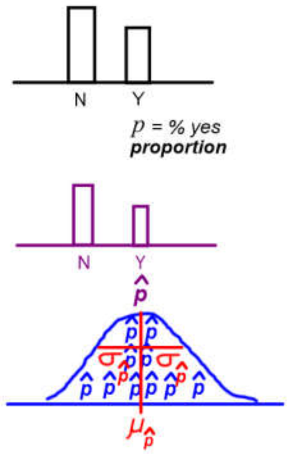
# AP Statistics – Lesson Notes - Chapter 20: Testing Hypotheses About Proportions

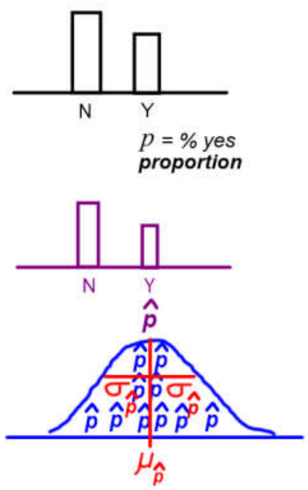



The 3 levels of inference:

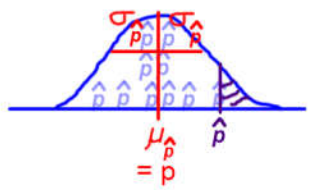
**Population**
has a (usually) unknown parameter

**Sample**
produces a statistic which is an estimate of the populaton parameter

**Sampling Distribution Model**
shows how the statistics from all possible samples would be distributed

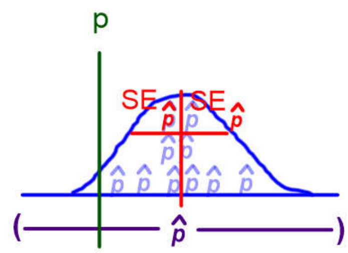



In Ch18...

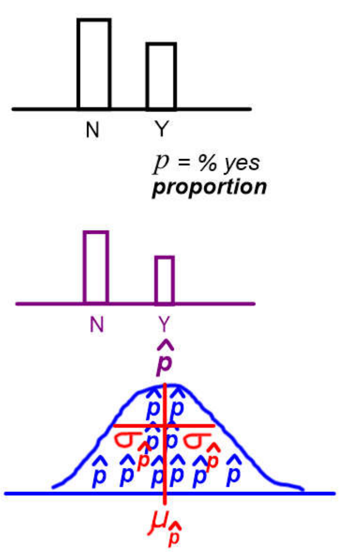



...we pretended we knew the population and were able to say how unusual a particular sample statistic would be.

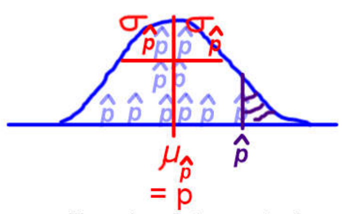
In Ch19...




...we acknowledged that we don't know the population parameter, so we centered everything at the statistic and added a margin of error around the statistic to state a confidence interval where we were confident the population parameter would be.




In Ch20...




...we will go back to centering at the population parameter, but because we don't know it, we'll ***assume*** what it is by stating something called a ***null hypothesis***.

We can then use the sampling distribution model to measure how unusual the sample statistic is, which will let us determine if this sample's statistic's is different enough from the null hypothesis to be called ***statistically significant***.

**What is a hypothesis?**

A ***hypothesis*** is a working assumption, something we believe might be true about a situation but which has not (yet) been demonstrated.

We then look at data, and see if the facts are consistent with the model proposed by the hypothesis.

## Can we prove a hypothesis is true? **No.**

We consider the facts:

If the facts are consistent with the hypothesis:
We believe the hypothesis is probably true **but we have not proved the hypothesis is true**.

If the facts are inconsistent with the hypothesis:
It depends upon how inconsistent. If it is inconsistent 'enough' then **we can reject the hypothesis**.

***It is much easier to disprove a hypothesis than to prove it, so we only use data to disprove (reject) hypotheses.***

Hypothesis: 'All cars are red'

## The Null Hypothesis and Alternative Hypothesis

In a statistical analysis, we believe something to be true. But we start by assuming the opposite: that what we believe to be true is *not true* and this statement is called **the Null Hypothesis**, $H_0$ .

What we originally believe might be true (the opposite of the Null Hypothesis) is called **the Alternative Hypothesis**, $H_a$ .

Once we've established the null hypothesis, we then look at the evidence and ask, "is the evidence convincing enough for us to reject the null hypothesis or must we fail to reject the null hypothesis?"

**We consider a situation/scenario.**

***There are two possible explanations...***

**$H_O$**: Actually, the noteworthy situation is false. *(**$H_O$**: null hypothesis is the "dull" hypothesis)*

**$H_a$**: The noteworthy situation we think may be true actually is true.

***What is our evidence?***

We assume the null hypothesis is true, consider the evidence, and decide if we can reject the null hypothesis.

***What can we conclude?***

**Evidence supports rejecting $H_o$**

We reject $H_o$.

We do have sufficient statistical evidence to conclude $H_a$.

**Evidence does not support rejecting $H_o$**

We fail to reject $H_o$.

We do not have sufficient statistical evidence to conclude $H_a$.

**Let's look at 3 examples...**

**We think all cars are red.**

***There are two possible explanations...***

**$H_O$**: Not all cars are red

**$H_a$**: All cars are red

***What is our evidence?***

We assume the null hypothesis is true, consider the evidence, and decide if we can reject the null hypothesis:

Assume all cars are red. We go to a parking lot and see a white car.

***What can we conclude?***

**Evidence supports rejecting $H_o$**

**Evidence does not support rejecting $H_o$**

We fail to reject $H_o$.
We do not have sufficient statistical evidence to conclude that all cars are red.

**We think the lottery used to select captains favored female pilots.**

***There are two possible explanations...***

**$H_O$**: The lottery was fair *($H_O$: null hypothesis is the "dull" hypothesis)*

**$H_a$**: The lottery favored females

***What is our evidence?***

We assume the null hypothesis is true, consider the evidence, and decide if we can reject the null hypothesis:

Assume the lottery was fair. We simulated conducting a fair lottery many times and 12% of the time, this many captains chosen were females. 12% is not that unusual, and could have occurred naturally by chance.

***What can we conclude?***

**Evidence supports rejecting $H_o$**

**Evidence does not support rejecting $H_o$**

***This wording has the correct level of 'strength'...it states what we believe about the scenario but not too strongly.***

We fail to reject $H_o$.
We do not have sufficient statistical evidence to conclude that the lottery favored females.

**We think a person committed murder.**

***There are two possible explanations...***

**$H_O$**: The person is not guilty *($H_O$: null hypothesis is the "dull" hypothesis)*

**$H_a$**: The person is guilty

***What is our evidence?***

We assume the null hypothesis is true, consider the evidence, and decide if we can reject the null hypothesis:

Assume the person is innocent (innocent until proven guilty). Let's say the prosecution provided the following evidence: the bullets matched a gun owned by the person and their fingerprints were the only ones on the gun. Also, a video recording shows the person shooting the victim with the gun.

***What can we conclude?***

**Evidence supports rejecting $H_o$**

**Evidence does not support rejecting $H_o$**

We reject $H_o$.
We do have sufficient evidence to conclude the person is guilty.

| We think all cars are red | We think the lottery used to pick captains favored females | We think a person committed murder |
|---|---|---|
| $H_O$: Not all cars are red | $H_O$: The lottery was fair | $H_O$: The person is not guilty |
| $H_a$: All cars are red | $H_a$: The lottery favored females | $H_a$: The person is guilty |
| **Evidence does not support rejecting $H_o$** | **Evidence does not support rejecting $H_o$** | **Evidence does support rejecting $H_o$** |
| We fail to reject $H_o$. We do not have sufficient statistical evidence to conclude that all cars are red. | We fail to reject $H_o$. We do not have sufficient statistical evidence to conclude that the lottery favored females. | We reject $H_o$. We do have sufficient evidence to conclude the person is guilty. |

*There are two possible explanations...*

***Things to notice about the conclusions...***

- ***We are always rejecting or not rejecting the null hypothesis...we are never favoring or proving a hypothesis.***
- ***The 1st sentence is always about the null hypothesis.***
- ***The 2nd sentence is always about the alternative hypothesis and always includes context wording.***
- ***The sentences include either both 'negative' or both 'positive' wording***

## The "p-value" (probability value)

The p-value is the probability that, if the null hypothesis were true, the evidence would be as different from the null hypothesis as we are seeing (or more different) just due to natural random sampling variation.

The p-value is the probability that allows us to judge how 'unusual' this result would be if the null hypothesis were true.

| **If the p-value is low (<0.05)** | **If the p-value is high (>0.05)** |
|---|---|
| We believe this result **is unusual**. | We believe this result **is not unusual**. |
| It would probably not occur randomly if the null hypothesis was true, so: | It could occur randomly if the null hypothesis was true, so: |
| **We reject the null hypothesis and conclude the alternate hypothesis is probably true** | **We fail to reject the null hypothesis and conclude the null hypothesis is probably true** |

***In our female pilots/captains activity, our p-value was about 0.12. This is above the cutoff of 0.05, so we concluded it is not that unusual to have this many captains be female just by chance (the null hypothesis). So we failed to reject Ho and state we do not have sufficient statistical evidence to conclude the lottery was unfair.***

How do calculate the p-value? In the 'red cars' and 'person is guilty' examples we didn't really quantify anything - there wasn't a p-value. But in the 'female pilots/captains' example, we conducted a simulation to imagine what would happen if the scenario happened a lot of times.

This allowed us to evaluate what did happen (the "evidence") in comparison to all the outcomes that might have happened.

By comparing our evidence to how we expect that evidence to naturally vary, we can determine if our evidence is "unusual" compared to what we expect.

## The "p-value" (probability value)

In the 'female pilots/captains' example, we produced a distribution of possible 'numbers of captains' under the assumption that the lottery was fair (that the null hypothesis was true). And we found that the evidence was not that "unusual"...it wasn't "convincing evidence" that something noteworthy occurred.

Our simulated distribution of possible 'numbers of captains' allowed us to compute what is essentially a probability, so the way p-value is usually determined to is compute a probability. But to do this, we need a distribution for how this 'evidence' is expected to naturally vary.

**If what we measure as evidence is a mean or a proportion from a sample, instead of a simulation, we can use what we learned in Chapter 18 about sampling distributions to obtain the distribution.**

## An example: DV seniors staying in state for college

Suppose that DV counseling department records show that, in the past, 78% of DV seniors stayed in state for college. We ask a SRS of 40 seniors this year whether they plan on staying in state for college, and supposed that 34 of them (85%) said they are staying in state. *Do we have reason to believe that more seniors are staying in state this year?*

### 1) State the hypotheses (both symbolically and in words)

Hypotheses:

define $p$ = the percentage of DV seniors staying in state for college.

Null hypothesis: $H_0:\ p = .78$ 78% of DV seniors stay in state for college.

Alternative hypothesis: $H_a:\ p > .78$ More than 78% of DV seniors stay in state for college.

***Notice…***

Hypotheses:

***$H_o$ is always 'equal'*** ***the 'status quo' value***

Null hypothesis: $H_0:\ p = .78$ 78% of DV seniors stay in state for college.

Alternative hypothesis: $H_a:\ p > .78$ More than 78% of DV seniors stay in state for college.

***this is a population parameter symbol*** ***this value is always the same for $H_o$ and $H_A$***

### 2) Check conditions

- **SRS**: Problem states a 'random sample' was taken.
- **Indep**: Assuming samples are independent from one another.
- **n<10% pop**: 40 < 10% of all DV seniors.
- **np** = (40)(.78) = 31.2
  **nq** = (40)(.22) = 8.8 ...this one is a little too low (should be at least 10).

We will proceed, but will note this condition is not met in our conclusion.

### 3) Conduct the test (find the p-value)

From our SRS we got $\hat{p} =$

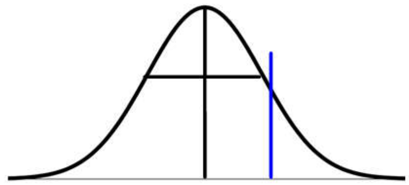
Every time we take a sample of n=40 and find the proportion staying in state for college we would expect that proportion to vary naturally from sample to sample. The sampling distribution of sample proportions shows what we expect this variation to look like:

$\mu_{\hat{p}} = p$ $\qquad \sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$

$\mu_{\hat{p}} =$

$\sigma_{\hat{p}} =$

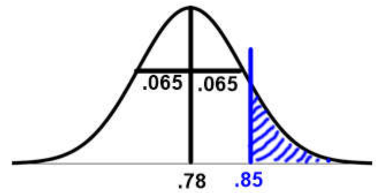
$\sigma_{\hat{p}} =$



***How unusual is it for this sample proportion to occur randomly if the null hypothesis is true (that there is no change in % seniors staying in state)?***

We can calculate the **p-value** (probability that this result occurs randomly due to natural sampling variation, if Ho is true):

$p-value = P\left(\hat{p} > .85\right) =$




$normalcdf\left(.85, 999, \; .78, .065\right) = .14$
lower upper mean SD

$p-value = .14$

### 4) Write the conclusion paragraph:

So what is our conclusion?

*If this year's population proportion really was still .78, it is 14% likely that, in a sample of 40 students, the sample's proportion would .85 or higher simply due to natural variation.*

.14 is not unusual (above .05), so we'll conclude that we do not have significant evidence that a higher proportion of seniors is staying in state this year. ***This means we <u>fail to reject</u> the null hypothesis,*** *and we believe that the proportion of DV seniors staying in state is probably still p=.78.*

*"With significance level of .05, the p-value=0.14 is high, so we <u>fail to reject Ho</u>. We <u>do not</u> have sufficient evidence to conclude that more seniors are staying in state this year compared to previous years."*

## Steps for conducting an hypothesis test...

### 1) State the hypotheses (both symbolically and in words and define all non-standard symbols)

define *p* = the percentage of DV seniors staying in state for college.

$H_0: \; p = .78$ The percentage of seniors staying in state for college is 78%.

$H_a: \; p > .78$ The percentage of seniors staying in state for college is greater than 78%.

### 2) Check conditions

- SRS: Problem states a 'random sample' was taken.
- Indep: Assuming samples are independent from one another.
- n<10% pop: Assuming one student staying in-state doesn't affect other students' decisions.
- np = (40)(.78) = 31.2

  nq = (40)(.22) = 8.8 ...this one is a little too low (should be at least 10).

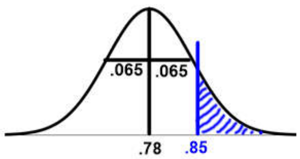
We will proceed, but note this condition is not met in our conclusion.

## 3) Conduct the test (find the p-value)

**by hand**

$$\hat{p} = \frac{34}{40} = .85$$

$$\sigma_{\hat{p}} = \sqrt{\frac{(.78)(.22)}{40}} = .065$$




$$normalcdf(.85, 999,\ .78, .065) = .14$$

**w/calculator**

Perform a 1-prop-ztest in a Ti-84 using

$p_0$: .78
x: 37
n: 40
prop>$p_0$

*Calculate*


z = 2.2138
p-value = 0.01342

## 4) Write the conclusion paragraph:

**1st sentence always written in terms of the null hypothesis (not in context):**

***If p-value < .05:***
*"With significance level of .05, the p-value=### is low so we reject Ho.*

***If p-value > .05:***
*"With significance level of .05, the p-value=### is high so we fail to reject Ho.*

**2nd sentence always written in terms of the alternate hypothesis and in the context of the problem:**

***If p-value < .05:***
*We do have sufficient statistical evidence to conclude (write out the alternate hypothesis)."*

***If p-value > .05:***
*We do not have sufficient statistical evidence to conclude (write out the alternate hypothesis)."*

*"With significance level of .05, the p-value=0.0134 is low so we reject Ho.*
*We do have sufficient statistical evidence to conclude that the percentage of seniors staying in state this year is greater than 78%.*
*(However, the sample size is slightly too low to meet conditions for inference, so we may not trust this result.)"*

Example 2: What if, when we sampled the 40 seniors, 37 had responded that they were staying in state?

Hypotheses are still:

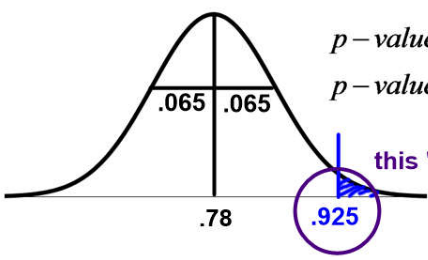
Null hypothesis: $H_0:\ p = .78$

Alternative hypothesis: $H_a:\ p > .78$

(define *p* = the percentage of DV seniors staying in state for college.)

But now:

$$\hat{p} = \frac{37}{40} = .925$$




$$p-value = normalcdf(.925, 999, .78, .065)$$
lower upper mean SD

$$p-value = .013$$

**this 'boundary' changes for each sample**

Now, the probability that this higher proportion occurs by chance is only about 1%.

1% is not likely, so this mean ***we reject the null hypothesis*** *in favor of the alternative hypothesis.* ***We have significant evidence to conclude that the proportion of this year's senior is higher than in the past.***

*"With significance level of .05, the p-value=0.013 is low so we reject Ho.*
*We do have sufficient evidence to conclude that more seniors are staying in state this year compared to previous years."*

## Why is p-value=.05 the cutoff value?

The truth is this is just a rule-of-thumb (although it is very frequently used). It means there is a 1/20 chance of making an error in the conclusion.

The cutoff should be whatever is convincing for the particular situation. Is a 5% chance that we conclude something that might be happening due to natural variation good enough? Consider these cases:

*A researcher claims that the proportion of college students who hold part-time jobs now is higher than in the past.*
Maybe even p=.10 would be enough to convince you that a change has occurred.

*A theoretical physicist's experiment is testing a new hypothesis in particle physics which, if true, will change our understanding of the universe.*
We might want to be really, really sure that a result is not due to chance and set the cutoff much lower than .05 (particle physics uses 5 $\sigma$)

**This 'cutoff value' has various names:**

**significance level, alpha level, standard of proof**

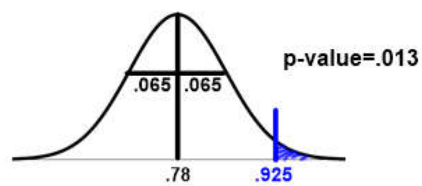
**(a lower significance (alpha) level = a higher standard of proof)**

# The 3 ways to shade: One- and two-sided alternatives

In the DV seniors in state college example, the wording of the problems asked *"Do we have reason to believe that **more** seniors are staying in state this year?"*
The word "more" required us to use the alternative hypothesis:

$H_A : p > .78$ The percentage of seniors in state has increased.

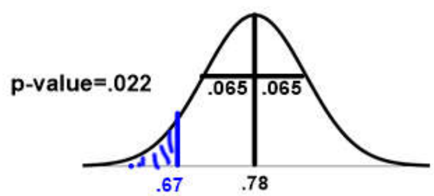
This is a **one-sided, upper hypothesis test**, because Ha is greater than, and the p-value would be the shaded region to the **right** of the statistic.




But if the questions had asked *"Do we have reason to believe the **less** seniors are staying in state this year?"* then we would need to use this alternative hypothesis:

$H_A : p < .78$ The percentage of seniors in state has decreased.

This is a **one-sided, lower hypothesis test**, because Ha is less than, and the p-value would be the shaded region to the left of the statistic.



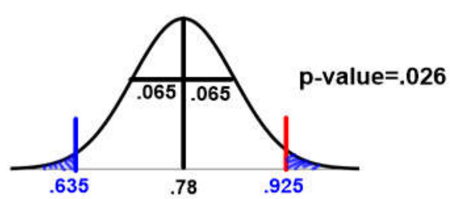

You can also have alternate hypothesis which do not specify a direction:

Suppose that DV counseling department records show that, in the past, 78% of DV seniors stayed in state for college. We ask a SRS of 40 seniors this year whether they plan on staying in state for college, and supposed that 37 of them (85%) said they are staying in state. *Do we have reason to believe that the percentage of seniors staying in state has changed?*

But if the questions had asked *"Do we have reason to believe the **less** seniors are staying in state this year?"* then we would need to use this alternative hypothesis:

$H_A : p \neq .78$ The percentage of seniors in state has changed.

This is a **two-sided hypothesis test**.



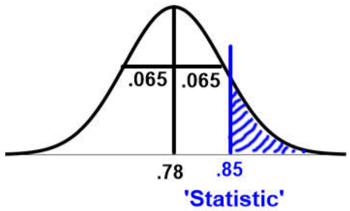

A two-sided hypothesis test is shaded is as follows:

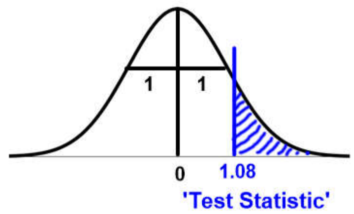
1) Locate the statistic on the distribution.
2) Shade from this statistic away from the mean (whichever side that is).
3) Compute the one-sided p-value on that side.
4) Double this value for the two-sided p-value.

## 'Statistic' vs. 'Test Statistic'

We could have figured out the z-score for this data value and used a standardized normal curve instead:

$$z = \frac{x-\mu}{\sigma} = \frac{.85-.78}{.065} = 1.08$$



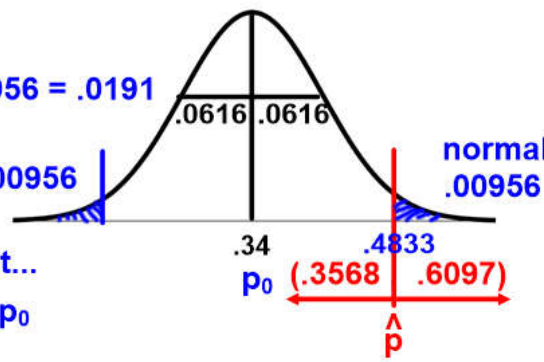




$$p-value = P\left(\hat{p} > .85 \mid H_o\ true\right)$$
$$= normalcdf\left(0.85, 999, 0.78, 0.065\right)$$
lower upper mean SD
$$= 0.14$$

$$p-value = P\left(Z > 1.08 \mid H_o\ true\right)$$
$$= normalcdf\left(1.08, 999, 0, 1\right)$$
lower upper mean SD
$$= 0.14$$

## How hypothesis tests and confidence intervals are related

p-value = 2 x .00956 = .0191




normalcdf(.4833, 999, .34, .0616)
.00956

**Hypothesis test...**

**...centered at $p_0$**

**...uses $p_0$ to compute $\sigma_{\hat{p}}$**

*With a significance level of .05, p-value = .0191 is low so we reject $H_0$. We do have sufficient statistical evidence to conclude that the percentage of DV seniors staying in state for college has changed.*

**Confidence interval...**

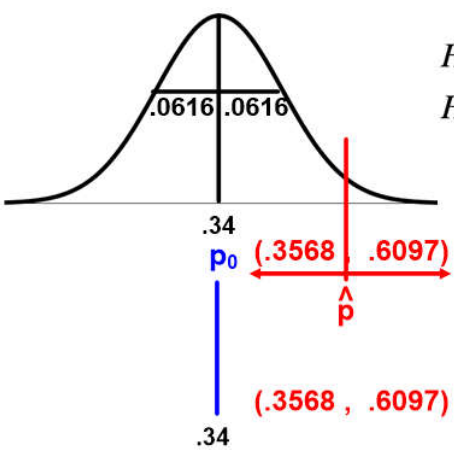
**$\hat{p} \pm (z^*)(SE_{\hat{p}})$**

**...centered at $\hat{p}$**

**...uses $\hat{p}$ to compute $SE_{\hat{p}}$**

*We are 95% confident that the percentage of all DV seniors staying in state for college is between 35.68% and 60.97%.*

## We can use a confidence interval to test an hypothesis




$H_0 : p = .34$ *No change in percentage.*

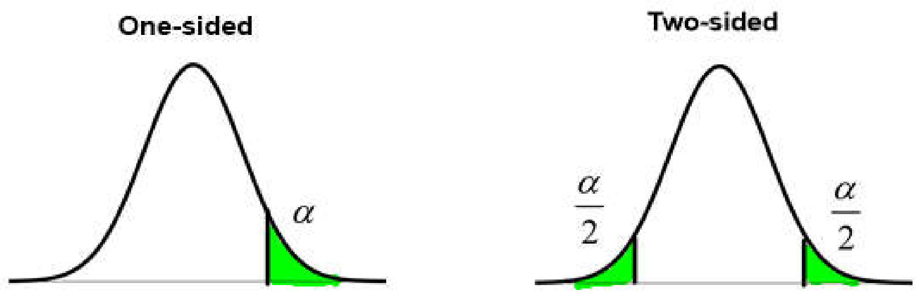
$H_A : p \neq .34$ *Is a change in percentage.*

**The percentage of students staying in state is likely not 34% because 34% is not within the 95% confidence interval.**
**(We can reject $H_0$ because it is not a 'likely' value).**

## 95% Confidence Interval ⟷ 2-sided z-test w/95% significance

A 95% Confidence Interval corresponds to a 2-sided hypothesis test with a significance level of 5%. The **'significance level'** is also called the **'alpha level'**, $\alpha = .05$

How this confidence level and alpha relate depends upon whether this is 1-sided or 2-sided:




## Standard wording formats you should memorize...

**Explain/interpret the confidence interval**

We are 90% confident that the true proportion of DV seniors staying in state for college is between 73% and 95%.

**Explain the confidence level / meaning of % confidence**

If we were to take many samples of size 40 seniors and compute confidence intervals for each, 90% of the confidence intervals would contain the true percentage of all DV seniors who are staying in state for college.

**Explain the results of a hypothesis test**

With significance level of 0.05, a p-value = 0.04 is low, so we reject Ho.
We do have sufficient statistical evidence to conclude that the proportion of DV seniors staying in state for college has increased.
---or---
With significance level of 0.05, a p-value = 0.08 is high, so we fail to reject Ho.
We do not have sufficient statistical evidence to conclude that the proportion of DV seniors staying in state for college has increase

**Explain the meaning of a p-value**

If 78% of DV seniors were actually still staying in state for college, our p-value of 0.09 means there is a 9% probability of this sample's results (85% staying in state) or higher occuring just due to chance (random sampling variation).