CHAPTER 9

CONFIDENCE INTERVALS AND HYPOTHESIS TESTING

OBJECTIVES

After completing this chapter, you should

- understand the derivation and characteristics of the sampling distribution of means and the related distribution of *t* scores.
- be able to compute confidence intervals and perform a one-sample *t* test.

CHAPTER REVIEW

The *sampling distribution of means* is derived by taking successive, same-sized random samples from some population, computing a mean for a measurable characteristic of each sample, and plotting the means on a frequency polygon. The *first property* of the sampling distribution is that its mean is the mean of the population or μ. The *second property,* which is a simplified version of the *central limit theorem,* is that the larger the size of each sample, the more nearly the sampling distribution will approximate the normal curve. The *third property* is that the larger the sample size, the smaller the standard deviation of the sampling distribution. The standard deviation is called the *standard error of the mean.*

When population parameters (μ and σ) are known, *z* scores are appropriate. However, we often do not know them, so they must be estimated from sample values. The sample mean (X) is an unbiased estimate of μ, and, as we defined it in Chapter 6, sample variance (s^2) is an unbiased estimate of population variance (σ^2). Recall that to obtain an unbiased estimate of σ^2, we divided the sum of squared deviations by $N-1$ rather than by *N*. This expression, $N-1$, is called *degrees of freedom* and is defined as the number of values free to vary after certain restrictions (e.g., the sum of the deviations equals 0) are placed on the data.

A *confidence interval* is a range of values within which the population mean almost certainly lies. The confidence intervals usually computed are the 95% and the 99% confidence intervals. Equations for the confidence intervals are derived in the chapter from the formula introduced in Chapter 6 to convert *z* scores to raw scores.

t scores are estimated *z* scores. They are used in place of *z* scores when population parameters are estimated from the sample. *t* scores correspond to the *t* distribution, and the *t* scores used in the confidence interval equations are determined from Table B (see Appendix 2), which contains values of *t* cutting off deviant portions of the distribution. In order to use the table of critical *t* scores, we need to know the *df*. For confidence intervals and the one-sample *t* test, $df = N - 1$. A confidence interval is an *interval estimate* of the population mean.

Another important use of the distribution of *t* is to test hypotheses about the value of μ. The seven-step procedure introduced for testing the null hypothesis is as follows:

1. **State the null hypothesis in symbols (H_0: $\mu = \mu_0$) and in words** in the context of the problem.
2. **State the alternative hypothesis in symbols (e.g., H_1: $\mu \neq \mu_0$) and in words**.
3. **Choose an α level,** the level at which you will reject or fail to reject the null hypothesis. Set $\alpha = .05$, if there are no specific instructions in the problem.
4. **State the rejection rule.** For example, the rule for a particular problem may be as follows: If $|t_{comp}|$ is $\geq t_{crit}$, then reject the null hypothesis. This is the rejection rule for a nondirectional hypothesis.
5. **Compute the test statistic.** In this chapter, the equation for the test statistic is $t_{\bar{X}} = \dfrac{\bar{X} - \mu}{s_{\bar{X}}}$.
6. **Make a decision by applying the rejection rule.**
7. **Write a conclusion statement in the context of the problem.**

For a directional test, H_0 is rejected if t_{comp} is of the same sign but is more extreme than t_{crit}.

A *Type I,* or α, *error* (a "false claim") is defined as rejecting the null hypothesis when it is really true. The probability of an α error is equal to α, the level at which we are trying to reject the null hypothesis.

Failing to reject H_0 when it is false is called a *Type II,* or β, *error* (a "failure of detection"). Lowering the value of α increases the probability of a β error.

The *power* of a statistical test is the probability that the test will detect a false null hypothesis. Factors affecting the power of a test are

1. The size of α. The smaller the α level, the less powerful the test will be.
2. The sample size. The larger the sample size, the greater the power of the test will be.
3. The distance between the hypothesized mean and the true mean. The greater the distance, the greater the power of the test will be.

In analyzing the results from large numbers of research studies, meta-analysis uses quantitative procedures to integrate the findings. Meta-analysis uses the *effect size*—the size of the difference between the null hypothesis and the alternative hypothesis in standardized units—results of studies rather than simply reporting whether or not the results were statistically significant.

Some researchers argue that hypothesis testing should be abandoned, claiming that the procedures are misleading and we should instead report confidence intervals and effect sizes. People opposed to hypothesis testing hold that many research studies don't have enough power to find what they are looking for, with a corresponding increase in Type II errors. This point is valuable if it forces experimenters to be more attentive to having sufficient power to detect an effect if it is present.

SYMBOLS

Symbol	Stands For
$\sigma_{\bar{x}}$	standard error of the mean
$\mu_{\bar{x}}$	mean of the sampling distribution of means, which equals μ
$z_{\bar{x}}$	z score for the sampling distribution of means
$s_{\bar{x}}$	estimated standard error of the mean
$t_{\bar{x}}$ or t	t score, which is an estimate of a z score
CI	confidence interval
df	degrees of freedom
$t_{.05}$ or $t_{.01}$	t scores from Table B cutting off deviant 5% or 1% of the distribution [occur with probability of .05(.01) or less]
H_0	null hypothesis
α	alpha level, the level at which we test H_0
H_1	alternative hypothesis
μ_0	specific value representing the "untreated" population mean
t_{comp}	computed t score
t_{crit}	critical t score from Table B
Type I, or α, error	rejecting true H_0
Type II, or β, error	failing to reject false H_0

FORMULAS

Formula 9-4. *Equation for estimated standard error of the mean*

$$s_{\bar{x}} = \frac{s}{\sqrt{N}}$$

The estimated standard error is found by dividing the sample standard deviation by the square root of sample size.

Formula 9-5. *Equation for $t_{\bar{x}}$, which is an estimate of $z_{\bar{x}}$*

$$t_{\bar{X}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

This equation is used to test hypotheses about the value of μ. It is the formula for the one-sample t test.

Formulas 9-6 and 9-7. *Equations for 95% and 99% confidence intervals*

95% CI $= \pm\, t_{.05} s_{\bar{X}} + \bar{X}$

99% CI $= ;\pm t_{.01} s_{\bar{X}} + \bar{X}$

$t_{.05}$ and $t_{.01}$ are the t scores cutting off the deviant 5% and 1% of the distribution of t, respectively. The values of t are found in Table B with $df = N - 1$.

TERMS TO DEFINE AND/OR IDENTIFY

sampling distribution of means

parent population

central limit theorem

standard error of the mean

degrees of freedom

estimated standard error of the mean

t distribution

confidence interval

critical values of *t*

interval estimate

null hypothesis

alternative hypothesis

nondirectional hypothesis

directional hypothesis

significant

alpha level

rejection rule

Type I error

Type II error

power

meta-analysis

effect size

FILL-IN-THE-BLANK ITEMS

Introduction

Intuitively, we understand that most of the statistics we are given are only (1) _____ because

they are based on a sample from the larger group of interest—the (2) _____. In this chapter,

we discuss the process of (3) _____ and how to determine the range within which our

(4) _____ should fall.

The Sampling Distribution of Means

The sample mean is an (5) _____ estimate of the population mean. The sampling

distribution of means is derived by extracting successive random samples, all with the same

(6) _____, from some population. For each sample, the mean of some characteristic is

computed and the (7) _____ are plotted on a (8) _____ polygon. The

resulting polygon is called the (9) _____ _____ _____

_____.

The properties of the sampling distribution of means are as follows:

1. The mean of the sampling distribution equals (10) _____.

2. The larger the size of each sample taken from the parent population, the more nearly the sampling

   distribution approximates the (11) _____ curve. This property is a simplified version of

   the (12) _____ _____ _____ .

3. The larger the size of each sample taken from the population, the smaller the

(13) _____ _____ of the sampling distribution. The standard deviation is

called the (14) _____ _____ of the mean and is symbolized by

(15) _____ .

The equation for a *z* score based on the sampling distribution of means is (16) _____ .

z scores obtained for a sample mean can be used in the same way as *z* scores for a

(17) _____ _____ .

Estimation and Degrees of Freedom

Often, we must estimate population values for μ and σ from our (18) _____ . We can use (19)

_____ to estimate the population mean and (20) _____ to estimate the

population standard deviation. As you recall from Chapter 6, in the equation for our unbiased estimate of

population variance, we divided the sum of squared deviations by (21) _____ rather than by

N because of the tendency of the equation with *N* in the denominator to (22) _____ either the

population variance or the population standard deviation.

$N - 1$ is referred to as (23) _____ _____ _____ ,

which is defined as the number of (24) _____ free to vary after certain

(25) _____ have been placed on the data.

A *t* score is an estimated (26) _____ and corresponds to a

(27) _____ distribution. The mathematics of the distribution were derived by William Sealy

(28) _____ , who published under the pseudonym (29) _____ .

Confidence Intervals

A (30) _____ _____ is a range of values around a sample mean within which

μ almost certainly lies. The confidence intervals usually computed are the 95% and the

(31) _____. The equations for the confidence intervals are derived from the formula used to

convert (32) _____ to raw scores. Instead of z scores, the confidence interval equation

requires (33) _____ obtained from Table (34) _____.

   For confidence intervals, df = (35) _____. The t distribution changes shape with

changes in (36) _____ _____. Rather than being an exact estimate of the

population mean, the confidence interval is an (37) _____ estimate.

Hypothesis Testing: One-Sample *t* Test

The one-sample t test is a procedure for testing the (38) _____ _____. The

null hypothesis, symbolized by (39) _____ , assumes a particular value for a population

parameter—in this case, for (40) _____, the mean of the sampling distribution of means. The

alternative to the null hypothesis is that the value of (41) _____ is something other than what

we have assumed it to be. If the alternative hypothesis, symbolized by (42) _____, doesn't

specify the direction in which H_0 will differ from μ, we say it is (43) _____. On the other

hand, an alternative hypothesis stating that μ will either be greater than H_0 or less than H_0 is called a

(44) _____ hypothesis.

   The seven-step procedure for testing the null hypothesis is as follows:

1. State the (45) _____ hypothesis in symbols and words.

2. State the alternative hypothesis in symbols and words.

3. Choose an (46) _____ level, which will always be set to .05 or .01 unless there are some

   special circumstances. Set α = (47) _____, if there are no other instructions in the

   problem.

4. State the (48) _____ rule.

5. Compute the (49) _____ statistic.

6. Make a (50) _____ by applying the rejection rule.

7. Write a (51) _____ statement in the (52) _____ of the problem.

Directional tests

For a directional test, t_{comp} should have the (53) _____ sign as t_{crit}. In addition, with a

directional test, t_{crit} should be (54) _____ extreme than for a nondirectional test, because all

of the probability is placed in (55) _____ _____ of the distribution. For this

reason, directional tests are (56) _____ powerful than nondirectional tests but hazardous if

the (57) _____ of the outcome cannot be predicted in advance.

Type I and Type II errors

The process of rejecting or failing to reject H_0 is sometimes called (58) _____

_____ _____. Rejecting H_0 when it is true is called a Type (59)

_____ or

(60) _____ error. The probability of committing this type of error is determined by the value

we set for (61) _____. Lowering the value of alpha will (62) _____ the

probability of this type of error.

Failing to reject a false null hypothesis is called a Type (63) _____ or

(64) _____ error. Although the probability of this type of error is unknown, it is increased by

(65) _____ in the value of alpha.

The power of a statistical test

The (66) _____ of a test is the probability that the test will detect a false hypothesis, given

by the equation (67) _____ = _____. Factors affecting power are the value of

(68) _____, the (69) _____ of the sample taken from the population, and the

distance between the hypothesized value of μ and the true value. Specifically, the smaller we set α, the

(70) _____ the power of the test will be. Also, the (71) _____ the sample

size, the greater the power of the test will be. Finally, the greater the distance between the hypothesized

value of μ and the true value, the (72) _____ the power of the test will be.

Meta-analysis

The magnitude of the difference between H_0 and H_1, called the (73) _____

_____, is the point of departure in the quantitative analysis of large numbers of research

studies using (74) _____. This form of analysis is more interested in

(75) _____ _____ than in whether a significant effect is present in a study.

Should hypothesis testing be abandoned?

Some researchers say we should (76) _____ hypothesis testing because the

(77) _____ _____ in psychology experiments is really much higher than most

researchers think. Anti-hypothesis testers claim that a large percentage of studies don't have enough

(78) _____ to detect an effect even when the effect is present. As a consequence, Type

(79) _____ errors are committed at high rates, sometimes as high as 60%. This point is valid

if it forces experimenters to be more attentive to having sufficient (80) _____ in their

experiments.

Troubleshooting Your Computations

When the confidence interval has been computed, look at it to be sure that it is (81) _____ in

the light of your data. For example, the confidence interval should contain the (82) _____ of

the sample. Be sure to use (83) _____ rather than N when finding the t score from Table B.

When computing t scores, the appropriate (84) _____ should be retained throughout the

computations. If the hypothesized mean is larger than the mean of the sample, the resulting value of t

should be (85) _____. Be sure that the absolute value of your computed t is larger than the

critical value of t from the table before (86) _____ H_0, if you're testing a nondirectional

hypothesis.

PROBLEMS

1. Find $s_{\bar{X}}$ for each of the following samples.

    a. $N = 37$, $s = 5.3$

    b. $N = 10$, $s = 2.5$

    c. $N = 93$, $\Sigma X = 1{,}032.3$, $\Sigma X^2 = 12{,}801.45$

    d. $df = 27$, $s^2 = 201.64$

    e. $N = 25$, $s = 10.75$

2. Use Table B to answer the following questions.

    a.   If $N = 10$, what t scores cut off the deviant 5% of the distribution?

    b.   If $df = \infty$, what t scores cut off the deviant 1% and 5% of the distribution? Are they similar to the z scores cutting off 1% and 5%? Explain.

    c.   For $N = 47$, what t scores cut off the deviant 5% and 1% of the distribution?

    d.   Why do the critical values of t decrease with increases in df?

e. According to the text, what critical values of t do you use when the exact df observed are not given in the table?

3. Find the 95% and the 99% confidence intervals for each of the following data sets.

   a. $N = 257$, $\Sigma X = 5{,}140$, $\Sigma X^2 = 106{,}912$

   b. $N = 26$, $\overline{X} = 10$, $s = 2$

   c. $N = 42$, $\Sigma X = 441$, $\Sigma X^2 = 4{,}914.42$

4. As part of its hiring procedure, a large company administers a standardized personality scale to job applicants. Fifty-four applicants for a quality control position have a mean score of 54.2, with $s = 16.1$, on the dimension of Conscientiousness. Assume that μ for Conscientiousness is 49.8.

   a. Determine whether applicants for the quality control position demonstrate higher Conscientiousness scores than the general population.

   b. Based on this sample of applicants, what is the 95% confidence interval for μ?

   c. What is the 99% confidence interval for μ?

5. A sample of 49 participants works at a perceptual task on which they have to correctly identify the shape of a stimulus after exposures of short duration. The number of correct identifications on 50 trials is recorded and the average is found to be 29.6, with $s = 7.3$.

   a. Construct the 95% confidence interval for μ.

   b. Construct the 99% confidence interval for μ.

6. Assume that the 327 students who have taken statistics at a large university constitute a population. Each student has been given a math achievement test with the following results: $\mu = 53.7$, $\sigma = 10.5$. On the basis of this information, answer the following questions.

   a. What is the standard error of the mean for samples of size $N = 25$?

   b. One sample of 25 students has been drawn from the population, and the average test score has been found to be 55.1, with $s = 8.5$. What is the estimated standard error?

   c. Test the null hypothesis using the sample described in part **b.**

   d. It is possible that you made an error in your decision in part **c.** If so, would it be a Type I or a Type II error?

7. The average rested worker at a calculator production plant can assemble 106 pocket calculators an hour. During the last hour of their shift, 26 workers assemble an average of 97.4 calculators, with $s = 17.2$. Is the performance of these workers significantly worse at the end of the shift?

8. In 20 years of coaching basketball, Coach Williams has kept records of her teams' performance making free throws during games. Her records indicate that the average player makes 71.1 shots out of 100. Not satisfied with that performance, Coach Williams hires a sports psychologist to work with the team to improve concentration and visualization at the foul line. At the end of the year, Coach Williams discovers that the 12 players on her team averaged 77.6 successful free throws per 100 attempts, with $s = 8.41$.

   a. What is $s_{\bar{X}}$ for the sample?

   b. Find the 95% confidence interval. Is 71.1 within the interval?

c. Test the hypothesis that $\mu = 71.1$.

d. Did working with the psychologist significantly improve free-throw shooting?

9. At a large high school, 537 seniors take the ACT with the following results: $\mu = 22.5$, $\sigma = 4.1$. Assuming that the 537 seniors constitute a population, answer the following questions.

a. Suppose that 50 samples of size $N = 10$ have been drawn with replacement from the population. The mean of the resulting sampling distribution is found to be 22.73, with $s = 4.05$. What is $s_{\bar{X}}$ for this sampling distribution?

b. What is $\sigma_{\bar{X}}$? How does it compare with $s_{\bar{X}}$ computed in part **a**?

c. Suppose we draw another sample of size $N = 10$ from the population and find its mean to be 20.85, with $s = 3.73$. Test the hypothesis that this sample was drawn from the original population with $\mu = 22.5$.

d. What is the 95% confidence interval for μ based on the sample in part **c**?

USING SPSS—EXAMPLE AND EXERCISE

SPSS has a specific, easy-to-use procedure for computing the one-sample t test.

Example: We will use SPSS to work Self-Test Exercise 7. The steps are as follows:

1. Start SPSS, enter the data, and name the variable **hypochon.**
2. *Analyze>Compare Means>One-Sample T Test.*
3. Move **hypochon** into the Test Variables box and enter 49.2 as the Test Value. Note that the Test Value is μ_0, the hypothesized value for μ for the null hypothesis.
4. Click *OK* and the solution should appear in the output Viewer window.
5. To obtain the 95% CI for μ, we must trick SPSS a bit and enter a Test Value of 0 and click *OK*.
6. Only the CI values should be read from this portion of the output.

Notes on Reading the Output

1. The column labeled "Sig. (2-tailed)" gives the exact p value for the computed $t = 2.425$. This means that $p = .034$, and we need not look up the critical values for t at the .05 or .01 levels. Our rule for rejecting H_0 can now be based on whether "Sig. (2-tailed)" or $p \leq .05$.
2. The 95% CI is the CI on the difference between the sample mean and the hypothesized mean. In order to obtain the correct CI, we must re-run the analysis with a Test Value set to 0. This portion of the output will give the correct CI, but the t value will not be correct and should be ignored.
   The solution output for the data of Self-Test Exercise 7 is as follows:

```
T-TEST
  /TESTVAL=49.2
  /MISSING=ANALYSIS
  /VARIABLES=hypochon
  /CRITERIA=CIN (.95) .
```

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
HYPOCHON	12	58.4167	13.1665	3.8008

Sig. (2-tailed) is the exact p value for the computed $t = 2.425$ and $p = .034$.

One-Sample Test

	Test Value = 49.2					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
HYPOCHON	2.425	11	.034	9.2167	.8511	17.5822

```
T-TEST
  /TESTVAL=0
  /MISSING=ANALYSIS
  /VARIABLES=hypochon
  /CRITERIA=CIN (.95) .
```

T-Test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
HYPOCHON	12	58.4167	13.1665	3.8008

Only the 95% CI is correct for the following output.

One-Sample Test

	Test Value = 0					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
HYPOCHON	15.369	11	.000	58.4167	50.0511	66.7822

Exercise Using SPSS

1. We have conducted a study of the verbal skills of females. The task was to unscramble 20 sentences within a 10-minute period. (*Example:* free are things best the life in—The best things in life are free.) Each of the 20 participants received a score indicating the number of sentences she unscrambled correctly. Several previous studies over the last 2 years have indicated that females averaged a score of 9.0 on the task. Using SPSS, test the hypothesis that this year's sample performed differently than in the past. Also provide a 95% CI for the population mean for this sample. The data are as follows: 15, 15, 14, 14, 13, 13, 13, 11, 11, 11, 11, 10, 10, 9, 9, 9, 8, 8, 6, 3. Write a brief conclusion for the hypothesis test. Also, interpret the confidence interval.

CHECKING YOUR PROGRESS: A SELF-TEST

1. The probability that an inferential test will detect a false null hypothesis is called the
    a. central limit theorem.
    b. power of a test.
    c. Type I error.
    d. Type II error.

2. If the null hypothesis is rejected when it shouldn't be, it is called a
    a. power.
    b. Type I error.
    c. Type II error.
    d. standard error.

3. If the null hypothesis is not rejected when it should be, it is called a
    a. power.
    b. Type I error.
    c. Type II error.
    d. standard error.

4. What are the properties of the sampling distribution of means?

5. A local school district employs a standardized reading test for all students entering the 9th grade. The mean score on this test is 27.4. Last year, the district instituted a reading program in the 6th through the 8th grades. The 217 students entering the 9th grade this year have a mean reading score of 28.2, with $s = 8.56$.

    a. Has the program improved reading?

    b. What is the 95% CI for μ?

6. A social psychologist finds that in a typical 15-minute conversation with a spouse, a person performs 32.1 nods of the head. In a sample of 13 couples experiencing marital difficulty, the average number of nods is 22.6, with $s = 7.6$.

    a. Did marital difficulty reduce nods?

    b. What is the 99% CI for μ?

7. The Hypochondriasis scale of the MMPI yields $\mu = 49.2$. The counseling center of a university routinely administers the MMPI to students seeking counseling. The hypochondriasis scores of the 12 students seeking counseling during the first week of the term are listed below.

$\underline{X}$
42
76
59
62
52
57
63
50
48
72
81
39

Find:

    a. $\overline{X}$

    b. s^2

    c. s

d. $s_{\bar{X}}$

e. What is the 95% CI for μ?

f. Do students seeking counseling evidence more hypochondriasis than would be expected from test norms?