# Stat 250 Final Review

## Confidence Intervals / Sample Sizes

# Confidence Interval Questions

- There are four different confidence interval questions you can be asked:
  - Calculations (including full, margin of error, standard error and critical value/multiplier)
  - Interpretations
  - Conditions
  - Impact of changing confidence level, sample size, p, or sigma

# Confidence Interval Types

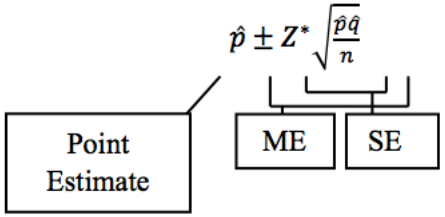
- For all of those types of questions, you first need to correctly identify the type of confidence interval being run. We have learned about six different confidence intervals:
  - One proportion (z)
  - Two proportions (z)
  - One mean
    - Given s (sample standard deviation), use t interval
  - Two means
    - Independent samples, unpooled – you are not responsible for the pooled version. Use t. Can use software for d.f. or smaller of two sample sizes minus 1.
    - Paired data, use t interval working with differences

# Calculation Questions

- Full calculation. Make sure the standard z critical values are on your cheat sheet (90% = 1.645, 95% = 1.96, 99% = 2.576) and that you can get t* values with Excel or calculator
  - Be able to do one prop and two prop. Graphing calculators have shortcuts you can use to double check (STAT>TESTS>One prop = 1-PropZInterval,Two prop = 2-PropZInterval)
  - Be able to do one mean, difference in means, and mean difference when given xbar and s values. Calculator shortcuts are Tinterval and 2-SampTInt

# Other Calculations

- Realize that all the calculations can break down into multiple parts: ($\widehat{q}$ is $(1 - \hat{p})$)




Point estimate +/- critical value * standard error

Note: In our class, we sometimes call the critical value the "multiplier."

1. In a random sample of 228 self-employed individuals, 27 reported having their tax returns audited by the IRS in the past year.

Which of the following is a 90% confidence interval for the population proportion of self-employed individuals who have been audited by the IRS in the past year?

A. (0.1055, 0.1313)

B. (0.0161, 0.2207)

C. (0.0832, 0.1536)

D. None of the above

1. In a random sample of 228 self-employed individuals, 27 reported having their tax returns audited by the IRS in the past year.

Which of the following is a 90% confidence interval for the population proportion of self-employed individuals who have been audited by the IRS in the past year?

A. (0.1055, 0.1313)

B. (0.0161, 0.2207)

**C. (0.0832, 0.1536)**

D. None of the above

2. Suppose that a 95 %confidence interval for a population proportion *p* is to be calculated based on a sample of 80 individuals. The multiplier to use is

- 1.645
- 1.96
- 1.99
- 1.66

2. Suppose that a 95 %confidence interval for a population proportion *p* is to be calculated based on a sample of 250 individuals. The multiplier to use is

- 1.645
- **1.96**
- 1.99
- 1.66

**Reasoning: One proportion intervals use Z, not T, so A and B are out. 95% refers to the middle percent, not a left tail. If 95% is in the middle, then 2.5% is in both tails. So the amount to the left of our standard normal is 95%+2.5% = 97.5%**

3. The mean typing speed for a sample of 100 students with even month birthdays was 53.25 with a standard deviation of 18.3. The mean typing speed for a sample of 101 students with odd month birthdays is 51.79 with a standard deviation of 17.39. Compute the standard error for the difference in means for the two groups.

3. The mean typing speed for a sample of 100 students with even month birthdays was 53.25 with a standard deviation of 18.3. The mean typing speed for a sample of 101 students with odd month birthdays is 51.79 with a standard deviation of 17.39. Compute the standard error for the difference in means for the two groups.

**Standard error is the portion of the margin of error without the multiplier. This is unpooled two means which has an**

**SE= sqrt($s_1^2/n_1 + s_2^2/n_2$) = sqrt ($18.3^2/100 + 17.39^2/101$) =2.519**

# Interpretations

- Our book offers two interpretations of a confidence interval:
  - You can be CL% confident that the calculated interval includes or covers the true value of the parameter of interest.
  - If many, many random samples were performed, approximately CL% of them will include the PARAMETER OF INTEREST.
- Two keys to these interpretations:
  - Like hypotheses, must be about parameters (population values).
  - Nothing too strongly worded here and nothing about probability of the parameter being in the interval.

4. Do people lie about voting? A survey was conducted to determine the proportion of eligible voters who said that they voted in the recent presidential election. Based on the response of 1002 randomly selected adults, a 96% confidence interval was calculated to be (0.671, 0.728). Which of the following is the correct interpretation for this confidence interval?

A. We are 96% confident that the proportion of adults in the sample who say they voted is between 0.671 and 0.728.

B. We are 96% confident that the population proportion of adults who say they voted is between 0.671 and 0.728.

C. The probability is 0.96 that between 67.1% and 72.8% of all adults voted.

D. We know that between 671.% and 72.8% of all adults say they voted.

4. Do people lie about voting? A survey was conducted to determine the proportion of eligible voters who said that they voted in the recent presidential election. Based on the response of 1002 randomly selected adults, a 96% confidence interval was calculated to be (0.671, 0.728). Which of the following is the correct interpretation for this confidence interval?

A. We are 96% confident that the proportion of adults in the sample who say they voted is between 0.671 and 0.728.

**B. We are 96% confident that the population proportion of adults who say they voted is between 0.671 and 0.728.**

C. The probability is 0.96 that between 67.1% and 72.8% of all adults voted.

D. We know that between 671.% and 72.8% of all adults say they voted.

**Reasoning: Must be about a population value!**

# Conditions

- This is the same as the section for hypothesis testing. Always – good data collection (representative of population)
  - Proportions: Need np and n(1-p) greater than or equal to 10 for any samples. For CIs, we can also think of this has the number of successes and number of failures both at least 10.
  - Means: Need n greater than or equal to 30 for any samples OR for the populations to be approximately normally distributed

# Changes to Confidence Interval / Sample Size

- You can decrease the width (margin of error) of a confidence interval by:
  - Increasing sample size
  - Decreasing confidence level
    - Note: Neither will impact the middle of the CI.
- You can decrease the necessary sample size for inference by:
  - Increasing the acceptable amount of error.
  - Decreasing the confidence level.

# Sample Size

- Be prepared to calculate sample size.

- We only learned two sample size formulas. One for proportions and one for means: (q is (1-p))

$$n = \frac{pq(z^*)^2}{(ME)^2} \qquad n = \left(\frac{z^*\sigma}{ME}\right)^2$$

Again, this comes down to having the common $z^*$ values on your cheat sheet and deciding which situation we're in.

When you don't have a previous study or estimate for p, use .5. When you don't have a previous study or estimate for sigma, use range/4.