AP STATS

Reading Logs
Chapt 19 notes
Confidence Intervals
These symbols all have very different meanings and it is important that you understand the differences and use the correct one. Define each:

- \( n \) sample size
- \( N \) actual population size
- \( \bar{x} \) sample mean
- \( \mu \) population mean
- \( s \) sample standard deviation
- \( \sigma \) population standard deviation
- \( \hat{p} \) sample proportion
- \( p \) population proportion
- \( \mu_{\bar{x}} \) mean of the sampling distribution of sample means
- \( \sigma_{\bar{x}} \) standard dev. of sampling distribution of sample means
- \( \mu_{\hat{p}} \) mean of the sample proportions
- \( \sigma_{\hat{p}} \) standard deviation of sample proportions
Chapter 19 – Confidence Intervals for Proportions

What's the difference between the 'standard deviation' and the 'standard error'?  
\[ SD_P, \text{ standard deviation of sampling distribution is } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ using population proportion} \]
\[ SE(\hat{p}), \text{ standard error of sample, } \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \text{ using sample proportion} \]

What is the purpose of a confidence interval?  
Confidence intervals give the range of values in which the true parameter value can be found.

What does “95% confidence” really mean?  
95% confidence - 95% of all data samples will yield intervals that contain the true parameter.

What happens to the margin of error as our confidence level changes?  
Confidence interval = estimate \pm ME  
As the confidence interval increases the ME increases  
As the confidence interval decreases the ME decreases

What are the assumptions/conditions needed for a confidence interval for proportions?  
Independence Assumptions: Flawless Independence 5 Randomization 5 Check conditions \( nP \geq 10 \) and \( nQ \geq 10 \)

Sample size Assumptions: Success/Failure \( nP \geq 10 \) \( nQ \geq 10 \)

What is the formula used for a 1-proportion Z-interval?  
Confidence interval for proportion \( p \) -
List and explain the steps given by the book to find a confidence interval for a proportion.

1. **Plan**
   - State problem & W’s
   - Identify the parameter
   - Identify the population
   - Choose confidence level

2. **Model**
   - Check conditions
     - Randomization
     - Plausible Independence
     - 10% Condition
     - Success/Failure Condition

3. **State sampling distribution**
   - Conditions are met, so I can use a Normal model to find one-proportion z-interval

4. **Mechanics: Construct confidence interval**
   a) Find Standard Error \( \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \)
   b) Find Margin of Error \( \hat{p} \cdot z^* \) (use 1.96 for 95% confidence)
   c) Write confidence interval

5. **State conclusion**

What is the general interpretation of a confidence interval?

I am \( % \) confident that the interval from \( \) to \( \) captures the true proportion of \( \).

What are two common ways to reduce your margin of error?

1) **Choose larger sample size**
2) **Reduce level of confidence.**

When solving for a minimum sample size, if you do not have an estimate for \( \hat{p} \), what should you use? **Use \( \hat{p} = .5 \)**

When solving for a minimum sample size, oftentimes you will get a decimal answer, what should you do? **Round up**

In order to cut your margin of error in half, you need a sample \( \frac{4}{1} \) times as large as the original sample.

*Write a brief summary of Chapter 19 here:*
AP STAT: Ch. 19 notes CONFIDENCE INTERVALS: For single sample proportions

- From Ch. 18.... Distribution of sample proportions (\( \hat{p} \)) follows the model: (if checks pass)
  \[ N(p, \sqrt{\frac{p(1-p)}{n}}) \]
- However, most of the time we don’t know the population proportion (p).
- We take samples and calculate \( \hat{p} \) in order to try and find p.
- *Since we don’t know p, we can’t find the standard deviation*
- \( \hat{p} \) is the estimate (estimator) for p.
- So... we can estimate the std. deviation with standard error:
  \[ SE_\hat{p} = \sqrt{\frac{\hat{p}(\hat{q})}{n}} \]

**CONFIDENCE INTERVALS... Intro**

**EXAMPLE:** It is election season. You open the newspaper and see a headline: 57% of the nation favors the Democratic candidate for President.

a) Does EXACTLY 57% of the nation favor the Democrat? Why?

\[ \text{You took a sample, not a census. So your } p \text{ may not be true.} \]

b) As you keep reading, you see the following: There is a margin of error of 5%. What does this mean about the TRUE PERCENT of the nation that favors the Democrat? Do you think that the pollsters have a really good idea of how many people will vote for the Democrat on Election Day?

\[ 57 \pm 5 \text{ The true percent is probably between 52\% to 62\%} \]

c) As you keep reading, you see the following: The results of this poll are given with 70% confidence. Do you think that their results are reliable? Why?

\[ \text{No, 70\% is not very high confidence} \]

d) What if they change their confidence to 95%? Do you think that their results are reliable? Why?

\[ \text{Yes, 95\% confidence is more reliable than 70\%} \]

e) What if the results are as follows: 57% will vote Democrat, margin of error of 15%, 99% confidence. What can you say about the results? Are they reliable? Why?

\[ \text{Now the the interval is } 57\% \pm 15\% \]
\[ 42 \text{ to } 72\% \]
\[ \text{This is large interval even though more reliable.} \]
CONFIDENCE INTERVALS... The basics:

- Based on \( \hat{p} \) and the sampling distribution of \( \hat{p} \).
- Start with \( \hat{p} \) and give ourselves a margin of error (MOE) on either side.
  - MOE = because many of our samples are not perfect
- Size of the interval (of the MOE) is based on sample size and level of confidence.
  - Larger sample size = smaller interval (more accurate)
  - Larger confidence level = larger (wider) interval (need more room for error)
- Basic Setup:
  Estimate ± MOE = \((a, b)\)
- Specifically, for 1-proportion sample:
  \[
  \hat{p} \pm \left( \frac{Z^*}{(SE)} \right) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = (a, b)
  \]
- \( Z^* \) = critical value
  - 90% of the data is between ± \( Z \) (confidence level)
- 3 main confidence levels:
<table>
<thead>
<tr>
<th>Confidence Level (C)</th>
<th>Tail Area</th>
<th>( Z^* )</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>0.05</td>
<td>1.645</td>
</tr>
<tr>
<td>95%</td>
<td>0.025</td>
<td>1.960</td>
</tr>
<tr>
<td>99%</td>
<td>0.005</td>
<td>2.576</td>
</tr>
</tbody>
</table>

- Other confidence levels... how do we find \( Z^* \)?
  - invNorm of the left tail gives you the \( -Z \)
  - For example, 90% confidence:
    - \( -Z = \text{invNorm}(0.05) = -1.645 \)

Interpreting Confidence Levels and Confidence Intervals

Interpret the confidence level: "\( \_\_\_\% \) of all possible samples will produce intervals that capture the value of the parameter".

Gives the probability that the interval will capture the true parameter in repeated samples.

Interpret the confidence interval: "We are \( \_\_\_\% \) confident that the interval from \( \_\_\_ \) to \( \_\_\_ \) captures the actual value of \( \_\_\_ \) (population parameter in context)".

Note: The higher the confidence level, the wider the interval.

Caution: The confidence is in the process used to generate the interval, not the interval itself. It does NOT mean "There is a \( \_\_\_\_\% \) chance (or probability) that the true population mean/proportion is between \( \_\_\_\_\_ \) and \( \_\_\_\_\_ \)." Either it is 'in' or it is 'out'.

Confidence Level & ME

Our sample estimate will not differ from the true population value by more than ME \( \_\_\_\% \) for \( \_\_\_\% \) of all samples.
 Specifies for Confidence Intervals for Proportions

- **State** – Identify the proportion in context you want to estimate and at what confidence level.
- **Plan** –
  - **Method**: 1 sample z confidence interval for proportion
  - **Check Conditions**:
    - **Random**: Data from a random sample or randomized experiment
    - Normal: \( np \geq 10 \) and \( n(1 - p) \geq 10 \) (number of successes/fails ≥ 10)
    - Independent observations: Check \( \text{Pop size} > 10 \times \text{sample size} \) if not sampling with replacement
  - **Do**
    \[
    \hat{p} \pm z \times \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}
    \]
  - \( z \) critical value, the z-score necessary to capture the confidence level \( \alpha \) of the standard normal distribution (other levels can be determined from z-table)
- **Conclude** – both sentences

On Calculator (use to check, must still must document everything) STAT→TESTS→1-PropZInt

Example: We want to know the real improvement rate for a new medication. We conduct an experiment and find that out of 53 subjects, 27% of them report improvement with the new medications. Create a 95% confidence interval (and interpret).

\[ \hat{p} = \frac{27}{53} = 0.51 \]

\[ n = 53 \]

\[ \alpha = 0.05 \]

\[ \text{SE}_p = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \]

\[ \text{SE}_p = \sqrt{\frac{0.51(0.49)}{53}} \]

\[ \text{SE}_p = 0.06 \]

\[ z = 1.96 \]

\[ ME = z \times SE_p = 1.96 \times 0.06 = 0.12 \]

\[ \hat{p} \pm ME = 0.51 \pm 0.12 \]

\[ (0.39, 0.63) \]

**C** – We are 95% confident that between 39% and 63% of people will improve with the new medication.

**OR** We are 95% confident that the true % improvement with medication is between 15% and 39%.
MORE ABOUT MARGIN OF ERROR:

Things that affect MOE (margin of error):
- Sample size (n)
- Confidence Level (Z*)

\[ \text{MOE} = Z^* \frac{SE}{\sqrt{n}} = Z^* \sqrt{\frac{p(1-p)}{n}} \]

*If we want a particular MOE, we can set a level of confidence and a sample size in order to attain that MOE.*

**Example:** Let's go back to our example about the improvement rate with a new medication. We found 27% improvement. How many subjects would we need in a new experiment to make a 98% confidence interval while still keeping a 5% MOE?

\[
\hat{p} = 0.27 \\
Z^* = \text{InvNorm}(1-0.05) \\
= 2.326 \\
ME = 0.05 = 2.326 \left( \sqrt{\frac{0.27(1-0.27)}{n}} \right) \\
\frac{0.05}{2.326} = \sqrt{\frac{0.27(1-0.27)}{n}} \\
0.05 = \frac{2.326}{2.326} \sqrt{\frac{0.27(1-0.27)}{n}} \\
n = \left( \frac{2.326}{0.05} \right)^2 \left( \frac{0.27(1-0.27)}{0.05} \right) \approx 426.546
\]

*Use 427 subjects for ME of 5% and CI of 98%.*

**NOTE:** If you are not given a value of \( \hat{p} \), you can use \( \frac{1}{2} \) as an estimate.

**Example:** What sample size must be used to estimate the outcome of a political election with a margin of error of 3% and 99% confidence?

\[
\hat{p}_{\text{estimated}} = 0.5 \\
ME = 0.3 = 2.576 \sqrt{\frac{0.5 \times 0.5}{n}} \\
n = \left( \frac{2.576}{0.3} \right)^2 \left( 0.5 \times 0.5 \right) = 184.327
\]

A sample size of 1844 should be used for a ME of 3% and 99% CI.

**Example:** What sample size must be used to estimate the true percent of left-handed people in the nation with 90% confidence and a margin of error of 8%? Assume that it has been shown in previous research that the percent of left-handed people was 38%.

\[
\hat{p} = 0.38 \\
Z^* = \text{InvNorm}(1-0.05) \\
= 1.645 \\
ME = 0.08 = (1.645) \sqrt{\frac{0.38 \times 0.62}{n}} \\
n = \left( \frac{1.645}{0.08} \right)^2 \left( 0.38 \times 0.62 \right) \approx 99.62
\]

A sample size of 100 is needed for CI of 90% and ME 8%.
Hershey’s Kisses and Confidence Intervals

Use this space to record the names of the other students in your group. You MUST make groups of FIVE!

1.

2.

3.

4.

“I got the instructions from my Statistics Professor. He was 80% confident that the true location of the restaurant was in this neighborhood.”

In this activity, we will estimate a confidence interval for how often a Hershey's kiss lands on its base as opposed to its side. To do this, we will drop Hershey's kisses, count how many land on their base, and calculate the confidence interval.

What is the population?

What is the sample?
Before you begin, make a guess about what will happen. If you could toss one Hershey’s Kiss over and over and over, what proportion of all tosses do you think the Kiss would settle on its base?

Toss your Hershey kiss 30 times. Record the result of each toss (S or B) in a table like the one shown. In the third column, calculate the proportion of base landings you have obtained so far.

<table>
<thead>
<tr>
<th>Toss</th>
<th>Outcome</th>
<th>Cumulative Probability of B's</th>
<th>Toss</th>
<th>Outcome</th>
<th>Cumulative Probability of B's</th>
<th>Toss</th>
<th>Outcome</th>
<th>Cumulative Probability of B's</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>S</td>
<td>0.04</td>
<td>11</td>
<td>S</td>
<td>0.12</td>
<td>21</td>
<td>S</td>
<td>0.20</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>0.06</td>
<td>12</td>
<td>S</td>
<td>0.13</td>
<td>22</td>
<td>S</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>0.12</td>
<td>13</td>
<td>S</td>
<td>0.14</td>
<td>23</td>
<td>S</td>
<td>0.25</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>0.14</td>
<td>14</td>
<td>S</td>
<td>0.15</td>
<td>24</td>
<td>S</td>
<td>0.26</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>0.16</td>
<td>15</td>
<td>S</td>
<td>0.17</td>
<td>25</td>
<td>S</td>
<td>0.27</td>
</tr>
<tr>
<td>6</td>
<td>B</td>
<td>0.18</td>
<td>16</td>
<td>S</td>
<td>0.19</td>
<td>26</td>
<td>S</td>
<td>0.29</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>0.20</td>
<td>17</td>
<td>S</td>
<td>0.21</td>
<td>27</td>
<td>S</td>
<td>0.30</td>
</tr>
<tr>
<td>8</td>
<td>B</td>
<td>0.22</td>
<td>18</td>
<td>S</td>
<td>0.23</td>
<td>28</td>
<td>S</td>
<td>0.33</td>
</tr>
<tr>
<td>9</td>
<td>B</td>
<td>0.24</td>
<td>19</td>
<td>S</td>
<td>0.25</td>
<td>29</td>
<td>S</td>
<td>0.35</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>0.26</td>
<td>20</td>
<td>S</td>
<td>0.27</td>
<td>30</td>
<td>S</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Make a scatterplot with the number of tosses on the horizontal axis and the cumulative proportion of B’s on the vertical axis. Connect consecutive points with a line segment. Does the overall proportion of B’s seem to be approaching a single value?
Compare your final sample proportion of bases with the others in your group. Did you all get the same answer?

Your result: \( \hat{p} = \frac{7}{30} = 0.2333 \)

The results of the others in your group:

Follow these steps to make a 95% confidence interval based on your result.

1. Calculate the sample standard deviation of your sample proportion:
   \[ SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \]

2. Calculate the Z-score with your calculator that corresponds to 95% confidence.
   \[ Z^* = \text{InvNorm}(\frac{1-CL}{2}) = 1.96 \]

3. The confidence interval is \( \hat{p} \pm Z^* \times SE \)
   your lower bound = \( 0.0819 \)
   your upper bound = \( 0.3847 \)

Use the lines below to roughly draw the confidence intervals of each person in your group.
(An example of a CI from 0.13 to 0.37 is given.)

Interpret the 95% confidence interval on your own (without your notes). Write your interpretation below.
Compare your interpretation with your group members' interpretation and come to an agreement on an appropriate interpretation. Write it below.

I am 95% confident that the true proportion for a Hershey kiss landing on its base is between ___________ and ___________.

Now merge your individual samples into one big sample, and make a new confidence interval. Remember that n is changing! Show work below.

\[ \hat{p} = \frac{32}{120} \]

\[ n = 120 \]

Group confidence interval: ___________

Draw this confidence interval on the above lines and label it “GROUP.”

Compare this confidence interval with the confidence intervals from your individual samples.

• Are the means different? If so, how?

• Are the standard deviations different? If so, how?

\[ \text{Standard Error is smaller because } n \text{ is larger.} \]

• Are the widths of the confidence intervals different? If so, how?

Yes, Smaller/narrower

How did increasing the sample size affect the width of the confidence interval?

Larger n, Smaller interval

What statistical concept or principle does this remind you of?

CLT, the larger n, closer to mean

When you are done, draw your group's confidence interval on the board.

Which of the following is true?

A. There is a 95% probability that the true mean will fall in our interval
B. There is a 95% probability that our interval will include the true mean.

Which of the following is true?

A. BEFORE we take the sample, there is a 95% probability that the confidence interval we will create WILL include the true mean.
B. AFTER we take the sample, there is a 95% probability that the confidence interval we created DOES include the true mean.

\[ \frac{1}{13} \]

\[ \frac{1}{3} \]

It does or does not