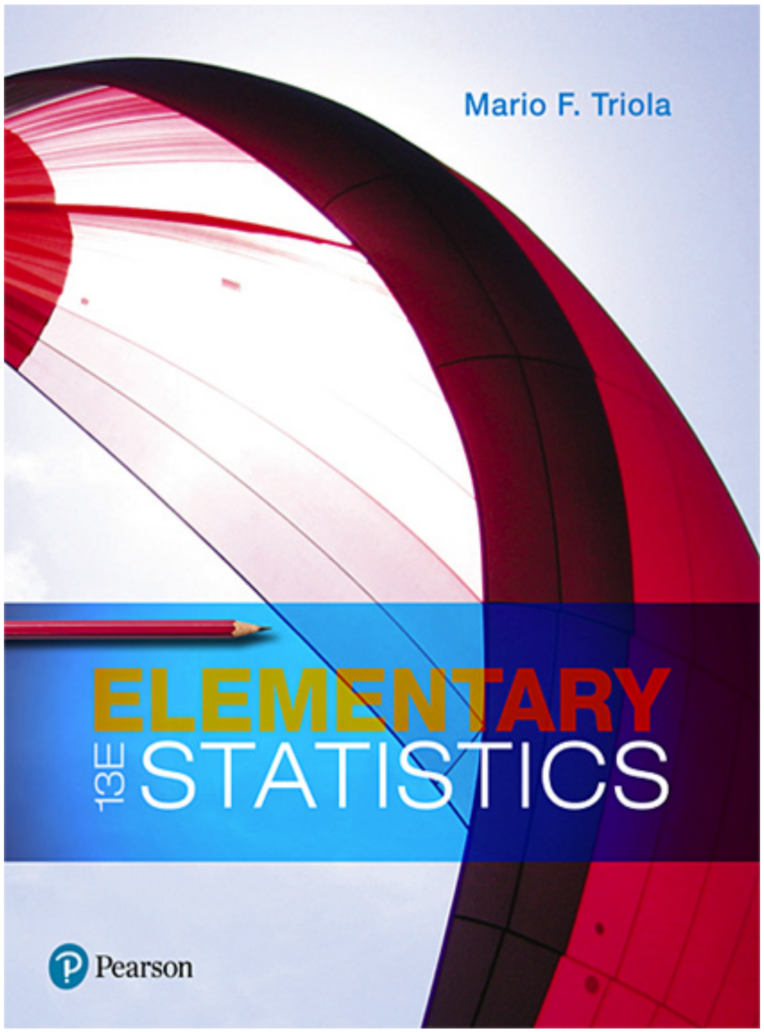
Elementary Statistics

Thirteenth Edition



Chapter 8

Hypothesis Testing

Hypothesis Testing

8-1 Basics of Hypothesis Testing

8-2 Testing a Claim about a Proportion

8-3 Testing a Claim About a Mean

8-4 Testing a Claim About a Standard Deviation or Variance

Key Concept

Testing a claim about a population mean is one of the most important methods presented in this book.

Part 1 of this section deals with the very realistic and commonly used case in which the population standard deviation σ is not known.

Part 2 includes a brief discussion of the procedure used when σ is known, which is very rare.

Testing Claims About a Population Mean with σ Not Known: Objective

Objective

Use a formal hypothesis test to test a claim about a population mean μ.

Testing Claims About a Population Mean with σ Not Known: Notation

Notation

n = sample size

$\bar{x}$ = **sample** mean

s = **sample** standard deviation

$\mu_{\bar{x}}$ = **population** mean

Testing Claims About a Population Mean with σ Not Known: Requirements

Requirements

1. The sample is a simple random sample.
2. Either or both of these conditions are satisfied: The population is normally distributed or $n > 30$.

Test Statistic for Testing a Claim About a Mean

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$$

- ***P*-values:** Use technology or use the Student *t* distribution (Table A-3) with degrees of freedom given by df = $n - 1$.
- **Critical values:** Use the Student *t* distribution (Table A-3) with degrees of freedom given by df = $n - 1$.

Requirement of Normality or $n > 30$

- If the original population is not itself normally distributed, we use the condition $n > 30$ for justifying use of the normal distribution.
- Sample sizes of 15 to 30 are sufficient if the population has a distribution that is not far from normal.
- In this text we use the simplified criterion of $n > 30$ as justification for treating the distribution of sample means as a normal distribution, regardless of how far the distribution departs from a normal distribution.

Important Properties of the Student *t* Distribution

1. The Student *t* distribution is different for different sample sizes.
2. The Student *t* distribution has the same general bell shape as the standard normal distribution; its wider shape reflects the greater variability that is expected when *s* is used to estimate σ.
3. The Student *t* distribution has a mean of $t = 0$.
4. The standard deviation of the Student *t* distribution varies with the sample size and is greater than 1.
5. As the sample size *n* gets larger, the Student *t* distribution gets closer to the standard normal distribution.

Example: Adult Sleep (1 of 8)

The author obtained times of sleep for randomly selected adult subjects included in the National Health and Nutrition Examination Study, and those times (hours) are listed below. Here are the unrounded statistics for this sample: $n = 12$, $\bar{x} = 6.83333333$ hours, $s = 1.99240984$ hours.

A common recommendation is that adults should sleep between 7 hours and 9 hours each night. Use the *P*-value method with a 0.05 significance level to test the claim that the mean amount of sleep for adults is less than 7 hours.

4 8 4 4 8 6 9 7 7 10 7 8

Example: Adult Sleep (2 of 8)

Solution

Requirement Check

(1) The sample is a simple random sample.

(2) The second requirement is that “the population is normally distributed or $n > 30$.” The sample size is $n = 12$, which does not exceed 30, so we must determine whether the sample data appear to be from a normally distributed population.

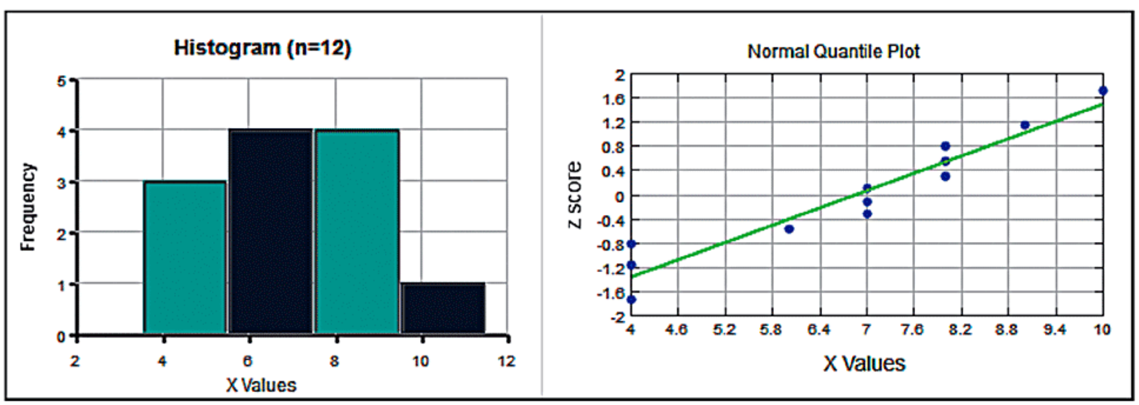
Example: Adult Sleep (3 of 8)

Solution

Requirement Check

The accompanying histogram and normal quantile plot, along with the apparent absence of outliers, indicate that the sample appears to be from a population with a distribution that is approximately normal. Both requirements are satisfied.

Statdisk



Pearson

Example: Adult Sleep (4 of 8)

Solution

Step 1: The claim that "the mean amount of adult sleep is less than 7 hours" becomes $\mu < 7$ hours.

Step 2: The alternative to the original claim is $\mu \geq 7$ hours.

Step 3: Because the statement $\mu < 7$ hours does not contain the condition of equality, it becomes the alternative hypothesis H_1. The null hypothesis H_0 is the statement that $\mu = 7$ hours.

- H_0: $\mu = 7$ hours (null hypothesis)
- H_1: $\mu < 7$ hours (alternative hypothesis and original claim)

Example: Adult Sleep (5 of 8)

Solution

Step 4: As specified in the statement of the problem, the significance level is $\alpha = 0.05$.

Step 5: Because the claim is made about the **population mean** μ, the sample statistic most relevant to this test is the **sample mean** $\bar{x}$, and we use the *t* distribution.

Example: Adult Sleep (6 of 8)

Solution

Step 6: The sample statistics of $n = 12$, $\bar{x} = 6.83333333$ hours, $s = 1.99240984$ hours are used to calculate the test statistic as follows, but technologies provide the test statistic of $t = -0.290$. In calculations such as the following, it is good to carry extra decimal places and not round.

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{6.83333333 - 7}{\frac{1.99240984}{\sqrt{12}}} = -0.290$$

Example: Adult Sleep (7 of 8)

Solution

***P*-Value with Technology**

We could use technology to obtain the *P*-value. The *P*-value is 0.3887 (rounded).

TI-83/84 Plus

NORMAL FLOAT AUTO REAL RADIAN MP

T-Test
$\mu<7$
t=-.2897748534
p=.3886888459
$\bar{x}$=6.833333333
Sx=1.99240984
n=12

Statdisk

t Test
Test Statistic, t: -0.2898
Critical t: -1.7959
P-Value: 0.3887

90% Confidence interval:
5.800414 < μ < 7.866252

Excel (XLSTAT)

Difference	-0.1667
t (Observed value)	-0.2898
t (Critical value)	-1.7959
DF	11
p-value (one-tailed)	0.3887
alpha	0.05

Minitab

Test of μ = 7 vs < 7

Variable	N	Mean	StDev	SE Mean	95% Upper Bound	T	P
Sleep	12	6.833	1.992	0.575	7.866	−0.29	0.389

JMP

Hypothesized Value	7
Actual Estimate	6.83333
DF	11
Std Dev	1.99241
t Test	
Test Statistic	-0.2898
Prob > |t|	0.7774
Prob > t	0.6113
Prob < t	0.3887

StatCrunch

Hypothesis test results:
μ : Mean of variable
$H_0 : \mu = 7$
$H_A : \mu < 7$

Variable	Sample Mean	Std. Err.	DF	T-Stat	P-value
Sleep	6.8333333	0.57515918	11	-0.28977485	0.3887

SPSS

	Test Value = 7					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
SLEEP	-.290	11	.777	-.16667	-1.4326	1.0993

Example: Adult Sleep (8 of 8)

Solution

Step 7: Because the *P*-value of 0.3887 is greater than the significance level of $\alpha = 0.05$, we fail to reject the null hypothesis.

Interpretation

Step 8: Because we fail to reject the null hypothesis, we conclude that there is not sufficient evidence to support the claim that the mean amount of adult sleep is less than 7 hours.

P-Value Method without Technology

If suitable technology is not available, we can use Table A-3 to identify a **range of values** containing the *P*-value.

In using Table A-3, keep in mind that it is designed for positive values of *t* and right-tail areas only, but left-tail areas correspond to the same *t* values with negative signs.

Example: Adult Sleep: *P*-Value Method without Technology

The previous example is a left-tailed test with a test statistic of $t = -0.290$ and a sample size of $n = 12$, so the number of degrees of freedom is $\text{df} = n - 1 = 11$.

Using the test statistic of $t = -0.290$ with Table A-3, examine the values of t in the row for df = 11 to see that 0.290 is less than all of the listed t values in the row, which indicates that the area in the left tail below the test statistic of $t = -0.290$ is greater than 0.10.

In this case, Table A-3 allows us to conclude that the *P*-value > 0.10, but technology provided the *P*-value of 0.3887. With the *P*-value > 0.10, the conclusions are the same as in the previous example.

Example: Adult Sleep: Critical Value Method (1 of 2)

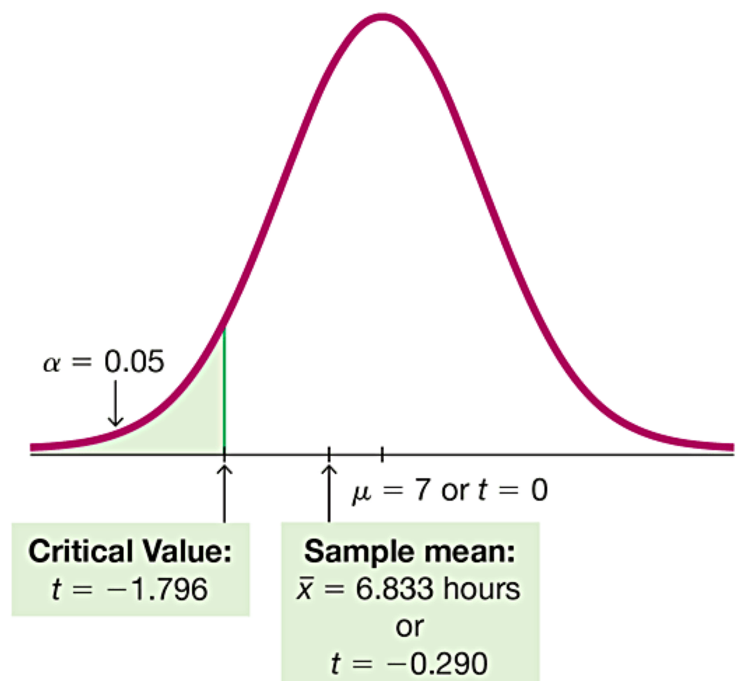
Example 1 is a left-tailed test with test statistic $t = -0.290$. The sample size is $n = 12$, so the number of degrees of freedom is df = $n - 1 = 11$.

Given the significance level of $a = 0.05$, refer to the row of Table A-3 corresponding to 11 degrees of freedom, and refer to the column identifying an "area in one tail" of 0.05. The intersection of the row and column yields the critical value of $t = 1.796$, but this test is left-tailed, so the actual critical value is $t = -1.796$.

Example: Adult Sleep: Critical Value Method (2 of 2)

The figure shows that the test statistic of $t = -0.290$ does not fall within the critical region bounded by the critical value $t = -1.796$, so we fail to reject the null hypothesis.

The conclusions are the same as those given in first example.



Example: Adult Sleep: Confidence Interval Method

The first example is a left-tailed test with significance level $\alpha = 0.05$, so we should use 90% as the confidence level.

For the sample data given in the first example, here is the 90% confidence interval estimate of μ: 5.80 hours < μ < 7.87 hours.

In testing the claim that $\mu < 7$ hours, we use H_0: $\mu = 7$ hours, but the assumed value of $\mu = 7$ hours is contained within the confidence interval limits, so the confidence interval is telling us that 7 hours could be the value of μ. We don't have sufficient evidence to reject H_0: $\mu = 7$ hours, so we fail to reject this null hypothesis and we get the same conclusions.

Example: Is the Mean Body Temperature Really 98.6°F? (1 of 8)

Data Set 3 "Body Temperatures" in Appendix B includes measured body temperatures with these statistics for 12 AM on day 2: $n = 106$, $\bar{x} = 98.20$°F, $s = 0.62$°F. Use a 0.05 significance level to test the common belief that the population mean is 98.6°F.

Example: Is the Mean Body Temperature Really 98.6°F? (2 of 8)

Solution

Requirement Check

(1) With the study design used, we can treat the sample as a simple random sample.

(2) The second requirement is that “the population is normally distributed or $n > 30$.” The sample size is $n = 106$, so the second requirement is satisfied and there is no need to investigate the normality of the data. Both requirements are satisfied.

Example: Is the Mean Body Temperature Really 98.6°F? (3 of 8)

Solution

Step 1: The claim that "the population mean is 98.6°F" becomes $\mu = 98.6°F$ when expressed in symbolic form.

Step 2: The alternative to the original claim is $\mu \neq 98.6°F$.

Step 3: Because the statement $\mu \neq 98.6°F$ does not contain the condition of equality, it becomes the alternative hypothesis H_1. The null hypothesis H_0 is the statement that $\mu = 98.6°F$.

H_0: $\mu = 98.6°F$ (null hypothesis and original claim)

H_1: $\mu \neq 98.6°F$ (alternative hypothesis)

Example: Is the Mean Body Temperature Really 98.6°F? (4 of 8)

Solution

Step 4: As specified in the statement of the problem, the significance level is $\alpha = 0.05$.

Step 5: Because the claim is made about the **population mean** μ, the sample statistic most relevant to this test is the **sample mean** $\bar{x}$. We use the t distribution because the relevant sample statistic is $\bar{x}$ and the requirements for using the t distribution are satisfied.

Example: Is the Mean Body Temperature Really 98.6°F? (5 of 8)

Solution

Step 6: The sample statistics are used to calculate the test statistic as follows, but technologies use unrounded values to provide the test statistic of $t = -6.61$.

$$t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{98.20 - 98.6}{\frac{0.62}{\sqrt{106}}} = -6.64$$

Example: Is the Mean Body Temperature Really 98.6°F? (6 of 8)

Solution

***P*-Value** The *P*-value is 0.0000 or 0 + (or “less than 0.01” if using Table A-3).

Critical Values: The critical values are ±1.983 (or ±1.984 if using Table A-3).

Confidence Interval: The 95% confidence interval is 98.08°F < μ < 98.32°F.

Example: Is the Mean Body Temperature Really 98.6°F? (7 of 8)

Solution

Step 7: All three approaches lead to the same conclusion: Reject H_0.

- ***P*-Value:** The *P*-value of 0.0000 is less than the significance level of $\alpha = 0.05$.
- **Critical Values:** The test statistic $t = -6.64$ falls in the critical region bounded by ±1.983.
- **Confidence Interval:** The claimed mean of 98.6°F does not fall within the confidence interval of $98.08°F < \mu < 98.32°F$.

Example: Is the Mean Body Temperature Really 98.6°F? (8 of 8)

Interpretation

Step 8: There is sufficient evidence to warrant **rejection** of the common belief that the population mean is 98.6°F.

Alternative Methods Used When Population Is Not Normal and $n \leq 30$

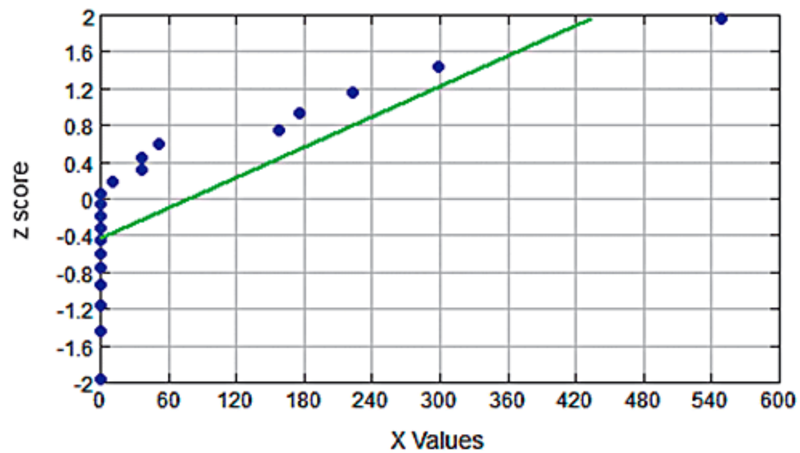
- **Bootstrap Resampling** Use the confidence interval method of testing hypotheses, but obtain the confidence interval using bootstrap resampling. Be careful to use the appropriate confidence level. Reject the null hypothesis if the confidence interval limits do not contain the value of the mean claimed in the null hypothesis.
- **Sign Test** See Section 13-2.
- **Wilcoxon Signed-Ranks Test** See Section 13-3.

Example: Bootstrap Resampling (1 of 3)

Listed below is a random sample of times (seconds) of tobacco use in animated children's movies. Use a 0.05 significance level to test the claim that the sample is from a population with a mean greater than 1 minute, or 60 seconds.

0	223	0	176	0	548	0	37	158			
51	0	0	299	37	0	11	0	0	0	0	

Example: Bootstrap Resampling (2 of 3)



Solution

Requirement Check

The *t* test described in Part 1 of this section requires that the population is normally distributed or $n > 30$, but we have $n = 20$ and the accompanying normal quantile plot shows that the sample does not appear to be from a normally distributed population. The *t* test should **not** be used.

Example: Bootstrap Resampling (3 of 3)

Solution

We use the bootstrap resampling method. After obtaining 1000 bootstrap samples and finding the mean of each sample, we sort the means. Because the test is right-tailed with a 0.05 significance level, we use the 1000 sorted sample means to find the 90% confidence interval limits of $P_5 = 29.9$ sec. and $P_{95} = 132.9$ sec. The 90% confidence interval is 29.9 seconds $< \mu <$ 132.9 seconds. Because the assumed mean of 60 seconds is contained within those confidence interval limits, we fail to reject H_0: $\mu = 60$ seconds. There is not sufficient evidence to support H_1: $\mu > 60$ seconds.

Testing a Claim about a Mean (When σ is Known): Test Statistic

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$$

- ***P*-value:** Provided by technology, or use the standard normal distribution (Table A-2) with the procedure in Figure 8-3.
- **Critical values:** Use the standard normal distribution (Table A-2).