Section 9.1 (Part 2 of 2)

Significance Tests:

- **Type 1 error, Type 2 error, and Power of a Test**

Learning Objectives

After this section, you should be able to…

✓ INTERPRET a Type I error and a Type II error in context, and give the consequences of each.

✓ DESCRIBE the relationship between the significance level of a test, *P*(Type II error), and power.

AP Stats Calculating Power, Type I and Type II Errors
WHAT YOU NEED TO KNOW !

Quote from AP Statistics Teacher Forum

"**Do not try to teach any calculations about Type II error or power**. Not only is that not required, it can be confusing and it distracts students from understanding the concepts. **They need to know what the two types of error are and what power is. They need to be able to explain them in the context of the questions. And they need to understand the interactions among the errors, power, sample size and effect size.** But **no calculations**!"

Type I and Type II Errors

When we draw a conclusion from a significance test, we hope our conclusion will be correct. But sometimes it will be wrong. There are two types of mistakes we can make. We can reject the null hypothesis when it's actually true, known as a **Type I error**, or we can fail to reject a false null hypothesis, which is a **Type II error**.

Definition:

If we reject H_0 when H_0 is true, we have committed a **Type I error**.

If we fail to reject H_0 when H_0 is false, we have committed a **Type II error**.

		Truth about the population	
		H_0 true	H_0 false (H_a true)
Conclusion based on sample	Reject H_0	**Type I error**	*Correct conclusion*
	Fail to reject H_0	*Correct conclusion*	**Type II error**

EXAMPLE #1: The O.J. Analogy for Understand Hypothesis Testing

Our jury system, you are innocent until proven guilty. This is how we set up our hypothesis statement

H_0 : p = O.J. innocent ← Null hypothesis. **We call H_o "H not."**
H_a : $p \neq$ O.J. guilty← Alternate hypothesis **This is called a 2-tail test. If we use > or <; it is a 1-tail test.**

The lawyers give evidence to prove their case (we will do the same).

- The jury comes back with the verdict. They do NOT say OJ is innocent. The jury says "NOT GUILTY (but we all know OJ was guilty)." They did not have enough evidence to convict O.J. *We will do the same… we NEVER accept the null hypothesis because we have a chance of making a mistake . If we do not have enough evidence, we "* ***Fail to reject H_o,*** *" (the null hypothesis).*
- Now if the jury had enough evidence **beyond a reasonable doubt.** The jury says "GUILTY." *We will do the same… If we have enough evidence, we "* ***REJECT Ho,*** *" the null hypothesis.*

Truth about the population (OJ)

Conclusion based on sample (trial evidence)		H_0 true OJ Not Guilty	H_0 false (H_a true) OJ Guilty
	Reject H_0 *OJ CONVICTED*	**Type I error (α)** **False Positive** *AN INNOCENT OJ CONVICTED*	*Correct conclusion* *A GUILTY OJ CONVICTED*
	Fail to Reject H_0 *OJ NOT CONVICTED*	*Correct conclusion* *AN INNOCENT OJ NOT CONVICTED*	**Type II error (β)** **False Negative** *A GUILTY OJ NOT CONVICTED*

Example #2 "Perfect Potatoes"

A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have "blemishes," the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips. To make the decision, a supervisor will inspect a random sample of potatoes from the shipment. The producer will then perform a significance test using the hypotheses

$$H_0 : p = 0.08$$
$$H_a : p > 0.08$$

where ***p* is the actual proportion of potatoes with blemishes in a given truckload**

Describe a Type I and a Type II error in this setting, and explain the consequences of each:

- A **Type I error** would occur if…

•*Consequence*:

- A **Type II error** would occur if…

•*Consequence*:

- **<u>Example</u>:** Perfect Potatoes
- ***Consequence of Type I or II Error***

• A **<u>Type I error</u>** would occur if the producer concludes that the proportion of potatoes with blemishes is greater than 0.08 when the actual proportion is 0.08 (or less).

•***Consequence*****: The potato-chip producer sends the truckload of acceptable potatoes away, which may result in lost revenue for the supplier.**

• A **<u>Type II error</u>** would occur if the producer does not send the truck away when more than 8% of the potatoes in the shipment have blemishes.

•***Consequence: The producer uses the truckload of potatoes to make potato chips. More chips will be made with blemished potatoes, which may upset consumers.***

■ **Example**: Perfect Potatoes
■ ***Measuring Type I or II Error***

We can assess the performance of a significance test by looking at the probabilities of the two types of error. That's because statistical inference is based on asking, "What would happen if I did this many times?"

For the truckload of potatoes we were testing (where *p* is the actual proportion of potatoes with blemishes).

$$H_0: p = 0.08$$
$$H_a: p > 0.08$$

Suppose that the potato-chip producer decides to carry out this test based on a random sample of 500 potatoes using a 5% significance level ($\alpha = 0.05$).

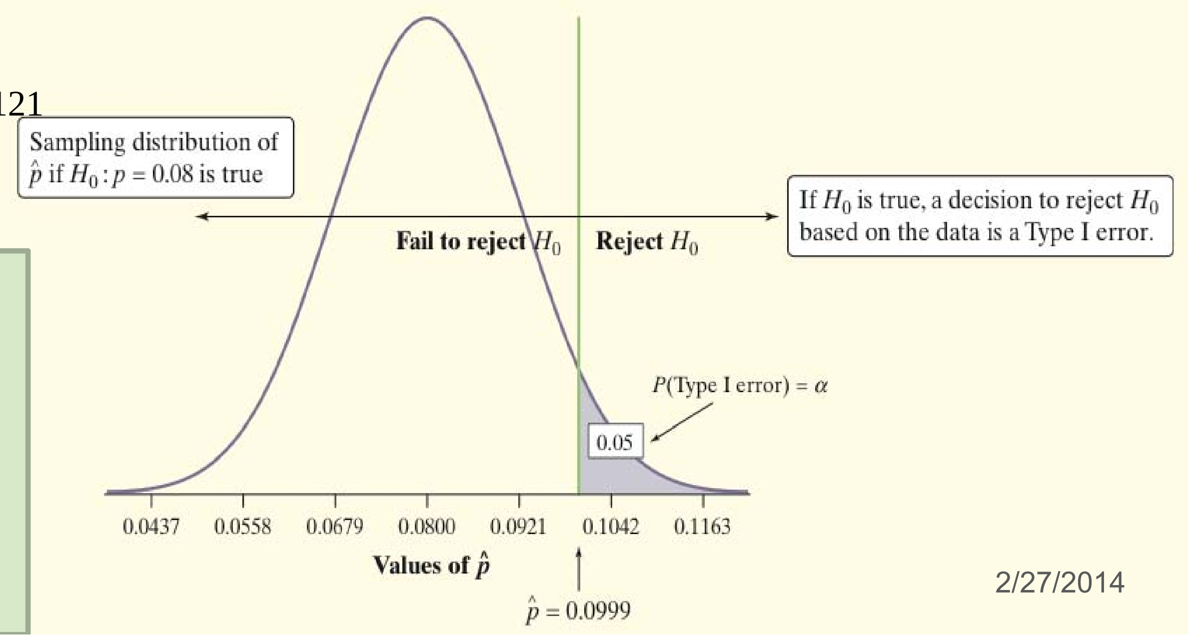
Assuming $H_0: p = 0.08$ is true, the sampling distribution of $\hat{p}$ will have:

Shape: Approximately Normal because 500(0.08) = 40 and 500(0.92) = 460 are both at least 10.

Center: $\mu_{\hat{p}} = p = 0.08$

Spread: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.08(0.92)}{500}} = 0.0121$

The shaded area in the right tail is 5%. Sample proportion values to the right of the green line at 0.0999 will cause us to reject H_0 even though H_0 is true. This will happen in 5% of all possible samples. That is, *P*(making a Type I error) = 0.05.



Error Probabilities

The probability of a Type I error is the probability of rejecting H_0 when it is really true. As we can see from the previous example, this is exactly the significance level of the test.

Significance and Type I Error

The significance level α of any fixed level test is the probability of a Type I error. That is, α is the probability that the test will reject the null hypothesis H_0 when H_0 is in fact true. Consider the consequences of a Type I error before choosing a significance level.

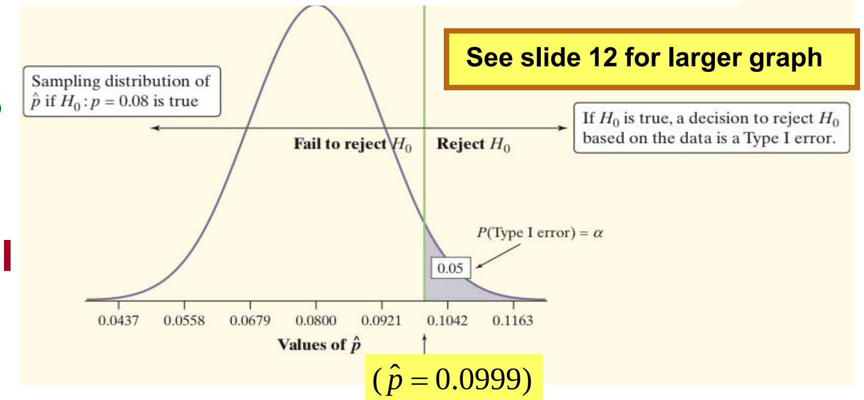
What about Type II errors? A significance test makes a Type II error when it fails to reject a null hypothesis that really is false. There are many values of the parameter that satisfy the alternative hypothesis, so we concentrate on one value. We can calculate the probability that a test *does* reject H_0 when an alternative is true. This probability is called the **power** of the test against that specific alternative.

Definition:

The **power** of a test against a specific alternative is the probability that the test will reject H_0 at a chosen significance level α when the specified alternative value of the parameter is true.

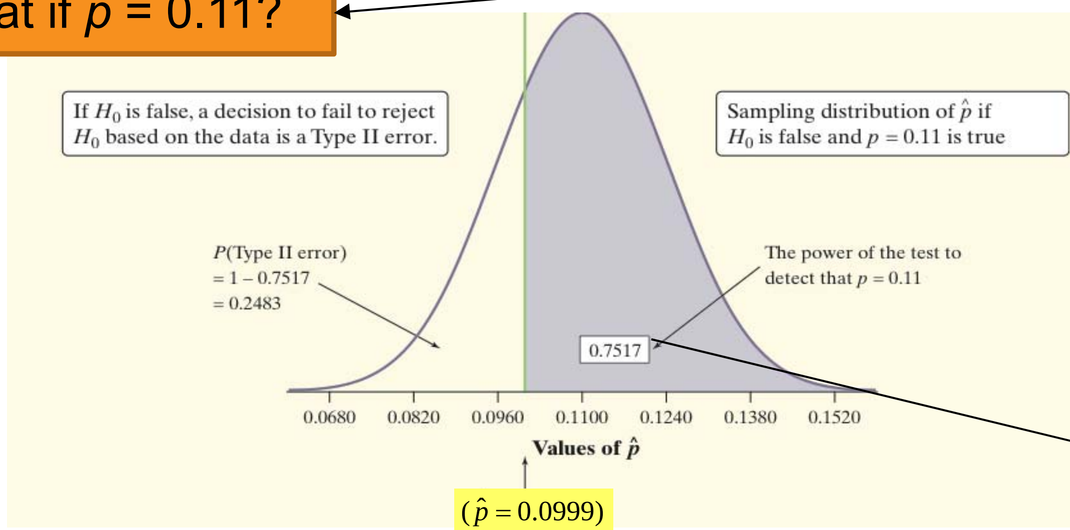
- **Example**: Perfect Potatoes
- ***Error Probabilities***

The potato-chip producer wonders whether the significance test of $H_0: p = 0.08$ versus $H_a: p > 0.08$ based on a random sample of 500 potatoes has enough power to detect a shipment with, say, 11% blemished potatoes. In this case, a particular Type II error is to fail to reject $H_0: p = 0.08$ when $p = 0.11$.



Earlier, we decided to reject H_0 at $\alpha = 0.05$ if our sample yielded a sample proportion to the right of the green line.

What if $p = 0.11$?



Since we reject H_0 at $\alpha = 0.05$ if our sample yields a proportion > 0.0999, we'd correctly reject the shipment about 75% of the time.

Power and Type II Error

The power of a test against any alternative is 1 minus the probability of a Type II error for that alternative; that is, power = 1 - β.

Planning Studies: The Power of a Statistical Test

How large a sample should we take when we plan to carry out a significance test? The answer depends on what alternative values of the parameter are important to detect.

Here are the questions we must answer to decide how many observations we need:

1.**_Significance level._** How much protection do we want against a Type I error — getting a significant result from our sample when H_0 is actually true?

2.**_Practical importance._** How large a difference between the hypothesized parameter value and the actual parameter value is important in practice?

3.**_Power._** How confident do we want to be that our study will detect a difference of the size we think is important?

Summary of influences on the question "How many observations do I need?"

- If you insist on a smaller significance level (such as 1% rather than 5%), you have to take a larger sample. A smaller significance level requires stronger evidence to reject the null hypothesis.
- If you insist on higher power (such as 99% rather than 90%), you will need a larger sample. Higher power gives a better chance of detecting a difference when it is really there.
- At any significance level and desired power, detecting a small difference requires a larger sample than detecting a large difference.

Power of a Test:

Summary

- A **Type I error** occurs if we reject H_0 when it is in fact true. **A Type II error** occurs if we fail to reject H_0 when it is actually false. In a fixed level α significance test, the probability of a Type I error is the significance level α.
- The power of a significance test against a specific alternative is the probability that the test will reject H_0 when the alternative is true. **Power** measures the ability of the test to detect an alternative value of the parameter. For a specific alternative, P(Type II error) = 1 - power.

APPENDIX:

Video #1 AP Stats Type I and Type II Errors

YouTube by Jerry Linch

http://www.youtube.com/watch?v=71-tKPgDEQU&feature=share&list=PLEYFL88U8S9H-QN_y0qHQ4hc3ZJC64h1o

- Take Notes on the YouTube Video
  - Potato example is very good at 23:44min
  - Skip Examples 26:55 to 31:45

Video #2 – AP Stats Calculating Power, Type I & Type II Errors

YouTube by Jerry Linch

http://www.youtube.com/watch?v=K8XwS74sSko&feature=share&list=PLEYFL88U8S9H-QN_y0qHQ4hc3ZJC64h1o

1) Take Notes on the YouTube Video

2) Here are the examples that will be reviewed:

A researcher selects a random sample of size 49 from a population with standard deviation $\sigma = 35$ in order to test at the 1% significance level the hypothesis:

H_0: $\mu = 680$

H_a: $\mu > 680$

What is the probability of committing a Type I error?

Bottles of a popular cola are suppose to contain 300 ml of cola. A consumer group believes the company is under-filling the bottles. (Assume $\sigma = 50$ with $n = 30$)

Find the power of this test against the alternative $\mu = 280$ ml. (Assume $\alpha = .05$)

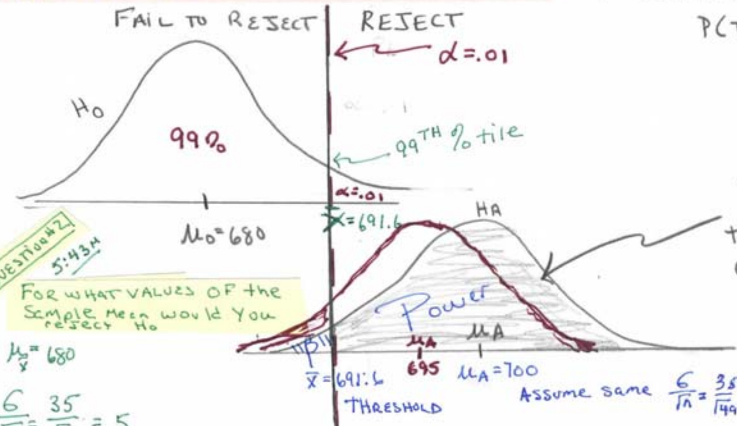
3) The following slides provide key points

A researcher selects a random sample of size 49 from a population with standard deviation $\sigma = 35$ in order to test at the 1% significance level the hypothesis:

$H_0: \mu = 680$

$H_a: \mu > 680$

(Q#1) What is the probability of committing a Type I error? → DEFINED BY SIGNIFICANCE LEVEL $P(\text{TYPE I ERROR}) = \alpha = .01$



Power is a measure that we made the correct decision, we reject Ho in favor of HA

Assume same $\frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{49}} = 5$

QUESTION #2 5:43M

FOR WHAT VALUES OF the Sample Mean would You reject Ho

$\mu_{\bar{x}} = 680$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{35}{\sqrt{49}} = 5$

$\bar{x}_{.99} = \text{InvNorm}(.99, 680, 5) = 691.632$

ANSWER we would reject Ho if the Sample mean ($\bar{x}$) WAS Greater than 691.632.

QUESTION #4 (15:59min) Suppose $\mu_a = 695$ $\alpha = .01$ $\beta =$ Power =

- SHIFT HA TO LEFT
- NOTICE α STAYS THE SAME $\alpha = .01$
- β GETS LARGER = $P(\text{Type II}) = \text{Normalcdf}(-E99, 691.63, 695, 5)$ (μ_A)

$\beta = .2502$

Power = .7498

QUESTION #3 (10:33MIN)

IF Ho is REJECTED, suppose that $\mu_a = 700$.

What is the probability of commiting a Type II ERROR

$\beta = \text{Normalcdf}(-\infty, 691.632, 700, 5) = .047$ (μ_A)

$\beta = .047$ $\alpha = .01$

WHAT IS THE POWER OF THE TEST?

Power $= 1 - \beta = 1 - .047 = .953$

$\alpha = P(\text{Type I}) = .001$

$\beta = P(\text{Type II}) = .047$

Power = .953

Bottles of a popular cola are suppose to contain 300 ml of cola. A consumer group believes the company is under-filling the bottles. (Assume σ = 50 with n = 30)

Find the power of this test against the alternative μ = 280 ml. (Assume α = .05)

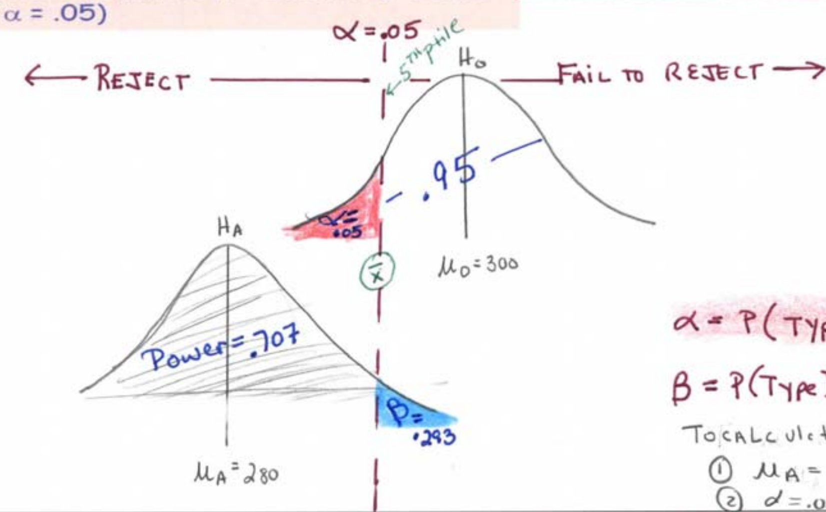
$H_0: \mu = 300ml$ $n = 30$

$H_A: \mu < 300ml$ $\sigma = 50$ (z inference)

FIND POWER

Assume $\mu_A = 280ml$

$\alpha = .05$



α = P(Type I error) = .05

β = P(Type II error)

To calculate β must be given
① $\mu_A = 280$ and
② $\alpha = .05$ (the 5TH %tile)

$\bar{X}$ = InvNorm(.05, 300, $50/\sqrt{30}$)

$\bar{X} = 284.985ml$ (This is considered to be an extreme value. We would Reject H_0 when $\bar{x} < 284.985$)

β = normalcdf(284.985, E99, 280, $50/\sqrt{30}$)

$\beta = .293$

Power = $1 - \beta = 1 - .293 = .707$ (Power ≥ .80 is Considered Good)

Conclusion: Best way to reduce errors and increase power is to increase sample size

n ↑	B ↓
Set α	Power ↑

7