In hypothesis testing we wish to make and then test a statement about a population.

The population may be a collection of people, apples, screws, soda cans, airplane arrivals, etc.

. . . and we might be interested in . . .

Data Values or *Proportions*

Chapter 10 - Introduction to Hypothesis Testing

Data Values

Family incomes in Nassau County

Blood pressures of people in sales jobs

Lengths of screws manufactured by

Weights of people in Manhattan

Arrival times of airplanes at JFK

Students' test scores at NCC

Proportions

Proportion of women who use perfume

Proportion of students who own Toyotas

Proportion of people who favor the death penalty

Proportion of "Family Guy" watchers who are between 25 and 35 years old.

Proportion of NY Times readers who earn more than $150,000 / year.

Chapter 10 - Introduction to Hypothesis Testing

In hypothesis testing, we will want to make a statement about a population data value and then test it.

Statements involving Data Values

Freshmen students at NCC are taller than the average of U.S. freshmen.

People between 5 and 6 feet tall in Manhattan weigh less than the national average for that height range.

Executives in sales have higher blood pressures than other executives.

NCC students had higher SAT scores than students at Hofstra.

Flights arriving at JFK have longer delays than flights arriving at Newark.

Chapter 10 - Introduction to Hypothesis Testing

In hypothesis testing, we will want to make a statement about a population proportion and then test it.

Statements involving Proportions

A greater proportion of women between 30 and 40 use perfume than the proportion of women between 20 and 30.

The proportion of people in Tennessee who favor the death penalty is greater than the national proportion.

The proportion of students at NCC who own Toyotas is greater than the proportion of students at Suffolk.

The proportion of 25 – 35 year olds who watch "Family Guy" is higher than the proportion of 35 – 45 year olds.

The proportion NY Times readers who earn over $150,000/year is higher than the proportion of Daily News readers who earn over $150,000.

Hypothesis Testing: The Basic Idea

In order to prove any of the preceding statements, the procedure we use is very simple.

We will make two statements which are mutually exclusive. They are called
1. ***the Null Hypothesis H_0 and***
2. ***the Alternative Hypothesis H_a***

We will design a test with the goal of proving that H_0 can not be true.

A LIE

Hypothesis Testing: The Basic Idea

In order to prove any of the preceding statements, the procedure we use is very simple.

We will make two statements which are mutually exclusive. They are called
1. ***the Null Hypothesis H_0 and***
2. ***the Alternative Hypothesis H_a***

We will design a test with the goal of proving that H_0 can not be true.

I WIN

If we are successful, we will be able to conclude that H_a must be true.

As in all statistical tests, if we state that H_0 is not true, there is always a probability that we are wrong. If that probability is small enough, we can remain confident in our conclusion.

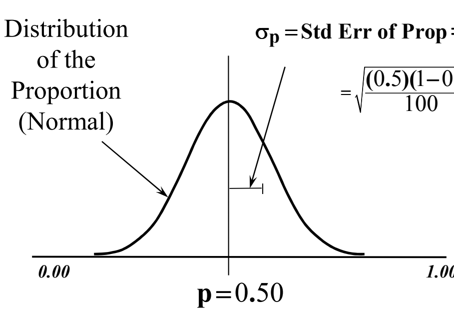
Hypothesis Test of Proportion

How can we test a coin to see if it is biased?

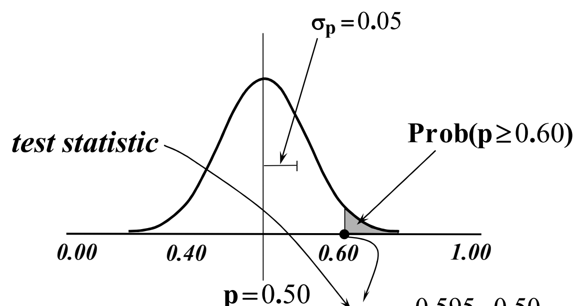
Ho: The coin is fair……P(Head) = 0.50
Ha: The coin is biased…P(Head) > 0.50

We assume (Null Hypothesis) that the coin is fair and we toss it 100 times expecting that the proportion of Heads will be not too different from 0.50.

$$\sigma_p = \text{Std Err of Prop} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.5)(1-0.5)}{100}} = 0.05$$



Suppose that the coin comes up Heads 60 times in 100 tosses?
(i.e. p = 60/100 = 0.60)
What is the probability that in 100 tosses, a fair coin would come up Heads with a proportion of 0.60 or more?



$$z_{0.60} = \frac{0.595 - 0.50}{0.05} = 1.9$$

$$\text{Prob}(p \geq 0.595) = \text{normalcdf}(1.9, E99) = 0.029$$

So there is only a 2.9% chance that a fair coin would come up Heads 60 or more times.

So there is only a 2.9% chance that fair coin would come up Heads 60 or more times.

Then what do we conclude about the coin?

Is it "fair" or "biased"?

If we conclude that the coin is "biased", what is the probability that we are wrong?

If we accuse the coin of being "biased" there is a 2.9% probability that we are wrong in doing so…

…because in fact, in 100 tosses, a "fair" coin <u>could</u> come up Heads 60 times or more.

So we…

Reject the null hypothesis
(and conclude the coin is biased)

because we are willing to be wrong with a probability 2.9%.

But ask yourself…

is a probability of being wrong of 2.9% small enough?

What if someone's life depended on it?

Consider the following.

The CDC (Center for Disease Control) says that of all the women in the US who had legal abortions in 2001, the proportion of women between the ages of 15 and 19 was 17.4%

$$p = 0.174$$

For that same year and age range, a sample of 15,000 NYC women resulted in a proportion of 16.3%

$$\hat{p} = 0.163$$

Can we conclude that the proportion of New York City women is less than the national average?

(Or could it be that the particular sample of women selected for NYC just happened to be lower? A different sample might have been higher.)

Consider what we know.

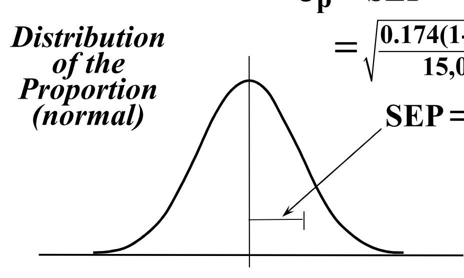
For the national group of women

$$p = 0.174$$

If a samples of size n = 15,000 were to be taken from that population, the sample proportions would distribute normally as shown.

$$\sigma_p = \text{SEP} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.174(1-0.174)}{15{,}000}}$$

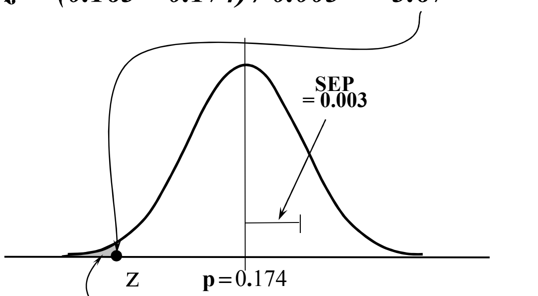
$$\text{SEP} = 0.003$$



$$p = 0.174$$

In the New York sample of women in that age range, the proportion was 0.163. What is the z-score of that proportion?

z = (0.163 – 0.174) / 0.003 = - 3.67



If the proportion of women in NYC were the same as the national average, what is the probability that a sample of NYC women would have a proportion with a

z-score of -3.67?

p-value = normalcdf(-E99,-3.67) = 0.00012

What does this mean?

If the proportion of NYC women in that category were typical of the national average, we would expect our sample of NYC women also to be about 0.174 … maybe plus or minus a standard error or two.

But since the NYC proportion was 3.67 standard errors <u>below</u> the national average, we reject the idea that the proportion of NYC women is the same as the national average. We conclude that for NYC, the proportion must be lower.

Why? Because, if the NYC proportion were the same as the national proportion, the probability that we would get a sample so far below the national mean proportion would be only 0.00012 . . . About 1 chance in 8333.

p-value = normalcdf(-E99,-3.67) = 0.00012

So we may reject the null hypothesis and conclude that for NYC women the proportion who have had abortions is less than the national average…

…and we may be confident that the probability that we are wrong in this conclusion is only 0.00012.

As small as it is, there is always this probability (p-value) that we are wrong when we reject the null hypothesis.

How small should the p-value be in order for us to reject the null hypothesis?

$$\alpha$$

α is the maximum probability of being wrong that you are willing to accept.

<u>Hypothesis Test of Mean</u>

<u>Commuting Time</u>

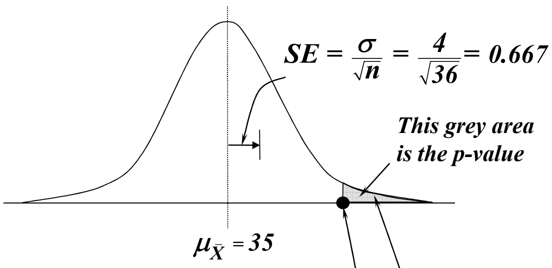
A commuter railroad claims that the mean commute time between two cities is 35 minutes with a standard deviation of 4 minutes.

A commuter riders group, suspecting that the trip takes longer, recorded the time of the trip from 36 randomly selected days during the year and found the mean trip time to be 37 minutes.

1. ***State the null and alternative hypotheses.***
2. ***Draw the sampling distribution of the mean.***
3. ***Indicate the test statistic.***
4. ***What is the probability that the sample mean trip time would be 37 minutes or more?***

Commuting Time

H_0: The mean commuter trip = 35 minutes.
Ha: The mean commuter trip > 35 minutes.



The sample mean trip time was 37 minutes.

So the Test Statistic, z is

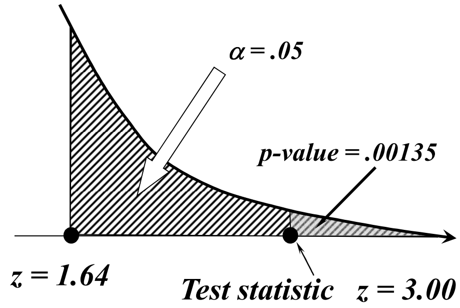
$$z = \frac{x - \mu}{\sigma} = \frac{37 - 35}{0.667} = 3.00$$

and the p-value is

$$\textbf{normalcdf}(3.00, \textbf{E}99) = 0.00135$$

Commuting Time

Let's take a closer look …



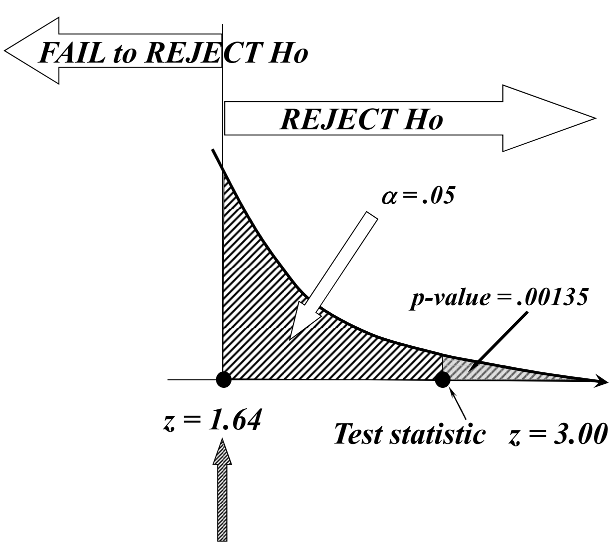
Recall, we only needed a p-value less than α = 5% (.05) to reject the null.

Our test statistic could have been only z = 1.64 and the p-value would have been less than the 5% we need to reject.

This z-score is called the "critical value," Because it tells us the smallest test statistic we need to reject the null hypothesis.

Commuting Time

So to add some labels …



Critical Value
(at which the test statistic is just big enough to reject at the given α.

Hypothesis Testing Terms

Null Hypothesis: Ho

A statement about the population parameter to be tested and always a statement of equality.

Alternative Hypothesis: Ha

An alternative statement to the Null hypothesis and always a statement of inequality, i.e. greater than, less than or not equal to.

The alternative hypotheses may be
- ***directional, i.e. imply direction "is greater than," or "is less than"***

or
- ***non-directional, "is not equal to."***

p-value: the probability of wrongly rejecting the null hypothesis.

α : the maximum probability of error you are willing to accept in rejecting H_0.

Hypothesis Testing Procedure

1: Formulate the Hypotheses: Ho and Ha.
Always trying to reject Ho.
If Ho is rejected it means Ha is true.
If we fail to reject Ho, we cannot conclude anything.

2: Determine the model to test the null.

3: Formulate the decision rule.
(i.e. if the p-value is less than α, reject H_0)

4: Analyze the sample data.

5: State the conclusion.

If the p-value is less than
0.01 (1%), the result is very significant
0.05 (5%), the result is significant.

If the p-value is greater than
0.10 (10%), the result is not significant.

Chapter 10 - Introduction to Hypothesis Testing

Pg 525/Ex 10.3 Null v. Alt - Directional
Pg 526/Ex 10.4 Null v. Alt - Non-Directional
Pg 527 Hypothesis Testing Procedure
1. **Formulate Hypotheses (Null & Alt)**
2. **Determine Model to Test Hypothesis**
3. **Formulate Decision Rule**
4. **Analyze Sample Data**
5. **State Conclusion**

Pg 528 - Definitions:
Type I Error - Reject a true null hypothesis
Type II Error - Fail to Reject a False null Hypothesis