Section 9.1 (Part 2) (pp. 547-551)

Type I and Type II Errors

Because we are basing our conclusion in a significance test on sample data, there is always a chance that our conclusions will be in error. It turns out that there are two types of errors that can be committed in a significance test:

- If we reject H_0 when H_0 is true, it is called a **Type I Error**
- If we fail to reject H_0 when H_0 is false, it is called a **Type II Error**

Memory aid: Type II error occurs when we fail **II** reject the null hypothesis when we really should.

Example - The manager of a fast-food restaurant wants to reduce the proportion of drive-through customers who have to wait for more than 2 minutes to receive their food once their order is placed. Based on store records, the proportion of customers who had to wait at least 2 minutes was $p = 0.63$. To reduce this proportion, the manager will collect a random sample of drive-through times and test the following hypotheses:

$H_0: p = 0.63$
$H_A: p < 0.63$

Where p is the true proportion of drive-through customers who have to wait more than 2 minutes after their order is placed to receive their food. Describe a Type I and a Type II error in this setting and explain the consequences of each.

Application - In the United States court system, defendants are "presumed innocent" and to be proven guilty, it must be done "beyond a shadow of a doubt."

In this situation, what constitutes a Type I error?

What constitutes a Type II error?

Which error is worse?

Error Probabilities - We can assess the performance of a significance test by looking at the probabilities of the two types of errors.

Example (cont) - In the fast-food example we were testing the following hypotheses

H_0: $p = 0.63$
H_A: $p < 0.63$

where *p* = the true proportion of drive-through customers who have to wait more than two minutes after their order is placed to receive their food. Suppose that the manager decided to carry out this test using a random sample of 250 orders with a significance level of $\alpha = 0.10$. What is the probability of making a Type I error?

A Type I error is to reject H_0 when _______________________________ .

If our sample results in a $\hat{p}$ that is much smaller than 0.63, we will ______________________ .

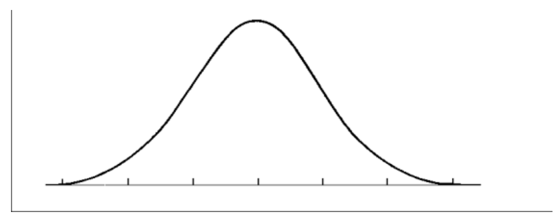
How small would $\hat{p}$ have to be? The 10% significance level tells us to count results that could happen less than 10% of the time by chance if ____.

Let's assume H_0: $p = 0.63$ is true. Then the sampling distribution of $\hat{p}$ will have:

Shape:

Center:

Spread:



From this example, what can we conclude about the probability of a Type I error?

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.

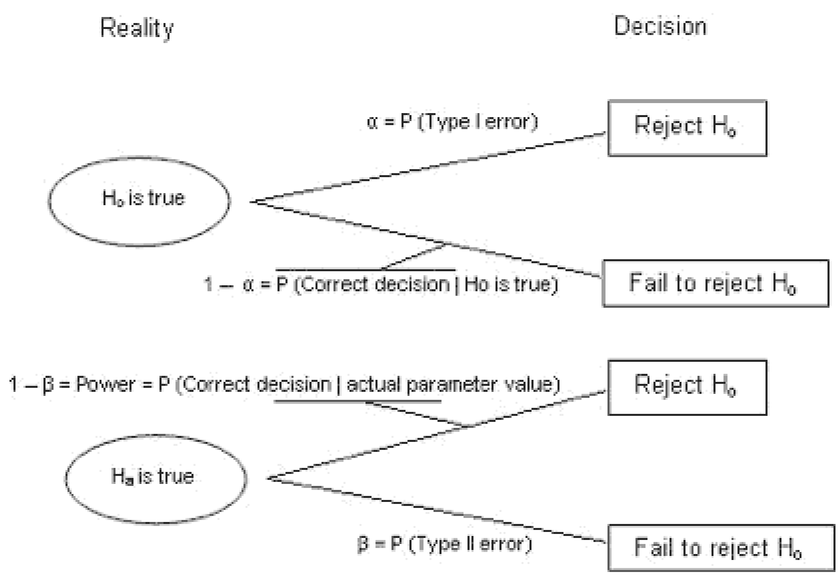
$$P(\text{Type I error}) = P(\text{Reject } H_0 \mid H_0 \text{ is true}) = \alpha$$

What about Type II errors?

A significance test makes a Type II error when ____ .

A high probability of a Type II error for a particular alternative hypothesis means that the test is not sensitive enough to usually detect the alternative. Rather than being concerned with the probability of failing to reject H_0 when H_0 is false we really want the probability of "getting it right." In other words, we want the probability of rejecting H_0 when H_0 is in fact false. This probability (of "getting it right") is referred to as the **power** of a significance test.

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. The **power** of a test against any alternative is *1 minus the probability of a Type II error for that alternative*, that is Power = $1 - \beta$ = P(Rejecting H_0 | H_A is true)



There are many ways to explain *power*:

- Power is the probability of rejecting the null hypothesis when in fact it is false.
- Power is the probability of making a correct decision (to reject the null hypothesis) when the null hypothesis is false.
- Power is the probability that a test of significance will pick up on an effect that is present.
- Power is the probability that a test of significance will detect a deviation from the null hypothesis, should such a deviation exist.
- Power is the probability of avoiding a Type II error.

Planning Studies: The Power of a Statistical Test - How large a sample should we take to carry out a significance test? The answer depends on what alternative values of the parameter are important to detect.

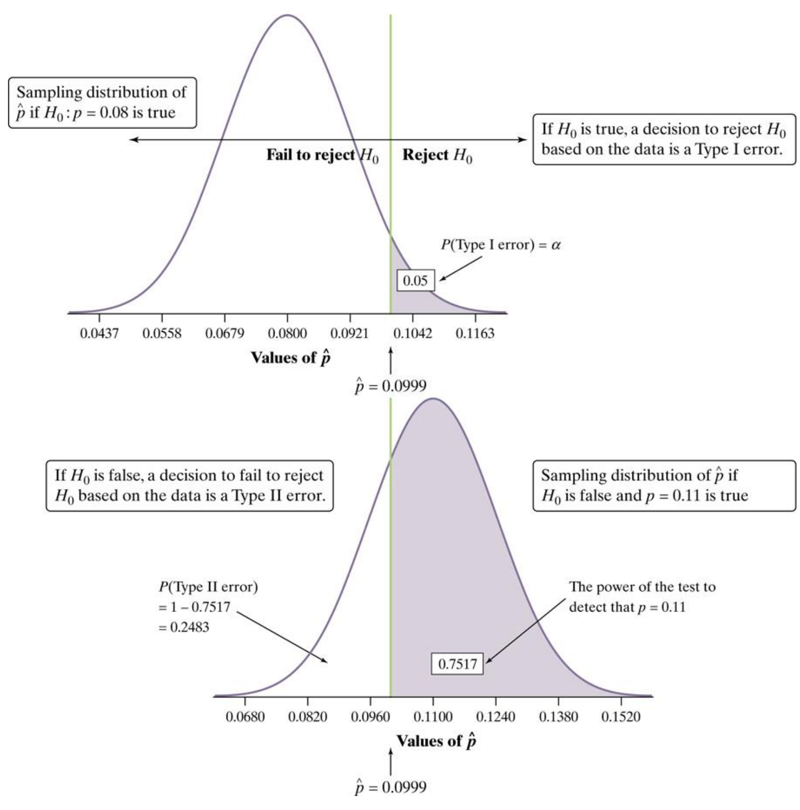
There are basically *four* things that affect power:

(1) The *significance level* α of the test.

(2) The *sample size n*.

(3) The inherent variability in the measured response variable.

(4) The difference between the hypothesized value of a parameter and its true value.



Example (cont) - Suppose the manager of the fast-food restaurant wants to change some aspects of his study about the proportion of drive-through customers who have to wait at least two minutes to receive their food after they place their order.

- Significance level: To reduce the possibility of Type I error and avoid the possibility of unnecessarily paying an extra employee, the manager reduces the significance level from 0.10 to 0.01.
- Practical importance: To justify the additional cost of the extra employee, the manager decides that the true proportion must be reduced to at most 0.53.
- To get faster results, the manager reduces the sample size from 250 to 100.

How will these changes affect the power of the test of H_0: $p = 0.63$ versus H_A: $p < 0.63$?

We will use the "Improved Batting Averages (Power) Applet" at www.rossmanchance.com to see.

1. Hypothesized value = 0.63 Alternative value = 0.53 n = 250, 10,000 samples, α = 0.10	3. Hypothesized value = 0.63 **Alternative value = 0.50** n = 250, 10,000 samples, α = 0.01
2. Hypothesized value = 0.63 Alternative value = 0.53 n = 250, 10,000 samples, **change** α = 0.01	4. Hypothesized value = 0.63 Alternative value = 0.53 **n = 100**, 10,000 samples, α = 0.01

Ways to Reduce Error

(1) If you insist on a smaller significance level, you have to take a larger sample. A smaller significance level requires stronger evidence to reject the null hypothesis.

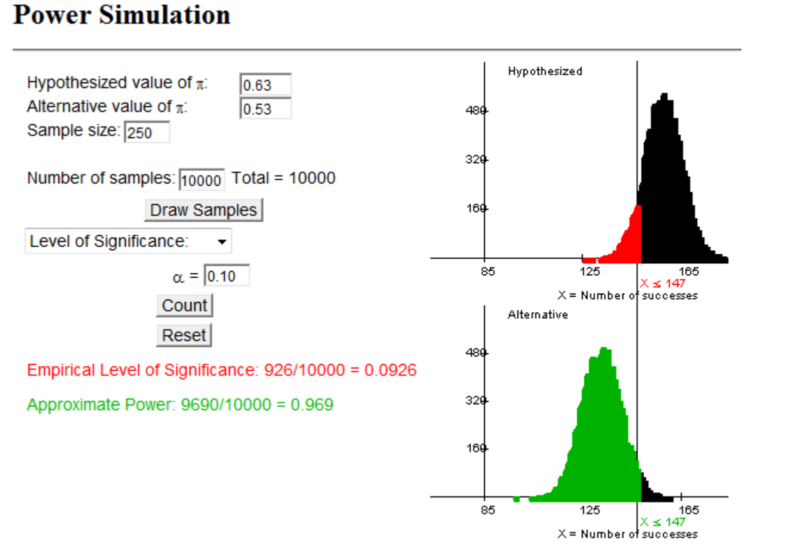
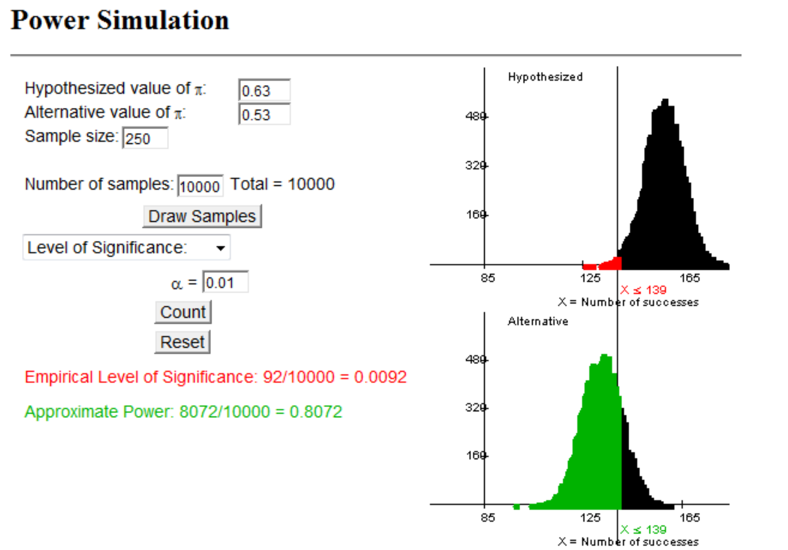
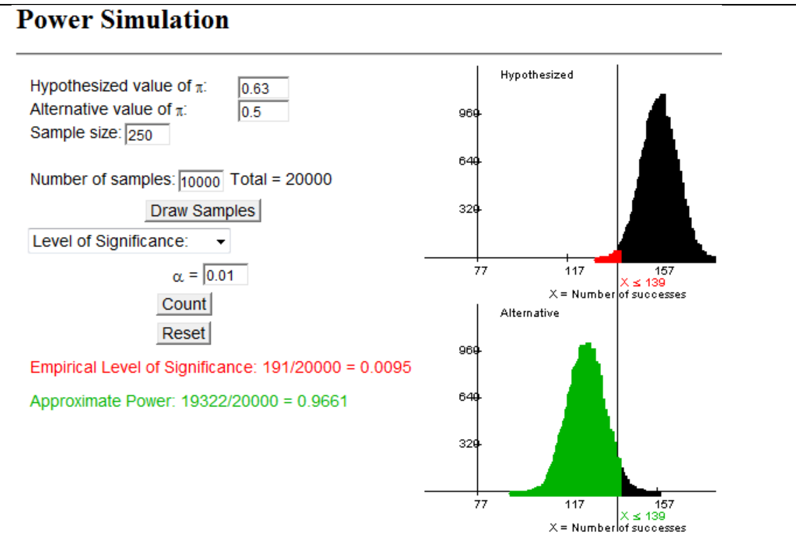
(2) If you insist on higher power, you will need a larger sample. Higher power gives a better chance of detecting a difference when it is really there.

(3) At any significance level and desired power, detecting small a small difference requires a larger sample than detecting a large difference.

The best advice for maximizing the power of a test is to choose as high an α level (Type I error probability) as you are willing to risk *and* as large a sample size as you can afford.

HW: Read pp. Read pp 547-551; problems 17, 21, 23, 25-30.

Improved Batting Averages Applet

<table>
<tr>
<td>Enter 0.63 for hypothesized value
Enter 0.53 for alternative value
Enter 250 for sample size
Enter 10,000 for number of samples
Press Draw Samples
Enter 0.1 for alpha, press count to verify that the power when p=0.53 is about 97%</td>
<td>
</td>
</tr>
<tr>
<td>Change alpha to 0.01 and press count</td>
<td>
</td>
</tr>
<tr>
<td>Change the alternative value to 0.50
Press draw samples
Choose level of significance and press count</td>
<td>
</td>
</tr>
</table>

Change the sample size to 100 Press draw samples Choose level of significance and press count