z-test for proportions

Objective

Hypothesis test for population proportion using the Z statistic.

Homework

- Discussion Question p433
- Ex 8-5 p434 5, 7, 8, 11, 12, 15, 19, 20

Z-test for Proportion

- Testing a hypothesis about population proportions

- Compares the proportion of a sample, $\hat{p}$, with the proportion of a **population**, **p**.

Z-test for Proportion

- A test of proportion is a binomial test, success or failure.

- Thus μ = **np**. **n** is the number of trials, **p** is the population proportion, or expected value, and $\sigma = \sqrt{npq}$

- When **np** and **nq** are ≥ 10 (or 5) the binomial distribution approaches normal, thus we can use the normal approximation for testing a hypothesis about the proportions.

Z-test for Proportion

$$\text{Test Statistic} = \frac{\text{Observed Value (proportion) - Expected Value}}{\text{standard error}}$$

$$Z = \frac{\text{Sample proportion - Population proportion}}{\text{adjusted standard deviation}}$$

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}, n = \text{sample size}$$

Steps to Hypothesis Testing

1. Formulate all **hypotheses**: H_0, H_a
2. Determine the **test statistic** and the **critical value of the test statistic** (based on α). (Currently t)

   2a. Draw, label, and appropriately shade the curve representing H_0
3. Find value(s) of the **test statistic** based on **the data** and the **p-value** for that statistic.
4. Make a **decision** to Reject or Fail to Reject H_0.
5. Write your **conclusion** in a complete sentence.

Calculator

- To find z when comparing proportions the calculator has a hypothesis test for proportions.

STAT ➢ TESTS ⋎ 5:1-PropZTest

p_0: population proportion

x: observed frequency ← Must be whole number

n: number of trials

This is H_a → prop: $\neq p_0$ $<p_0$ $>p_0$

Calculate

- Sometimes the problem gives you the count (number of successes) but asks for proportion. Sometimes the problem gives you the proportion. **Read the problems carefully!**

Example the First

- A female college professor feels that girls are out-performing the college average in her math class. The college has 52% passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

$p = .52, \hat{p} = 16/27 = .5926, n = 27, \alpha = .01$

1. Hypotheses

$\alpha = .01$

$H_0: \rho \leq .52$ and $H_a: \rho > .52$

- A female college professor feels that girls are out-performing the college average in her math class. The college has 52% passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

$p = .52$, $\hat{p} = 16/27 = .5926$, $n = 27$, $\alpha = .01$

TI-84

STAT ➢ TESTS ⋎ 5:1-PropZTest

p_0: .52
x: 16
n: 27
prop: $\neq p_0$ $<p_0$ $>p_0$
Calculate

prop>.52
z=.7550086004
p=.2251218385
p^= .5925925926
n: 27

- A female college professor feels that girls are out-performing the college average in her math class. The college has **52%** passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

$p = .52$, $\hat{p} = 16/27 = .5926$, $n = 27$, $\alpha = .01$

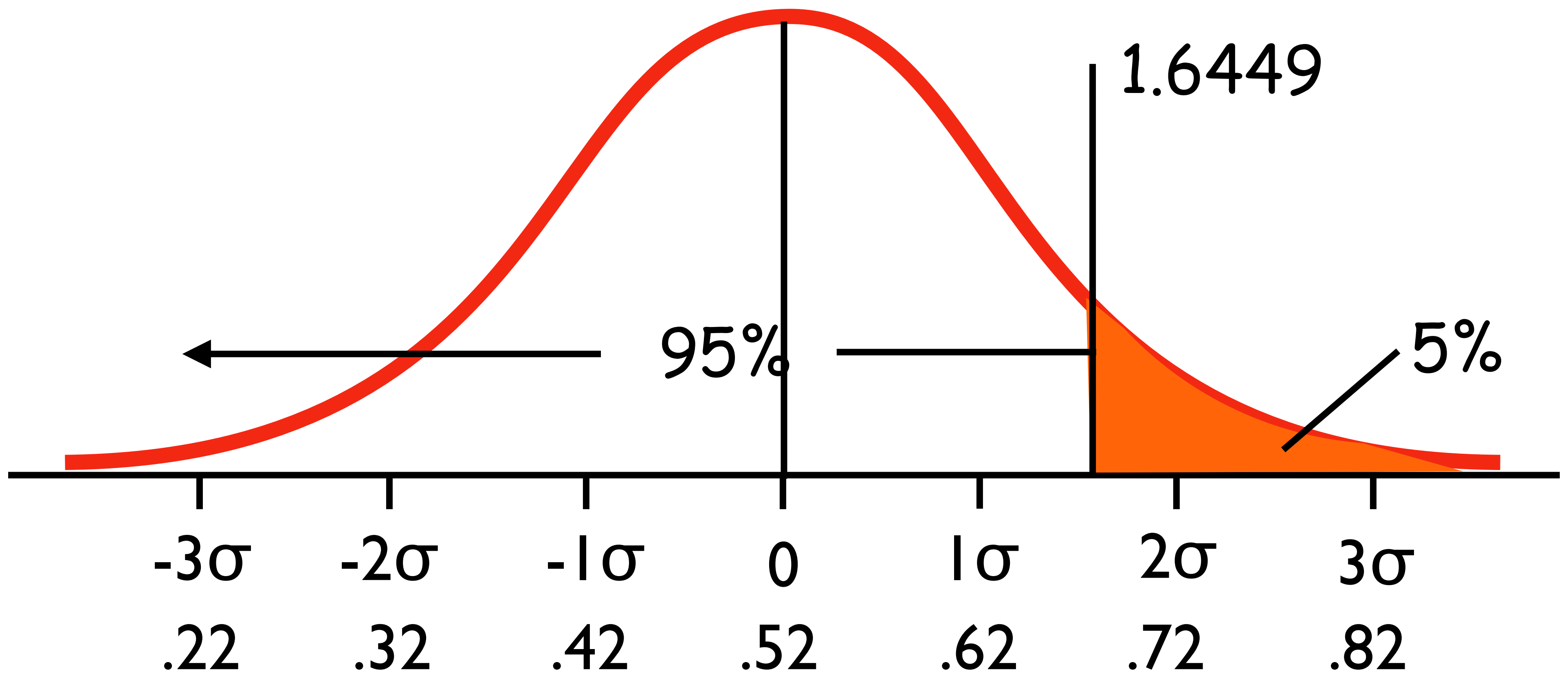
prop>.52
z=.7550086004
p=.2251218385
p^= .5925925926
n: 27

2. Critical Value

We are given the level of significance of .01. A one-tailed test means all the critical area is above the mean.

The critical value of the z-statistic corresponding to $\alpha = .01$ is…

Invnorm(.99, 0, 1) = 2.3263



- A female college professor feels that girls are out-performing the college average in her math class. The college has 52% passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

$p = .52,\ \hat{p} = 16/27 = .5926,\ n = 27,\ \alpha = .01$

prop>.52
z=.7550086004
p=.2251218385
p^= .5925925926
n: 27

3. Calculate z & p

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}},\ n = \text{sample size}$$

$$\hat{p} = \frac{16}{27} = .5926 \qquad n = 27$$

$$z = \frac{.5926 - .52}{\sqrt{\frac{(.52)(.48)}{27}}} = \frac{.0726}{.0961} = .7551$$

p(z > .7551) = normalcdf(.7551, 99, 0, 1) = .2251

- A female college professor feels that girls are out-performing the college average in her math class. The college has 52% passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

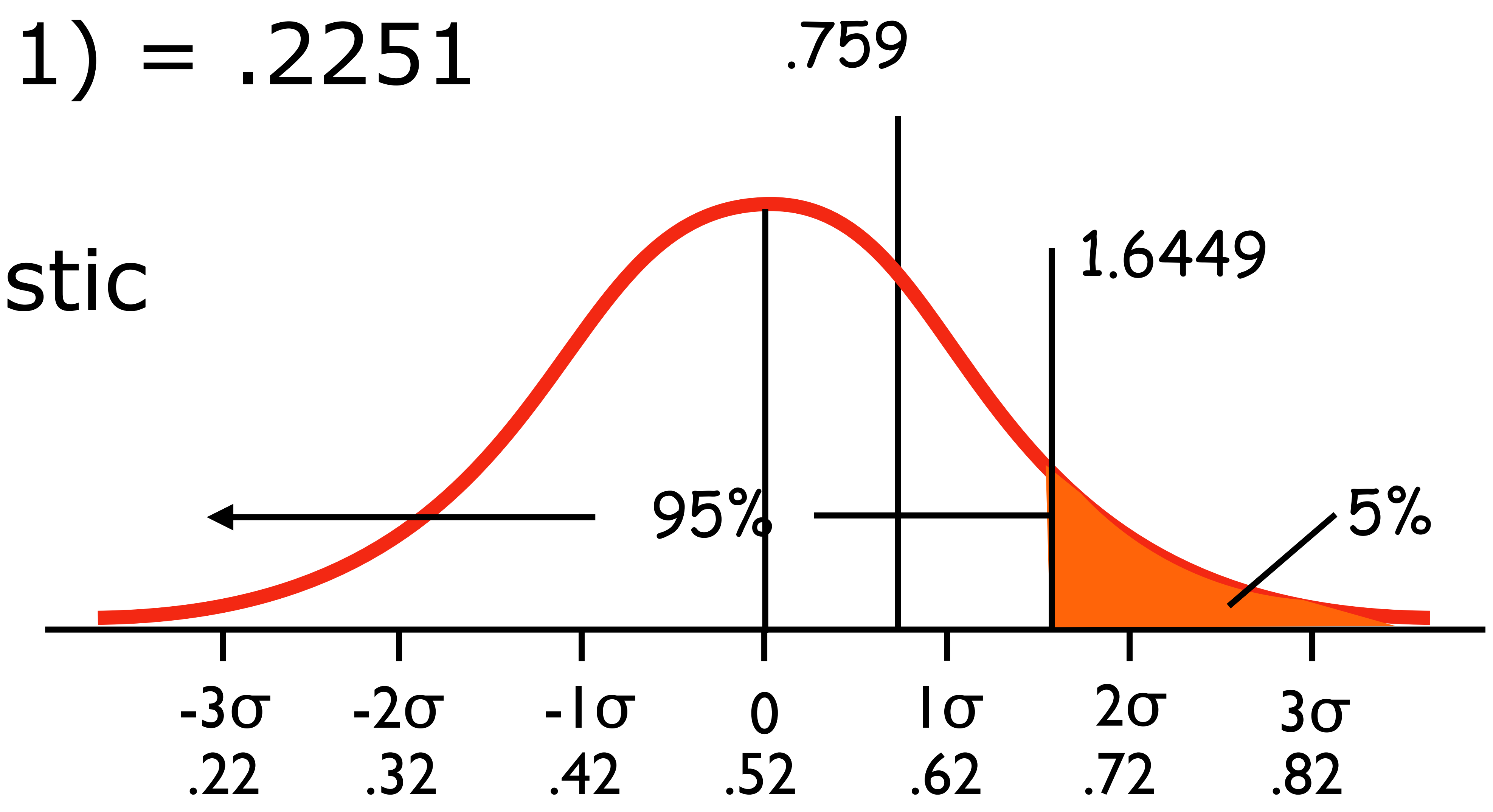
$p = .52$, $\hat{p} = 16/27 = .5926$, $n = 27$, $\alpha = .01$

prop>.52
z=.7550086004
p=.2251218385
p^= .5925925926
n: 27

4. Decision

$p(z > .7551) = \text{normalcdf}(.7551, 99, 0, 1) = .2251$

- $.7551 < 2.33$ and $.2251 > .01$, our z statistic does not reach the rejection region.



We Fail to Reject H_0.

- A female college professor feels that girls are out-performing the college average in her math class. The college has 52% passing rate in that particular course. If 16 out of 27 females pass her course, is the professor correct in believing females pass at a higher rate than on average? Test at the 0.01 level.

$p = .52$, $\hat{p} = 16/27 = .5926$, $n = 27$, $\alpha = .01$

prop>.52
z=.7550086004
p=.2251218385
p^= .5925925926
n: 27

5. Conclusion

- There is **not sufficient evidence** to support (it is not plausible to believe) the suggestion that girls in the female teacher's class are scoring higher than the general population.

Example the 2nd

- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At α = .05, is there enough evidence to reject the educator's claim.

$p = .11,\ \hat{p} = 25/200 = .125,\ n = 200,\ \alpha = .05$

1. Hypotheses

The null hypothesis is that the proportion of seniors dropping out is **not different** from 11% The alternate hypothesis is the proportion of seniors dropping is **different** than 11%. Thus this is a two-tailed test.

α = .05

$H_0: \rho = .11$ and $H_a: \rho \neq .11$

- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At α = .05, is there enough evidence to reject the educator's claim.

$p = .11,\ \hat{p} = 25/200 = .125,\ n = 200,\ \alpha = .05$

Run the Test First

Calculator

STAT ➢ TESTS ⋎ 5:1-PropZTest

p_0: .11
x: 38
n: 200
prop: $\neq p_0$ $<p_0$ $>p_0$
Calculate

Results

prop ≠.11
z= .6779769221
p=.4977861816
p^= .125
n: 200

- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At α = .05, is there enough evidence to reject the educator's claim.

prop ≠.11
z= .6779769221
p=.4977861816
p^= .125
n: 200

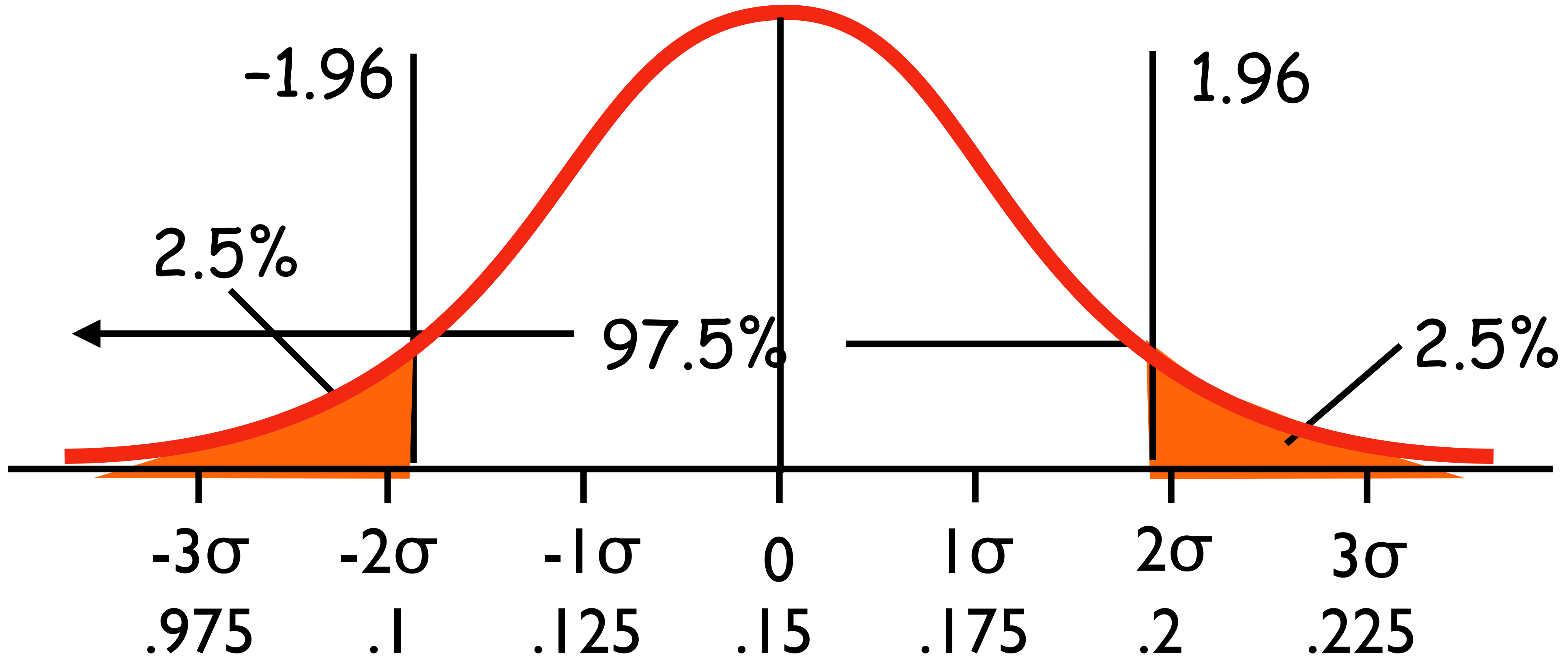
$p = .11,\ \hat{p} = 25/200 = .125,\ n = 200,\ \alpha = .05$

2. Critical Value

We are given the level of significance of .05. A two tailed test means the critical area is both above and below the mean.

The critical value of the z-statistic corresponding to α = .05 is 1.96.

Invnorm(.975, 0, 1)



- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At $\alpha = .05$, is there enough evidence to reject the educator's claim.

prop ≠.11
z= .6779769221
p=.4977861816
p^= .125
n: 200

$p = .11,\ \hat{p} = 25/200 = .125,\ n = 200,\ \alpha = .05$

3. Calculate z & p

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}},\ n = \text{sample size}$$

$$\hat{p} = \frac{25}{200} = .125 \qquad n = 200$$

$$z = \frac{.125 - .11}{\sqrt{\frac{(.11)(.89)}{200}}} = .6780$$

p(z > .6787) = normalcdf(.6780, 10^99, 0, 1) = .2489 **x 2** = .4978

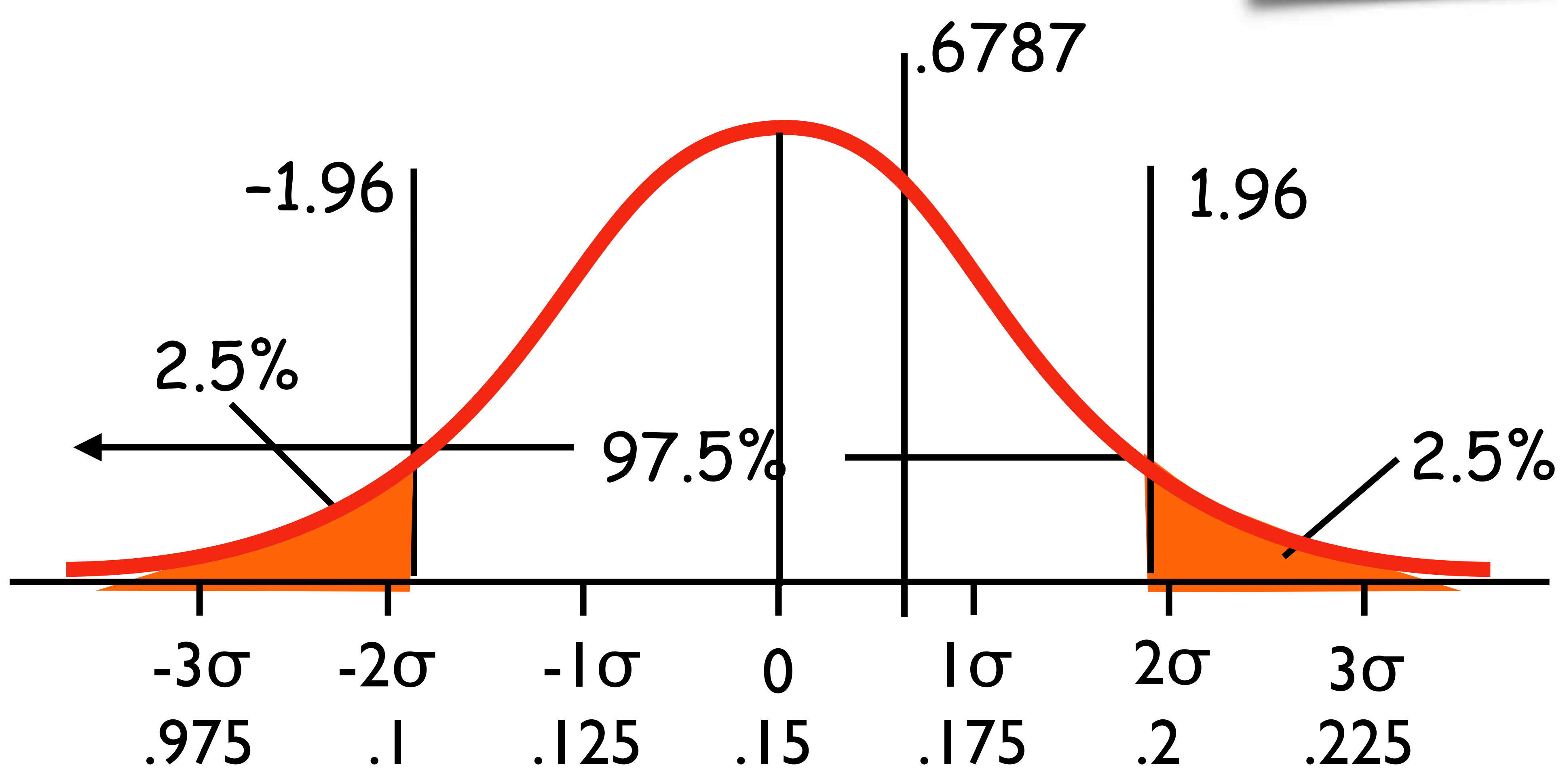
- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At α = .05 is there enough evidence to reject the educator's claim.

prop ≠.11
z= .6779769221
p=.4977861816
p^= .125
n: 200

$p = .11$, $\hat{p} = 25/200 = .125$, $n = 200$, $\alpha = .05$

4. Decision

- .4974 > .05, .6779 < 1.96 so our z statistic does not reach the rejection region.



- **We fail to reject the null.**

- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At α = .05, is there enough evidence to reject the educator's claim.

prop ≠.11
z= .6779769221
p=.4977861816
p^= .125
n: 200

$p = .11,\ \hat{p} = 25/200 = .125,\ n = 200,\ \alpha = .05$

5. Conclusion

- **We fail to reject the null.**

- There is **not sufficient evidence** to reject the educator's claim of a 11% dropout rate. In other words, there is **not sufficient evidence** to claim the dropout rate is different than 11%.

Confidence Interval

- Now we tie together the idea of hypothesis testing and confidence intervals.

- When we **fail to reject H_0**, we are essentially suggesting that the population parameter will fall within the confidence interval.

- When we **reject H_0** we believe that the population parameter falls outside the confidence interval.

3 Cases for a Confidence Interval

- For a population **mean** …

  a. σ known (or n ≥ 30).

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

  b. σ not known and n < 30.

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}}$$

- Confidence interval for a **proportion.**

$$\hat{p} - z_{\alpha/2}\frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}} < p < \hat{p} + z_{\alpha/2}\frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}}$$

Objective: Students will perform hypothesis tests for population proportions using the z statistic.

Statistics 8-5

- An educator estimates that the dropout rate for seniors at high schools in California is 11%. Last year 25 seniors from a random sample of 200 California seniors withdrew. At $\alpha = .05$, is there enough evidence to reject the educator's claim.

$p = .11$, $\hat{p} = 25/200 = .125$, $n = 200$, $\alpha = .05$

Confidence Interval

- The confidence interval for proportions when $\alpha = .05$, **with a two-tailed test**, is: **C = .95**

$$95\%CI = \hat{p} \pm z_{\alpha/2} \frac{\sqrt{\hat{p}\hat{q}}}{\sqrt{n}}$$

$$95\%CI = .125 \pm 1.96 \frac{\sqrt{(.125)(.875)}}{\sqrt{200}}$$

$$.0792 < p < .1708$$

STAT ➢ TESTS A:1-PropZInterval

x: 25
n: 200
C-Level: .95
Calculate

(.07917, .17083)

The hypothesized proportion of .11 does, in fact, fall within the confidence interval, thus confirming our hypothesis conclusion.

Example the 3rd

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90,\ \hat{p} = 376/400 = .94,\ n = 400,\ \alpha = .05$

Is this evidence that the local rate is higher than the national rate?

Is this "convincing" evidence that the local rate is higher?

Did the ad campaign **raise** the local level above the 90% national rate?

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

1. Hypotheses

The null hypothesis is that the proportion of homes is **not higher** than 90% The alternate hypothesis is the proportion of homes is **higher** than 90%. Thus this is a one-tailed test.

$\alpha = .05$

H_0: $\rho \le .90$ and **H_a**: $\rho > .90$

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90,\ \hat{p} = 376/400 = .94,\ n = 400,\ \alpha = .05$

Run the Test First

Calculator

STAT ➢ TESTS ∀ 5:1-PropZTest

p_0: .90
x: 376
n: 400
prop: $\neq p_0$ $< p_0$ $> p_0$
Calculate

Results

prop $> p_0$
z=2.666666667
p=.0038304251
p^=.94
n: 400

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

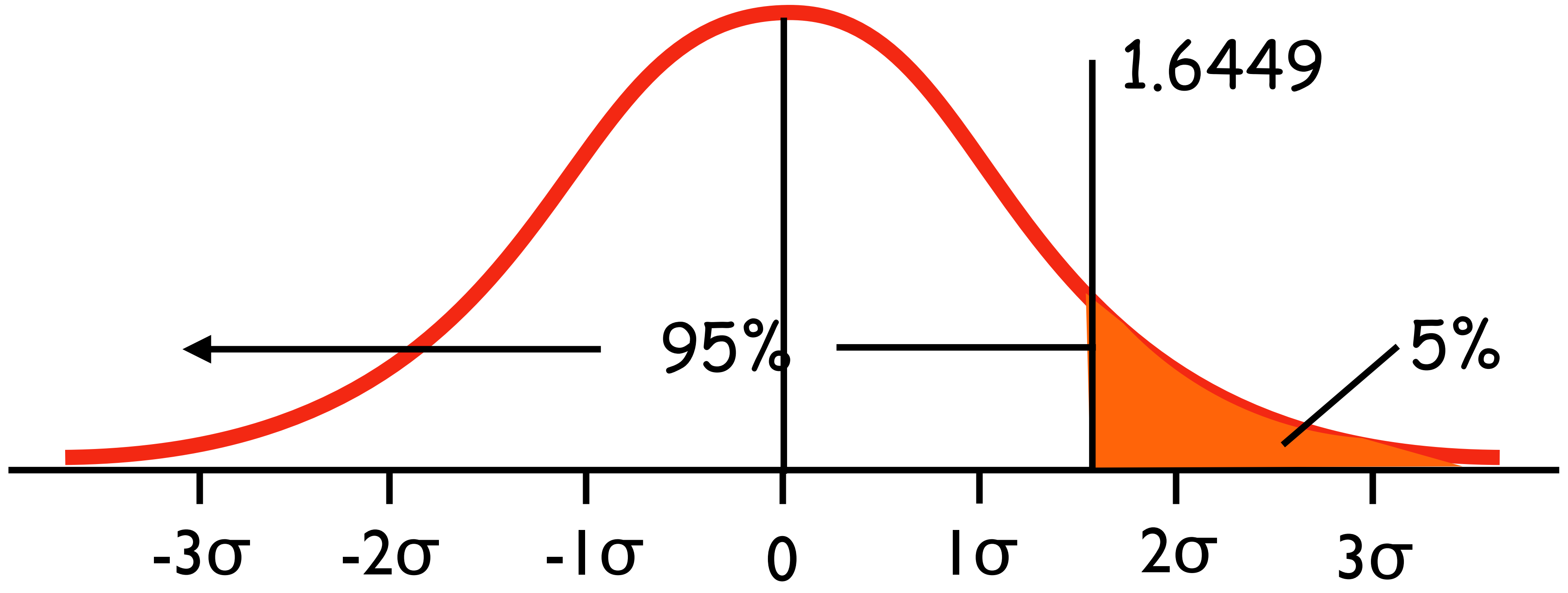
prop $>p_0$
z=2.666666667
p=.0038304251
p^=.94
n: 400

2. Critical Value

The level of significance is .01. A one tailed test means all the critical area is above the mean.

The critical value of the z-statistic corresponding to $\alpha = .05$ is 1.6449.

Invnorm(.95, 0, 1) = 1.6449



- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

prop >p_0
z=2.666666667
p=.0038304251
p^=.94
n: 400

3. Calculate z & p

$$z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}, n = \text{sample size}$$

$$\hat{p} = \frac{376}{400} = .94 \quad n = 400$$

$$z = \frac{.94 - .9}{\sqrt{\frac{(.9)(.1)}{400}}} = 2.666666667$$

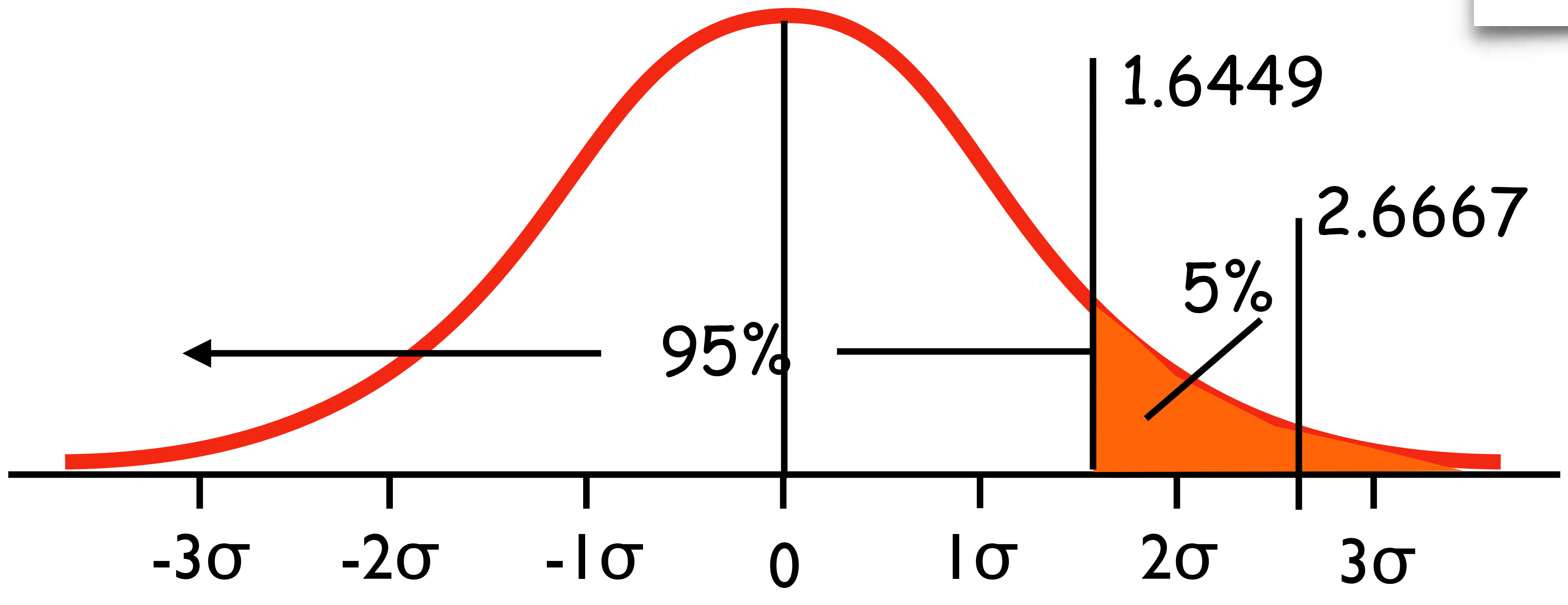
p(z > 2.6667) = normalcdf(2.6667, 99, 0, 1) = .0038

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

prop >p_0
z=2.666666667
p=.0038304251
p^=.94
n: 400

4. Decision



- .0038 ≤ .05, and 2.6667 > 1.6449 so our z statistic falls within the critical region.

- We **reject the null**.

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90,\ \hat{p} = 376/400 = .94,\ n = 400,\ \alpha = .05$

prop >p_0
z=2.666666667
p=.0038304251
p^=.94
n: 400

- We **reject the null**.

5. Conclusion

There is sufficient evidence to suggest that local compliance is greater than the 90% national rate.

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

Confidence Interval

Since we **reject the null** we can nail down the conclusion by creating an interval within which we expect to find the actual compliance rate.

STAT ➢ TESTS ∀ A:1-PropZInt

x: 376
n: 400
C-Level: **.90**
Calculate

(.92047, .95953)
$\hat{p}$=.94
n=400

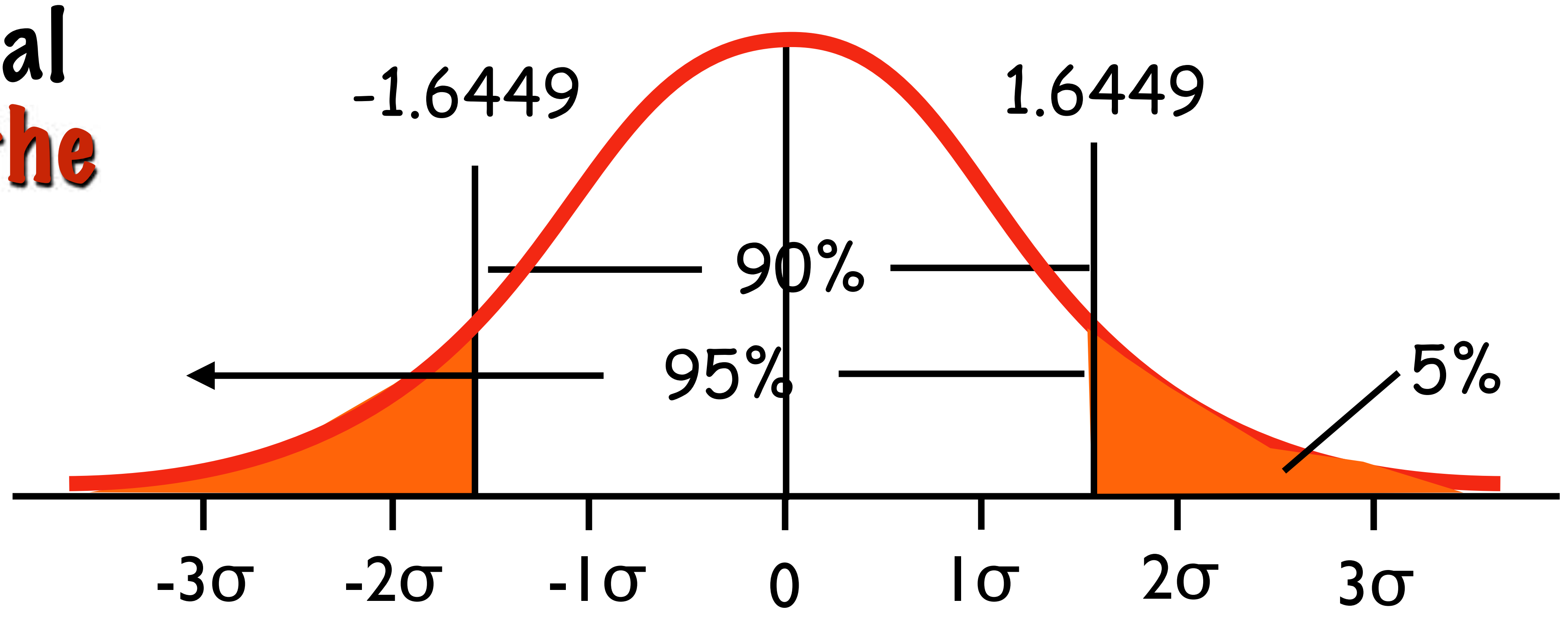
Note the confidence level.

- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90,\ \hat{p} = 376/400 = .94,\ n = 400,\ \alpha = .05$

(.92047, .95953)
$\hat{p} = .94$
$n = 400$

When we create our confidence interval, remember that the hypothesis test was a one tail test. The confidence interval is always two-tailed. To use the **same critical value of the test statistic**, we must change the confidence level.



- A 1996 report from the U.S. Consumer Product Safety Commission claimed that at least 90% of all American homes have at least one smoke detector. A city's fire department has been running a public safety campaign consisting of posters, billboards, and ads on radio and TV and in the newspaper. The city wonders if this concerted effort has raised the local level above the 90% national rate. Building inspectors visit 400 randomly selected homes and find that 376 have smoke detectors.

$p = .90$, $\hat{p} = 376/400 = .94$, $n = 400$, $\alpha = .05$

(.92047, .95953)
$\hat{p}=.94$
$n=400$

Confidence Interval

$$SD(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{400}} = \sqrt{\frac{(.94)(.06)}{400}} = .0119$$

Invnorm(.95, 0, 1) = 1.6449

The margin of error (ME) = $(z^*)\,SE(\hat{p}) \approx 1.6449(0.0119) \approx$ **0.0196**

98% CI = **.94 ± .0196** = (.9204, .9596)

Based on a sample of 400 homes, we are 95% confident the true proportion of homes with detectors is within **0.94 ± 0.0196 or between 92% and 96%.**