Chapter 2 Review Sheet with Answers

Topics to Study

- Confidence Interval Key Terms
- Statistics and Parameters
- Sampling Distributions
- Central Limit Theorem
- Know how to interpreting confidence intervals
- Standard Normal Z Distribution, the Student T-distribution, and Z and T Critical Values
- Table summarizing critical value, standard error, margin of error and confidence intervals
- Confidence Interval Assumptions
- Bootstrapping
- Two-population confidence intervals

Chapter 2 Vocabulary Terms to Study

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Sampling Distribution: Take many random samples from a population, calculate a sample statistic like a mean or percent from each sample and graph all of the sample statistics on the same graph. The center of the sampling distribution is a good estimate of the population parameter.

Sampling Variability: Random samples values and sample statistics are usually different from each other and usually different from the population parameter.

Point Estimate: When someone takes a sample statistic and then claims that it is the population parameter.

Margin of Error: Total distance that a sample statistic might be from the population parameter. For normal sampling distributions and a 95% confidence interval, the margin of error is approximately twice as large as the standard error.

Standard Error: The standard deviation of a sampling distribution. The distance that typical sample statistics are from the center of the sampling distribution. Since the center of the sampling distributions is usually close to the population parameter, the standard error tells us how far typical sample statistics are from the population parameter.

Confidence Interval: Two numbers that we think a population parameter is in between. Can be calculated by either a bootstrap distribution or by adding and subtracting the sample statistic and the margin of error.

Critical Value Z-score (Z_c): The number of standard errors the sample proportion should be from the population proportion based on the confidence level and the standard normal distribution.

Critical Value T-score (T_c): The number of standard errors the sample mean should be from the population mean based on the confidence level, degrees of freedom and the student T distribution.

95% Confident: 95% of confidence intervals contain the population value and 5% of confidence intervals do not contain the population value.

90% Confident: 90% of confidence intervals contain the population value and 10% of confidence intervals do not contain the population value.

99% Confident: 99% of confidence intervals contain the population value and 1% of confidence intervals do not contain the population value.

Bootstrapping: Taking many random samples values from one original real random sample with replacement.

Bootstrap Sample: A simulated sample created by taking many random samples values from one original real random sample with replacement.

Bootstrap Statistic: A statistic calculated from a bootstrap sample.

Bootstrap Distribution: Putting many bootstrap statistics on the same graph in order to simulate the sampling variability in a population, calculate standard error, and create a confidence interval. The center of the bootstrap distribution is the original real sample statistic.

Table summarizing critical value, standard error, margin of error and confidence intervals

	Critical Value	times	**Standard Error**	equals	**Margin of Error**		**Confidence Interval**
	(Z or T)		S / square root (n)				(Statistic + or - Margin of Error)
Conf Level Decreases (95% => 90%)	Gets Smaller	times	Stays Same	equals	Gets Smaller		Gets Narrower
	(1.96 to 1.645)						10+ or - 4 = (6,14)
							10 + or - 1 = (9,11)
Conf Level Increases (95% => 99%)	Gets Larger	times	Stays Same	equals	Gets Larger		Gets Wider
	1.96 to 2.576						10 + or - 1 = (9,11)
							10+ or - 4 = (6,14)
Sample Size Decreases (n=100 to n=25)	Z score stays Same	times	Gets Larger!!	equals	Gets Larger		Gets Wider
(Assume 95% conf level)	T scores get Larger		30/sq root 100 = 3				10 + or - 1 = (9,11)
			30/sq root 25 = 6				10 + or - 4 = (6,14)
Sample Size Increases (n=25 to n=100)	Z score stays Same	times	Gets Smaller	equals	Gets Smaller		Gets Narrower
(Assume 95% conf level)	T scores get close to Z		30/sq root 25 = 6				10+ or - 4 = (6,14)
			30/sq root 100 = 3				10 + or - 1 = (9,11)

Assumptions (Conditions) for making a Confidence Interval

Population Mean Average Confidence Interval

- **Random Sample(s)**
- **Individuals Independent**
- **Data is normal (bell shaped) or the sample size (n) is at least 30**

Population Proportion (Percentage) Confidence Interval

- **Random Sample(s)**
- **Individuals Independent**
- **Data has at least 10 successes and at least 10 failures**

Bootstrap Confidence Intervals

- **Random Sample(s)**
- **Individuals Independent**

Chapter 2 Review Practice Problems

1. Determine if each of the following symbols are a mean, standard deviation, proportion, slope, or correlation coefficient. Also, decide if it is a sample statistic or a population parameter.

(N , n , π , p̂ , μ , x̄ , σ , s , ρ , r , β_1 , b_1, σ^2, s^2)

2. For each number determine the symbol used from the following list and if it is a statistic or a parameter.

(N , n , π , p̂ , μ , x̄ , σ , s , ρ , r , β_1 , b_1, σ^2, s^2)

a) "We tested a sample of incoming college freshman and found that their sample mean average IQ was 101.9, a sample standard deviation of 14.8 and a sample variance of 219.04. We think the population mean IQ is 100, the population standard deviation for IQ scores is 15, and the population variance is 225."

b) "We want to check and see if the population correlation coefficient could be zero and the population slope could be about 20 pounds per degree Fahrenheit. The sample correlation coefficient was 0.338 and the sample slope was 13.79 pounds per degree Fahrenheit."

c) "Our study found that of the people tested in the sample, only 3% showed side effects to the medication. We think the population percentage of side effects is closer to 1.5%".

d) "We took a random sample of 238 people from a population of about 5 million people."

3. List the assumptions that need to be checked before you make a one-population mean confidence interval.

4. List the assumptions that need to be checked before you make a one-population variance or standard deviation confidence interval.

5. List the assumptions that need to be checked before you make a one-population proportion confidence interval.

6. List the assumptions that need to be checked before you make a two-population mean confidence interval.

7. List the assumptions that need to be checked before you make a two-population proportion confidence interval.

8. List the assumptions for a bootstrap confidence interval.

9. Define the following terms: Population, Census, Sample, Statistic, Parameter, Sampling Distribution, Sampling Variability, Point Estimate, Margin of Error, Standard Error, Confidence Interval, Critical Value Z-score, Critical Value T score, 95% Confident, 90% Confident, 99% Confident, Bootstrapping, Bootstrap Sample, Bootstrap Statistic, Bootstrap Distribution

10. Write a sentence to explain the following confidence intervals. Assume the confidence intervals came from unbiased random sample data that met all of the assumptions.

a) Explain the one-population mean confidence interval (55.6 pounds, 69.4 pounds). *Confidence Level = 99%*

b) Explain the one-population proportion confidence interval (0.352 , 0.411). *Confidence Level = 90%*

c) Explain the one-population standard deviation confidence interval (3.1 pounds, 4.7 pounds). *Confidence Level = 95%*

d) Explain the one-population variance confidence interval (461.8 square inches, 591.3 square inches). *Confidence Level = 99%*

e) Explain the two-population mean confidence interval (+13.2 kg, +14.8 kg). *Confidence Level = 95%* Is there a significant difference between the populations? Explain why.

f) Explain the two-population mean confidence interval (−\$3.79, +\$4.13). *Confidence Level = 90%* Is there a significant difference between the populations? Explain why.

g) Explain the two-population proportion confidence interval (−0.024, +0.017). *Confidence Level = 95%* Is there a significant difference between the populations? Explain why.

h) Explain the two-population proportion confidence interval (−0.072, −0.057). *Confidence Level = 99%* Is there a significant difference between the populations? Explain why.

11. Explain what a sampling distribution is and how we can use it to find the population parameter, standard error and better understand sampling variability.

12. Explain how a critical value Z-score or T-score and standard error can be used to calculate the margin of error. How can we use margin of error to make the confidence interval.

13. In one-population variance confidence intervals, how does the computer use the chi-squared critical values, the degrees of freedom and the sample variance to make the confidence interval?

14. Answer the following questions about the T-distribution.

a) Who invented the T-distribution?
b) What company did he work for?
c) Why did he invent T-scores?
d) Why did he have to publish the T-distribution discovery under a pseudonym?
e) What pseudonym did he use?
f) When are T-scores significantly larger than Z-scores?
g) When are T-scores and Z-scores almost the same?
h) What types of confidence intervals use Z-scores?
i) What types of confidence intervals use T-scores?
j) How is degrees of freedom usually calculated for one quantitative data set?

15. Explain the ideas behind the Central Limit Theorem.

16. Explain the process of bootstrapping and how a bootstrap distribution may be used to calculate a confidence interval without a formula. What assumptions are necessary to make a bootstrap confidence interval? How is a bootstrap distribution different from a sampling distribution?

Chapter 2 Review Sheet Answers

1.

N: Parameter describing the size of a population.

n: Statistic describing the size of a sample.

π or p: Parameter describing a population proportion.

$\hat{p}$: Statistic describing a sample proportion.

μ: Parameter describing a population mean average.

$\bar{x}$: Statistic describing a sample mean average.

σ: Parameter describing a population standard deviation.

s: Statistic describing a sample standard deviation.

σ^2: Parameter describing a population variance.

s^2: Statistic describing a sample variance.

ρ: Parameter describing a population correlation coefficient.

r: Statistic describing the correlation coefficient from sample data.

β_1: Parameter describing a population slope.

b_1: Statistic describing a slope from sample data.

2.

a)

$\bar{x}$ = 101.9 (statistic)
s = 14.8 (statistic)
s^2 = 219.04 (statistic)

μ = 100 (parameter)
σ = 15 (parameter)
σ^2 = 225 (parameter)

b)

ρ = 0 (parameter)
β_1 =

β_1=20 (parameter)

r = 0.338 (statistic)
$b_1 = 13.79\ (statistic)$

c)

p̂ = 0.03 (statistic)
π = 0.015 (parameter)

d)

n = 238 (statistic)
N = 5,000,000 (parameter)

3.

<u>One-population Mean Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample size should be at least 30 or have a nearly normal shape.

4.

<u>One-Population Variance or Standard Deviation Assumptions</u>

- The quantitative sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- The sample data must be normal.

5.

<u>One-population Proportion Assumptions</u>

- The categorical sample data should be collected randomly or be representative of the population.

- Data values within the sample should be independent of each other.
- There should be at least ten successes and at least ten failures.

6.

Two-population Mean Assumptions (Matched Pair)

- The quantitative ordered pair sample data should be collected randomly or be representative of the population.
- Data values within the sample should be independent of each other.
- There should be at least thirty ordered pairs or the differences should have a nearly normal shape.

Two-population Mean Assumptions (Not Matched Pair)

- The two quantitative samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the two samples should be independent of each other.
- The sample sizes should be at least 30 or have a nearly normal shape.

7.

Two-population Proportion Assumptions

- The two categorical samples should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- Data values between the samples should be independent of each other.
- There should be at least ten successes and at least ten failures.

8.

Bootstrap and Randomized Simulation Assumptions

- The sample data should be collected randomly or be representative of the population.
- Data values within each sample should be independent of each other.
- If multiple samples were collected that were not matched pair, then the data values between the samples should be independent of each other.

9.

Vocabulary

Population: The collection of all people or objects to be studied.

Census: Collecting data from everyone in a population.

Sample: Collecting data from a small subgroup of the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Sampling Distribution: Take many random samples from a population, calculate a sample statistic like a mean or percent from each sample and graph all of the sample statistics on the same graph. The center of the sampling distribution is a good estimate of the population parameter.

Sampling Variability: Random samples values and sample statistics are usually different from each other and usually different from the population parameter.

Point Estimate: When someone takes a sample statistic and then claims that it is the population parameter.

Margin of Error: Total distance that a sample statistic might be from the population parameter. For normal sampling distributions and a 95% confidence interval, the margin of error is approximately twice as large as the standard error.

Standard Error: The standard deviation of a sampling distribution. The distance that typical sample statistics are from the center of the sampling distribution. Since the center of the sampling distributions is usually close to the population parameter, the standard error tells us how far typical sample statistics are from the population parameter.

Confidence Interval: Two numbers that we think a population parameter is in between. Can be calculated by either a bootstrap distribution or by adding and subtracting the sample statistic and the margin of error.

Critical Value Z-score (Z_c): The number of standard errors the sample proportion should be from the population proportion based on the confidence level and the standard normal distribution.

Critical Value T-score (T_c): The number of standard errors the sample mean should be from the population mean based on the confidence level, degrees of freedom and the student T distribution.

95% Confident: 95% of confidence intervals contain the population value and 5% of confidence intervals do not contain the population value.

90% Confident: 90% of confidence intervals contain the population value and 10% of confidence intervals do not contain the population value.

99% Confident: 99% of confidence intervals contain the population value and 1% of confidence intervals do not contain the population value.

Bootstrapping: Taking many random samples values from one original real random sample with replacement.

Bootstrap Sample: A simulated sample created by taking many random samples values from one original real random sample with replacement.

Bootstrap Statistic: A statistic calculated from a bootstrap sample.

Bootstrap Distribution: Putting many bootstrap statistics on the same graph in order to simulate the sampling variability in a population, calculate standard error, and create a confidence interval. The center of the bootstrap distribution is the original real sample statistic.

10.

a)

We are 99% confident that the population mean weight is between 55.6 pounds and 69.4 pounds.

b)

We are 90% confident that the population proportion is in between 0.352 and 0.411.

OR

We are 90% confident that the population percentage is in between 35.2% and 41.1%.

c)

We are 95% confident that the population standard deviation is between 3.1 pounds and 4.7 pounds.

d)

We are 99% confident that the population variance is between 461.8 square inches and 591.3 square inches.

e)

Since the confidence interval is (positive, positive), this indicates that population 1 is significantly larger than population 2. The confidence interval indicates that population 1 is between 13.2 kg and 14.8 kg larger.

Sentence: We are 95% confident that the population mean average weight of population 1 is between 13.2 kg and 14.8 kg larger than population 2.

f)

Since the confidence interval is (negative, positive), this indicates that there is no significant difference between the population mean averages for population 1 and population 2. They are very close and we cannot tell which population is larger.

Sentence: We are 90% confident that there is no significant difference between the population mean averages for population 1 and population 2.

g)

Since the confidence interval is (negative, positive), this indicates that there is no significant difference between the population proportions for population 1 and population 2. The population percentages are very close and we cannot tell which population is larger.

Sentence: We are 95% confident that there is no significant difference between the population proportions for population 1 and population 2.

OR

We are 95% confident that there is no significant difference between the population percentages for population 1 and population 2.

h)

Since the confidence interval is (negative, negative), this indicates that population 1 is significantly smaller than population 2. The confidence interval indicates that population 1 is between 0.057 (5.7%) and 0.072 (7.2%) smaller than population 2.

Sentence: We are 99% confident that the population proportion for population 1 is between 0.057 and 0.072 less than population 2.

OR

We are 99% confident that the population percentage for population 1 is between 5.7% and 7.2% less than population 2.

11.

A sampling distribution is created by taking hundreds or thousands of random samples from a population. We then calculate a sample statistic like the mean, standard deviation or proportion and put all of the thousands of random sample statistics on the same graph. The center of the sampling distribution is often very close to the actual population parameter. The standard error is the standard deviation of the sampling distribution and helps us understand how far typical statistics may be from the center (population parameter). Overall the sampling distribution teaches us about sampling variability. This is the principle that random samples are usually very different from each other and are often very far off from the population parameter.

12.

If the sampling distribution is normal, then we can multiply the critical value Z-score or T-score times the standard error to get the margin of error. If we take the original random sample statistic and add and subtract the margin of error, we will get the upper and lower limits of the confidence interval.

13.

$\frac{s^2(n-1)}{\chi^2_{upper}}$ < Population Variance (σ^2) < $\frac{s^2(n-1)}{\chi^2_{lower}}$

To create a confidence interval for the population variance, multiply the sample variance s^2 by the degrees of freedom n-1 and then divide by the chi-squared critical values.

14.

a)

William Gosset

b)

Guiness Beer

c)

Needed better accuracy for smaller data sets.

d)

Guiness Beer had a strict no publishing policy and would have fired him.

e)

“student”

f)

When the sample size is small, the T-scores are significantly greater than the Z-scores. This accounts for more variability in smaller data sets.

g)

When the sample size is large, the T-scores and Z-scores are about the same.

h)

One and two-population proportion (percentage) confidence intervals use Z-scores.

i)

One and two-population mean average confidence intervals use T-scores.

j)

For one quantitative data set with a sample size “n”, the degrees of freedom is “n – 1”.

15.

The central limit theorem states that a sampling distribution made of sample means will be normal (bell shaped) if the sample size is sufficiently large. The mean and standard deviation of a sampling distribution are important calculations in inferential statistics but are only accurate if the sampling distribution is normal. The central limit theorem discusses what conditions need to be met for the sampling distributions to be normal. These are the foundations of our assumptions for the confidence intervals. For quantitative sample data that is not normal, we will need a sample size of at least 30 to ensure the sampling distribution will look nearly normal. For quantitative sample data that is already normal, we can use a sample below 30. For categorical data, we will need at least 10 successes and failures for the sampling distribution for proportions to look normal.

16.

Suppose we want to make a confidence interval, but the data does not pass the assumptions for our sampling distribution to look normal. This would mean our traditional formulas involving standard error, Z-scores and T-scores may not be very accurate. Bootstrapping is a technique for calculating a confidence interval directly without the traditional formula. In a bootstrap, we take thousands of random samples with replacement from the one random sample. We then calculate the statistic from each of the bootstrap samples and put all of the thousand bootstrap statistics on the bootstrap distribution. In a sense, we have created a simulated sampling distribution but not from the population. Find the middle 95%, 90% or 99% markers from the distribution and you have your confidence interval. Bootstrapping still requires the data to be collected randomly and individual observations should be independent, but it does not require the same central limit theorem assumptions. The bootstrap distribution does not have to look normal since we are calculating the middle 95%, 90% or 99% directly.
