MAT1372, Classwork17, Fall2025

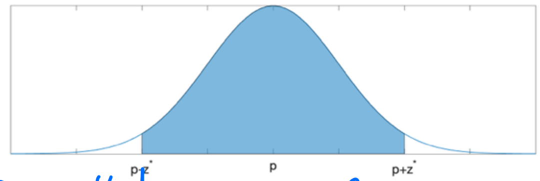
5.2 Confidence Intervals for a Proportion (Conti.)

5. Confidence Interval Using Any Confidence Level.

If a point estimate closely follows a normal distribution with standard error SE, then a confidence interval for the population parameter is

point estimate $\pm$ $z^* \times$ SE

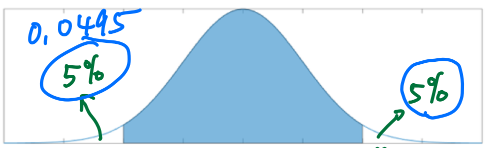
where $z^\star$ corresponds to the confidence level selected.



Margin of Error: In a confidence interval, $z^* \times SE$ is called margin of error

6. Find the confidence levels $z^\star$ for 90%, 95%, and 99% confidence intervals.

Find z^* for 90% confidence level



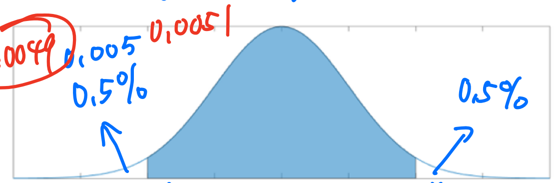
$-z^*$ $\quad$ z^*

$P(Z < -z^*) = 0.05$ $\quad$ $P(Z < z^*) = 95\%$

$-z^* = -1.65$ $\quad$ $z^* = 1.65$

90%: point estimate $\pm$ $1.65 \times SE$

Find z^* for 99% confidence level



$-z^*$ $\quad$ z^*

$P(Z < -z^*) = 0.005$

$-z^* = -2.58 \Rightarrow z^* = 2.58$

99% : point estimate $\pm 2.58 \times SE$.

7. Confidence Interval for a Single Proportion

Once you've determined a one-proportion confidence interval would be helpful for an application, there are four steps to constructing the interval:

90%, 95%, 99%

Prepare. Identify $\hat{p}$ (sample proportion) and n (sample size), and what confidence level you want.

Check. Verify the conditions to ensure $\hat{p}$ is nearly normal. For one-proportion confidence level, use $\hat{p}$ in place of p to check the success-failure condition (Central Limit Thm)

Calculate. If conditions hold, compute SE using $\hat{p}$, find z^* and construct interval

Conclude. Interpret the confidence interval in the context of the problem

8. In Section 5.1 we learned about 88.7% of a random sample of 1000 American adults supported expanding the role of solar power. Compute and interpret 90% and 99% confidence intervals for the population proportion

① $\hat{p} = 0.887$, $n = 1000$

② Check success-failure condition

$np = 1000 \cdot 0.887 = 887 \geq 10$

$n(1-p) = 113 \geq 10$

$\hat{p}$ follows normal distr.

③ Find SE

$SE_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} \approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.01$

④ Construct 90%

$\hat{p} \pm 1.65 \times SE$

$0.887 \pm 1.65 \times 0.01$

$(0.887 - 0.0165,\ 0.887 + 0.0165)$

$\Rightarrow (0.8705,\ 0.9034)$

99% CI

$\hat{p} \pm 2.58 \times SE$

$0.887 \pm 2.58 \times 0.01$

$(0.887 - 0.0258,\ 0.887 + 0.0258)$

$\Rightarrow (0.8612,\ 0.9128)$

Interval (smaller number, larger number)

5.3 Hypothesis Testing of a Proportion

0. Sample proportion and Sample mean.

What is a sample proportion $\hat{p}$? often sampling is done in order to estimate the proportion of population that has a specific characteristic

Example: The bar exam passing rate is p and the passing rate of sampling n people from takers $\hat{p} = \frac{\text{\# of the people who pass the exam from the sample}}{n}$

What is a sample mean $\bar{x}$? The average of a sample from the sampling of population

Example: The average height of 1000 American adults

1. *How many of the world's 1 year-old children have been vaccinated against disease: 20%, 50%, or 80%?*

2. Null and Alternative Hypotheses.

The null hypothesis (H_0) often represents a skeptical perspective or a perspective of "no difference" or a claim to be tested

The alternative hypothesis (H_A) represents an alternative claim under consideration and is often represented by a range of possible parameter values, it is usually stronger

Our job as data scientists is to play the role of a skeptic: before we buy into the H_A, we need to see strong supporting evidence.

3. We're interested in understanding how much people know about world health and development. If we take a multiple choice world health question, then we might like to understand if

H_0: People never learn these particular topic and their respones are simply equivalent to random guesses.

H_A: People have knowledge that helps them do better than random guessing, or perhaps, they have false knowledge that leads them to actually do worse than random guessing.

These competing ideas are called hypothesis.

4. If we're interested in the proportion that people pick the correct answer of 1. While it's helpful to put these hypotheses in words, it can be useful to write them using mathematical notation:

H_0: $p = 0.333$ (or $\frac{1}{3}$) (random guessing)

H_A: $p \neq 0.333$

In this hypothesis setup, we want to make a conclusion about the population proportion. The value we are comparing the parameter to is called the null value, which in this case is $p_0 = 0.333$.

5. It may seem impossible that the proportion of people who get the correct answer is exactly 33.3%. If we don't believe the null hypothesis, should we simply reject it?

NO. While we may not buy into the notion that $\hat{p} = \frac{1}{3}$, the hypothesis testing framework requires a stronger evidence before we reject the H_0