AP Statistics
Unit 06 – Day 02 Notes
Confidence Intervals: Proportions

Name Key
Period

Conditions for Constructing A Confidence Interval About a Proportion

- **Random Condition:** The data come from a well-designed random sample or randomized experiment.
- **10% Condition:** When sampling without replacement, check that the sample is less than 10% of the population. *This allows us to use the standard deviation equation.*
- **Large Counts Condition:** $n\hat{p} \geq 10$, $n\hat{q} \geq 10$. *This allows us to do Normal calculations.*

Constructing a Confidence Interval for p

When the conditions are met.....

- Sampling distribution of $\hat{p}$ will be approximately Normal
- $\mu_{\hat{p}} = p$
- standard deviation $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- statistic $\pm$ (critical value)·(standard deviation of statistic)

Standard Error: When the standard deviation of a statistic is estimated from data, we call it standard error and calculate it with

- $SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
- describes how close the sample proportion will typically be to the population proportion

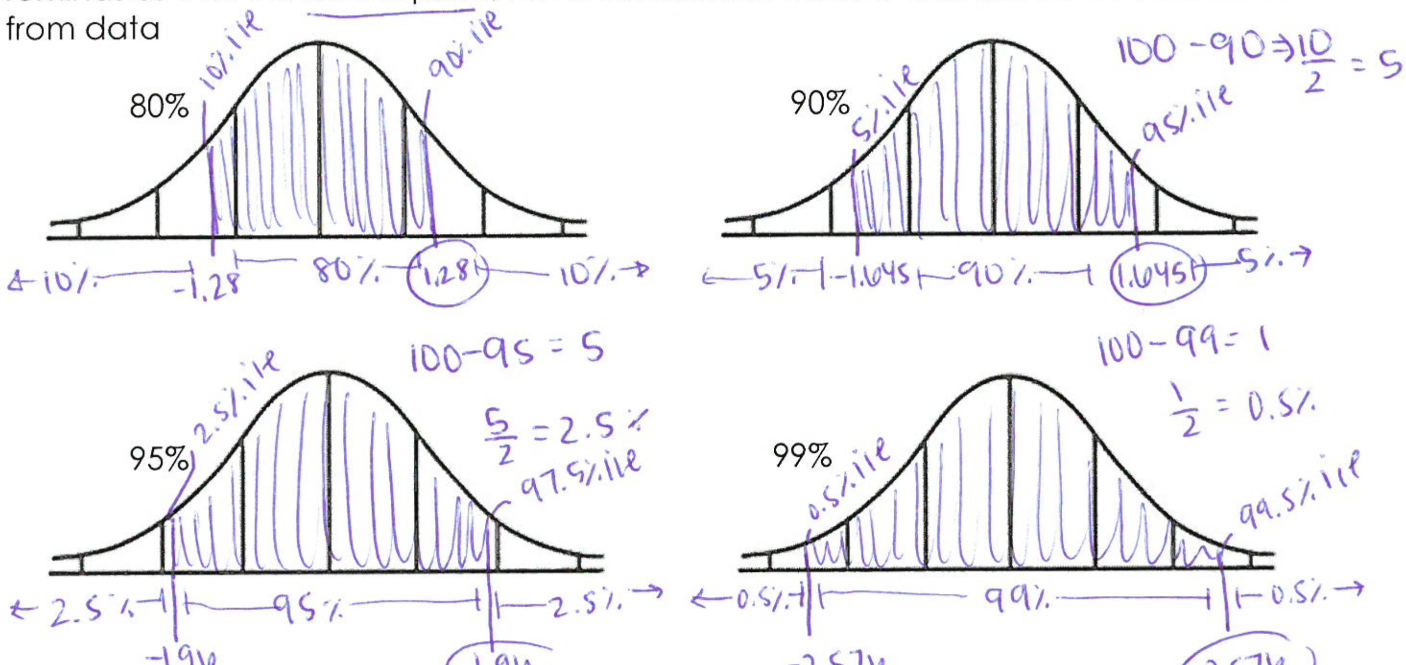
Critical Value: the multiplier that makes the interval wide enough to have the stated capture rate. Depends on the confidence level and the sampling distribution of the statistic.

z^* or t^*

z critical value

How to find z^* by hand:

- the number of standard deviations or critical values from where the sample proportion is
- reminds us that it is critical point, not a standardized score that has been calculated from data



How to find z^* in the calculator:

1. 2nd Vars
2. invNorm(percentile corresponding to correct confidence interval)
3. ENTER

Example:

95% confidence interval → area: 0.025 or 0.975 (plug into invNorm)

Answer: 1.96

How to find z^*on the t-table:

- since the conditions are met and it is a normal distribution you can use the last row on the t-table where the df are ∞.

Example 1:

Use technology to find the critical value of z^* for an 92% confidence interval.

100 - 92 = 8%

8/2 = 4% on either side

4th %ile or 96%ile

invNorm(0.96) = 1.751

Confidence Interval Equation for Proportions

ONE-SAMPLE Z INTERVAL FOR A POPULATION PROPORTION

statistic ± (critical value)·(standard deviation of statistic)

$$\boxed{\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Example 2:

Mrs. De Marre's class took an SRS of number of people at BHS that have lockers. Out of 1655 students total, only 210 had a locker.

a) Calculate and interpret a 90% confidence interval for p. (Just these steps).

$\hat{p} = 210/1655 = 0.1269$ $\quad z^* = 1.645$ (calc) Option A:

DO: $0.1269 \pm 1.645 \sqrt{\frac{(0.1269)(1-0.1269)}{1655}} = \boxed{(0.11343, 0.14035)}$

Conclude: we are 90% confident that the interval from 0.11343 to 0.14035 captures the true proportion of students at BHS with a locker

b) Mrs. De Marre claims that at least a quarter of the students at BHS have a locker. Use your result from part (a) to comment on this claim.

0.25 or 25% is not in our interval of plausible values so Mrs. D is probably wrong.
(her claim is probably not true)

Confidence Intervals: A FOUR-STEP PROCESS

1. **State**: What parameter do you want to estimate, and at what confidence level?
   we want to find p with _% confidence. $\hat{p}$=
2. **Plan**: Identify the appropriate inference method. Check conditions.
   Random, 10% condition, Large counts AND we will construct a 1 proportion z interval to estimate p.
3. **Do**: If the conditions are met, perform calculations.
   z^*, CI w/ #s plugged in, (,)
4. **Conclude**: Interpret your interval in the context of the problem.
   sentence frame

Example 3:

Alcohol abuse has been described by college presidents as the number one problem on campus, and it is an important cause of death in young adults. How common is it? A survey of 10,904 randomly selected U.S. college students collected information on drinking behavior and alcohol-related problems. The researches defined "frequent binge drinking" as having five or more drinks in a row three or more times in the past two weeks. According to this definition, 2486 students were classified as frequent binge drinkers.

n

x

a. Identify the parameter of interest.

State: p = the true proportion of US college students who are classified as frequent binge drinkers.

$\hat{p} = 2486/10904 = 0.2280$

b. Check conditions for constructing a confidence interval for the parameter.

Plan: Random: randomly selected US college students ✓

10% condition: n = 10904 109040 < all US college students ✓

Large counts: $n\hat{p} \geq 10$ $n\hat{q} \geq 10$

$10904 \cdot 0.2280 = 2486 \geq 10$ ✓ $10904(1-0.2280) = 8418 \geq 10$ ✓

Because our conditions are met, we will calculate/construct a 1-proportion z interval. (or 1-sample z interval for proportions).

c. Find the critical value for a 99% confidence interval. Show your method. ~~Draw a picture~~. Then calculate the interval.

DO: $z^* = 2.576$

$$0.2280 \pm 2.576\sqrt{\frac{(0.2280)(1-0.2280)}{10904}} = \boxed{(0.21764, 0.23834)}$$

d. Interpret the interval in context.

Conclude: We are 99% confident that the interval from 0.21764 to 0.23834 captures the true proportion (p) of US college students who are classified as frequent binge drinkers.

Example 4:

The Gallup Youth Survey asked a random sample of 439 U.S. teens aged 13 to 17 whether hey thought young people should wait to have sex until marriage. Of the sample, 246 said "Yes." Construct and interpret a 95% confidence interval for the proportion of all teens who would say "Yes" if asked this question.

State: we want to estimate the true proportion (p) of all teens who think young people should wait to have sex until marriage with 95% confidence. $\hat{p} = 246/439 = 0.5604$

Plan: Random: random sample of 439 US teens ✓

10% condition: $n = 439$ $4390 <$ all US teens aged 13-17 ✓

Large counts: $n\hat{p} = 246 \geq 10$ ✓ $n\hat{q} = 193 \geq 10$ ✓

Because our conditions are met, we will construct a 1-proportion z-interval to estimate p.

DO: $z^* = 1.96$

$$0.5604 \pm 1.96\sqrt{\frac{(0.5604)(1-0.5604)}{439}} = \boxed{(0.51393, 0.60679)}$$

conclude: we are 95% confident that the interval from 0.51393 to 0.60679 captures the true proportion of all teens who think young people should wait to have sex until marriage.

Confidence Intervals in the Calculator

STAT → TESTS → 1-PropZInt (this is option A)

x = number of successes

n = total number in sample

C-level =

Choosing a Sample Size

- In doing an experiment, we may want to choose a sample size that allows us to estimate a population proportion within a given margin of error
- Because the margin of error involves the sample proportion $\hat{p}$, we have to guess the values of $\hat{p}$ when choosing n. Here are two ways to do that:
    1. Use a guess for $\hat{p}$ from a previous study
    2. Use $\hat{p} = 0.5$ as a guess. The guess is conservative because if we end up with another $\hat{p}$, we will get a margin of error smaller than planned

- $ME \geq z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Example 5:

A company has received complaints about its customer service. The managers intend to hire a consultant to carry out a survey of customers. Before contacting the consultant, the company president wants some idea of the sample size that she will be required to pay for. One critical question is the degree of satisfaction with the company's customer service, measured on a 5-point scale. The president wants to estimate the proportion p of customers who are satisfied (that is, who choose either "somewhat satisfied" or "very satisfied", the two highest levels on the 5-point scale). She decides that she wants ~~to~~ her estimate to be within 3% at the 95% confidence level.

a. How large a sample is needed?

$$\frac{n\left(\frac{0.03}{1.96}\right)^2}{\left(\frac{0.03}{1.96}\right)^2} \geq \frac{\cancel{1.96}}{\cancel{1.96}} \frac{\left(\sqrt{\frac{(0.5)(0.5)}{\cancel{n}}}\right)^2 \cdot \cancel{n}}{\left(\frac{0.03}{1.96}\right)^2}$$

$n \geq 1067.07$

$\boxed{n \geq 1068}$ customers

b. In the company's prior-year survey, 80% of customers surveyed said they were "somewhat satisfied" or "very satisfied." Using this value as a guess for $\hat{p}$, find the sample size needed for a margin of error of 3% at a 95% confidence level.

$$0.03 \geq 1.96\sqrt{\frac{(0.8)(0.2)}{n}}$$

$n \geq 682.93$

$\boxed{n \geq 683 \text{ customers}}$

c. What if the company president demands 99% confidence instead? Determine how this would affect your answer in part (a).

an increased confidence level requires an increased sample size if we want the same margin of error.

$$0.03 \geq 2.576\sqrt{\frac{(0.5)(0.5)}{n}}$$

$n \geq 1843.03$

$n \geq 1844$ customers

our sample size increased