Stat 201: Introduction to Statistics

Standard 27: Significance Tests - Proportions

Confidence Intervals to Testing

- As we’ve we can come up with interesting observations from the confidence intervals we found earlier

- Next we will learn how to formally test whether or not the population proportion is a particular value based off our sample proportion

Vocabulary of Testing

- A **Hypothesis** is a proposition assumed as a premise in an argument, i.e. we assume it to be true. It's a statement regarding a characteristic of one or more populations.

- **Hypothesis testing** is a procedure based on evidence found in a sample to test hypothesis – to see if we have enough evidence to suggest the alternative.

Vocabulary of Testing

- The **null hypothesis** (H_0) is the hypothesis we conclude to be true unless we have data that is sufficient to suggest otherwise – think "innocent until proven guilty"

- The **alternative hypothesis** (H_a) is the hypothesis that we conclude to be true if we have data that is sufficient to suggest the null hypothesis is not true

Hypothesis

1. *Two-tailed test*
   - $H_0: parameter = some\ value$
   - $H_1: parameter \neq some\ value$
2. *Left-tailed test*
   - $H_0: parameter \geq some\ value$
   - $H_1: parameter < some\ value$
3. *Right-tailed test*
   - $H_0: parameter \leq some\ value$
   - $H_1: parameter > some\ value$

- **Your book always has the $H_0: parameter = some\ value$

Watch These!

- Intro with funny accent*:
  - https://www.youtube.com/watch?v=0zZYBALbZgg
  - The example is a little advanced for now but the explanation is VERY good!

- P-value:
  - https://www.youtube.com/watch?v=eyknGvncKLw

Hypothesis Test for Proportions: Step 1

- State Hypotheses to some value we're interested in, p_o- it's usually easier to start with H_a
  - **Null hypothesis**: that the population proportion equals some p_o
    - $H_o: p \leq p_o$ (one sided test)
    - $H_o: p \geq p_o$ (one sided test)
    - $H_o: p = p_o$ (two sided test)
  - **Alternative hypothesis:** What we're interested in
    - $H_a: p > p_o$ (one sided test)
    - $H_a: p < p_o$ (one sided test)
    - $Ha: p \neq p_o$ (two sided test)

Hypothesis Test for Proportions: Step 2

- **Check the assumptions:**
  1. The variable must be categorical
  2. The data should be obtained using randomization
  3. The sample size is sufficiently large where p_o is the testing value (note we use $\rho = \mathrm{p}_0$)
     - $np_o \geq 15$
     - $n(1 - p_o) \geq 15$
     - It is safe to assume the distribution of p_o has a bell shaped distribution if both are ≥ 15

Hypothesis Test for Proportions: Step 3

- **Calculate Test Statistic, z***
  - The test statistic measures how different the sample proportion we have is from the null hypothesis
  - We calculate the z-score by assuming that p_o is the population proportion (we use $\rho = \mathrm{p}_0$)

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

Hypothesis Test for Proportions: Step 4

- **Determine the P-value**
  - The P-value describes how unusual the sample data would be if H_o were true, which is what we're assuming ($\rho = \mathrm{p}_0$).
  - z* is the test statistic from step 3

Alternative Hypothesis	Probability	Formula for the P-value
$H_a: \rho > p_o$	Right tail	P(Z>z*)=1-P(Z<z*)
$H_a: \rho < p_o$	Left tail	P(Z<z*)
$H_a: \rho \neq p_o$	Two-tail	2*P(Z<-|z*|)

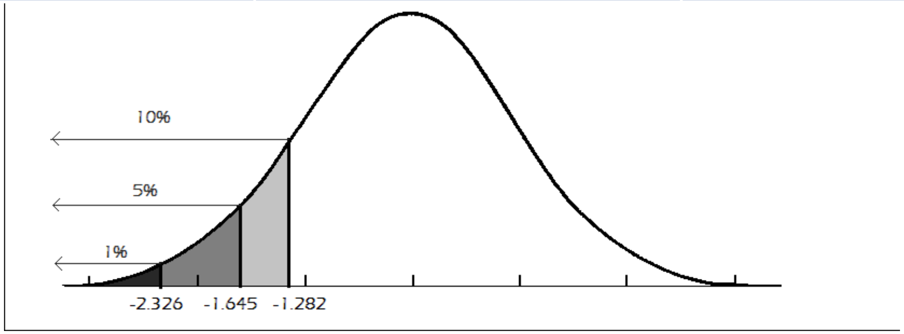
Hypothesis Test for Proportions: Step 5

- 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

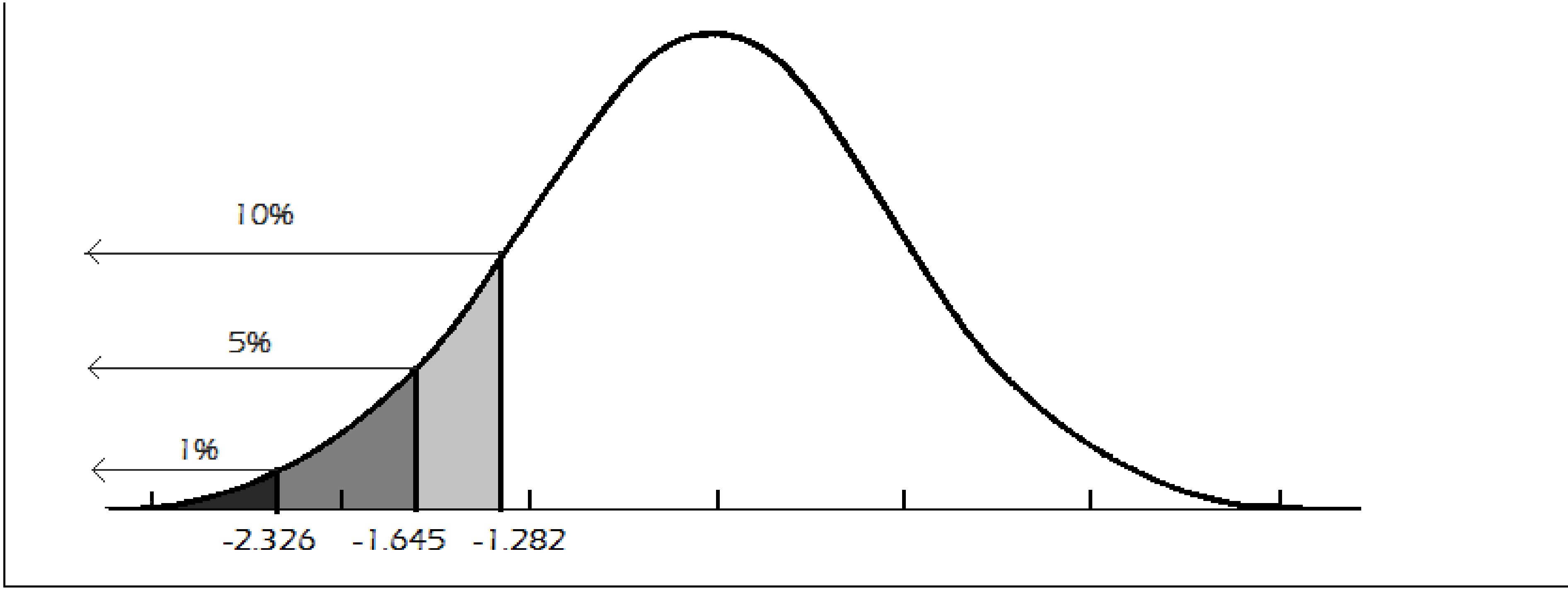
Hypothesis Test for Proportions: Step 5 with Pictures

- For a left tailed test: $H_a: \rho < p_0$ → We have rejection regions for H_o are as follows
  - Note: all of the rejection region is in the left tail, where $\hat{p}$ is much smaller than p_0

Confidence	Reject (test stat)	Reject (p-value)
0.90	Test-stat<-1.282	P-value<.1
0.95	Test-stat<-1.645	P-value<.05
0.99	Test-stat<-2.326	P-value<.01



Zoom In

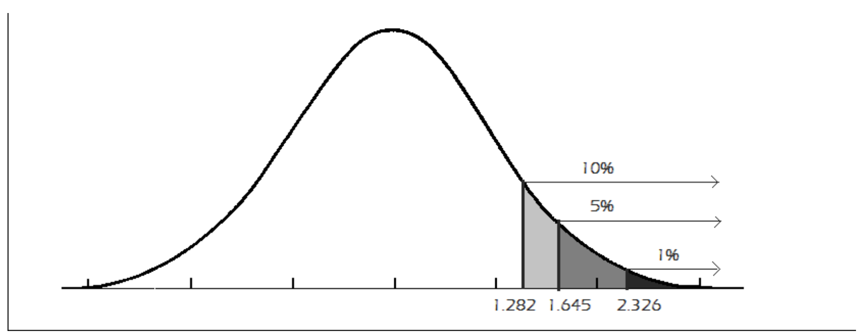


Confidence	Reject (test stat)	Reject (p-value)
0.90	Test-stat<-1.282	P-value<.1
0.95	Test-stat<-1.645	P-value<.05
0.99	Test-stat<-2.326	P-value<.01

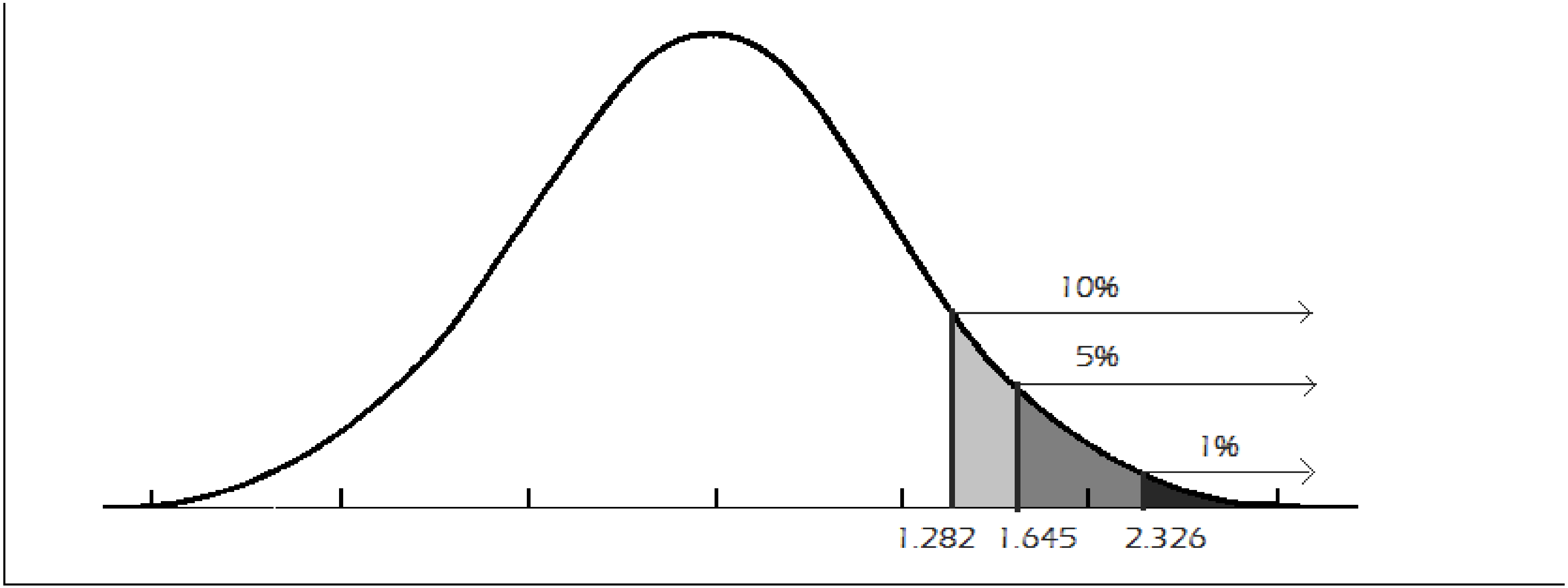
Hypothesis Test for Proportions: Step 5 with Pictures

- For a right tailed test: $H_a: \rho > p_0$ → We have rejection regions for H_o are as follows
  - Note: all of the rejection region is in the right tail, where $\hat{p}$ is much larger than p_0

Confidence	Reject (test stat)	Reject (p-value)
0.90	Test-stat>1.282	P-value<.1
0.95	Test-stat>1.645	P-value<.05
0.99	Test-stat>2.326	P-value<.01



Zoom In

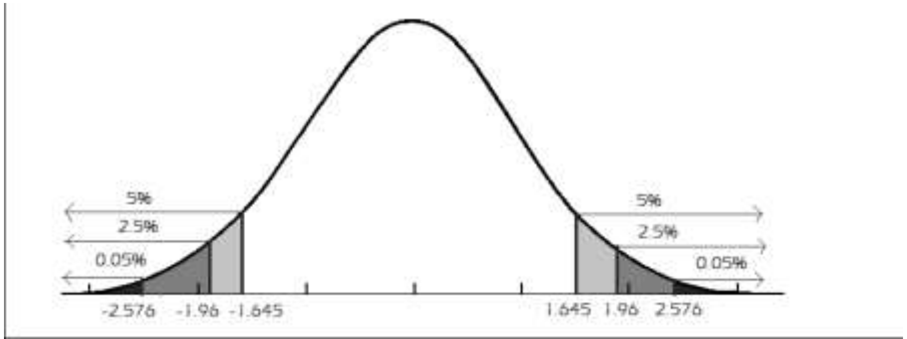


Confidence	Reject (test stat)	Reject (p-value)
0.90	Test-stat>1.282	P-value<.1
0.95	Test-stat>1.645	P-value<.05
0.99	Test-stat>2.326	P-value<.01

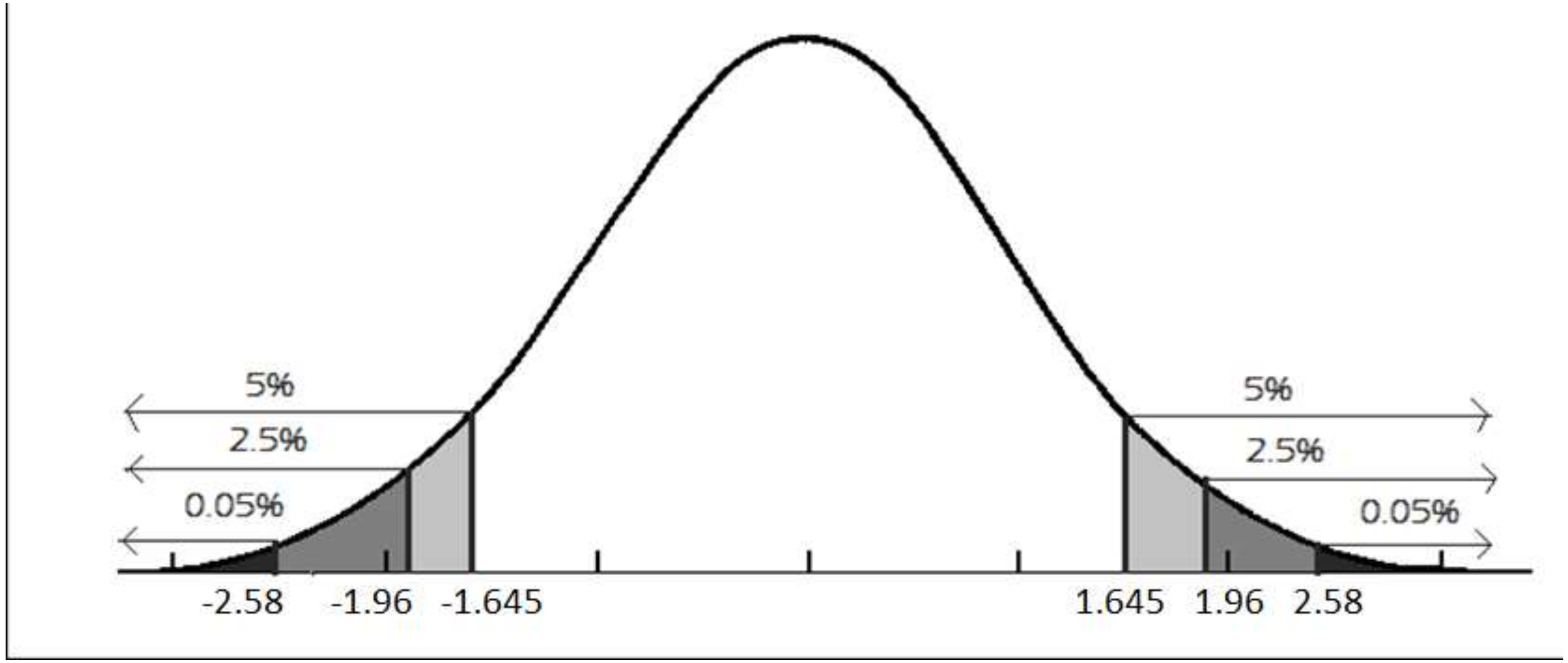
Hypothesis Test for Proportions: Step 5 with Pictures

- For a two tailed test: $H_a: \rho \neq p_0$ → We have rejection regions for H_o are as follows
  - Note: here we split the rejection region into both tails, where $\hat{p}$ is very different from p_0

Confidence	Reject (test stat)	Reject (p-value)
0.90	|Test-stat|<1.645	P-value<.1
0.95	|Test-stat|<1.960	P-value<.05
0.99	|Test-stat|<2.576	P-value<.01



Zoom In



Hypothesis Test for Proportions: Step 5 with Pictures

- The idea is – if our z^* is in the rejection region, our sample $\hat{p}$ is too unusual for the null hypothesis to be true so the data shows sufficient evidence against the null suggesting the alternative might be true.

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- At the .01 level of significance (99% confidence) is there evidence that there is a home field advantage?
- $\hat{p} = \frac{1335}{2429} = .5496$

Example – Step One

- State the Hypotheses: we are interested in whether or not there was a home field advantage, whether or not the population proportion of home games won by the home team is **greater than .50**
  - $H_o: \rho \leq .5$
  - $H_a: \rho > .5$

Example – Step Two

- Check Assumptions
  - The variable is categorical
    - Either the home team won or they didn't
  - The data was collected randomly
  - $np_o = 2429(.5) = 1214.5 \geq 15$
  - $n(1 - p_o) = 2429(.5) = 1214.5 \geq 15$
  - So, it is safe to assume the distribution of p_o has a bell shaped distribution

Example – Step Three

- Calculate the test statistic:

$$z^* = \frac{(\hat{p} - p_o)}{\sqrt{\frac{p_o(1 - p_o)}{n}}} = \frac{(.5496 - .5)}{\sqrt{\frac{.5(1 - .5)}{2429}}} = 4.89$$

Example – Step Four

- Determine P-value
  - From the table $pvalue = 1 - P(Z < z*)$

$$
\begin{aligned}
pvalue &= 1 - P(Z < 4.89) \\
&= 1 - \text{pnorm}(4.89,0,1) \\
&= .000000504179
\end{aligned}
$$

Z-table:

$$pvalue = 1 - P(Z < 4.89) \approx 1 - 1 = 0$$

Example – Step Five

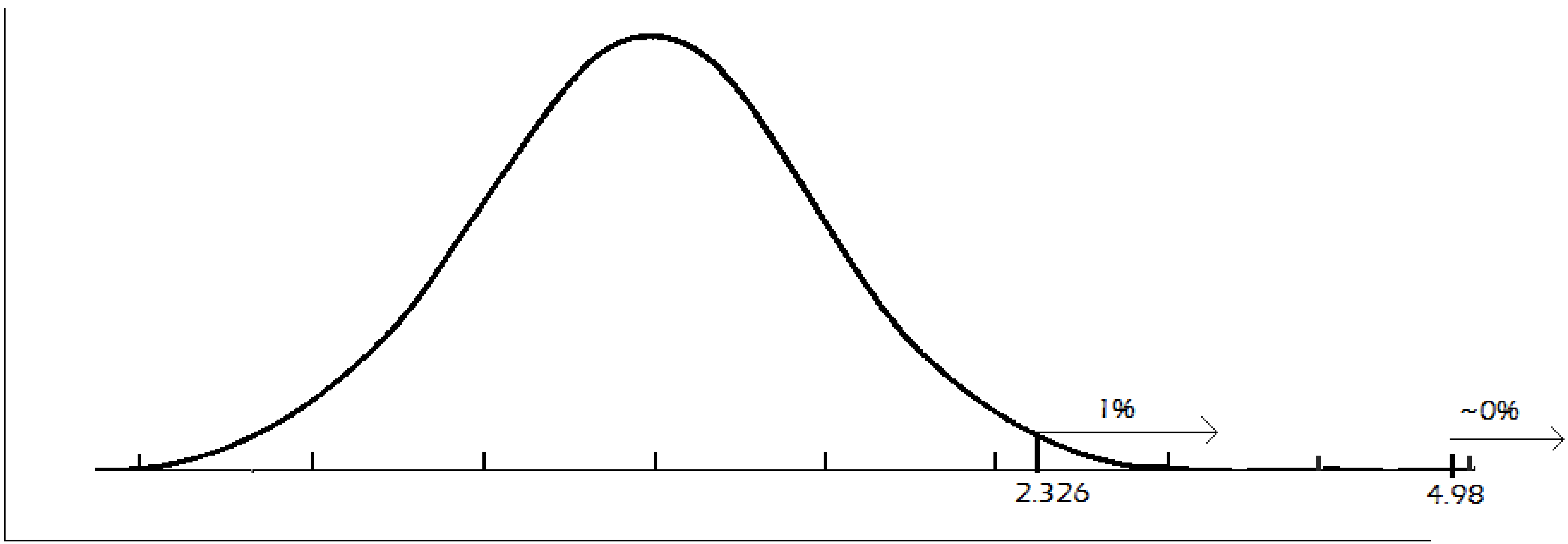
- State Conclusion
  - Since $.0000005041799 < .01$ we reject H_o

  At the .01 level of significance, or 99% confidence level, there is sufficient evidence to suggest that there is a home field advantage (the alternative)

Example – Step Five

- State Conclusion: We reject H_o for any of the following reasons
  - By P-value:
    - $.0000005041799 < .01$
  - By Z-statistic:
    - $|4.89| > 2.575829$
  - By $\hat{p}$:
    - $.5496 > x = z\sigma_{\hat{p}} + \mu_{\hat{p}} = 2.575829\sqrt{\frac{.5(1-.5)}{2429}} + .5 = .526132$

Example – Step Five



Hypothesis Testing for Proportions on your TI Calculator

- Hypothesis testing for proportions
  - https://www.youtube.com/watch?v=Y5wK1zHQ_OI

Hypothesis Testing for Proportions on your TI Calculator

- **INPUT:**
  1. Press STAT
  2. Press → to TESTS
  3. Scroll down using ↓ to highlight '5: 1-PropZTest'
  4. Press ENTER
  5. Enter the we're interested in next to 'p_0:'
  6. Enter the number of the total that had the behavior we're looking for next to 'x:'
  7. Enter the total number observations next to 'n:'
  8. Select the appropriate alternative hypothesis on the 'prop' line by highlighting the correct inequality and pressing ENTER
  9. Highlight 'Calculate'
  10. Press ENTER

Hypothesis Testing for Proportions on your TI Calculator

- **OUTPUT:**
  - z = our test statistics
  - p = our p-value for the test
    - We make our decision based on this
  - $\hat{p}$ = the sample proportion for the problem
  - n = the sample size and should match the number you entered in step 6 above

Confidence Intervals for Proportions

- **StatCrunch Commands w/ data**
  - Stat→Proportion Stats→One Sample →with data (if you have the a list of data)→Choose the column→type the success value into the success box→ choose hypothesis→ enter the correct hypothesis→ Compute

- **StatCrunch Commands w/ summaries**
  - Stat→Proportion Stats→One Sample →with summary (if you have the count) → enter the number of success and total observations→ enter the correct hypothesis→ Compute

Confidence Intervals

Assumptions	Point Estimate	Margin of Error	Margin of Error
1. Random Sample *2.* $n\hat{p} \geq 15$ And $n(1-\hat{p}) \geq 15$	$\hat{p}$	$z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

- We are --% confident that the true population proportion lays on the confidence interval.

Hypothesis Testing

Step One:	1. $H_0: p = p_0$ & $H_a: p \neq \mathrm{p}_0$ 1. $H_0: p \geq p_0$ & $H_a: p < \mathrm{p}_0$ 2. $H_0: p \leq p_0$ & $H_a: p > \mathrm{p}_0$
Step Two:	1. Categorical 2. Random 3. $\mathrm{n}p_o \geq 15$ & $\mathrm{n}(1 - p_o) \geq 15$
Step Three:	$z^* = \dfrac{(\hat{p} - p_o)}{\sqrt{\dfrac{p_o(1 - p_o)}{n}}}$
Step Four:	$H_a: p \neq \mathrm{p}_0$→ p-value = 2*P(Z<-|z*|) $H_a: p < \mathrm{p}_0$→ p-value = P(Z<z*) $H_a: p > \mathrm{p}_0$→ 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