$\hat{p}$ by 0.5 and replace $\hat{q}$ by 0.5, which is shown in the second formula.

When an estimate $\hat{p}$ is known: $(E)^2 = \left(z_{\alpha/2}\sqrt{\frac{\hat{p}\hat{q}}{n}}\right)^2$

$\cancel{n}\, E^2 = (z_{\alpha/2})^2 \frac{\hat{p}\hat{q}}{\cancel{n}} \cdot n$, $\div E^2$

$$\boxed{n = \frac{(z_{\alpha/2})^2\hat{p}\hat{q}}{E^2}}$$

When no estimate $\hat{p}$ is known:

$$n = \frac{(z_{\alpha/2})^2 0.25}{E^2}$$

ROUND-OFF RULE FOR DETERMINING SAMPLE SIZE

If the computed sample size n is not a whole number, round the value of n up to the next larger whole 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).

$\alpha = 0.05$

$z_{.05/2} = z_{.025} = 1.96$

$E = 2\% = 0.02$

$\hat{p} = 10\% = 0.1$

$\hat{q} = 1 - \hat{p} = 0.9$

$n = \frac{(z_{\alpha/2})^2\hat{p}\hat{q}}{E^2}$

$n = \frac{(1.96)^2(0.1)(0.9)}{(0.02)^2}$

$\boxed{n \approx 865}$

7.3 ESTIMATING A POPULATION MEAN: SIGMA KNOWN

Key Concept…

In this section we present methods for estimating a population mean. In addition to knowing the values of the sample data or statistics, we must also know the value of the population standard deviation, σ. Here are three concepts that should be learned in this section.

* not very real-world to know σ.

1. We should know that the sample mean $\bar{x}$ is the best point estimate of the population mean μ.

2. We should learn how to use sample data to construct a confidence interval for estimating the value of a population mean, and we should know how to interpret such confidence intervals.

3. We should develop the ability to determine the sample size necessary to estimate a population mean.

* very similar to 7.2, but instead of estimating pop. proportion, we're estimating pop. mean.

POINT ESTIMATE

The sample mean $\bar{x}$ is an unbiased estimator of the population mean μ, and for many populations, sample means tend to vary less than other measures of center, so the sample mean $\bar{x}$, is usually the best point estimate of the population mean μ.

*Due to Central limit theorem (6.5)

KNOWLEDGE OF SIGMA

The methods of this section require that we know σ, but in 7.4 we will learn methods to estimate a population mean without knowledge of the value of σ. *more realistic

NORMALITY REQUIREMENT

The population must either be normally distributed or $n > 30$. If $n \leq 30$, the population does not need to have a distribution that is exactly normal as long as it is loosely normal. As long as there are no outliers and if a histogram of the sample data is not dramatically different from being bell-shaped, the normality requirement is satisfied.

SAMPLE SIZE REQUIREMENT

The minimum sample size actually depends on how much the population distribution departs from a normal distribution. Sample sizes of 15 to 30 are sufficient if the population has a distribution that is not far from normal, but some other populations have distributions that are extremely far from normal and sample sizes greater than 30 might be necessary.

CONFIDENCE LEVEL

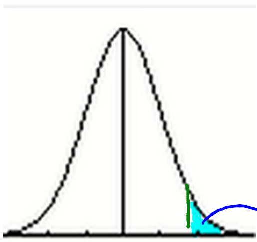
The confidence interval is associated with a confidence level, such as 95% or 0.95. The confidence level gives us the success rate of the procedure used to construct the confidence interval. Remember ~~the~~ that α is the complement of the confidence level.

95% level of confidence → $1-\alpha = 95\%$

so $\alpha = 0.05$

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

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



$1-\alpha = 98\%$

$\alpha = 0.02$

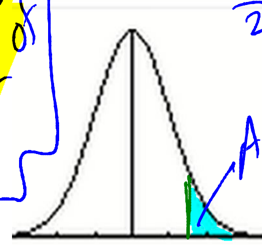
$\alpha/2 = 0.01$

$A = 0.01$ so $\boxed{z_{0.01} = 2.33}$

critical value that corresponds to CI of 98%

b. $\alpha = .04$

$\frac{\alpha}{2} = 0.02$



$A = 0.02$

in between $z = 2.05$ and $z = 2.06$

$\boxed{z_{0.02} = 2.055}$

critical value that corresponds to CI of 96%

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

1. Verify that the requirements are satisfied.
2. Refer to table A2 or use technology to find the critical value $z_{\alpha/2}$ that corresponds to the desired confidence level.
3. Evaluate the margin of error $E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$.
4. Using the value of the calculated margin of error E and the value of the sample mean $\bar{x}$, find the values of the confidence interval limits: $\bar{x} - E$ and $\bar{x} + E$. Substitute those values in the general format for the confidence interval: $\bar{x} - E < \mu < \bar{x} + E$ or $\bar{x} \pm E$ or $(\bar{x} - E, \bar{x} + E)$.
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 original set of data to construct a confidence interval, round the confidence interval limits to one more decimal place than is used for the original set of data.
2. When the original set of data is unknown and only the summary statistics ($n, \bar{x}, s$) are used, round the confidence interval limits to the same number of digits as the sample 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.

$\boxed{\bar{x} = \$415,953}$

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

$1-\alpha = 95\%$

$\alpha = 0.05$

$\alpha/2 = 0.025$

$A = 0.025$ so $z_{0.025} = 1.96$

critical value

$z_{0.025}$

$\bar{x} - E < \mu < \bar{x} + E$

$\boxed{272355.0 < \mu < 559551.0}$

$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$

$E = 1.96 \cdot \frac{463364}{\sqrt{40}} \approx 143597.99$

```
            143597.9913
415953-143597.99
              272355.01
415953+143597.99
              559550.99
```

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

$\boxed{\text{Yes.}}$

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.

1-Var Stats
$\bar{x}=3.9$
$\Sigma x=39$
$\Sigma x^2=227$
$Sx=2.884826203$
$\sigma x=2.736786437$
$\downarrow n=10$

$n=10$

$\bar{x}=3.9$

$\sigma=2.87$

$z_{\alpha/2}=z_{.025}=1.96$

$E=z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}$

$E=1.96\cdot\frac{2.87}{\sqrt{10}}$

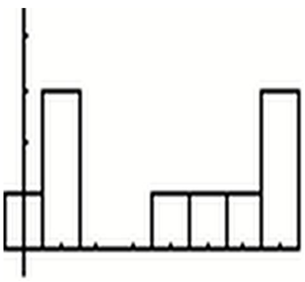
$E\approx 1.78$

$\bar{x}-E<\mu<\bar{x}+E$

$3.9-1.78<\mu<3.9+1.78$

$\boxed{2.1<\mu<5.7}$

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.



No. The histogram of the sample data is far from bell-shaped and $n\le 30$, so the requirements aren't met. The CI is not a good estimate for the population mean → Garbage in, garbage out!

FINDING THE SAMPLE SIZE REQUIRED TO ESTIMATE A POPULATION MEAN

Objective: Determine how large a sample should be in order to estimate the population mean μ.

Notation:

μ is pop. mean

σ is pop. standard deviation

$\bar{x}$ is sample mean

E is desired margin of error

$z_{\alpha/2}$ is critical value

Requirements:

Sample must be a SRS, σ must be known.

$(E)^2=\left(z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}}\right)^2 \rightarrow \frac{n\cancel{E^2}}{\cancel{E^2}}=\frac{(z_{\alpha/2})^2\cdot\sigma^2}{\cancel{n}}\cdot\frac{\cancel{n}}{E^2} \rightarrow n=\frac{(z_{\alpha/2})^2\sigma^2}{E^2}$ or $n=\left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$

ROUND-OFF RULE FOR SAMPLE SIZE *n*

If the computed sample size n is not a whole number, round the value of n up to the next larger whole number.

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.

$1-\alpha = 90\%$

$\alpha = 0.1$

$\alpha/2 = 0.05$

$z_{0.05} = 1.645$

$\sigma = 0.88$

$E = 0.1$

$$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$$

$$n = \left(\frac{1.645 \cdot 0.88}{0.1}\right)^2$$

$$\boxed{n \approx 210}$$

7.4 ESTIMATING A POPULATION MEAN: SIGMA NOT KNOWN

Key Concept…

In this section, we present methods for estimating a population mean when the population standard deviation σ is not known. With σ unknown, we use the student t distribution instead of a normal distribution, assuming the relevant requirements are satisfied. The student t distribution 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 allow the publication of research results so he published under the pseudonym "Student". In real circumstances,

σ is typically unknown, which makes the methods of this section realistic and practical.

POINT ESTIMATE

The sample mean $\bar{x}$ is an unbiased estimator of the population mean μ.

STUDENT *t* DISTRIBUTION

If a population has a normal distribution, then the distribution is a student t distribution for all samples of size n. A student t distribution is referred to as a t distribution. Because we do not know the value of the population standard deviation σ, we estimate it with the value of the sample standard deviation s, but this introduces another source of unreliability, especially with small samples. In order to maintain a desired confidence level, we compensate for this additional unreliability by making the confidence interval wider: we use critical values $t_{\alpha/2}$ that are larger than the critical values of $z_{\alpha/2}$ from the normal distribution. A critical value of $t_{\alpha/2}$ can be found using technology or Table A-3.