STAT 201 Chapter 8.1-8.2

Confidence Intervals for Proportion

Recall Proportion Sampling Distribution

- The **mean** of the sampling distribution for a sample proportion will always equal the population proportion: $\boldsymbol{\mu_{\hat{p}} = p}$
- The **standard error**, the standard deviation of the sample proportion, is:

$$\boldsymbol{\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}}$$

Point Estimate

- Often, we do not know the population proportion.
- We use our sample proportions to make **inference** on the population parameter

- $\hat{p}$ is our **point estimate** for the population proportion
  - Our ‘best’ guess for the true population proportion

Example

- Suppose you are in a sports company. Tom is your direct boss, and Jerry is the President of the company.

- Jerry wants to know the winning rate for MLB home games. He asks Tom, and Tom asks you.

Example

- You do not know the truth (population population) and decide sample 2429 games. You find that home teams won 1335 of 2429 games.
- Now, you know sample proportion is $\hat{p} = \frac{1335}{2429} = .549$ and tell your boss "the winning rate for MLB home games is 54.9%"

Example

- Your boss, Tom, is pretty satisfied and tells this results to his boss, Jerry. However, Jerry is skeptical.
- He has some basic knowledge of statistics and he samples 5000 games by himself and find the MLB home game winning rate is 53%. From then on, Jerry teases Tom everyday and Tom decides to fire you.
- Any smarter way to answer Tom’s question?

Confidence Intervals

- We use our sample proportions to make inference on the population proportion in the sense of interval

$$\hat{p} \pm z_{\frac{\alpha}{2}}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$$

- $\hat{p}$ is our **point estimate** for the population proportion
- $z_{\frac{\alpha}{2}}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**

Confidence Intervals Bounds

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$$

Lower Bound = $\hat{p} - z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$

Upper Bound = $\hat{p} + z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$

Confidence Intervals Assumption

- **Assumption:**
  - Data must be obtained through randomization
  - We **MUST** make sure that $n\hat{p} \geq 15$ and $n(1-\hat{p}) \geq 15$. This ensures that $\hat{p}$ follows a bell shaped distribution

Confidence Intervals – Common Z's

- We choose z based on our desired confidence level
  - **Level of confidence** = (1-α) * 100%
  - **Level of significance = Error Probability** = α= 1- Level of confidence

Level of Confidence	Error Probability (α)	Z
.9 (90%)	.1 (10%)	1.645
.95 (95%)	.05 (5%)	1.96
.99 (99%)	.01 (1%)	2.58

- Our interval will get larger when the margin of error increases
  1) When we increase confidence → increase z
  2) When we decrease n

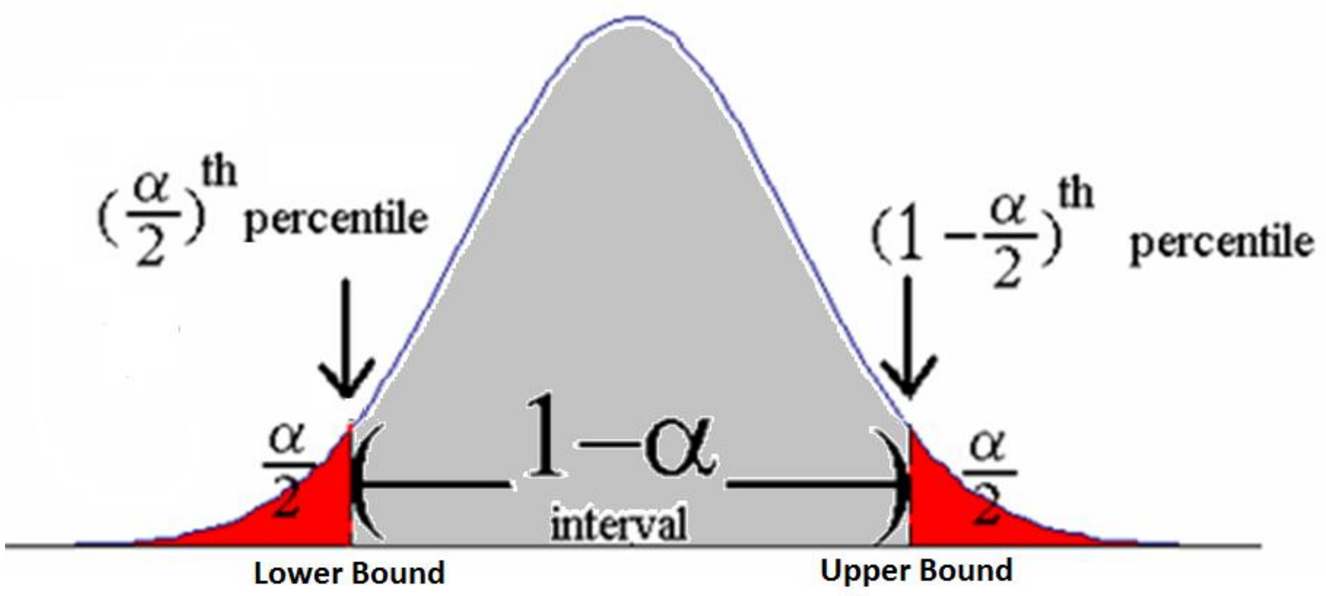
Confidence Intervals: Margin of Error

- $Z_{\frac{\alpha}{2}}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**
  - **As n increases**, the margin of error decreases causing the width of the confidence interval to narrow
  - **As n decreases**, the margin of error increases causing the width of the confidence interval to grow wider

Confidence Intervals: Margin of Error

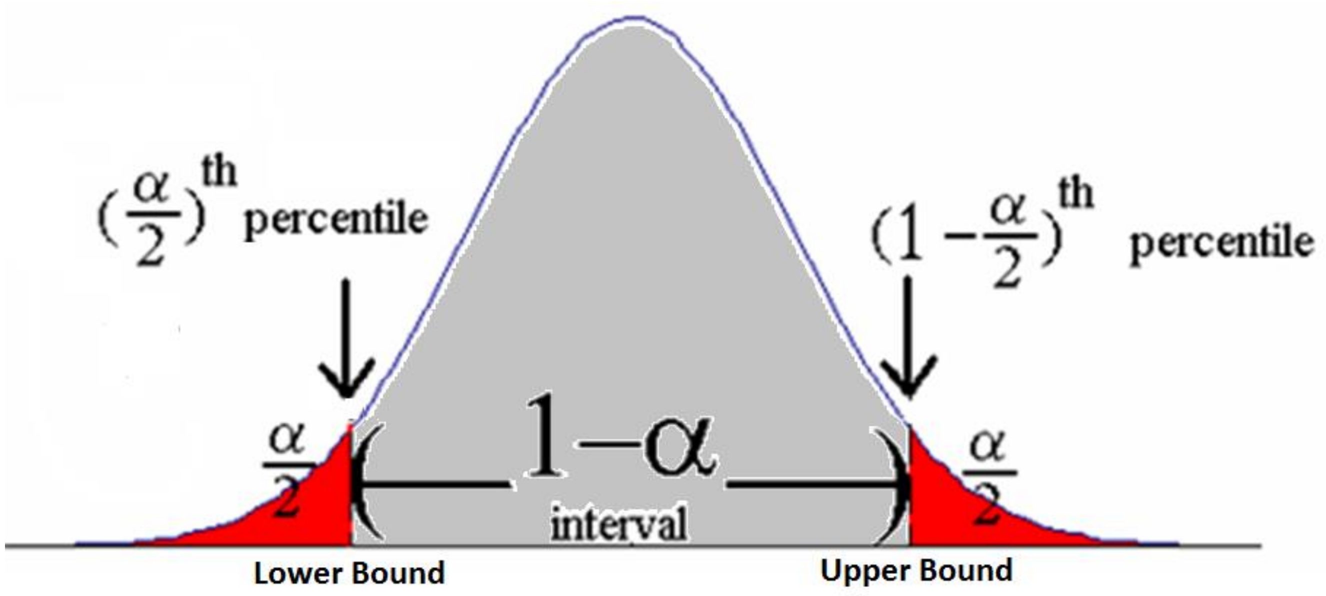
- $z_{\frac{\alpha}{2}}\sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$ is our **margin of error**
    - **As the confidence level decreases**, z decreases causing the margin of error to decrease, causing the width of the confidence interval to narrow
    - **As the confidence level increases**, z increases causing the margin of error to increase, causing the width of the confidence interval to grow wider

Confidence Intervals



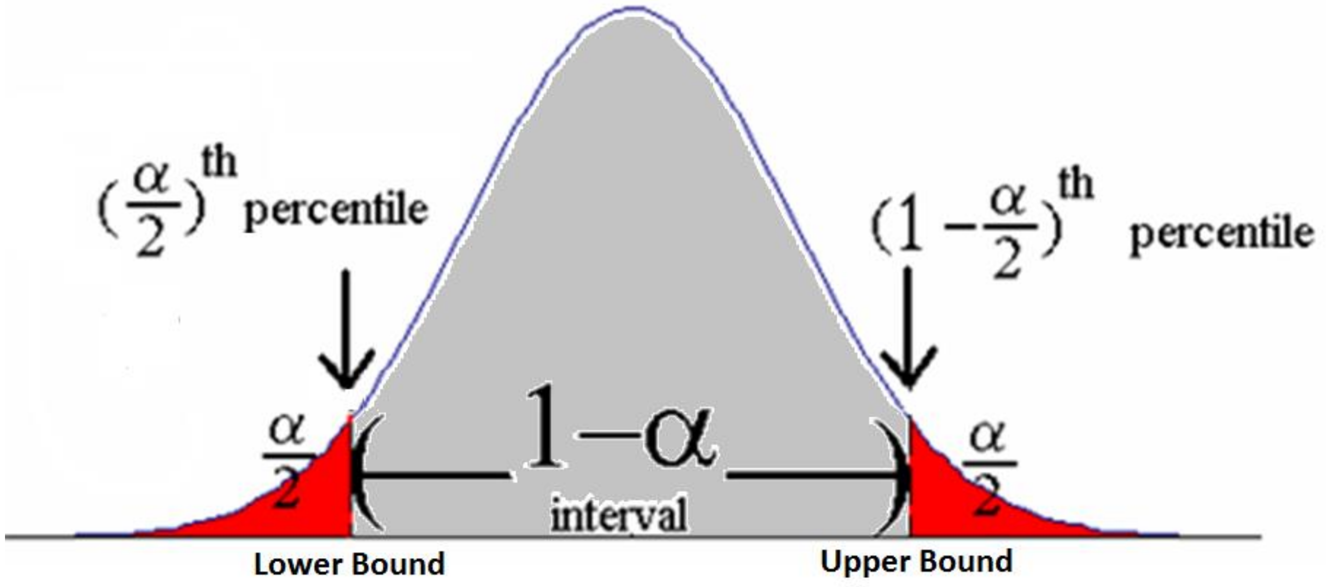
- We choose our values such that
  - Our point estimate is the mean, the 50th percentile
  - Our lower bound is the $\frac{\alpha}{2}^{\text{th}}$ percentile
  - Our upper bound is the 1-$\frac{\alpha}{2}^{\text{th}}$ percentile

How We Found the Common Z's: 90%



- For a 90% confidence interval lower bound, we need to find the z with a percentile of
$$\frac{\alpha}{2} = \frac{1 - confidence}{2} = \frac{1 - .90}{2} = \frac{.10}{2} = .0500$$
- If we look this up in the z-table we see that a z-score of -1.645 gives us a value very close to .0500

How We Found the Common Z's: 90%



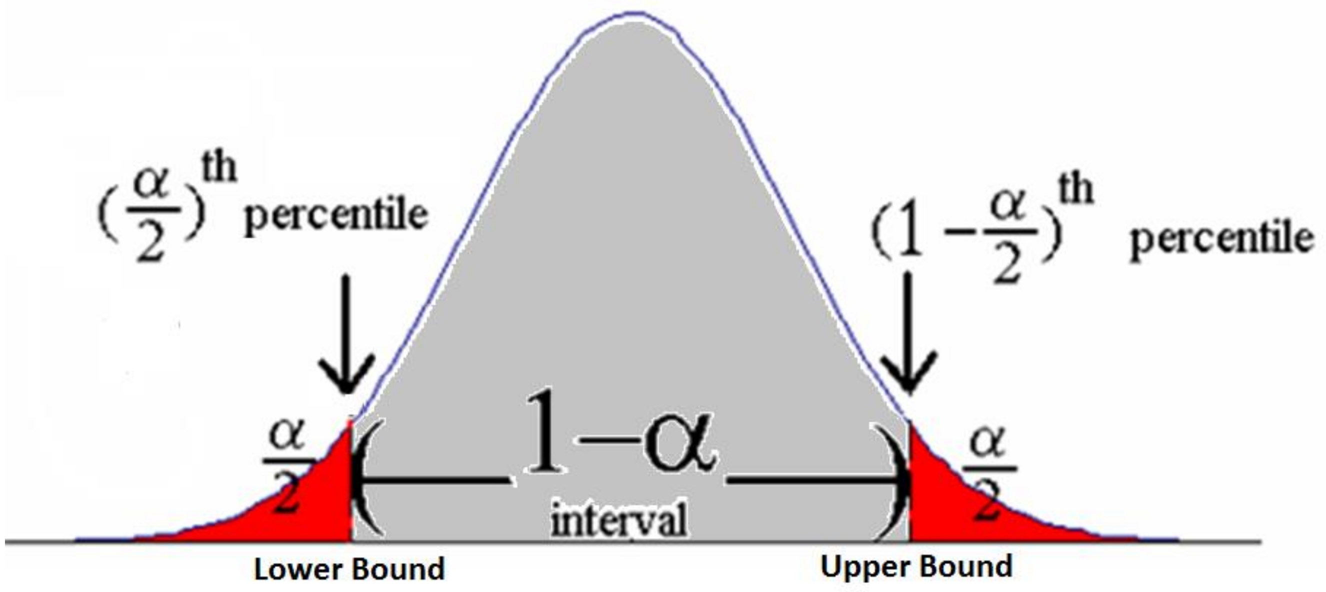
- For a 90% confidence interval upper bound, we need to find the z with a percentile of
$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.90}{2}=1-\frac{.10}{2}=.9500$$
- If we look this up in the z-table we see that a z-score of 1.645 gives us a value very close to .9500

How We Found the Common Z's: 90%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -1.645 gives us a value very close to .0500
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 1.645 gives us a value very close to .9500

- This is why we have plus or minus z=1.645 for a 90% confidence interval

How We Found the Common Z's: 95%

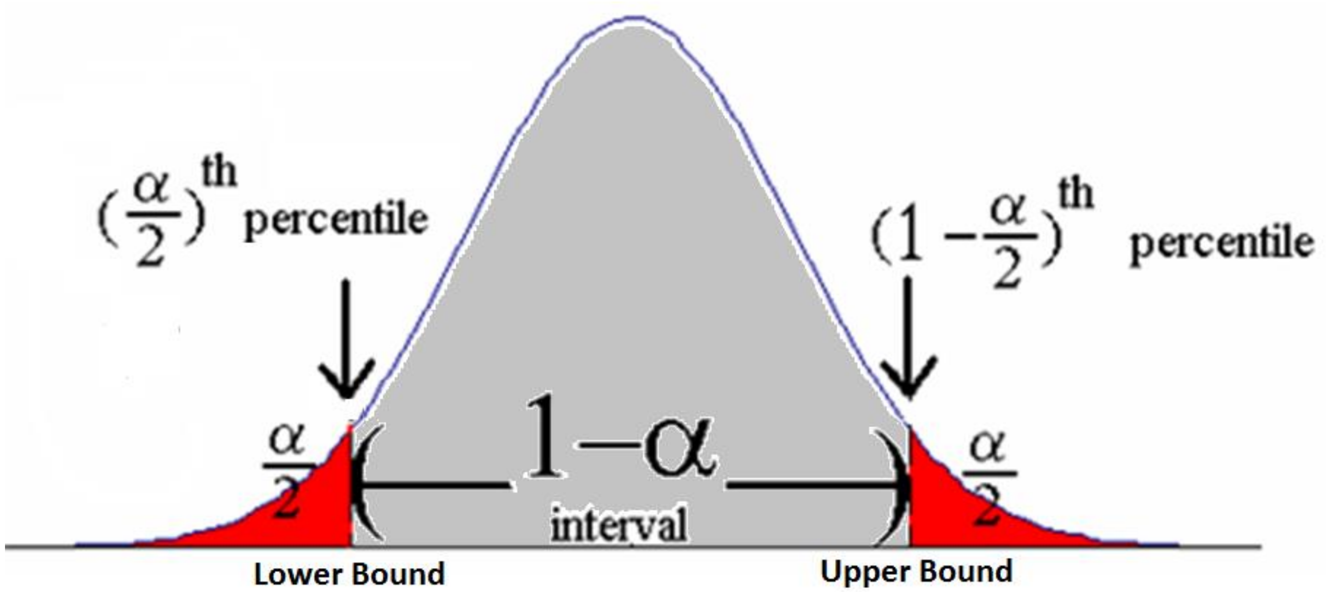


- For a 95% confidence interval lower bound, we need to find the z with a percentile of

$$\frac{\alpha}{2} = \frac{1 - confidence}{2} = \frac{1 - .95}{2} = .00250$$

- If we look this up in the z-table we see that a z-score of -1.96 gives us a value very close to .0250

How We Found the Common Z's: 95%



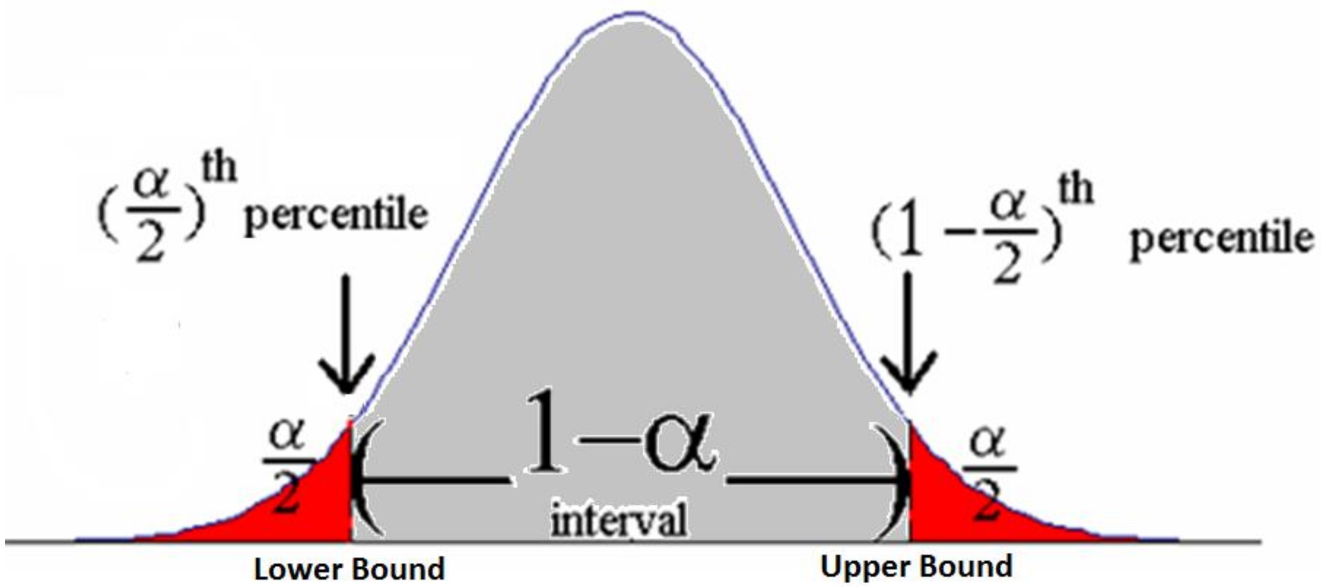
- For a 95% confidence interval upper bound, we need to find the z with a percentile of $1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.95}{2}=1-\frac{.05}{2}=.9750$
- If we look this up in the z-table we see that a z-score of 1.96 gives us a value very close to .9750

How We Found the Common Z’s: 95%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -1.96 gives us a value very close to .0250
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 1.96 gives us a value very close to .9750

- This is why we have plus or minus z=1.96 for a 95% confidence interval

How We Found the Common Z's: 99%

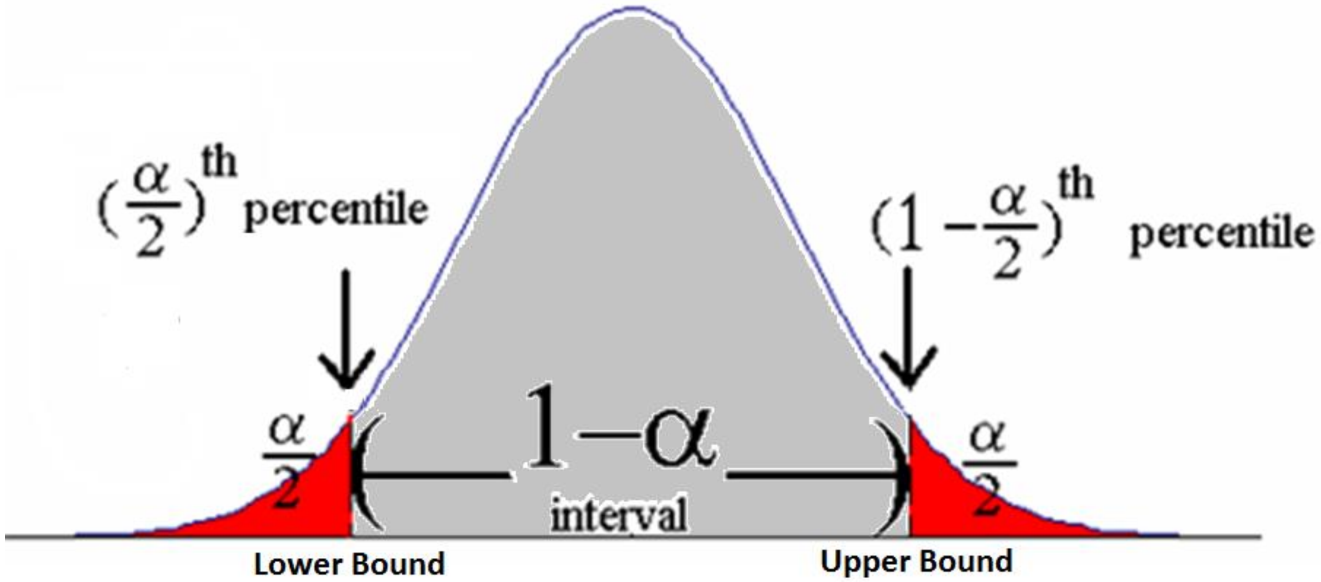


- For a 99% confidence interval lower bound, we need to find the z with a percentile of

$$\frac{\alpha}{2} = \frac{1 - confidence}{2} = \frac{1 - .99}{2} = .0050$$

- If we look this up in the z-table we see that a z-score of -2.58 gives us a value very close to .0050

How We Found the Common Z's: 99%



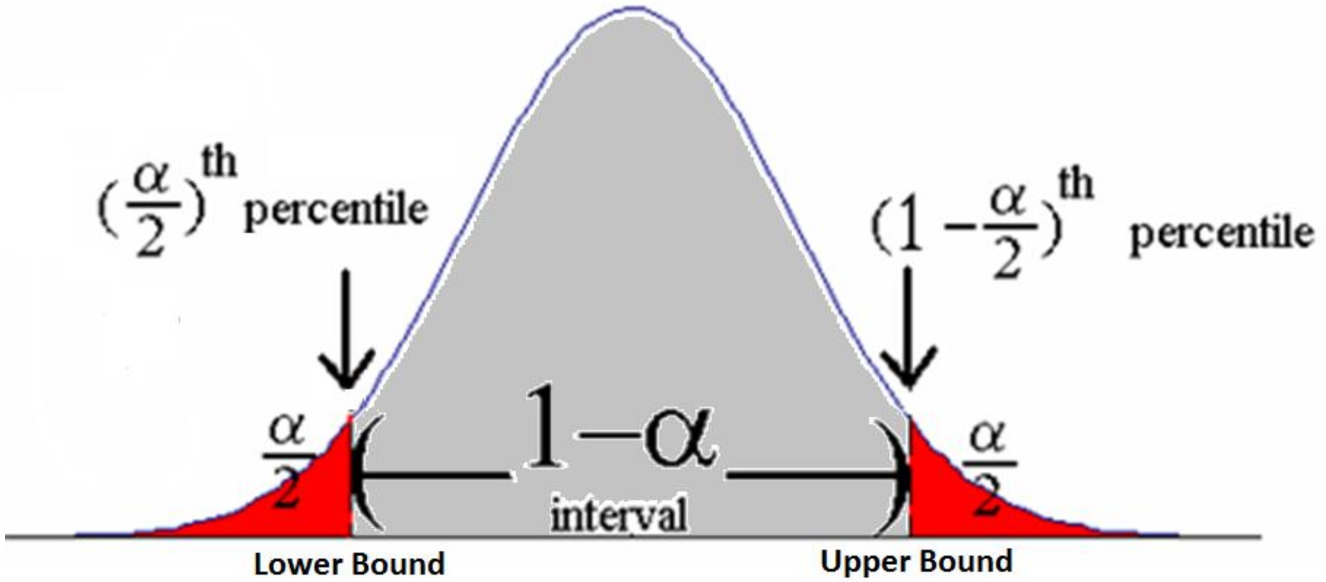
- For a 99% confidence interval upper bound, we need to find the z with a percentile of
$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.99}{2}=1-\frac{.01}{2}=.9950$$
- If we look this up in the z-table we see that a z-score of 2.58 gives us a value very close to .9950

How We Found the Common Z’s: 99%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -2.58 gives us a value very close to .0050
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 2.58 gives us a value very close to .9950

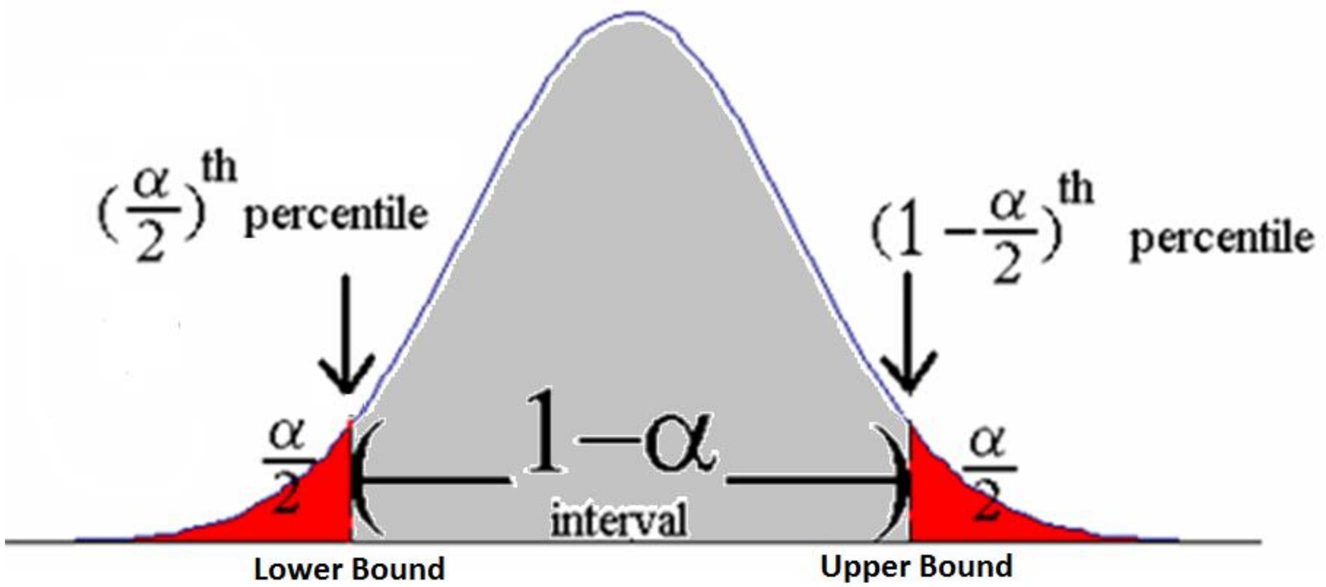
- This is why we have plus or minus z=2.58 for a 99% confidence interval

How We Find an Uncommon Z: 98%



- For a 98% confidence interval lower bound, we need to find the z with a percentile of $\frac{\alpha}{2} = \frac{1 - confidence}{2} = \frac{1-.98}{2} = \frac{.02}{2} = .0100 \rightarrow 1\%$
- If we look this up in the z-table we see that a z-score of -2.33 gives us a value very close to .0100

How We Found the Common Z's: 98%



- For a 98% confidence interval upper bound, we need to find the z with a percentile of
$$1-\frac{\alpha}{2}=1-\frac{1-confidence}{2}=1-\frac{1-.98}{2}=1-\frac{.02}{2}=.9900$$
- If we look this up in the z-table we see that a z-score of 2.33 gives us a value very close to .9900

How We Found the Common Z’s: 98%

- **Lower Bound:** If we look this up in the z-table we see that a z-score of -2.33 gives us a value very close to .0100
- **Upper Bound:** If we look this up in the z-table we see that a z-score of 2.33 gives us a value very close to .9900

- This is why we have plus or minus z=2.33 for a 98% confidence interval

Example

- Suppose we don't know the winning rate, say p, of MLB home games and want to make an inference (estimation).
- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- Our sample proportion $\hat{p} = \frac{1335}{2429} = .549$

Example

- A random sample of MLB home games showed that the home teams won 1335 of 2429 games.
- Our sample proportion $\hat{p} = \frac{1335}{2429} = .549$
- Check Assumptions:
  - $n * \hat{p} = 2429 * .549 = 1333.521 \geq 15$
  - $n * (1 - \hat{p}) = 2429 * .451 = 1095.479 \geq 15$
  - It is safe to assume the distribution of $\hat{p}$ has a bell shaped distribution
  - The data is from a random sample

Example

- 95% CI:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$$

$$.549 \pm (1.96)\sqrt{\frac{.549(.451)}{2429}} = (.529, .569)$$

- We are 95% confident that the true population proportion of home team wins is between 52.9 and 56.9 percent.
- We see here that there is a small home field advantage because all of the values in our 95% CI are above 0.5.

Example

- 99% CI:

$$\hat{p} \pm z_{\frac{\alpha}{2}} \sqrt{\frac{\left(\hat{p}(1-\hat{p})\right)}{n}}$$

$$.549 \pm (2.58)\sqrt{\frac{.549(.451)}{2429}} = (.523, .575)$$

- We are 99% confident that the true population proportion of home team wins is between 52.3 and 57.5 percent.
- Still, we see here that there is a small home field advantage but we note the interval is larger