Stat 201: Introduction to Statistics

Standard 32: Significance Tests – for Proportion Differences

Telling Which Parameter We’re After

- As statisticians, or data scientists, it’s our job to hear a problem and decide what we’re after
  - We call the parameter of interest the **target parameter**

Parameter	Point Estimate	Key Phrase	Type of Data
$\mu_1 - \mu_2$	$\overline{x_1} - \overline{x_2}$	Mean Difference	Quantitative
$p_1 - p_2$	$\widehat{p_1} - \widehat{p_2}$	Difference of Proportion, percentage, fraction, rate	Qualitative (Categorical)

Difference of Proportions

- In the frame of Chapter 9, **hypothesis tests**, we need to find complete the same five step process as before with different formulas to find the p-value and make decisions.

Hypothesis Test for Proportion Differences: Step 1

- State Hypotheses:
  - **Null hypothesis**: that the population proportion equals some p_o
    - $H_o\colon \rho_d = \rho_1 - \rho_2 \le p_0$ (one sided test)
    - $H_o\colon \rho_d = \rho_1 - \rho_2 \ge p_0$ (one sided test)
    - $H_o\colon \rho_d = \rho_1 - \rho_2 = p_0$ (two sided test)
  - **Alternative hypothesis:** What we're interested in
    - $H_a\colon \rho_d = \rho_1 - \rho_2 > p_o$ (one sided test)
    - $H_a\colon \rho_d = \rho_1 - \rho_2 < p_o$ (one sided test)
    - $Ha\colon \rho_d = \rho_1 - \rho_2 \ne p_o$ (two sided test)

Hypothesis Test for Proportion Differences: Step 2

- Check the assumptions
  - The variable must be categorical
  - The data are obtained using randomization
  - We want at least 5 of each category within each group.
    - Five that have the attribute and five that don't have the attribute for each group, at least

Hypothesis Test for Proportion Differences: Step 3

- **Calculate Test Statistic**
  - The test statistic measures how different the sample proportion we have is from the null hypothesis
  - We calculate the z-statistic by assuming that ρ_{d_0} is the population proportion difference

$$z^* = \frac{((\hat{p}_1 - \hat{p}_2) - p_{d_o})}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

Hypothesis Test for Proportion Differences: Step 4

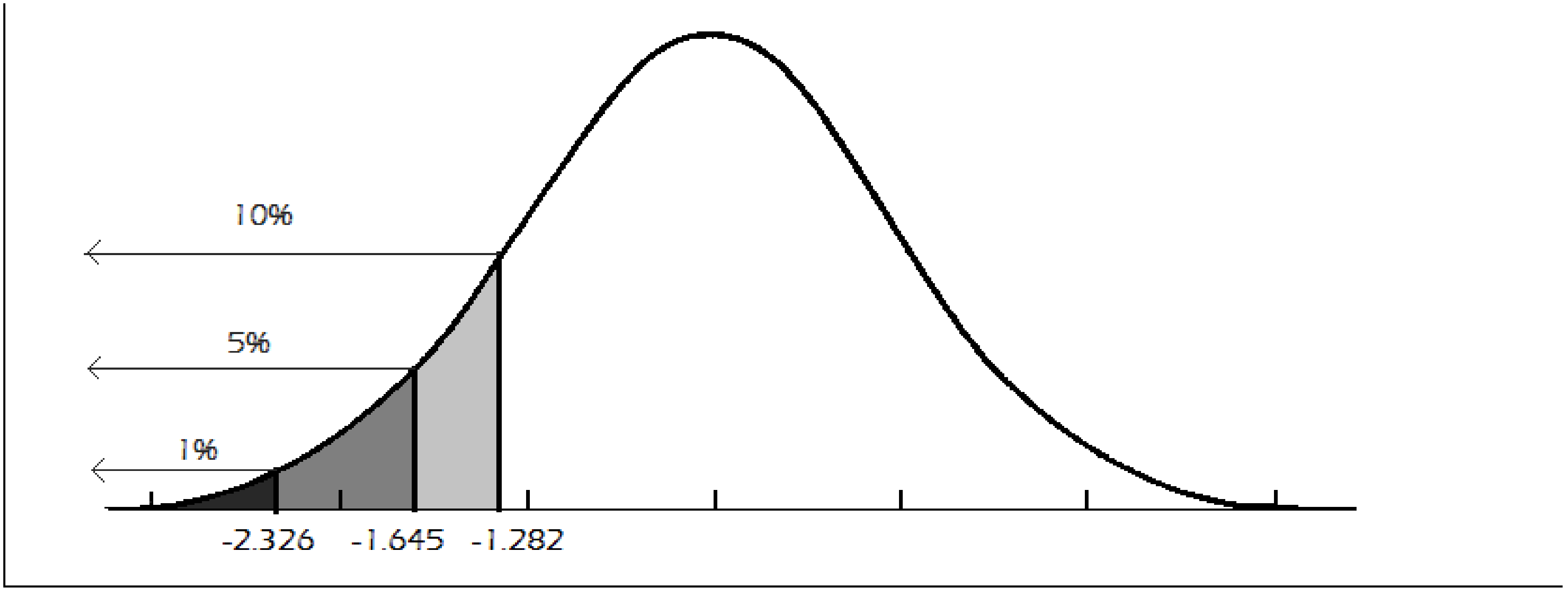
- Determine the P-value
  - The P-value describes how unusual the sample data would be if H_o were true.

<table>
<tr><th>Alternative Hypothesis</th><th>Probability</th><th>Formula for the P-value</th></tr>
<tr><td>$H_a: \rho_1 - \rho_2 > p_o$</td><td>Right tail</td><td>P(Z>z*) = 1-P(Z<z*)</td></tr>
<tr><td>$H_a: \rho_1 - \rho_2 < p_o$</td><td>Left tail</td><td>P(Z<z*)</td></tr>
<tr><td>$H_a: \rho_1 - \rho_2 \neq p_o$</td><td>Two-tail</td><td>2*P(Z<-|z*|)</td></tr>
</table>

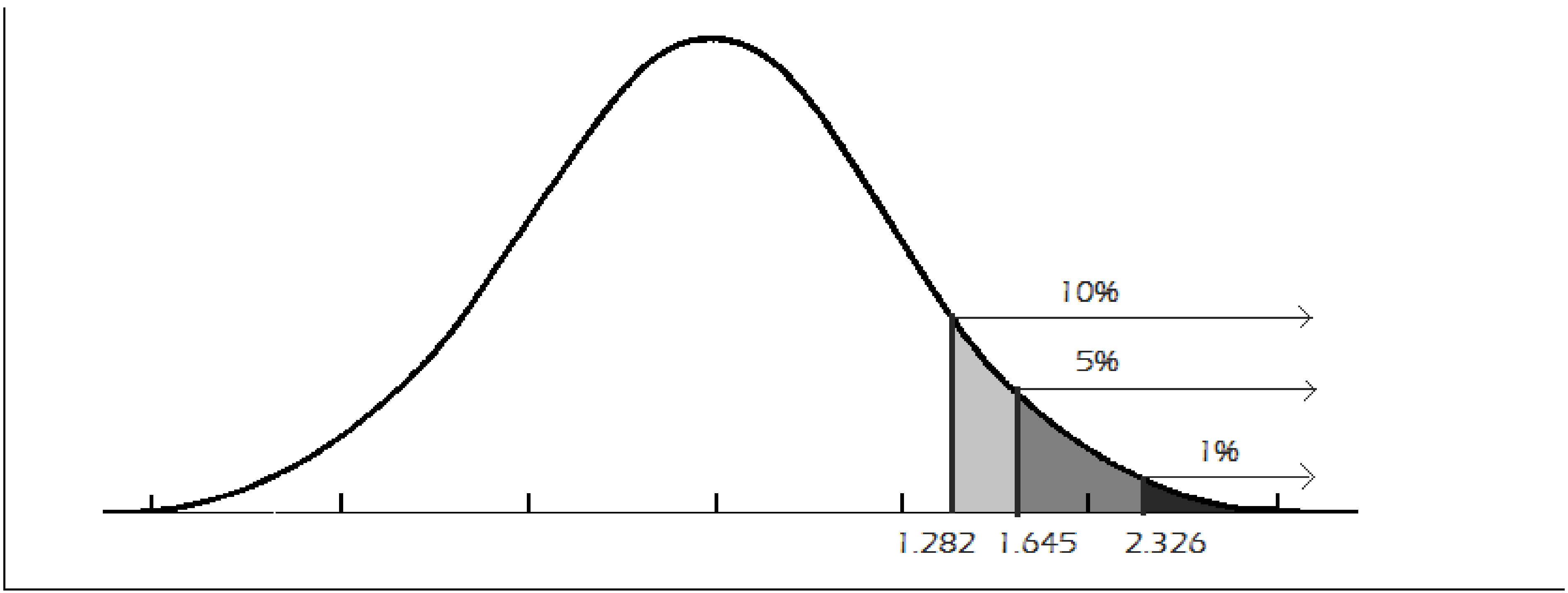
Hypothesis Test for Proportion Differences: Step 5

- Summarize the test by reporting and interpreting the P-value
  - Smaller p-values give stronger evidence against H_o
- If p-value$\leq (1 - confidence) = \alpha$
  - Reject H_o, with a p-value = ____, we have sufficient evidence that the alternative hypothesis might be true
- If p-value$> (1 - confidence) = \alpha$
  - Fail to reject H_o, with a p-value = ____, we do not have sufficient evidence that the alternative hypothesis might be true

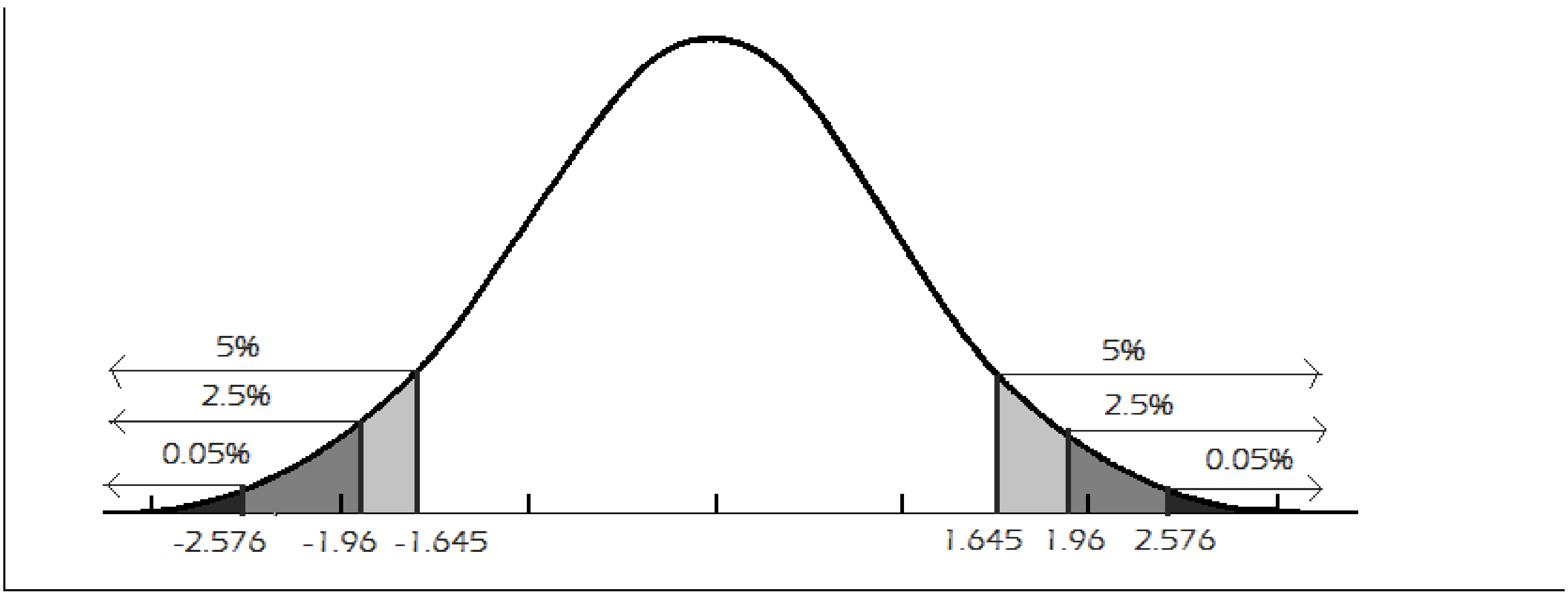
Zoom In



Zoom In



Zoom In



Example

- 6,450 transgender and gender non-conforming study participants were asked about whether or not they maintained their family bonds.
- 2773 maintained their family ties of which 887 had attempted suicide
- 3677 experienced rejection from their family of which 1,875 had attempted suicide

Example

- 2773 maintained their family ties of which 887 had attempted suicide

$$\widehat{p_1} = \frac{887}{2773} = .31987$$

- 3677 experienced rejection from their family of which 1,875 had attempted suicide

$$\widehat{p_2} = \frac{1875}{3677} = .50993$$

Example

- Test, with 99% confidence, that the population proportion of transgender and gender non-conforming people who are rejected by their family and those that maintained their family ties are not equally likely to attempt suicide

$$H_o: \rho_d = \rho_1 - \rho_2 = 0$$
$$Ha: \rho_d = \rho_1 - \rho_2 \neq 0$$

Example

- Check the assumptions
  - The variable is categorical
  - The data are obtained using randomization
  - We have at least 5 of each category within each group.

Example

- **Calculate Test Statistic**

$$z^* = \frac{((\hat{p}_1 - \hat{p}_2) - p_0)}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$$

$$= \frac{((.31987 - .50993) - 0)}{\sqrt{\frac{.31987(1-.31987)}{2773} + \frac{.50993(1-.50993)}{3677}}}$$

$$= \frac{((.31987 - .50993) - 0)}{\sqrt{\frac{.31987(1-.31987)}{2773} + \frac{.50993(1-.50993)}{3677}}} = -15.70704$$

Example

- Determine the P-value:

$$2P(Z < -|-15.70704|)$$

$$=2P(Z < -15.70704)$$

$$\approx 0$$

Example

- Summarize the test by reporting and interpreting the P-value:

If $0 \leq (1 - .99) = .01$

  – Reject H_o, we have sufficient evidence that the alternative hypothesis might be true - the population proportion of transgender and gender non-conforming people who are rejected by their family and those that maintained their family ties are not equally likely to attempt suicide

Summary!

Sampling Distribution for the Sample Proportion Summary

Shape of sample	Center of sample	Spread of sample
The shape of the distribution is bell shaped if $n * p \geq 15$ $\boldsymbol{and}$ $n * (1 - p) \geq 15$	$\mu_{\hat{p}_d} = p_1 - p_2$	$\sigma_{\hat{p}_d} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$

Hypothesis Testing

Step One:	*1.* $H_0: p_d = p_{d_0}$ & $H_a: p_d \neq p_{d_0}$ *1.* $H_0: p_d \geq p_{d_0}$& $H_a: p_d < p_{d_0}$ *2.* $H_0: p_d \leq p_{d_0}$& $H_a: p_d > p_{d_0}$
Step Two:	1. Categorical 2. Random 3. 5 in each group 4. $\mathrm{n}p_{d_0} \geq 15$ & $\mathrm{n}(1 - p_{d_0}) \geq 15$
Step Three:	$z^* = \dfrac{((\hat{p}_1 - \hat{p}_2) - p_{d_o})}{\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}}$
Step Four:	$H_a: p_d \neq p_{d_0}$→ p-value = 2*P(Z<-|z*|) $H_a: p_d < p_{d_0}$→ p-value = P(Z<z*) $H_a: p_d > p_{d_0}$→ p-value = P(Z>z*) = 1-P(Z<z*)
Step Five:	If p-value$\leq (1 - confidene) = \alpha$ → Reject H_0 If p-value$> (1 - confidence) = \alpha$ → Fail to Reject H_0