INTRODUCTION TO CHAPTER 7

<u>Inferential Statistics</u>: We use sample data to make generalizations or inferences about a population.

<u>In Chapter 7</u>: We will use sample data to estimate the value of the population parameters proportion and mean.

We will also present methods for determining the sample size necessary to estimate those parameters.

<u>In Chapter 8</u>: We will use sample data to test some claims, or hypotheses, about a population.

Read the Chapter Problem on page 327.

7.2 Estimating a Population Proportion

In this section, we are going to estimate the proportion of times that a person believes the earth is warming.

We use the sample of 1501 trials and consider the sample proportion $=$ 70% as the best **point estimate** of the population proportion.

Since we have no indication of how good our best estimate is, instead of using a single value, .70, we may use a range of values (or interval) that is likely to contain the true value of the population proportion. This is called a **confidence interval**.

With a confidence interval is associated a **degree of confidence**.

The degree of confidence tells us the percentage of times that the confidence interval actually does contain the population parameter (e.g. proportion or mean), assuming taht the estimation process is repeated a large number of times.

We will be working under the following three assumptions:

ASSUMPTIONS

1. The sample is a simple random sample.
2. The conditions for a binomial distribution are satisfied. That is: there is a fixed number of trials, the trials are independent, there are two cateogries of outcomes, and the probabilities remain constant for each trial.
3. The normal distribution can be used to approximate the distribution of sample proportions because np $\geq$ 5 and nq $\geq$ 5 are both satisfied.

NOTATION FOR PROPORTIONS

p is the population proportion (percent of those who have the "quality" under discussion)

$x =$ # of successes (# of those who have the "quality" under discussion) in a sample of size n.

$\widehat{p}$ (read "p-hat") is the sample proportion (percent of those in the sample who have the "quality" under discussion) $\widehat{p} = \frac{x}{n}$

$\widehat{q}$ (read "q-hat") is the percent of those in the sample who do not have the "quality" under discussion) $\widehat{q} = 1 - \widehat{p}$

SOME DEFINITIONS:

ESTIMATE: is a specific value or range of values used to approximate a population parameter.

POINT ESTIMATE: a single value (or point) used to approximate a population parameter. *(It corresponds to a point on the number scale)*

BEST POINT ESTIMATE OF POPULATION PROPORTION p:

Use $\widehat{p} = \frac{x}{n}$ as best point estimate

STANDARD DEVIATION OF SAMPLE PROPORTIONS: $\sigma_{\text{sample prop}} = \sqrt{\frac{\widehat{p}\widehat{q}}{n}}$

(Another type of estimate that reveals how good the point estimate is, is the:)
CONFIDENCE INTERVAL (or interval estimate): a range (or interval) of values that is *likely* to contain the true value of the population parameter. (A way of fine-tuning our estimate.)

(A confidence interval is associated with a degree of confidence which is a measure of how certain we are that our interval contains the population parameter)
DEGREE OF CONFIDENCE (or level of confidence, or confidence coefficient): the probability 1-α that the confidence interval contains the true value of the population parameter. (*The value of α is the complement of the degree of confidence)*

Often used percentages: 90% (α=.10), 95% (α=.05), 99% (α=.01)

Example: The 0.95 (or 95%) degree of confidence interval estimate of the population proportion p is $0.676 < p < 0.724$

INTERPRETING A CONFIDENCE INTERVAL
In our example, we are 95% confident that the interval from 0.676 to 0.724 actually does contain the true value of p. This means that if we were to select many different samples of size 1501 and construct the confidence intervals, 95% of them would actually contain the value of the population proportion p.

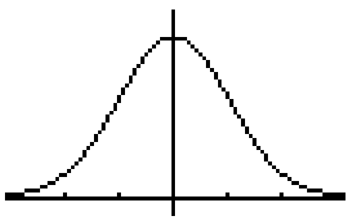
NOTE: It is incorrect to say that there is a 95% chance that the true population proportion will fall between 0.6766 and 0.724. (Why? Because p is a constant, not a random variable. p has already occurred, we just don't know what it is.)

CRITICAL VALUES: Borderline numbers that are the z-scores that separate sample means that are likely to occur from those that are unlikely to occur.

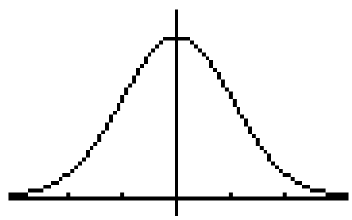
- $z_{\alpha/2}$ is the positive z value separating an area of $\alpha/2$ in the right tail of the standard normal distribution.

- $-z_{\alpha/2}$ separates an area of $\alpha/2$ in the left tail.

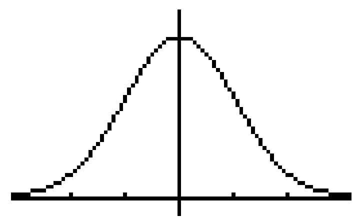
EXAMPLES Look up $z_{\alpha/2}$ in Table A-2:



for 90%: $z_{\alpha/2} =$

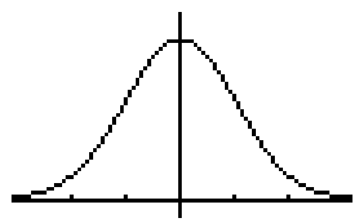


for 95%: $z_{\alpha/2} =$



for 99%: $z_{\alpha/2} =$

*****#8, pg. 340**



MARGIN OF ERROR (E) (or maximum error of the estimate): maximum likely difference between the observed sample proportion $\hat{p}$ and the true value of the population proportion p. It is calculated by multiplying the critical value and the standard deviation of the sample proportion:

$$E = z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

*****#18, pg. 340**

CONFIDENCE INTERVAL FOR THE POPULATION PROPORTION p:

$$\widehat{p} - E < p < \widehat{p} + E,$$

where $E = z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}}$

PROCEDURE FOR CONSTRUCTING A CONFIDENCE INTERVAL FOR *p*:

1. Find the critical value $z_{\alpha/2}$.
2. Evaluate the margin of error E.
3. Then using E and the sample proportion $\widehat{p}$, the confidence interval is:

$$\widehat{p} - \boldsymbol{E} < \boldsymbol{p} < \widehat{\boldsymbol{p}} + \boldsymbol{E}$$

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

3 significant digits (just like probabilities)

******#22, pg. 340***

******#30, pg. 341***

USING THE TI-83 TO CONSTRUCT CONFIDENCE INTERVALS FOR p:

STAT>>TESTS choose A:1-propZInt. Then enter x, n and α and calculate.

******#30, pg. 341***

FINDING THE POINT ESTIMATE AND E FROM A CONFIDENCE INTERVAL

Point estimate of p"

$$\widehat{p} = \frac{(upper\, confidence\, interval\, limit) + (lower\, confidence\, interval\, limit)}{2}$$

Margin of Error:

$$E = \frac{(upper\, confidence\, interval\, limit) - (lower\, confidence\, interval\, limit)}{2}$$

*****#14, pg. 340**

If you solve $E = z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}}$ for n, you get a formula which is used to determine the appropriate sample size that is needed to estimate p.

DETERMINING SAMPLE SIZE:

If we have an estimate of $\widehat{p}$: $\boldsymbol{n = \frac{[z_{\alpha/2}]^2\widehat{p}\widehat{q}}{E^2}}$

If no estimate of $\widehat{p}$ is known: $\boldsymbol{n = \frac{[z_{\alpha/2}]^2 0.25}{E^2}}$

ROUND-OFF RULE FOR SAMPLE SIZE, n: Use the computed size if it is a whole number. If it is not a whole number, round it up to the next higher whole number.

*****#26, pg. 340**

*****#28, pg. 340**

MORE PRACTICE

*****#32, pg. 341**

*****#34, pg. 341**