SOC 3811/5811: BASIC SOCIAL STATISTICS

Hypothesis Testing

Hypothesis Testing

LAST WEEK

Confidence intervals use sample data to specify a range of values within which we are confident that the population parameter falls

Example: “With 95% certainly we conclude that the population proportion of people who own a car is between 0.401 and 0.420”

Example: “With 99% certainty we conclude that the mean income in the population is somewhere between $31,200 and $32,000”

Hypothesis Testing

THIS WEEK

Hypothesis testing (or *significance testing*) uses sample data to test particular claims about the value of a population parameter

Example: "Do our sample data support the assertion that more than 41% of people in the population own cars?"

Example: "Do our sample data support the assertion that the mean income in the population is greater than $31,900?"

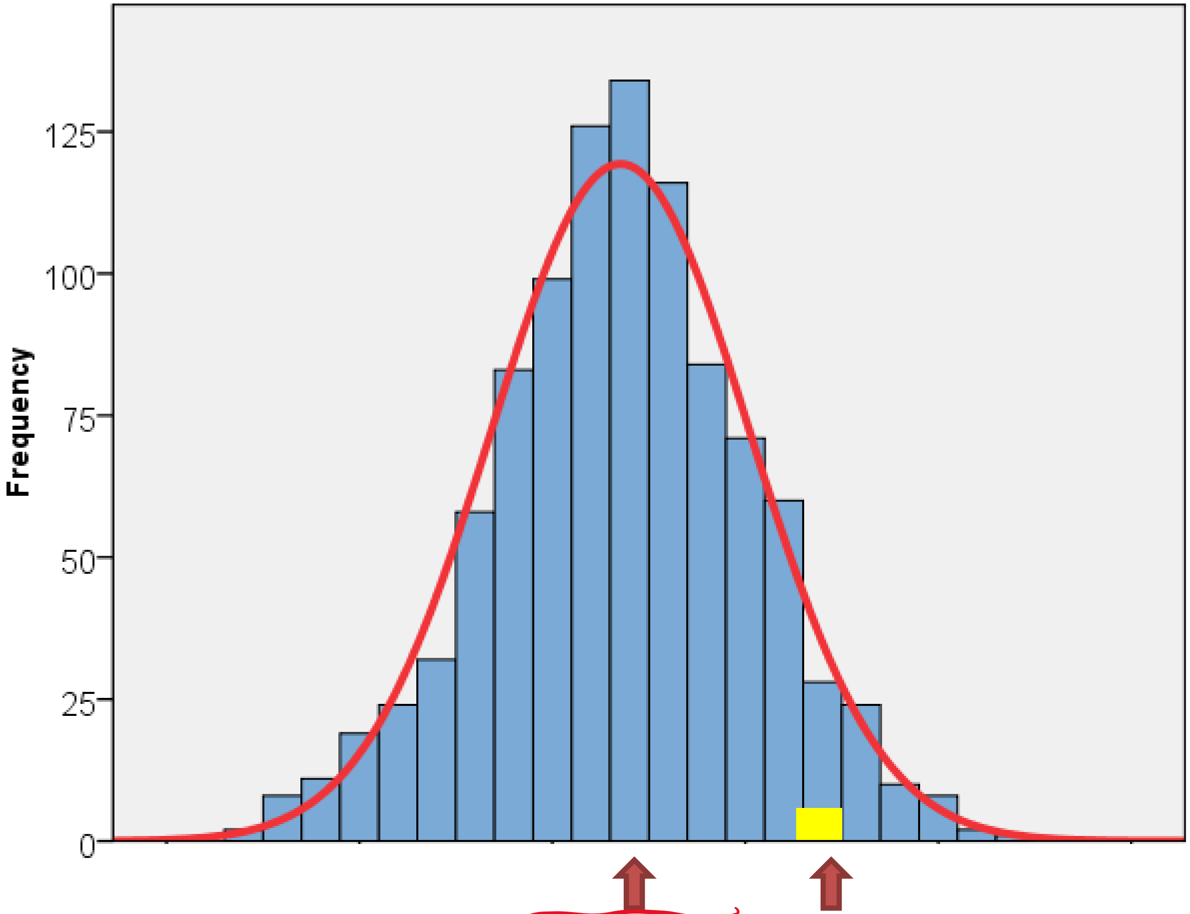
Hypothesis Testing

In 2012 the U.S. National Transportation Safety Board set a 5-year goal of having more than 95% of all American drivers use their seatbelts

To see whether they are on target for meeting that goal, they randomly sampled 1,000 American drivers in 2014

They found that 962—or 96.2%—of the 1,000 drivers they sampled use their seatbelts

Is this evidence sufficient to confidently conclude that more than 95% of *all* American drivers use their seatbelts?



p=0.950 0.962

If this is true… …then what is the probability of observing a value this extreme?

Hypothesis Testing

We will consider hypothesis tests for…

…single population proportions (p)

…single population means (μ_Y)

…differences in population proportions ($p_1 - p_2$)

…differences in population means ($\mu_1 - \mu_2$)

As with confidence intervals, the logic is the same for each type of hypothesis test

Hypothesis Testing

Hypothesis tests involve two competing (mutually exclusive and exhaustive) hypotheses; what these hypotheses are depends on the question at hand

Examples:

H_0 *Hyp. 1*: More than 95% of all U.S. drivers use their seat belts Null

H_A *Hyp. 2*: 95% or fewer of all U.S. drivers use their seat belts Alt

Hyp. 1: There are 12.1 penguins per square mile

Hyp. 2: The number of penguins per square mile is not 12.1

Hypothesis Testing

These two hypotheses are typically called…

…the "null" hypothesis (H_0) and

…the "alternative" hypothesis (H_1 or H_A … H_1 in this class)

The **null hypothesis** often says that nothing is happening or that something is not true

The **alternative hypothesis** usually says that something is happening or that something is true

Hypothesis Testing

Null Hypothesis	**Alternative Hypothesis**
pH of Ocean Water is 7.8	pH of Ocean Water is not 7.8
$\mu = 7.8$	$\mu \neq 7.8$
The Average Skink Weighs an Ounce or Less	The Average Skink Weighs More than an Ounce
$\mu \leq 1$ oz	$\mu > 1$ oz
10% of People Carry Guns	The % Who Carry Guns is Not 10%
$p = 0.10$	$p \neq 0.10$
Coke and Pepsi Taste the Same	Coke and Pepsi Taste Different
$T_P = T_C$ (or, $T_P - T_C = 0$)	$T_P \neq T_C$ (or, $T_P - T_C \neq 0$)

Pro Tip: The equal sign always goes in the null hypothesis!

Hypothesis Testing

Depending on the substantive situation, we might distinguish between **one-sided** and **two-sided** tests

If we are asking whether a population value is higher than a certain threshold (say, 0), then we use a one-sided test

Example: H_0: $\mu \leq 0$ H_1: $\mu > 0$

If we are asking whether a population value is lower than a certain threshold (say, 0), then we also use a one-sided test

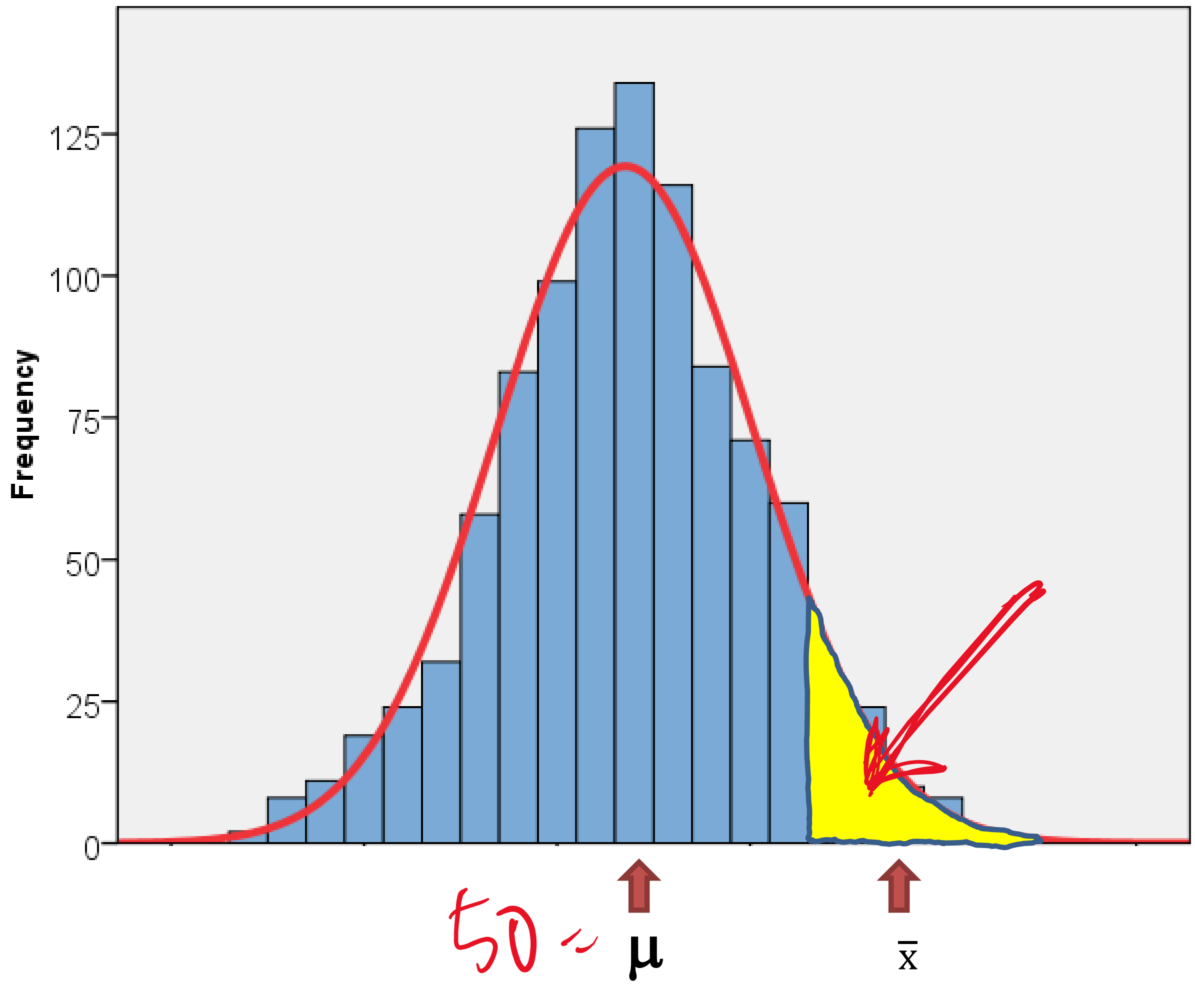
Example: H_0: $\mu \geq 0$ H_1: $\mu < 0$

If we are asking whether a population value is simply different than a certain threshold (say, 0) then we use a two-sided test

Example: H_0: $\mu = 0$ H_1: $\mu \neq 0$

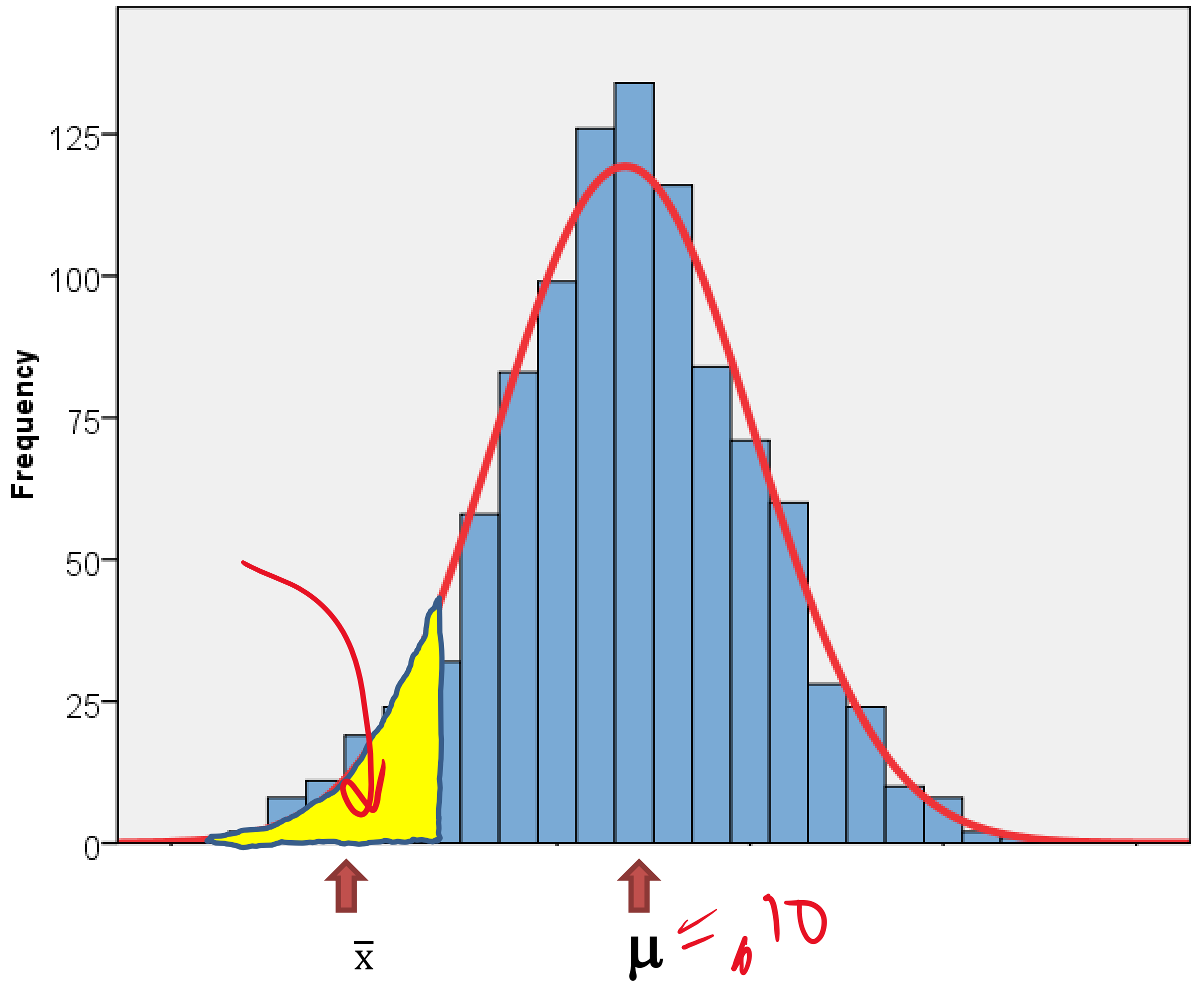
One-sided (upper tail) hypothesis test

Assuming that μ is true, what is the probability of observing one sample mean that far **above** μ?



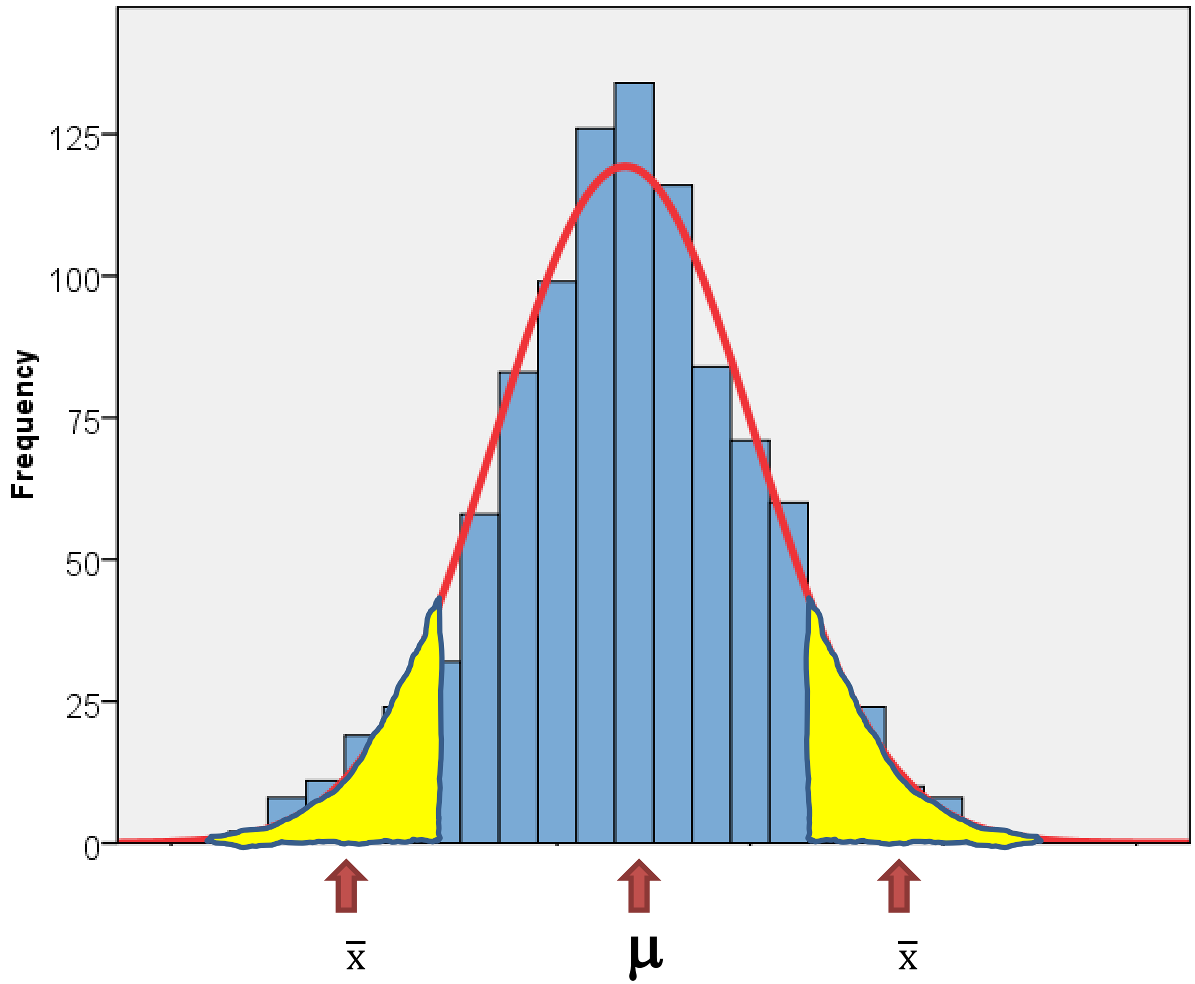
One-sided (lower tail) hypothesis test

Assuming that μ is true, what is the probability of observing one sample mean that far **below** μ?



Two-sided hypothesis test

Assuming that μ is true, what is the probability of observing one sample mean that far **above or below** μ?



Hypothesis Testing

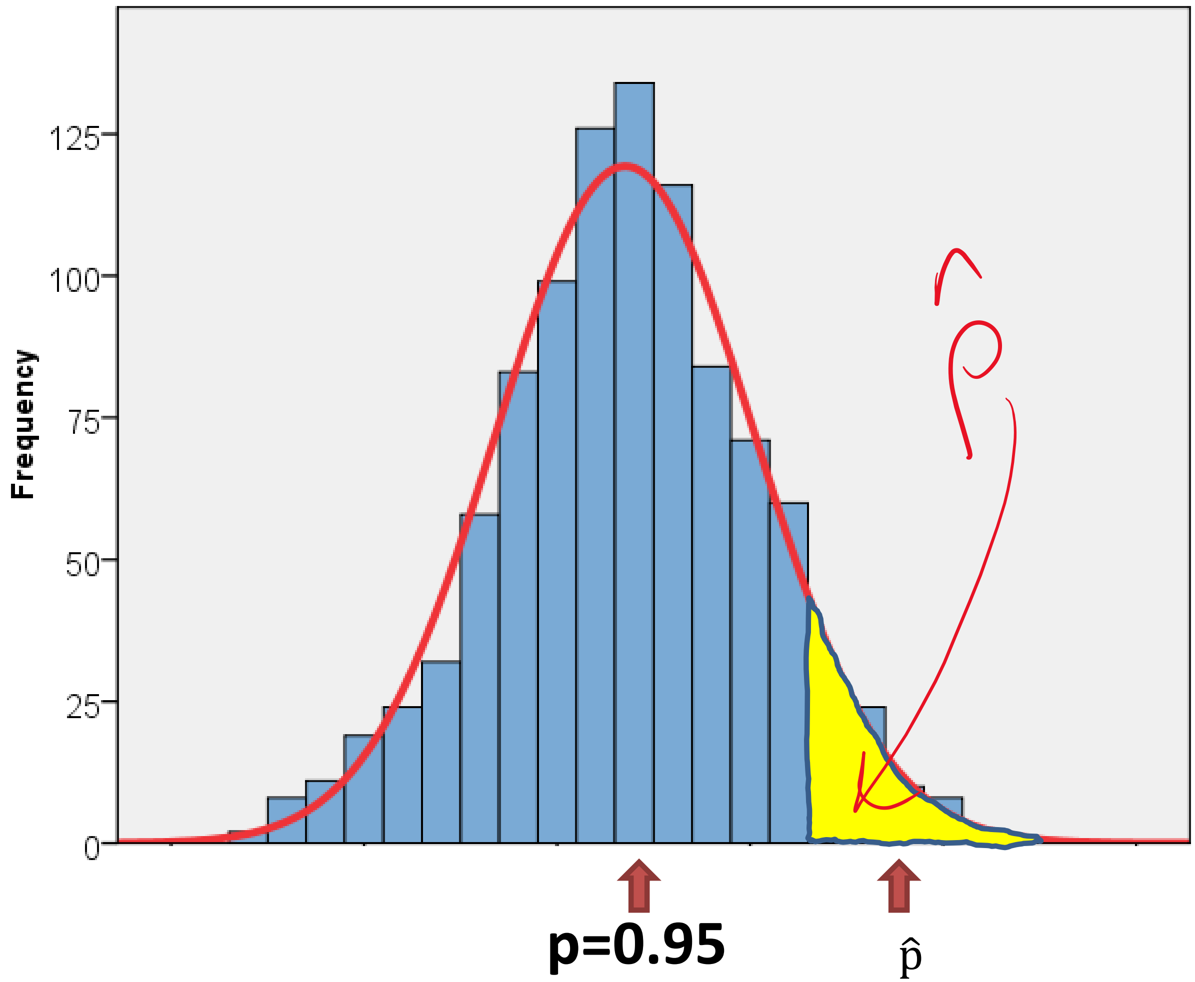
Do more than 95% of all American drivers use seat belts?

We will answer “yes” if we conclude that <u>more than</u> 95% use their seat belts. So, we employ a one-sided hypothesis test:

H_o: $p \leq 0.95$ H_a: $p > 0.95$

One-sided (upper tail) hypothesis test

Assuming that p=0.95 or less, what is the probability of observing one sample mean that far **above** p?



Hypothesis Testing

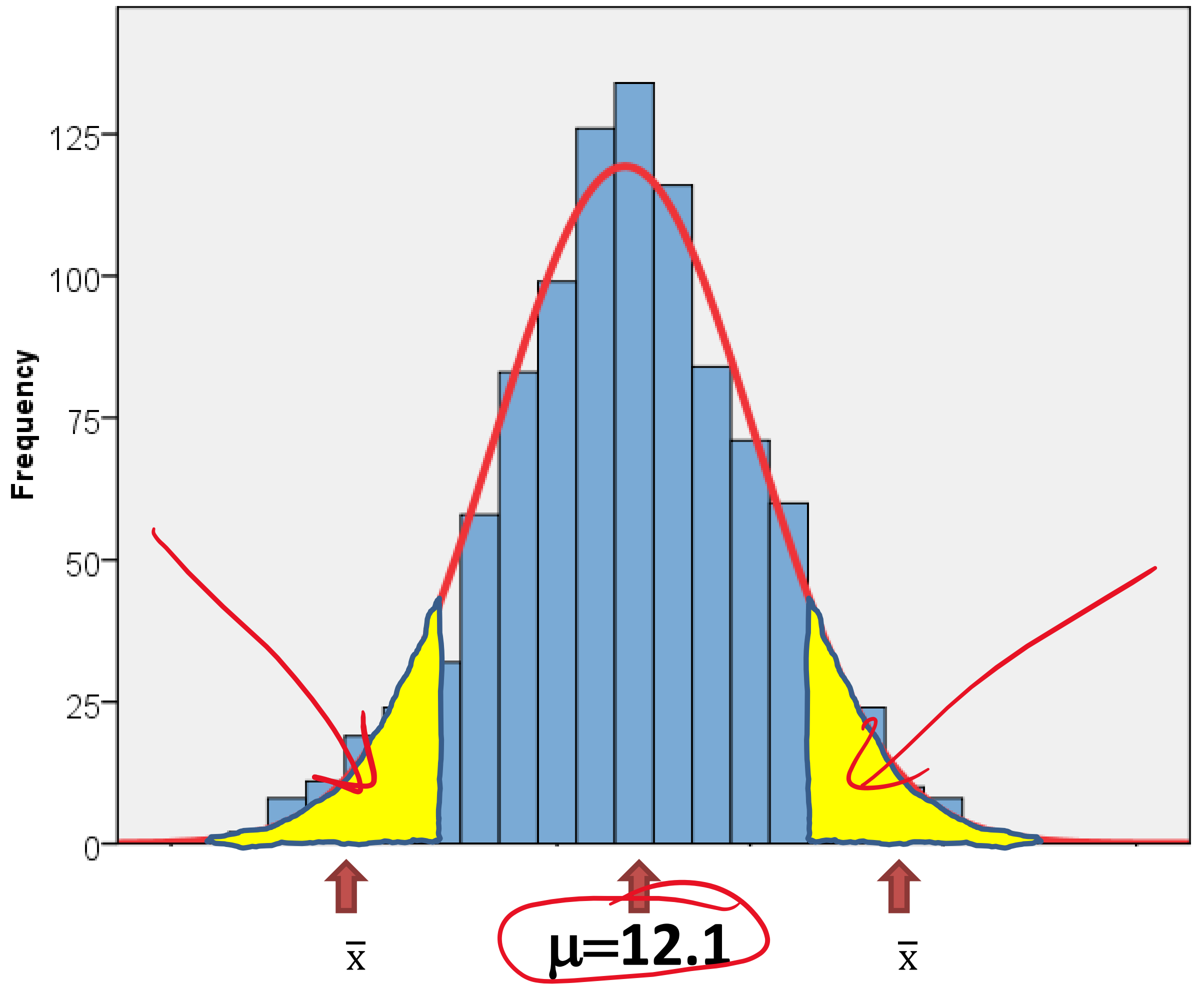
Are there 12.1 penguins per square mile in Antarctica?

We will answer "no" if we conclude that the mean number of penguins per square mile is greater than 12.1 or less than 12.1. So, we employ a two-sided hypothesis test:

H_o: $\mu=12.1$ H_a: $\mu\neq12.1$

Two-sided hypothesis test

Assuming that μ is 12.1, what is the probability of observing one sample mean that far **above or below** 12.1?



Worksheet

Does the average dog weigh less than 20 pounds ?

H_0: H_a:

Do Americans typically have 3 credit cards?

H_0: H_a:

Are more than 10% of women in abusive relationships ?

H_0: H_a:

Is it true that 5% of kids are in street gangs ?

H_0: H_a:

Hypothesis Testing

1. State the null (H_0) and alternative (H_1) hypotheses
2. Check that the sample data conform to basic assumptions; if they do not, then do not go any further
3. Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis
4. Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level
5. Calculate the test statistic t, z
6. Compare the test statistic to the critical value
   - If the test statistic is as large as the critical value, then reject H_0 (with probability of α of doing so even though H_0 should not actually be rejected)
   - If the test statistic is less than or equal to the critical value, then do not reject H_0 (with probability of β of doing so even though H_0 should be rejected)

Hypothesis Testing

We always operate under the assumption that the null hypothesis is true in the population

Given that the null hypothesis is assumed to be correct …

…if the probability of observing the sample data (that is, the observed sample proportion, the observed sample mean, etc.) is very low then we reject the null

“Rejecting” and “failing to reject” the null hypothesis are the only possible outcomes of a hypothesis test

We always operate under the assumption that the null hypothesis is true in the population

The only possible outcomes are:

1. Reject the null hypothesis (H_0)
2. Fail to reject to null hypothesis (H_0)

Hypothesis Testing

We always operate under the assumption that the null hypothesis is true in the population

Given that the null hypothesis is assumed to be correct …

…if the probability of observing the sample data (that is, the observed sample proportion, the observed sample mean, etc.) is very low then we reject the null

"Rejecting" and "failing to reject" the null hypothesis are the only possible outcomes of a hypothesis test

Example 1

In 2007 the U.S. National Transportation Safety Board set a 5-year goal of having more than 95% of all American drivers use their seatbelts

To see whether they are on target for meeting that goal, they randomly sampled 1,000 American drivers in 2012

They found that 962—or 96.2%—of the 1,000 drivers they sampled use their seatbelts

Is this evidence sufficient to confidently conclude that more than 95% of <u>all</u> American drivers use their seatbelts?

Example 1

1. State the null (H_0) and alternative (H_1) hypotheses

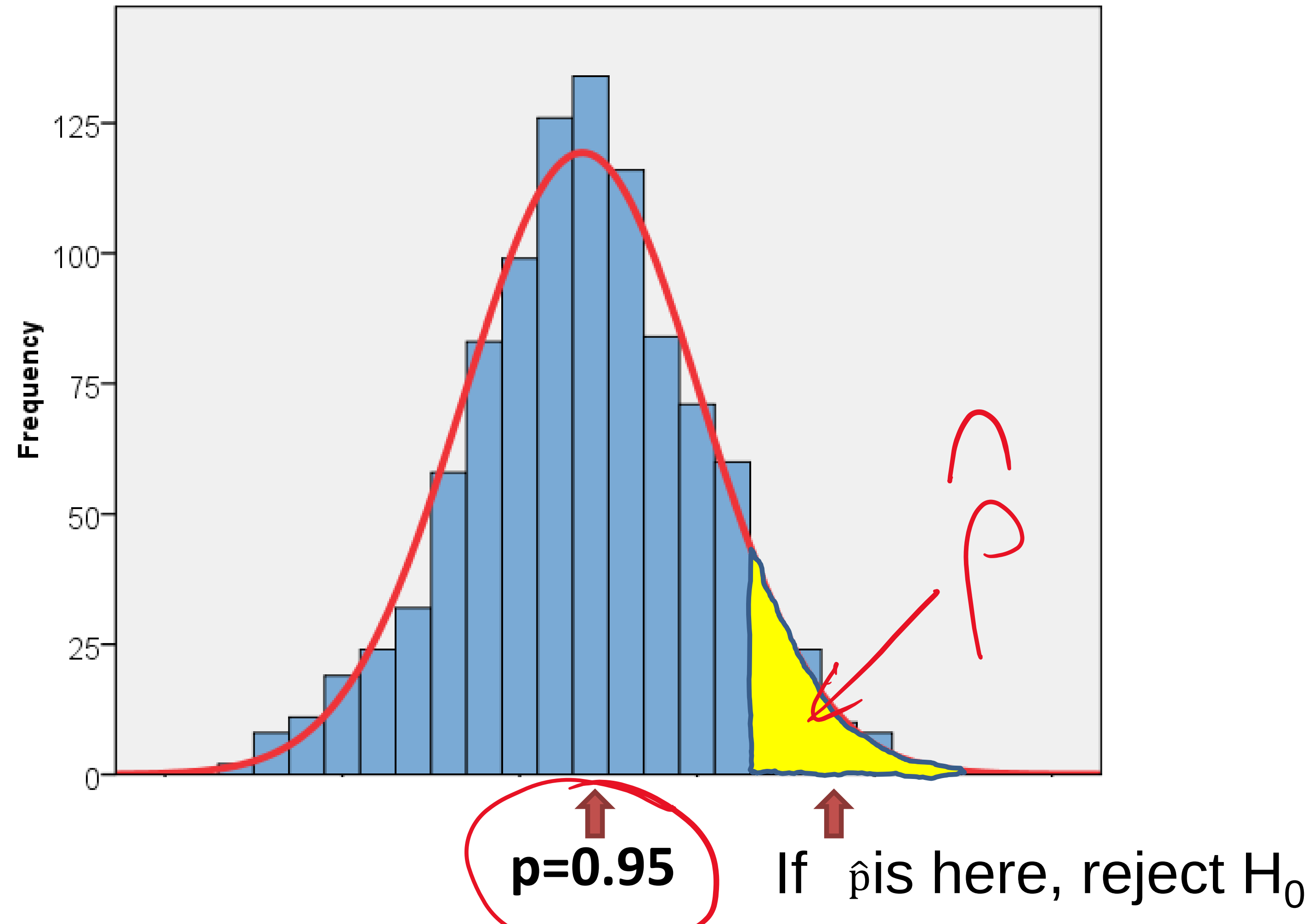
Because we are asking whether the proportion exceeds a particular value (instead of just being different from it), this is a one-sided hypothesis test

H_0: $p \leq 0.95$

H_1: $p > 0.95$

One-sided (upper tail) hypothesis test

Assuming that p=0.95 or less, what is the probability of observing one sample proportion that far **above** p?



Example 1

2. Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The sample must have been randomly selected from the population … (True)

np_0 and $n(1-p_0)$ must both be at least 10 (where p_0 is the proportion we assume to be correct based on the null hypothesis)

$np_0 = 1{,}000 \times 0.95 = 950$ and $n(1-p_0) = 1{,}000 \times 0.05 = 50$

(True)

Example 1

3. Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

The precise α level we select is arbitrary, and varies by discipline (and sometimes by journal or even by author)

0.05 is most common; 0.01 is common, too

Let's go with α=0.05 in this example

Example 1

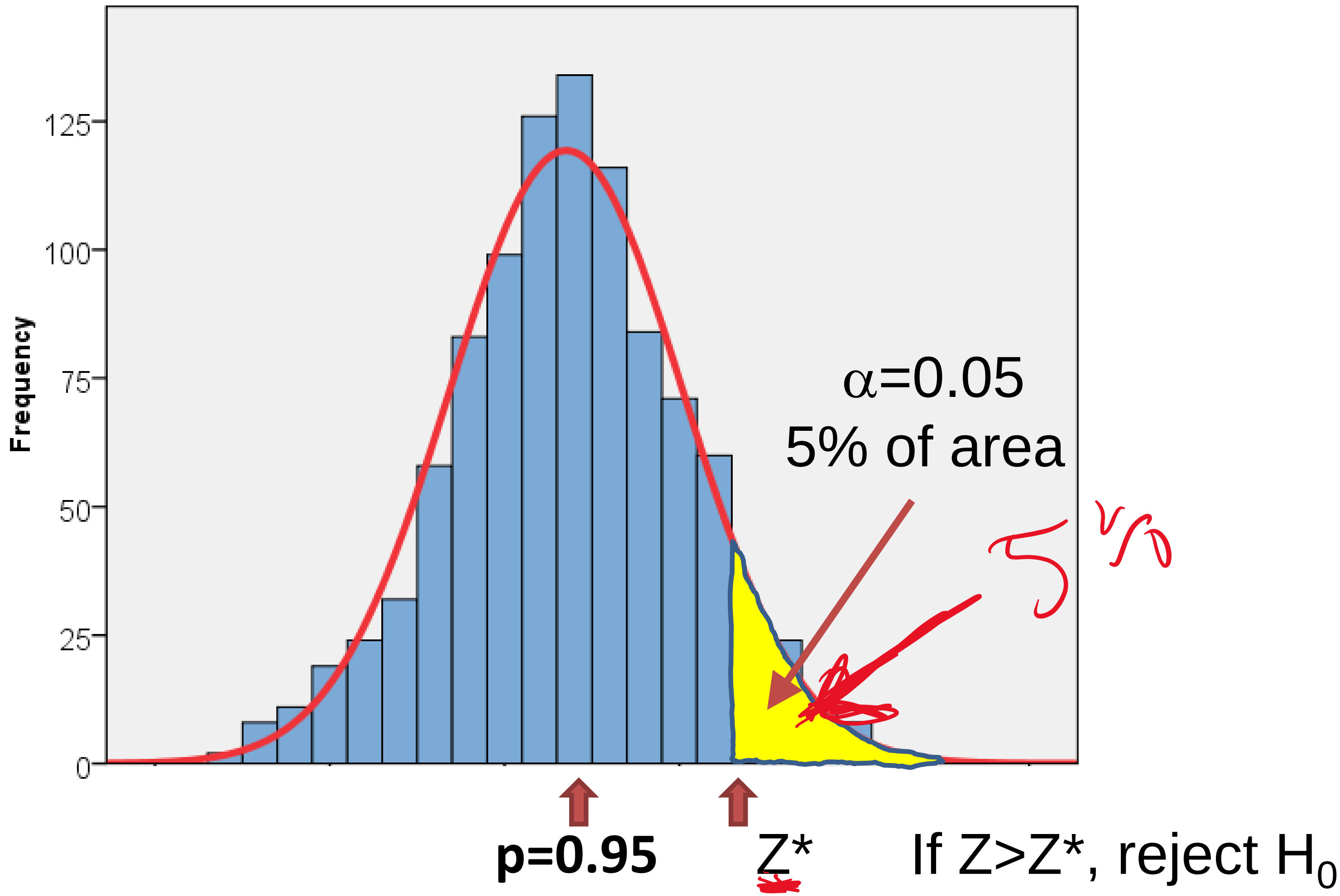
4. Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

How large must the critical value Z* be in order to have the area under the standard normal curve that exceeds Z* be less than α?

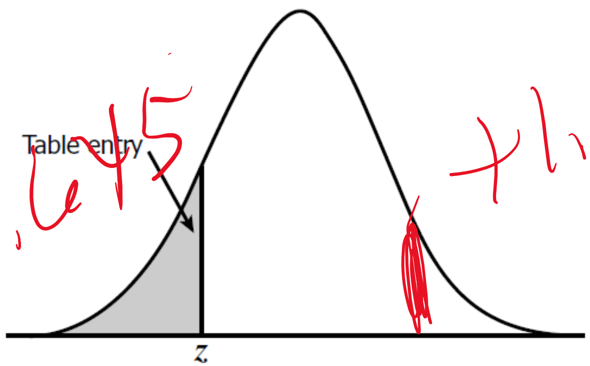
It can then be helpful to re-write the hypotheses in terms of the Z-score(s)

One-sided (upper tail) hypothesis test

Assuming that p=0.95 or less, what is the probability of observing one sample proportion that far **above** p?



Standard Normal Probabilities



−1.645

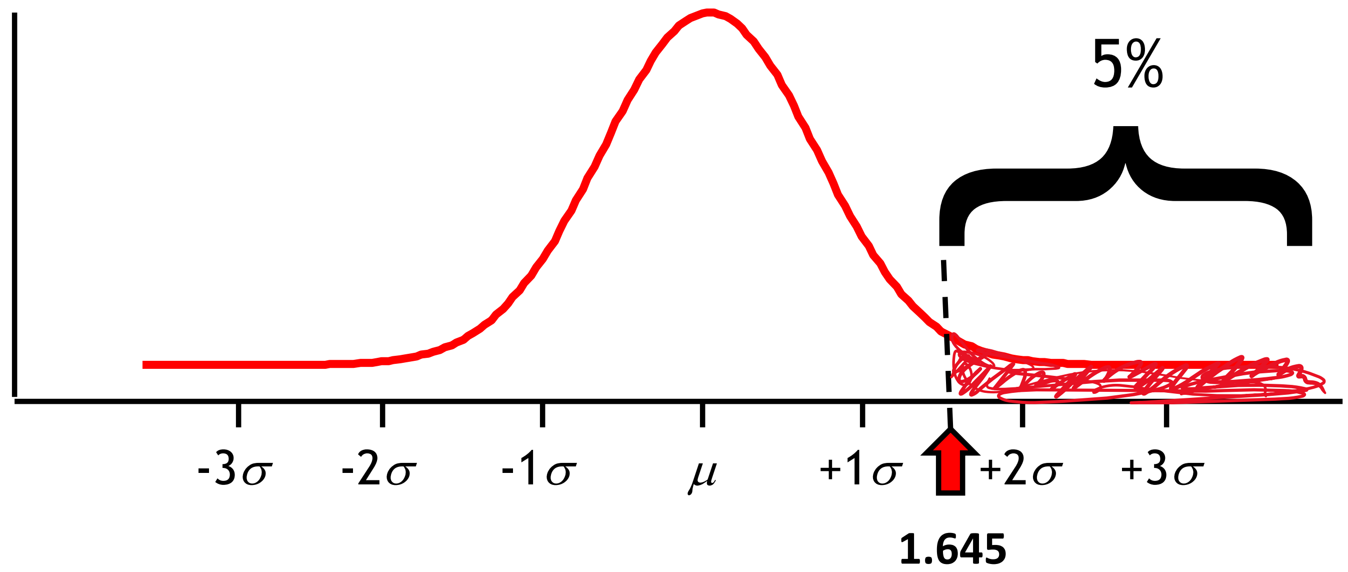
+1.645

Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681

Example 1

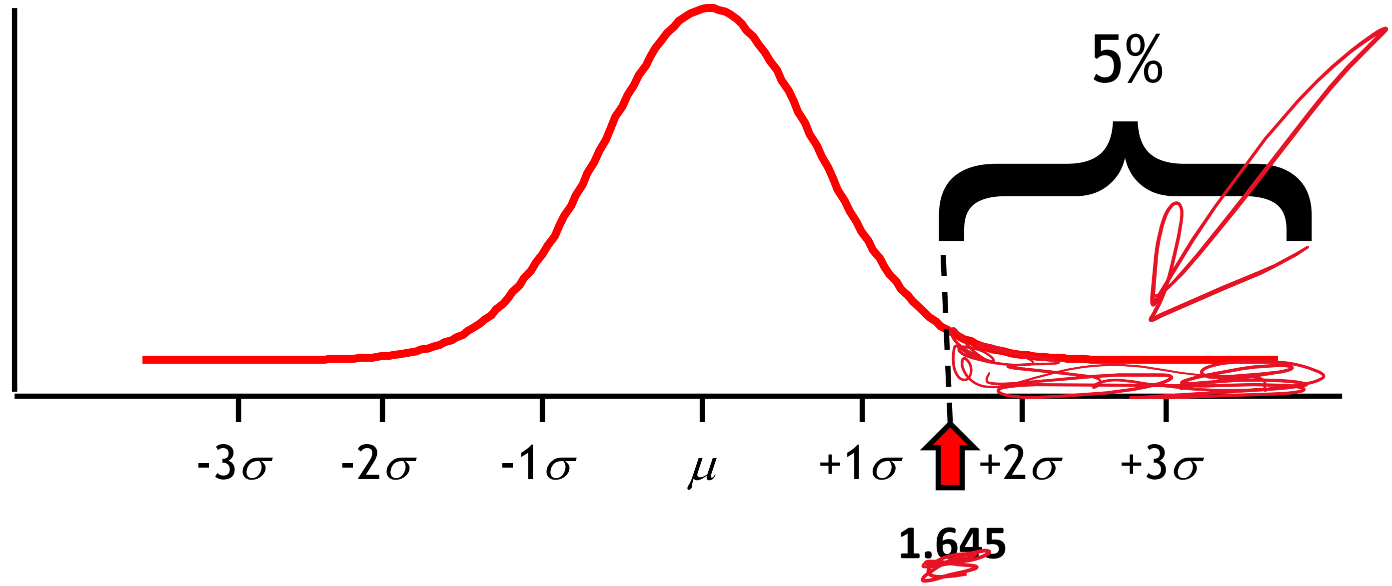
With α=0.05 and a one-sided test …



… a Z-score that is 1.645 in absolute value has 5% of the area under the curve beyond it

Example 1

We are using α=0.05 and a one-sided test …



… so our critical value Z* is +1.645

H_0: p ≤ 0.95 … fail to reject H_0 if test statistic Z ≤ 1.645

H_1: p > 0.95 … reject H_0 if test statistic Z > 1.645

Example 1

5. Calculate the test statistic

All test statistics for hypothesis testing…

…subtract the population value that is assumed to be true under the null from the observed (sample) value

…divide that figure by the standard deviation of the sampling distribution of the statistic in question

For proportions:

$$Z = \frac{\text{sample estimate} - \text{null value}}{\text{null standard deviation}} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$$

$p_0 = p_{null}$

Example 1

In our example:

H_0: $p \leq 0.95$

H_1: $p > 0.95$

$$n = 1{,}000$$

$$\hat{p} = \frac{962}{1000} = .962$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{N}}} = \frac{.962 - .95}{\sqrt{\frac{.95(1 - .95)}{1000}}} = 1.74$$

Example 1

In our example:

H_0: $p \leq 0.95$

H_1: $p > 0.95$

$$n = 1{,}000$$

$$\hat{p} = \frac{962}{1{,}000} = 0.962$$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1 - p_0)}{N}}} = \frac{0.962 - 0.950}{\sqrt{\frac{0.950(1 - 0.950)}{1{,}000}}} = 1.74$$

Example 1

6. Compare the test statistic to the critical value

- If the test statistic is larger than the critical value, then reject H_0 (with probability of α of doing so even though H_0 should not actually be rejected)
- If the test statistic is less than or equal to the critical value, then do no reject H_0 (with probability of β of doing so even though H_0 should be rejected)

In our example, our critical value Z* is +1.645

We obtained a test statistic of 1.74

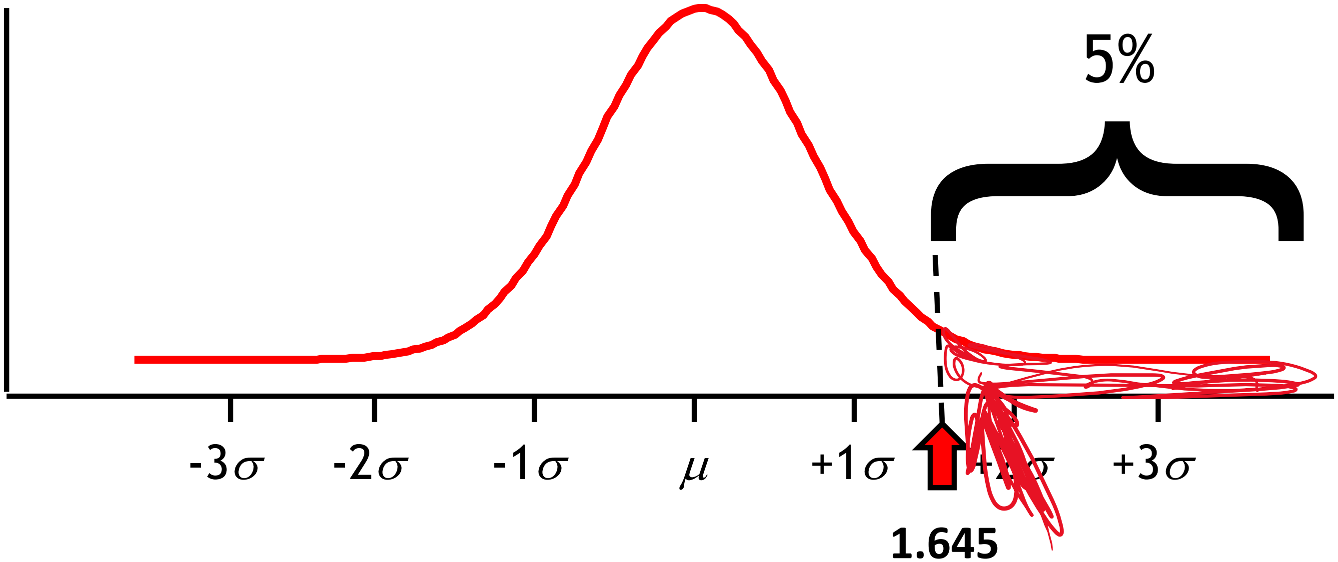
H_0: $p \leq 0.95$ … fail to reject H_0 if test statistic $Z \leq 1.645$

H_1: $p > 0.95$ … reject H_0 if test statistic $Z > 1.645$

Conclusion? Reject H_0

Example 1

We are using α=0.05 and a one-sided test …



… so our critical value Z* is +1.645

H_0: p ≤ 0.95 … fail to reject H_0 if test statistic Z ≤ 1.645

H_1: p > 0.95 … reject H_0 if test statistic Z > 1.645

Example 1

In 2007 the U.S. National Transportation Safety Board set a 5-year goal of having more than 95% of all American drivers use their seatbelts

> To see whether they are on target for meeting that goal, they randomly sampled 1,000 American drivers in 2012
>
> They found that 962—or 96.2%—of the 1,000 drivers they sampled use their seatbelts

Is this evidence sufficient to confidently conclude that more than 95% of <u>all</u> American drivers use their seatbelts?

YES

Example 2

A veterinarian claims that 6% of cats have FIDS (Feline Immune Deficiency Syndrome)

- To evaluate this claim, researchers randomly sampled 320 cats
- They found that 26—or 8.1%—of the 320 cats have FIDS

Is this evidence sufficient to confidently conclude that the population proportion of cats who have FIDS is different from 0.06?

Example 2

1. State the null (H_0) and alternative (H_1) hypotheses

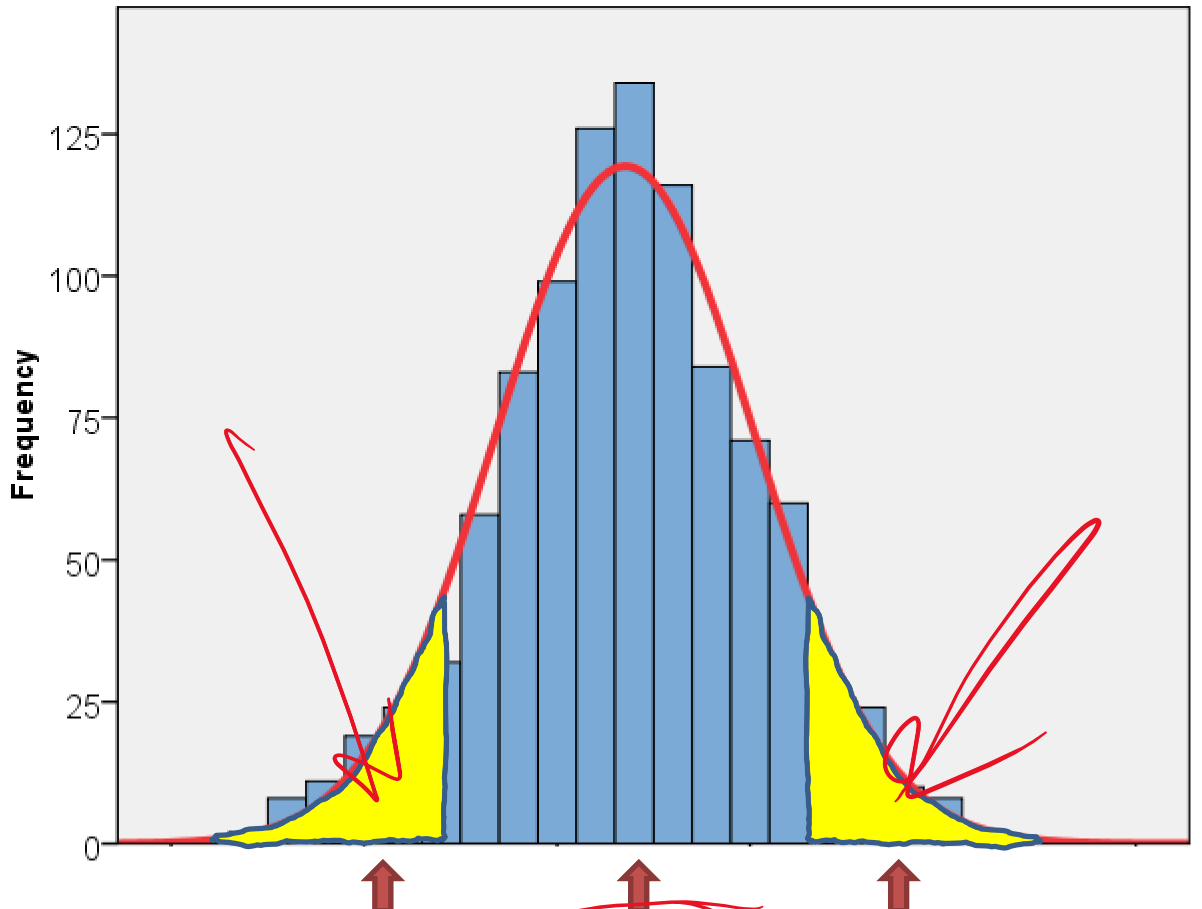
Because we are asking whether the proportion is different from a particular value (regardless of whether it is higher or lower than that value) this is a two-sided hypothesis test

H_0: $p = 0.06$

H_1: $p \neq 0.06$

One-sided (upper tail) hypothesis test

Assuming that p=0.95 or less, what is the probability of observing one sample proportion that far **above** p?



If $\hat{p}$ is here, reject H_0 p=0.06 If $\hat{p}$ is here, reject H_0

Example 2

2. Check that the sample data conform to basic assumptions; if they do not, then do not go any further

The sample must have been randomly selected from the population … (True)

np_0 and $n(1-p_0)$ must both be at least 10 (where p_0 is the proportion we assume to be correct based on the null hypothesis)

$np_0=320\times0.06=19.2$ and $n(1-p_0) = 320\times0.94=300.8$

(True)

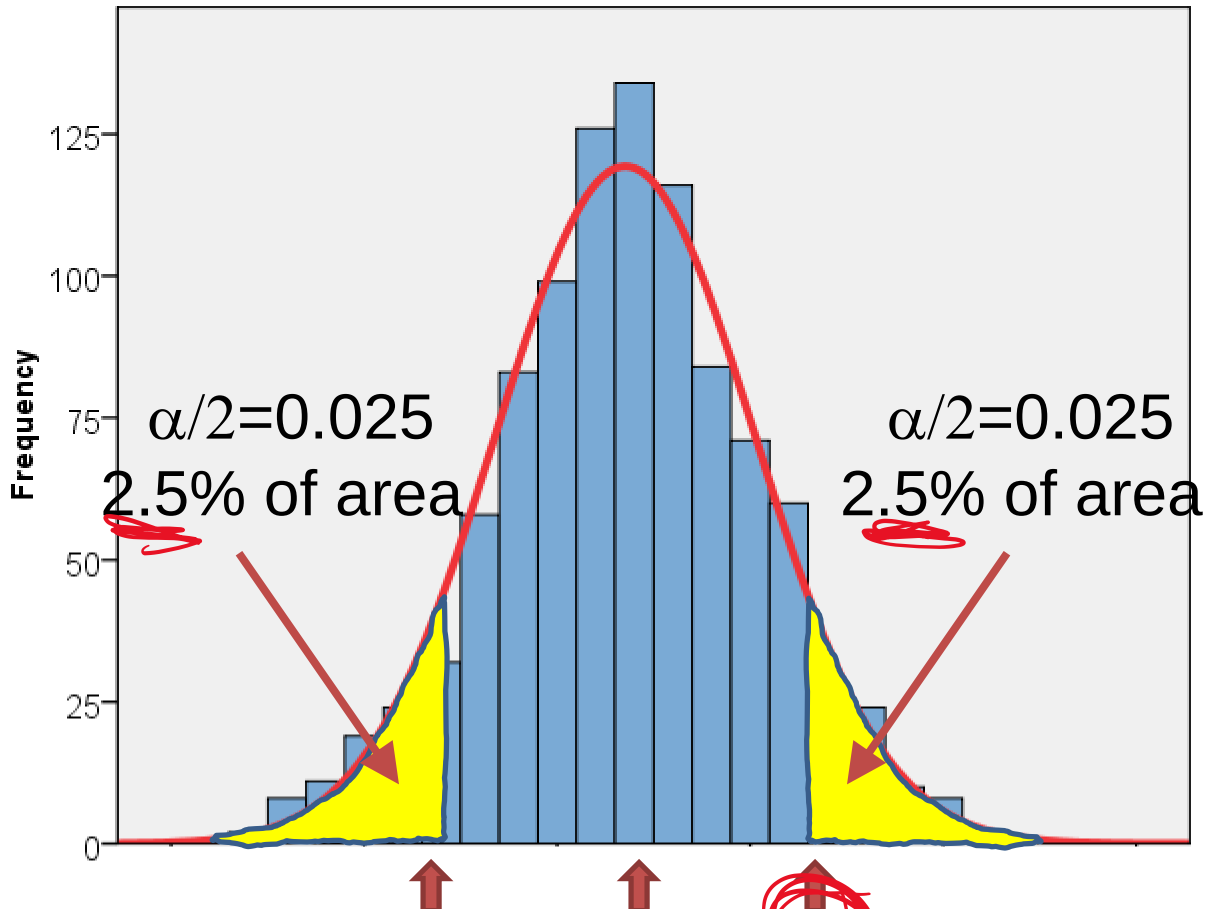
Example 2

3. Choose an α probability level … that is, a probability associated with incorrectly rejecting the null hypothesis

Let’s go with α=0.05 in this example

One-sided (upper tail) hypothesis test

Assuming that p=0.95 or less, what is the probability of observing one sample proportion that far **above** p?



If $Z<-Z^*$, reject H_0 $-Z^*$ **p=0.06** Z^* If $Z>Z^*$, reject H_0

Example 2

4. Determine the "critical value" … that is, how large the test statistic must be in order to reject the null hypothesis at the given α level

What critical value Z* must we choose so that 5%/2=2.5% of the area under the normal curve lies above +Z* and so that 5%/2=2.5% of the area under the normal curve lies below –Z*?

1.96

Standard Normal Probabilities

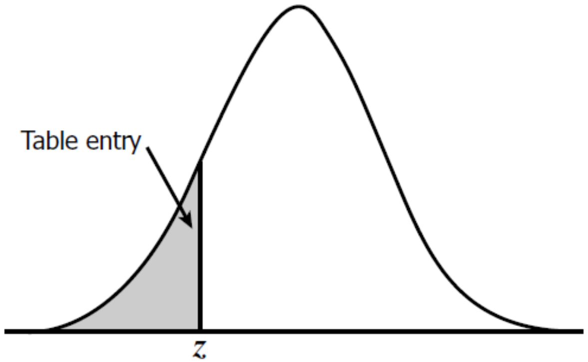
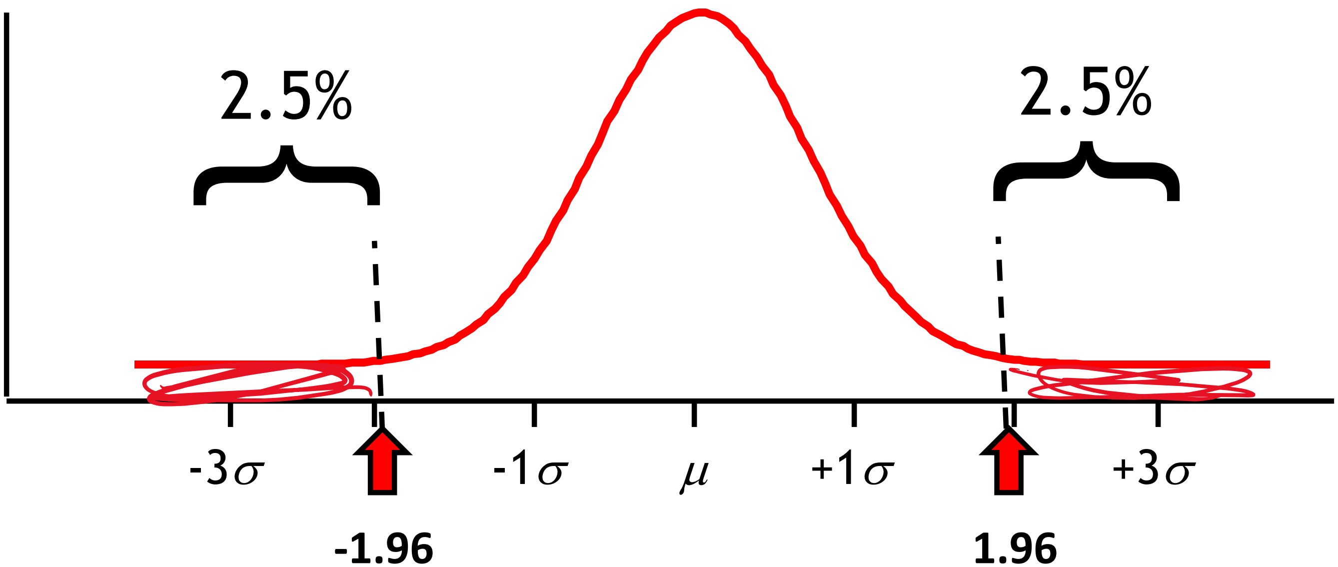


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681

Example 2

With α=0.05 and a two-sided test …



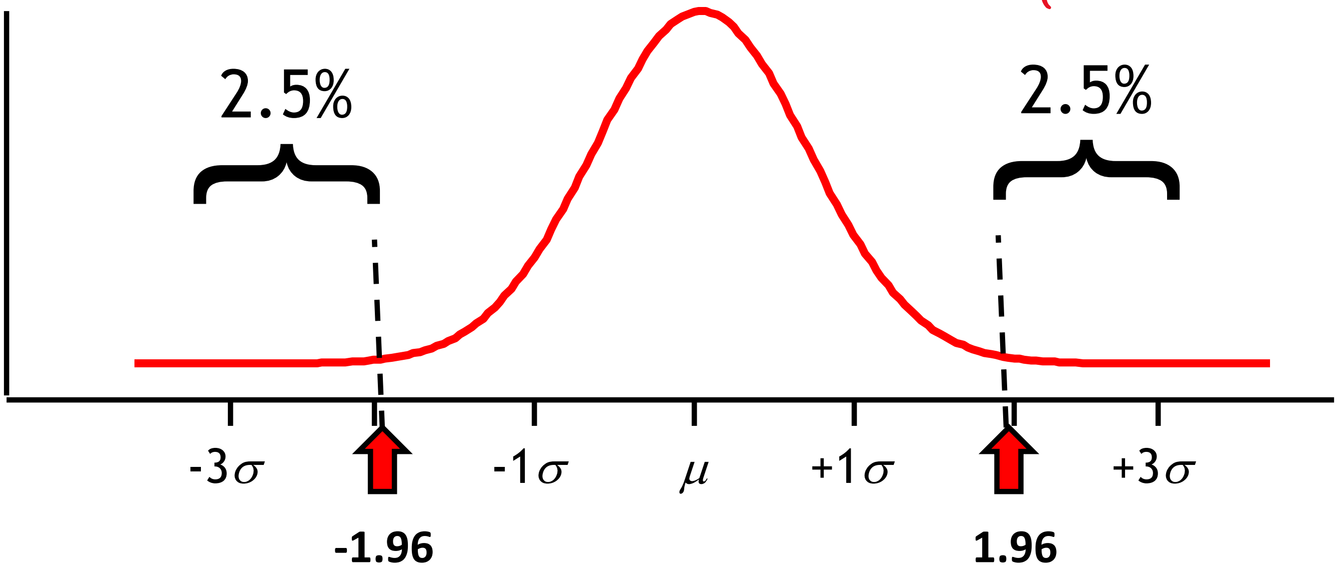
… a Z-score that is 1.96 in absolute value has 2.5% of the area under the curve beyond it in both directions

Example 2

$|2| = 2$

$|-2| = 2$

With α=0.05 and a two-sided test …



… so our critical value Z* is 1.96

H_0: p = 0.06 … fail to reject H_0 if $|Z| \leq 1.96$

H_1: p ≠ 0.06 … reject H_0 if $|Z| > 1.96$

Example 2

In our example:

$H_0: p = 0.06$

$H_1: p \neq 0.06$

$n = 320$

$\hat{p} = \frac{26}{320} = .081$

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.081 - .06}{\sqrt{\frac{.06(.94)}{320}}} = 1.601$$

Example 2

In our example:

H_0: $p = 0.06$

H_1: $p \neq 0.06$

$$\mathrm{n} = 320$$

$$\hat{\mathrm{p}} = \frac{26}{320} = 0.08125$$

$$\mathrm{Z} = \frac{\hat{\mathrm{p}} - \mathrm{p}_0}{\sqrt{\frac{\mathrm{p}_0(1 - \mathrm{p}_0)}{\mathrm{n}}}} = \frac{0.08125 - 0.06}{\sqrt{\frac{0.06(1 - 0.06)}{320}}} = 1.601$$

Example 2

6. Compare the test statistic to the critical value

In our example, our critical value Z* is 1.96

We obtained a test statistic of 1.601

H_0: $p = 0.06$ … fail to reject H_0 if $|Z| \leq 1.96$

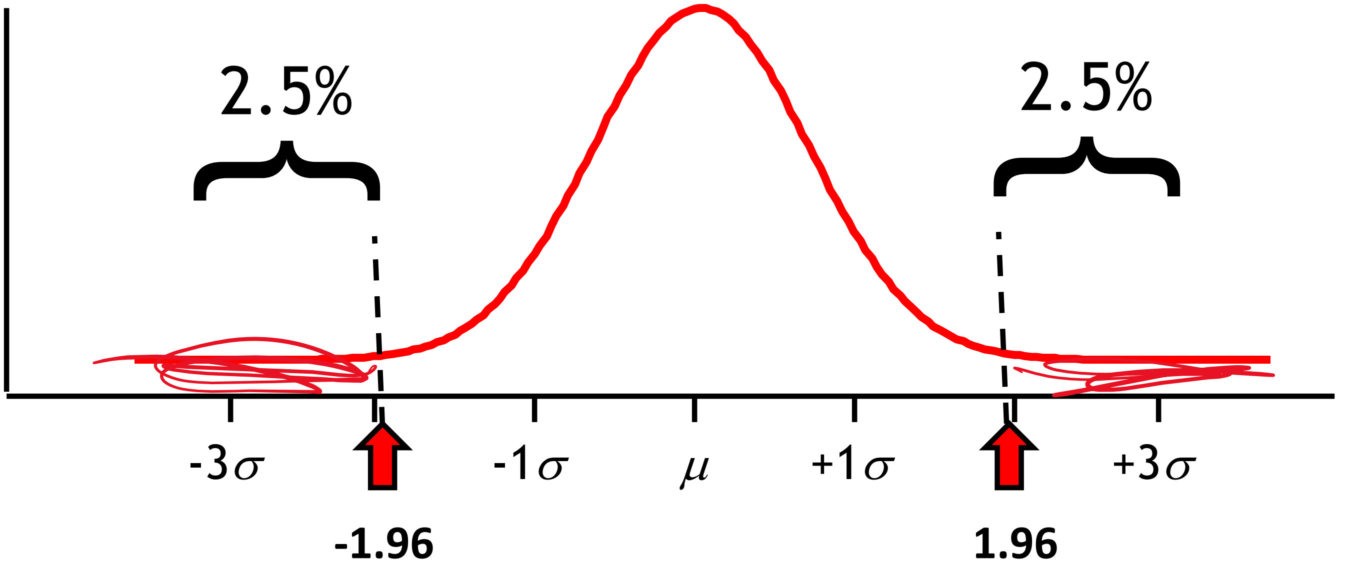
H_1: $p \neq 0.06$ … reject H_0 if $|Z| > 1.96$

Conclusion? Fail to Reject H_0

(But what if we had used α=0.10?)

Example 2

With α=0.05 and a two-sided test …



… so our critical value Z* is 1.96

H_0: p = 0.06 … fail to reject H_0 if $|Z| \leq 1.96$

H_1: p ≠ 0.06 … reject H_0 if $|Z| > 1.96$

Example 2

A veterinarian claims that 6% of cats have FIDS (Feline Immune Deficiency Syndrome)

- To evaluate this claim, researchers randomly sampled 320 cats
- They found that 26—or 8.1%—of the 320 cats have FIDS

Is this evidence sufficient to confidently conclude that the population proportion of cats who have FIDS is different from 0.06?

NO

Worksheet

I think that 5% of people in Minneapolis own RVs

To see whether this is true, I randomly sampled 2,310 people in Minneapolis and asked them whether they own an RV

I found that 110—or 4.76%—of the 2,310 people own RVs

Is this evidence sufficient to confidently (use α=0.01) conclude that the population proportion of people in Minneapolis who own RVs is different from 0.05?

Worksheet

I think the poverty rate is more than 20%

To see whether this is true, I randomly sampled 1,100 people and determined whether they were below the poverty line

I found that 242—or 22.0%—of the 1,100 people were poor

Is this evidence sufficient to confidently (use α=0.05) conclude that the population proportion of people who are poor is more than 0.20?

Errors in Hypothesis Testing

When we test a hypothesis, we can come to one of two conclusions

Reject the null hypothesis (H_0)

Fail to reject the null hypothesis (H_0)

Regardless of what we decide, we risk making an error

If we reject H_0, it could be that H_0 is actually true … such that we should not have rejected it

If we fail to reject H_0, it could be that H_0 is actually false … such we should have rejected it

Errors in Hypothesis Testing

A drug company has developed a new drug that they think will reduce rates of depression

The drug company will conclude that its new drug is effective if fewer than 10% of people who take it suffer from depression after taking the drug

To determine whether the new drug is effective the drug company might thus test the hypotheses:

$H_0: p \geq 0.10$

$H_1: p < 0.10$

Errors in Hypothesis Testing

The drug company risks making one of two errors:

First, the company could conclude that the drug works when in fact it does not … this is called a <u>Type I</u> error

They would reject the null when in fact the null is true

Consequence: They mistakenly claim that they have an effective drug to combat depression

Errors in Hypothesis Testing

The drug company risks making one of two errors:

Second, the company could conclude that the drug does not work when in fact it does … this is a Type II error

They would fail to reject H_0 when H_0 should be rejected

Consequence: They withhold an effective depression treatment from people who might benefit from it

Errors in Hypothesis Testing

		THE TRUTH IS…	
		H_0 is True	H_0 is False
WE CONCLUDE…	Fail to Reject H_0	✓	✗ Type II Error P(Type II Error)=β
	Reject H_0	✗ Type I Error P(Type I Error)=α	✓

Want More?

David Lane's Books

- http://onlinestatbook.com/2/logic_of_hypothesis_testing/logic_hypothesis.html
- http://davidmlane.com/hyperstat/logic_hypothesis.html

Lowry's Book (Chapter 7)

- http://vassarstats.net/textbook/

Dallal's Book

- http://www.jerrydallal.com/LHSP/sigtest.htm

Stat Trek's Discussion

- http://stattrek.com/hypothesis-test/hypothesis-testing.aspx