CHAPTER PROBLEM

Is the "Freshman 15" real, or is it a myth?

There is a popular belief that college students typically gain 15 lb (or 6.8 kg) during their freshman year. This 15 lb weight gain has been deemed the "Freshman 15". Reasonable explanations for this phenomenon include the new stresses of college life (not including a statistics class which is just plain fun), new eating habits, increased levels of alcohol consumption, less free time for physical activities, cafeteria food with an abundance of fat and carbohydrates, the new freedom to choose among a variety of foods (including sumptious pizzas that are just a phone call away), and a lack of sleep that results in lower levels of leptin, which helps regulate appetite and metabolism. But is the Freshman 15 real, or is it a myth that has been perpetuated through anectodal evidence and/or flawed data?

Several studies have focused on the credibility of the Freshman 15 belief. We will consider results from one reputable study with results published in the article "Changes in Body Weight and Fat Mass of Men and Women in the First Year of College: A Study of the 'Freshman 15'" by Daniel Hoffman, Peggy Policastro, Virginia Quick, and Soo-Kyung Lee, *Journal of American College Health,* Vol.55, No. 1. The authors of that article have provided the data from their study, and much of it is listed in Data Set 3 in Appendix B. If you examine the weights in Data Set 3, you should note the following:

- The weights in Data Set 3 are in kilograms, not pounds, and 15 lb is equivalent to 6.8 kg. The "Freshman 15 (pounds)" is equivalent to the "Freshman 6.8 kilograms."
- Data Set 3 includes 2 weights for each of the 67 study subjects. Each subject was weighed in September of the freshman year, and again in April of the freshman year. These 2 measurements were made at the beginning and end of the seven months of campus life that passed between the measurements. It is important to recognize that each individual pair of before and after measurements is from the same student, so the lists of 67 before weights and 67 after weights constitute **paired** data from the 67 subjects in the study.
- Because the "Freshman 15" refers to weight gained, we will use weight changes in this format:

  (April weight) – (September weight)

  If a student does gain 15 lb, the value is positive 15. If a student loses 15 pounds, the value is -15.

- The publshed article about the "Freshman 15" study includes some limitations, including these:
  - All subjects volunteered for the study.
  - All of the subjects were attending Rutgers, The State University of New Jersey.

The "Freshman 15" constitutes a **claim** made about the population of college students. If we use μ_d to denote the mean of the weight gain differences for college students during their freshman year, the "Freshman 15" is the claim the $\mu_d = 15$ lb or $\mu_d = 6.8$ kg. Because the sample weights are measured in kilograms, we will consider the claim to be $\mu_d = 6.8$ kg. Later in this chapter, a formal hypothesis test will be used to test this claim. We will then be able to reach one of two possible conclusions: Either there is sufficient evidence to warrant rejection of the claim that $\mu_d = 6.8$ kg (so the "Freshman 15" is rejected) or we will conclude that there is not sufficient evidence to warrant rejection of the claim that $\mu_d = 6.8$ kg (so the "Freshman 15" cannot be rejected). We will then be able to determine whether or not the Freshman 15 is a myth.

	MATH 103 CHAPTER 9 HOMEWORK
9.2	1-5, 8, 10, 11, 17, 18, 25, 26, 33, 34
9.3	1-6, 9, 10, 11, 12, 15, 16, 21, 22, 25, 26, 32, 37
9.4	1-5, 7, 11, 12, 15, 16, 17, 18, 19

9.1 REVIEW AND PREVIEW

In Chapters 7 and 8 we introduced methods of ___________________________

___________________________. In Chapter 7 we presented methods of

____________________________ ____

interval __________________________ of _________________________________

_____________________________. In Chapter 8 we presented methods of testing ______________________ made about ___________________________

_____________________________. Chapters 7 and 8 both involved methods

for dealing with a ______________________ from a _____________________

________________________. The objective of this chapter is to

___________________ the methods for ____________________________

values of the ___________________________ _______________________ and

the methods for _______________________ __________________________ to

situations involving ____________ sets of _________________ ___________

instead of just _____________.

9.2 INFERENCES ABOUT TWO PROPORTION

Key Concept…

In this section we present methods for (1) _________________ a

__________________ made about the _________________________

________________________ and (2) constructing a

_______________________ _______________________

__________________ of the _______________________ between the

two __________________________ _________________________. This

section is based on ______________________, but we can use the same

methods for dealing with ________________________ or the

____________________ _______________________ of

____________________________.

OBJECTIVES

NOTATION FOR TWO PROPORTIONS

p_1

$\hat{p}_1 = \frac{}{}$

$n_1 =$

$\hat{q}_1 =$

$x_1 =$

The corresponding notations $p_2, n_2, x_2, \hat{p}_2$, and $\hat{q}_2$ apply to population 2.

POOLED SAMPLE PROPORTION

The __________________ ___________________ _____________________ is

denoted by _____ and is given by:

REQUIREMENTS

1. The ________________ _______________________ are from __________ ____________ ___________ samples that are ___________________________.

2. For each of the ____________ samples, the number of ________________ is ____ ______________ ___________ and the number of ____________________ is at __________________ _____. That is, __________________ and ____________________ for each of the two samples.

TEST STATISTIC FOR TWO PROPORTIONS (WITH $H_0: p_1 = p_2$)

$z = \underline{\hspace{6cm}}$

$P-$value:

Critical values:

CONFIDENCE INTERVAL ESTIMATE OF $p_1 = p_2$

The confidence interval estimate of the ___________________ _______________ is:

where the __________________ of ____________________ ______ is given by

Rounding: Round the confidence interval limits to ___________ significant digits.

HYPOTHESIS TESTS

For tests of hypotheses made about ____________ population

_______________________, we consider only tests having a ____________

hypothesis of _______________. Note that under the assumption of __________

________________________, the best estimate of the _____________________

________________________ is obtained by ____________________ __________

samples into __________________ __________________ __________________,

so that _______ is the __________________________ of the ______________

_________________________ __________________________.

CAUTION!!! When testing a claim about ___________ population proportions,

the ____________________ method and the _______________________

method are equivalent, but they _________ ______ equivalent to the

__________________________ ________________________ method!!! If you

want to ____________ a claim about _____________

_______________________ ________________________, use the ___________

method or the __________________________ method; if you want to

_________________________ the _________________________ between

_______ _______________________ _______________________, use a

___________________________ __________________________.

Example 1: In a 1993 survey of 560 college students, 171 said they used illegal drugs during the previous year. In a recent survey of 720 college students, 263 said that they used illegal drugs during the previous year (based on data from the National Center for Addiction and Substance Abuse at Colombia University).

a. Use a 0.05 significance level to test the claim that the proportion of college students using illegal drugs in 1993 was less than it is now.

Example 2: Among 2739 female atom bomb survivors, 1397 developed thyroid diseases. Among 1352 male atom bomb survivors, 436 developed thyroid diseases (based on data from "Radiation Dose-Response Relationships for Thyroid Nodules and Autoimmune Thyroid Diseases in Hiroshima and Nagasaki Atomic Bomb Survivors 55-58 Years After Radiation Exposure," by Imaizumi, et al., *Journal of the American Medical Association,* Vol. 295, No. 9).

a. Use a 0.01 significance level to test the claim that the female survivors and male survivors have different rates of thyroid diseases.

b. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.01 significance level.

c. What conclusion does the confidence interval suggest?

9.3 INFERENCES ABOUT TWO MEANS: INDEPENDENT SAMPLES

Key Concept…

In this section, we present methods for using ____________________

____________________ from _________ ___________________________

samples to ___________________ hypotheses made about ____________

________________________ ___________________ or to construct

__________________________ _______________________ estimates of the

_________________________ between _______________ population

____________________.

PART I: INDEPENDENT SAMPLES WITH σ_1 AND σ_2 UNKNOWN AND NOT ASSUMED EQUAL

DEFINITION

Two _______________________ are **independent** if the _______________

____________________ from one population _________ __________

____________________ ____ or somehow ___________________ ____________

or _____________________ with the ____________________ ______________

from the other population.

Two ____________________ are **dependent** if the sample values are

___________.

Inferences about Means of Two Independent Populations, With σ_1 and σ_2 Unknown and Not Assumed to be Equal

OBJECTIVES

NOTATION

Population 1:

$\mu_1 =$ $s_1 =$

σ_1

$\bar{x}_1 =$ $n_1 =$

The corresponding notations for ______, ______, ______, ______,and ______ apply

to population ______.

REQUIREMENTS

1. ______ and ______ are ________________ and it is not ______________ that

______ and _______ are ___________.

2. The ________ samples are __________________.

3. Both samples are _____________________ ___________________ __________________.

4. Either or both of these conditions are satisfied: The two ______________

____________ are both __________ (with __________ and ___________) or both samples come from populations having ____________________ __________________.

HYPOTHESIS TEST STATISTIC FOR TWO MEANS: INDEPENDENT SAMPLES

$t = \overline{\qquad\qquad\qquad\qquad\qquad\qquad}$

Degrees of Freedom: When finding __________________ __________________ or

_________________, use the following for determining the number of degrees of

freedom.

1. In this book we use the conservative estimate: df = ____________ of ________

and ___________.

2. Statistical software packages typically use the more accurate but more difficult estimate given below:

$$\text{df} = \frac{(A+B)^2}{\dfrac{A^2}{n_1 - 1} + \dfrac{B^2}{n_2 - 1}}, \quad A = \frac{s_1^{\,2}}{n_1}, \quad B = \frac{s_2^{\,2}}{n_2}$$

***P*-values and critical values:** Use Table A-3.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$: INDEPENDENT SAMPLES

The confidence interval estimate of the difference ______________ is

and the number of degrees of freedom df is as described above for hypothesis tests.

EQUIVALENCE OF METHODS

The _________________ method of hypothesis testing, the ________________ method of hypothesis testing, and _________________ _________________ all use the same _____________________ and __________________ __________, so they are ___________________ in the sense that they result in the _______ ________________________. A null hypothesis of _________________ or ___________________ can be tested using the ________________ method, the ____________________ method, or by determining whether the _________________ ___________________ includes _______.

Example 1: Determine whether the samples are independent or dependent.

a. To test the effectiveness of Lipitor, cholesterol levels are measured in 250 subjects before and after Lipitor treatments.

b. On each of 40 different days, the author measured the voltage supplied to his home and he also measured the voltage produced by his gasoline powered generator.

Example 2: Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal. A simple random sample of 13 four-cylinder cars is obtained, and the braking distances are measured. The mean braking distance is 137.5 feet and the standard deviation is 5.8 feet. A SRS of 12 six-cylinder cars is obtained and the braking distances have a mean of 136.3 feet with a standard deviation of 9.7 feet (based on Data Set 16 in Appendix B).

a. Construct a 90% CI estimate of the difference between the mean braking distance of four-cylinder cars and six-cylinder cars.

b. Does there appear to be a difference between the two means?

c. Use a 0.05 significance level to test the claim that the mean braking distance of four-cylinder cars is greater than the mean braking distance of six-cylinder cars.

PART 2: ALTERNATIVE METHODS

Part 1 in this section dealt with situations in which the two ________________

______________________ _________________ are _________________ and

____________ assumed to be ______________. In Part 2 we address two other

situations: (1) The two __________________ __________________

_______________ are both _____________; (2) the two _________________

__________________ ____________________ are _________________ but

_____________________ to be _________________.

ALTERNATIVE METHOD WHEN σ_1 AND σ_2 ARE KNOWN

In reality, the population standard deviations are almost _______________

______________, but if they are known, the ______________ _____________

and ________________ ____________________ are based on the ____________

____________________ instead of the _____ _________________.

Inferences about Means of Two Independent Populations, With σ_1 and σ_2 Known
REQUIREMENTS
1. The two population standard deviations ____ and ____ are both _________. 2. The two ____________ are ____________________. 3. Both samples are _____________ ______________ _______________.

4. Either or both of these conditions is satisfied: The two sample sizes are both large, with _______________ and _______________ or both samples come from _________________ having _____________ __________________.

HYPOTHESIS TEST

Test statistic:

________________ and ______________ _______________: Refer to Table _____.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$

ALTERNATIVE METHOD: ASSUME THAT $\sigma_1 = \sigma_2$ AND POOL THE SAMPLE VARIANCES

Even when the specific values of ______ and _____ are ________ ____________, if it can be assumed that they have the _____________ value, the sample variances _______ and ______ can be _______________ to obtain an _________________ of the ___________________ population __________________ _________. The _______________ __________________ of ________ is denoted _________ and is a ________________ ______________ of ________ and ________, which is

shown below.

Inferences about Means of Two Independent Populations, Assuming that $\sigma_1 = \sigma_2$

REQUIREMENTS

1. The two population standard deviations are ____ _______, but they are

   assumed to be ________. That is, _______________.

2. The two ___________ are ___________________.

3. Both samples are _____________ ______________ ______________.

4. Either or both of these conditions is satisfied: The two sample sizes are

   both large, with ______________ and ______________ or both samples

   come from ________________ having ____________ _________________.

HYPOTHESIS TEST

Test statistic:

where

and the number of degrees of freedom is given by ___________________.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$

Confidence interval:

where

and __________ is as given in the above test statistic and the number of degrees of freedom is ________________.

TI-83/84 PLUS

Example 3: Assume that the two samples are independent simple random samples selected from normally distributed populations. Also assume that the population standard deviations are equal. The mean tar content of a simple random sample of 25 unfiltered king size cigarettes is 21.1 mg, with a standard deviation of 3.2 mg. The mean tar content of a SRS of 25 filtered 100 mm cigarettes is 13.2 mg with a standard deviation of 3.7 mg (based on Data Set 4 in Appendix B).

a. Construct a 90% CI estimate of the difference between mean tar content of unfiltered king size cigarettes and the mean tar content of filtered 100 mm cigarettes.

b. Does the result suggest that 100 mm filtered cigarettes have less tar than unfiltered king size cigarettes?

c. Use a 0.05 significance level to test the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. What does the reslt suggest about the effectiveness of cigarette filters?

9.4 INFERENCES FROM DEPENDENT SAMPLES

Key Concept…

In this section we present methods for testing hypotheses and constructing confidence intervals involving the _______________ of the _________________ of the ______________ of two _______________ _________________. With _______________ samples, there is some _________________ whereby each value in one sample is ____________ with a ________________ value in the other sample. Here are two typical examples of dependent samples:

- Each pair of sample values consists of two measurements from the ____________ subject
- Each pair of sample values consists of a ____________ _________.

Because the hypothesis test and CI use the same ________________ and _______________ ______________, they are ______________ in the sense that they result in the _____________ __________________.

Consequently, the _____________ hypothesis that the ____________ ______________ _______________ _____ can be tested by determining whether the ________________ ________________ includes ______.

There are no exact procedures for dealing with _________________ samples, but the _____ ___________________ serves as a reasonably good approximation, so the following methods are commonly used.

Inferences about Means of Two Dependent Populations

OBJECTIVES

NOTATION

$d =$ $s_d =$

μ_d

$\bar{d} =$ $n =$

REQUIREMENTS

1. The ____________ data are ____________________.

2. The samples are _______________ _________________ _________________.

3. Either or both of these conditions are satisfied: The number of ___________ of ______________ is ______________ (__________) or the pairs of values have __________________ that are from a population that is approximately __________________.

HYPOTHESIS TEST FOR DEPENDENT SAMPLES

$t = \frac{\qquad}{\qquad}$

Degrees of Freedom: ____________
***P*-values and critical values:** Use Table A-3.

CONFIDENCE INTERVALS FOR DEPENDENT SAMPLES

where

and

Example 1: Assume that the paired sample data are SRSs and that the differences have a distribution that is approximately normal.

a. Listed below are BMIs of college students.

April BMI	20.15	19.24	20.77	23.85	21.32
September BMI	20.68	19.48	19.59	24.57	20.96

i. Use a 0.05 significance level to test the claim that the mean change in BMI for all students is equal to 0.

ii. Construct a 95% CI estimate of the change in BMI during freshman year.

iii. Does the CI include zero, and what does that suggest about BMI during freshman year?

b. Listed below are systolic blood pressure measurements (mm Hg) taken from the right and left arms of the same woman. Use a 0.05 significance level to test for a difference between the measurements from the two arms. What do you conclude?

Right arm	102	101	94	79	79
Left arm	175	169	182	146	144