CHAPTER PROBLEM

How do we interpret a poll about global warming?

Global warming is the increase in the mean temperature of air near the surface of the earth and the increase in mean temperature of the oceans. Scientists generally agree that global warming is caused by increased amounts of carbon dioxide, methane, ozone, and other gases that result from human activity.

Global warming is believed to be responsible for the retreat of glaciers, the reduction in the arctic region, and a rise in sea levels. It is feared that continued global warming will result in even higher sea levels, flodding, draught, and more severe weather.

Because global warming appears to have the potential for causing dramatic changes in our environment, it is critical that we recognize that potential. Just how much do we all recognize global warming? In a Pew Research Center poll, respondents were asked "From what you've read and heard, is there solid evidence that the average temperature on earth has been increasing over the past few decardes, or not?" In response to that question, 70% of 1501 randomly selected adults in the United States answered "yes". Therefore, among those polled, 70% believe in global warming. Although the subject matter of this poll has great significance, we will focus on the interpretation and anlysis of the results. Some important issues that relate to this poll are as follows:

- How can the poll results be used to estimate the percentage of all adults in the United States who believe that the earth is getting warmer?
- How accurate is the result of 70% likely to be?
- Given that only 1501/225,139,000 or 0.0007% of the adult population in the United States were polled, is the sample size too small to be meaningful?
- Does the method of selecting the people to be polled have much of an effect on the results?

We can answer the last question based on the sound sampling methods discussed in Chapter 1. The method of selecting the people to be polled most definitely has as effect on the results. The results are likely to be poor if a convenience sample or some other nonrandom sampling method is used. If the sample is a simple random sample, the results are likely to be good.

Our ability to understand polls and to interpret the results is crucial for our role as citizens. As we consider the topics of this chapter, we will learn more about polls and surveys and how to correctly interpret and present results.

MATH 103 CHAPTER 7 HOMEWORK
7.2 1-27 odd, 30, 33, 36, 37, 41, 45
7.3 1-7, 10, 11, 12, 13, 16, 17-20, 21, 25, 26, 31, 33
7.4 1-13, 16, 17, 20, 23, 25, 27, 29, 30

7.1 REVIEW AND PREVIEW

In Chapters 2 and 3 we used "_____ _____" when we _____ data using tools such as _____, and statistics such as the _____ and _____ _____. We use "_____ _____" when we use _____ data to make inferences about _____ _____. Two major activities of _____ statistics are (1) to use _____ data to _____ values of _____ _____, and (2) to test _____ or _____ made about _____ parameters. In this chapter we begin working with the true core of _____ statistics as we use sample data to _____ values of population _____.

7.2 ESTIMATING A POPULATION PROPORTION

Key Concept…

In this section we present methods for using a _____ _____ to estimate a _____ _____. There are three main ideas:

1. The _____ _____ is the best _____ _____ of the _____ _____.
2. We can use a _____ _____ to construct a _____ _____ to estimate the _____ _____ of a population _____, and we should know how to _____ such confidence intervals.
3. We should know how to find the _____ _____ necessary to _____ a population _____.

DEFINITION

A **point estimate** is a _____ value (or _____) used to _____ a _____ parameter.

The _____ _____ _____ is the best

_____ _____ of the _____

_____ _____.

DEFINITION

A **confidence interval (aka _____ _____)** is a _____ (or an _____) of _____ used to _____ the _____ value of a _____ _____. *A* _____ _____ is often abbreviated as CI.

A confidence interval is associated with a _____ level, such as 0.95 (or 95%). The confidence level gives us the _____ _____ of the _____ used to _____ the _____ _____. The _____ level is often expressed as the _____ or _____ _____, where _____ is the _____ of the _____ level.

DEFINITION

The **confidence level** is the _____ _____

(often expressed as the equivalent percentage value) that the _____

_____ actually does _____ the

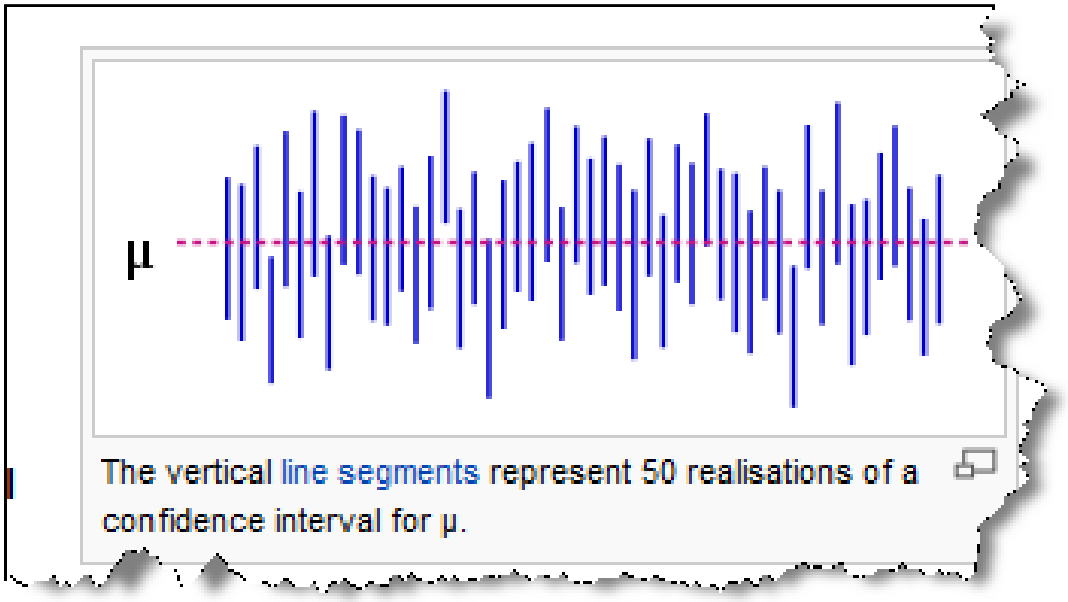
_____ _____, assuming that the

_____ process is _____ a _____

number of times. (The _____ _____ is also

called the _____ of _____, or the _____

_____).



CRITICAL VALUES

The methods of this section (and many others) include a reference to a

_____ _____ that can be used to _____

between _____ _____ that are _____

to _____ and those that are _____ to _____.

Such a _____ is called a _____ _____.

Critical values are based on the following observations:

1. Under certain conditions, the _____ distribution of sample _____ can be _____ by a _____ distribution.
2. A _____ associated with a _____ _____ has a probability of _____ of falling in the _____ _____.
3. The _____ separating the _____ region is commonly denoted by _____ and is referred to as a _____ value because it is on the _____ separating _____ from _____ _____ that are _____ to occur from those that are unlikely to occur.

DEFINITION

A **critical value** is the _____ on the _____ separating _____ _____ that are likely to occur from those that are _____ to occur. The number _____ is a _____ _____ that is a _____ with the property that it _____ an _____ of _____ in the _____ tail of the _____ _____ distribution.

Example 1: An interesting and popular hypothesis is that individuals can temporarily postpone their death to survive a major holiday or important event such as a birthday. In a study of this phenomenon, it was found that in the week before and the week after Thanksgiving, there were 12,000 total deaths, and 6062 of them occurred in the week before Thanksgiving.

a. What is the best point estimate of the proportion of deaths in the week before Thanksgiving to the total deaths in the week before and the week after Thanksgiving?

b. Construct a 95% confidence interval estimate of the proportion of deaths in the week before Thanksgiving to the total deaths in the week before and the week after Thanksgiving.

c. Based on the result, does there appear to be any indication that people can temporarily postpone their death to survive the Thanksgiving holiday? Why or why not?

Example 2: In a study of 420,095 cell phone users in Denmark, it was found that 135 developed cancer of the brain or nervous system. Prior to this study of cell phone use, the rate of such cancer was found to be 0.0340% for those not using cell phones.

a. Use the sample data to construct a 95% confidence interval estimate of the percentage of cell phone users who develop cancer of the brain or nervous system.

b. Do cell phone users appear to have a rate of cancer of the brain or nervous system that is different from the rate of such cancer among those using cell phones? Why or why not?

DEFINITION

When the data from a _____ _____ sample are used to _____ a _____ _____ ___, the **margin of error**, denoted by ____, is the _____ likely _____ (with probability _____) between the _____ _____ _____ ____

and the _____ _____ of the _____

_____ _____. The _____ of _____

_____ is also called the _____ _____ of the

_____ and can be found by _____ the

_____ value and the _____

_____ of _____ _____ as

shown in the formula below:

ROUND-OFF RULE FOR CONFIDENCE INTERVAL ESTIMATES OF *p*

Round the confidence interval _____ for _____ to _____

_____ _____.

DETERMINING SAMPLE SIZE

Suppose we want to _____ data in order to _____

some _____ _____. How do we know how

many sample items must be obtained? If we solve the _____ for

_____ of _____ _____ for _____, we get the first

formula below. Note that this formula requires _____. If no such estimate is

known, we replace _____ by _____ and replace _____ by _____,

which is shown in the second formula.

When an estimate ____ is known:

When no estimate ____ is known:

ROUND-OFF RULE FOR DETERMINING SAMPLE SIZE

If the computed sample size _____ is not a _____ _____,

round the value of ____ _____ to the next _____ _____

number.

Example 3: As your text was being written, former NYC mayor Rudolph Giuliani announced that he was a candidate for the presidency of the United States. If you were a campaign worker and needed to determine the percentage of people that recognized his name, how many people should you have surveyed to estimate that percentage? Assume that you wanted to be 95% confident that the sample percentage was in error by no more than 2 percentage points, and also assume that a recent survey indicated that Giuliani's name is recognized by 10% of all adults (based on data from a Gallup poll).

7.3 ESTIMATING A POPULATION MEAN: SIGMA KNOWN

Key Concept…

In this section we present methods for _____ a _____ _____. In addition to knowing the values of the _____ data or _____, we must also know the value of the _____ _____ _____, _____. Here are three concepts that should be learned in this section.

1. We should know that the _____ _____ _____ is the best _____ _____ of the _____ _____ _____.
2. We should learn how to use _____ _____ to construct a _____ _____ for _____ the value of a _____ _____, and we should know how to _____ such _____ _____.
3. We should develop the ability to _____ the _____ _____ necessary to _____ a _____ _____.

POINT ESTIMATE

The _____ _____ _____ is an _____

estimator of the _____ _____ _____, and for

many populations, _____ _____ tend to _____

less than other measures of _____, so the _____

_____ _____, is usually the best _____

_____ of the _____ _____ _____.

KNOWLEDGE OF SIGMA

The methods of this section require that we know _____, but in 7.4 we will learn

methods to _____ a _____ _____

without knowledge of the value of _____.

NORMALITY REQUIREMENT

The population must either be _____ _____

or _____. If _____, the population does not need to have a

_____ that is _____ _____ as

long as it is _____ _____. As long as there

are no _____ and if a _____ of the

_____ _____ is not _____

different from being _____, the _____

requirement is satisfied.

SAMPLE SIZE REQUIREMENT

The _____ sample size actually depends on how much the

_____ _____ departs from a

_____ _____. Sample sizes of _____ to

_____ are sufficient if the population has a _____

that is not far from _____, but some other populations have

_____ that are extremely far from _____ and

_____ _____ greater than _____ might be

necessary.

CONFIDENCE LEVEL

The _____ _____ is associated with a

_____ _____, such as _____ or

_____. The _____ _____ gives us the

_____ _____ of the _____ used

to construct the confidence interval. Remember the _____ is the

_____ of the _____ _____.

Example 1: Find the indicated critical value $z_{\alpha/2}$.

a. Find the critical value that corresponds to a 98% confidence level.

b. $\alpha = .04$

PROCEDURE FOR CONSTRUCTING A CONFIDENCE INTERVAL FOR μ WITH KNOWN σ.

1. Verify that the _____ are _____.
2. Refer to table _____ or use _____ to find the _____ _____ _____ that corresponds to the desired _____ _____.
3. Evaluate the _____ of _____ _____.
4. Using the value of the _____ _____ of _____ _____ and the value of the _____ _____, find the values of the _____

_____ _____: _____

and _____. Substitute those values in the _____

_____ for the _____ _____:

_____ or _____ or

_____.

5. Round the resulting values by using the following round-off rule.

ROUND-OFF RULE FOR CONFIDENCE INTERVALS USED TO ESTIMATE μ

1. When using the _____ set of _____ to _____ a confidence _____, round the _____ _____ _____ to _____ _____ _____ place than is used for the _____ set of data.
2. When the _____ set of data is _____ and only the _____ _____ (_____) are used, round the _____ _____ limits to the same number of digits as the _____ mean.

Example 2: A simple random sample of 40 salaries of NCAA football coaches has a mean of $415,953. Assume that $\sigma = \$463,364$.

a. Find the best point estimate of the mean salary of all NCAA football coaches.

b. Construct a 95% confidence interval estimate of the mean salary of an NCAA football coach.

c. Does the confidence interval contain the actual population mean of $474,477?

Example 3: Polling organizations typically generate the last digits of telephone numbers so that people with unlisted numbers are included. Listed below are digits randomly generated by STATDISK. Such generated digits are from a population with a standard deviation of 2.87.

1 1 7 0 7 4 5 1 7 6

a. Use the methods of this section to construct a 95% confidence interval estimate of the mean of all such generated digits.

b. Are the requirements for the methods of this section all satisfied? Does the confidence interval from part (a) serve as a good estimate for the population mean? Explain.

FINDING THE SAMPLE SIZE REQUIRED TO ESTIMATE A POPULATION MEAN

Objective:

Notation:

Requirements:

ROUND-OFF RULE FOR SAMPLE SIZE *n*

If the _____ sample size ___ is _____ a _____

_____, round the value of _____ _____ to the next

_____ _____ _____.

Example 4: A researcher wants to estimate the mean grade point average of all current college students in the United States. She has developed a procedure to standardize scores from colleges using something other than a scale from 0 and 4. How many grade point averages must be obtained so that the sample mean is within 0.1 of the population mean. Assume that a 90% confidence level is desired. Also assume that a pilot study showed that the population standard deviation is estimated to be 0.88.

7.4 ESTIMATING A POPULATION MEAN: SIGMA NOT KNOWN

Key Concept…

In this section, we present methods for _____ a

_____ _____ when the

population _____ _____ _____ is not known. With _____ unknown, we use the _____ ___ _____ instead of a _____ _____, assuming the relevant _____ are satisfied. The _____ _____ _____ was developed by William Gosset (1876-1937). William Gosset was a Guinness Brewery employee. He needed a distribution that could be used with small samples. The brewery where he worked did not the publication of research results so he published under the pseudonym "_____". In real circumstances, _____ is typically _____, which makes the methods of this section _____ and _____.

POINT ESTIMATE

The _____ _____ _____ is an _____ estimator of the _____ _____ ____.

STUDENT *t* DISTRIBUTION

If a population has a _____ distribution, then the distribution of

is a _____ _____ _____ for all samples of size _____. A _____ _____ _____ is referred to as a _____ _____. Because we _____ _____ know the value of the _____ _____ _____ _____, we _____ it with the value of the _____ _____ _____ _____, but this introduces another source of _____, especially with _____ _____. In order to maintain a desired _____ _____, we compensate for this additional unreliability by making the _____ _____ _____: we use _____ _____ _____ that are _____ than the _____ _____ of _____ from the _____ _____. A _____ _____ of _____ can be found using _____ or _____ _____.

DEFINITION

The number of **degrees of freedom** for a collection of _____ _____ is the _____ of _____ _____ that can _____ after certain restrictions have been _____ on all data values. The number of _____ of _____ is often abbreviated as _____.

For example: If 10 students have quiz scores with a mean of 80, we can freely assign values to the first _____ scores, but the _____ score is then _____. The _____ of the 10 scores must be _____so the _____ score must be _____ _____ the _____ of the _____ _____ scores. Because the first 9 scores can be _____ selected to any values, we say there are _____ _____ of _____ _____. For the applications of this section, the number of degrees of freedom is simply the _____ _____ _____ _____.

Example 1: A sample size of 21 is a simple random sample selected from a normally distributed population. Find the critical value $t_{\alpha/2}$ corresponding to a 95% confidence level.

PROCEDURE FOR CONSTRUCTING A CONFIDENCE INTERVAL FOR μ WITH UNKNOWN σ.

1. Verify that the _____ are _____.
2. Using _____ _____ of _____, refer to table _____ or use _____ to find the _____ _____ _____ that corresponds to the desired _____ _____. For the _____ _____, refer to the "_____ in _____ _____".
3. Evaluate the _____ of _____ _____.
4. Using the value of the _____ _____ of

_____ _____ and the value of the _____

_____ _____, find the values of the _____

_____ _____: _____

and _____. Substitute those values in the _____

_____ for the _____ _____.

5. Round the resulting values by using the following round-off rule.

ROUND-OFF RULE FOR CONFIDENCE INTERVALS USED TO ESTIMATE μ

1. When using the _____ set of _____ to _____ a confidence _____, round the _____ _____ _____ to _____ _____ _____ place than is used for the _____ set of data.
2. When the _____ set of data is _____ and only the _____ _____ (_____) are used, round the _____ _____ limits to the same number of digits as the _____ mean.

TABLE A-3 *t* Distribution: Critical *t* Values

	Area in One Tail				
	0.005	0.01	0.025	0.05	0.10
Degrees of Freedom	Area in Two Tails				
	0.01	0.02	0.05	0.10	0.20
1	63.657	31.821	12.706	6.314	3.078
2	9.925	6.965	4.303	2.920	1.886
3	5.841	4.541	3.182	2.353	1.638
4	4.604	3.747	2.776	2.132	1.533
5	4.032	3.365	2.571	2.015	1.476
6	3.707	3.143	2.447	1.943	1.440
7	3.499	2.998	2.365	1.895	1.415
8	3.355	2.896	2.306	1.860	1.397
9	3.250	2.821	2.262	1.833	1.383
10	3.169	2.764	2.228	1.812	1.372
11	3.106	2.718	2.201	1.796	1.363
12	3.055	2.681	2.179	1.782	1.356
13	3.012	2.650	2.160	1.771	1.350
14	2.977	2.624	2.145	1.761	1.345
15	2.947	2.602	2.131	1.753	1.341
16	2.921	2.583	2.120	1.746	1.337
17	2.898	2.567	2.110	1.740	1.333
18	2.878	2.552	2.101	1.734	1.330
19	2.861	2.539	2.093	1.729	1.328
20	2.845	2.528	2.086	1.725	1.325
21	2.831	2.518	2.080	1.721	1.323
22	2.819	2.508	2.074	1.717	1.321
23	2.807	2.500	2.069	1.714	1.319
24	2.797	2.492	2.064	1.711	1.318
25	2.787	2.485	2.060	1.708	1.316
26	2.779	2.479	2.056	1.706	1.315
27	2.771	2.473	2.052	1.703	1.314
28	2.763	2.467	2.048	1.701	1.313
29	2.756	2.462	2.045	1.699	1.311
30	2.750	2.457	2.042	1.697	1.310
31	2.744	2.453	2.040	1.696	1.309
32	2.738	2.449	2.037	1.694	1.309
34	2.728	2.441	2.032	1.691	1.307
36	2.719	2.434	2.028	1.688	1.306
38	2.712	2.429	2.024	1.686	1.304
40	2.704	2.423	2.021	1.684	1.303
45	2.690	2.412	2.014	1.679	1.301
50	2.678	2.403	2.009	1.676	1.299
55	2.668	2.396	2.004	1.673	1.297
60	2.660	2.390	2.000	1.671	1.296
65	2.654	2.385	1.997	1.669	1.295
70	2.648	2.381	1.994	1.667	1.294
75	2.643	2.377	1.992	1.665	1.293
80	2.639	2.374	1.990	1.664	1.292
90	2.632	2.368	1.987	1.662	1.291
100	2.626	2.364	1.984	1.660	1.290
200	2.601	2.345	1.972	1.653	1.286
300	2.592	2.339	1.968	1.650	1.284
400	2.588	2.336	1.966	1.649	1.284
500	2.586	2.334	1.965	1.648	1.283
750	2.582	2.331	1.963	1.647	1.283
1000	2.581	2.330	1.962	1.646	1.282
2000	2.578	2.328	1.961	1.646	1.282
Large	2.576	2.326	1.960	1.645	1.282

Example 2: In a study designed to test the effectiveness of acupuncture for treating migraine, 142 subjects were treated with acupuncture and 80 subjects were given a sham treatment. The numbers of migraine attacks for the acupuncture treatment group had a mean of 1.8 and a standard deviation of 1.4. The numbers of migraine attacks for the sham treatment group had a mean of 1.6 and a standard deviation of 1.2.

a. Construct a 95% confidence interval estimate of the mean number of migraine attacks for those treated with acupuncture.

b. Construct a 95% confidence interval estimate of the mean number of migraine attacks for those given a sham treatment.

c. Compare the two confidence intervals. What do the results suggest about the effectiveness of acupuncture?

IMPORTANT PROPERTIES OF THE STUDENT *t* DISTRIBUTION

1. The Student *t* distribution is _____ for different _____ _____.
2. The Student *t* distribution has the _____ general _____ _____ as the _____ _____ distribution, but it reflects the greater _____ (with _____ distributions) that is expected of _____ _____.
3. The Student *t* distribution has a mean of _____ (just as the _____ _____ distribution has a mean of _____).
4. The standard _____ of the Student *t* distribution _____ with the _____ size, but is _____ than _____ (unlike the _____ _____ distribution, which has _____).
5. As the _____ _____ _____, the Student *t* distribution gets _____ to the _____ _____ _____.

CHOOSING THE APPROPRIATE DISTRIBUTION

It is sometimes difficult to decide whether to use the _____

_____ _____ _____ or the

_____ _____ _____.

METHOD	CONDITIONS
Use normal (z) distribution	σ _____ and _____ distributed population or σ known and _____
Use *t* distribution	σ _____ _____ and _____ distributed population or σ _____ _____ and _____
Use a nonparametric method or bootstrapping	Population is _____ _____ distributed and _____

Example 3: Choosing distributions. You plan to construct a confidence interval for the population mean μ. Use the given data to determine whether the margin of error E should be calculated using a critical value of $z_{\sigma/2}$ from the normal distribution, $t_{\sigma/2}$ from a t distribution, or neither (methods of this chapter cannot be used).

a. $n = 7,\ \bar{x} = 80,\ s = 8$, and the population has a very skewed distribution

b. $n = 150,\ \bar{x} = 23.5,\ \sigma = 0.2$, and the population has a skewed distribution

c. $n=10,\ \bar{x}=65,\ s=12$, and the population has a normal distribution

d. $n=13,\ \bar{x}=5,\ \sigma=3$, and the population has a normal distribution

e. $n=92,\ \bar{x}=20.7,\ s=2.5$, and the population has a skewed distribution

FINDING A POINT ESTIMATE AND *E* FROM A CONFIDENCE INTERVAL

The _____ _____ _____ is the value

_____ between the _____

_____ _____. The _____

of _____ _____ is _____ _____ the

_____ between those _____.

Point estimate of μ:

Margin of error:

USING CONFIDENCE INTERVALS TO DESCRIBE, EXPLORE, OR COMPARE DATA

In some cases, we might use a _____ _____ to achieve an ultimate goal of _____ the _____ of a _____ _____. In other cases, _____ _____ might be among the different _____ used to _____, _____, or _____ data sets. When two or more data sets have _____ confidence intervals, one could _____ conclude that there does not appear to be a significant difference between the estimated _____.

TI-83/84 PLUS

Example 4: In a sample of seven cars, each car was tested for nitrogen-oxide emissions (in grams per mile) and the following results were obtained: 0.06, 0.11, 0.16, 0.15, 0.14, 0.08, 0.15 (based on data from the EPA).

a. Assuming that this sample is representative of the cars in use, construct a 98% confidence interval estimate of the mean amount of nitrogen-oxide emissions for all cars.

b. If the EPA requires that nitrogen-oxide emissions be less than 0.165 g/mi, can we safely conclude that this requirement is being met?

Example 5: Listed below are 12 lengths (in minutes) of randomly selected movies from Data Set 9 in Appendix B.

110 96 125 94 132 120 136 154 149 94 119 132

a. Construct a 99% confidence interval estimate of the mean length of all movies.

b. Assuming that it takes 30 minutes to empty a theater after a movie, clean it, allow time for the next audience to enter, and show previews, what is the minimum time that a theater manager should plan between start times of movies, assuming that this time will be sufficient for typical movies?