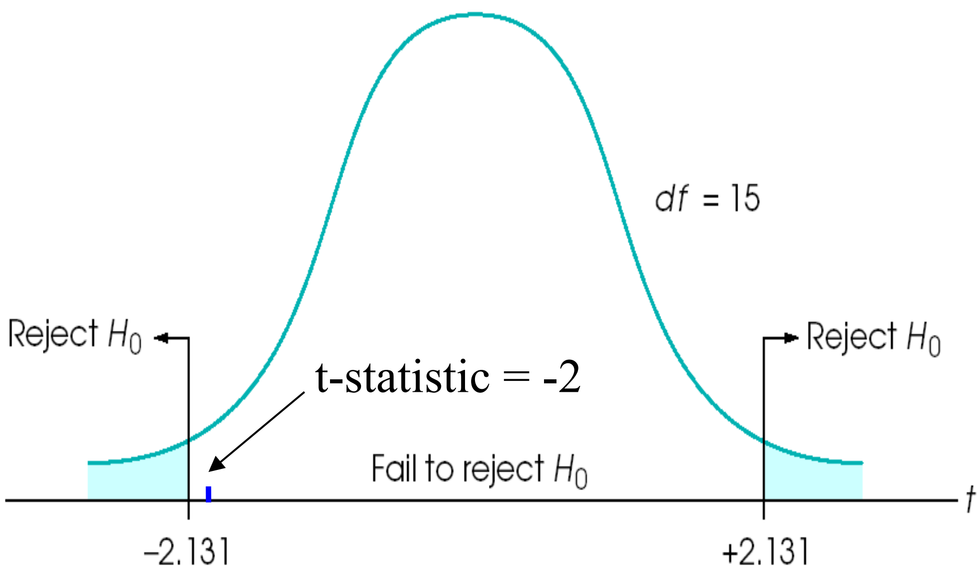
$$s^2 = \frac{SS}{n-1} = \frac{240}{15} = 16$$

$$t = \frac{M-\mu}{s_M} = \frac{84-86}{1} = -2$$

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{16}{16}} = 1$$

T-STATISTIC

STEP 4: Make a decision



The calculated t statistic (-2.0) is less than the critical t value (-2.131). Thus, we fail to reject the null hypothesis.

T-STATISTIC

EFFECT SIZE

Cohen's d = |mean difference|/ sample standard deviation

	Evaluation of Effect Size
$0<d<.2$	Small effect (mean difference less than .2 standard deviation
$.2<d<.8$	Medium effect
$d>.8$	Large effect (mean difference greater than .8 standard deviation

T-STATISTIC

EFFECT SIZE

Cohen’s d = mean difference / sample standard deviation

For the previous example, Cohen’s d can be calculated as:

$$Cohen's\ d = \frac{|84 - 86|}{4} = 0.5$$

According to standards suggested by Cohen, this is a medium effect.

T-STATISTIC

Reporting the results of a t-test

In this study, it was predicted that the average test score of children from low SES families is not different from that of the population. A one-sample t-test was used to test this hypothesis, and indicates that there is no significant difference in the average test scores between children in general and those from low SES families (t (15)=-2.0, p <.05, d = .5).

T-STATISTIC

Example 2. A psychologist has prepared an "Optimism Test" that is administered to graduating college seniors. The test measures how each graduating class feel about its future. Last year's class had a mean score of $\mu = 15$. A sample of $n = 9$ seniors from this year's class was selected and tested. Psychologist is interested in whether this year's class has a different level of optimism than last year's class?

T-STATISTIC

STEP 1: State the hypotheses

The *null hypothesis* $H_0: \mu = 15$

The *alternative hypothesis* $H_1: \mu \neq 15$

T-STATISTIC

STEP 2: Set the criteria for the decision

We will use t-test rather than z, because σ is unknown.

$df = n - 1 = 9 - 1 = 8$

For a two-tailed test with $\alpha = .05$ and $df = 8$, the critical t-values are ± 2.306.

T-STATISTIC

STEP 3: Collect data and compute test statistic

Sample data: 7, 12, 11, 15, 7, 8, 15, 9, 6

For the t formula, we need to compute sample mean and estimated standard error

$$t = \frac{M - \mu}{s_M}$$

T-STATISTIC

STEP 3: Collect data and compute test statistic

$$M = \frac{\sum X}{n} = \frac{90}{9} = 10$$

$$s^2 = \frac{SS}{n-1} = \frac{94}{8} = 11.75$$

$$s_M = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{11.75}{9}} = 1.14$$

$$t = \frac{M - \mu}{s_M} = \frac{10 - 15}{1.14} = -4.39$$

T-STATISTIC

STEP 4: Make a decision

The calculated t statistic (-4.39) is smaller than the critical t value (-2.306). Thus, we reject the null hypothesis and conclude that there is a significant difference in the level of optimism between this year's and last year's graduating classes, $t\ (8) = -\ 4.39$, $p < .05$, two-tailed. The effect size was large, $d = 1.46$.

T-STATISTIC

ASSUMPTIONS OF THE *t* TEST

- Random sampling
- Independent observation
- Normality – the population sampled must be normal