Section 9.1 (Part 2) (pp. 538-549)

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.

	H_0 TRUE	H_0 FALSE (H_A TRUE)
REJECT H_0	TYPE I	CORRECT
FAIL TO REJECT H_0	CORRECT	TYPE II

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: $p < 0.63$

TYPE I - REJECT $p = 0.63$ WHEN THAT IS TRUE

TYPE II - FAIL TO REJECT $p = 0.63$ WHEN $p < 0.63$

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.

WHICH IS WORSE?

I: TRUE PROPORTION OF CUSTOMERS HAS BEEN REDUCED WHEN IT REALLY HAS NOT BEEN. HE WOULD NOT HAVE ADD'L EMPLOYEES ON DUTY.

II: MGR DID NOT DECIDE TRUE PROPORTION OF CUSTOMERS HAS BEEN REDUCED WHEN IN FACT IT HAS. MGR WOULD HAVE PEOPLE ON DUTY UNNECESSARILY.

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."

	INNOCENT	GUILTY
REJECT INNOCENT	TYPE I	✓
FAIL TO REJECT INNOCENT	✓	TYPE II

In this situation, what constitutes a Type I error?

FINDING GUILTY WHEN INNOCENT.

What constitutes a Type II error?

FINDING INNOCENT WHEN GUILTY

Which error is worse?

(DISCUSS)

H_0: PRESUMED INNOCENT

H_1: GUILTY.

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 H_0 IS TRUE.

If our sample results in a $\hat{p}$ that is much smaller than 0.63, we will REJECT H_0.

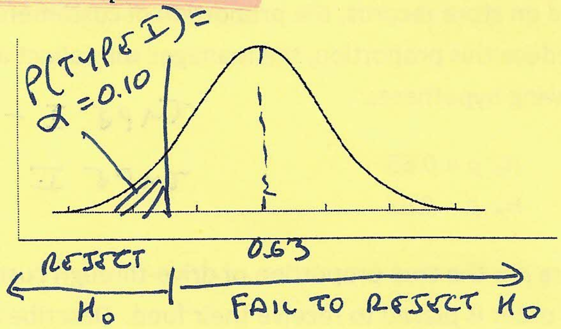
How small would $\hat{p}$ have to be? The 10% significance level tells us to count results that could happen less than 10% of the time by chance if H_0 IS TRUE AS EVIDENCE H_0 IS FALSE.

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

Shape: APPROX NORMAL SINCE
$(250)(0.63) = 157.5$, $(250)(1-.63) = 92.5$
BOTH ≥ 10

Center:
$\mu_{\hat{p}} = p = 0.63$

Spread:
$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.63)(0.37)}{250}} =$
0.0305



From this example, what can we conclude about the probability of a Type I error? IN THIS CASE, A TYPE I ERROR OCCURS WHEN THE TRUE PROPORTION OF CUSTOMERS WHO HAVE TO WAIT AT LEAST 2 MIN REMAINS $p = 0.63$, BUT WE GET A VALUE OF $\hat{p}$ SMALL ENOUGH THAT THE P-VALUE IS LESS THAN 0.10. WHEN H_0 IS TRUE, THIS WILL HAPPEN 10% OF THE TIME. IN OTHER WORDS P(TYPE I) = α.

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(Type I error) = P(Reject H_0 | H_0 is true) = α

What about Type II errors?

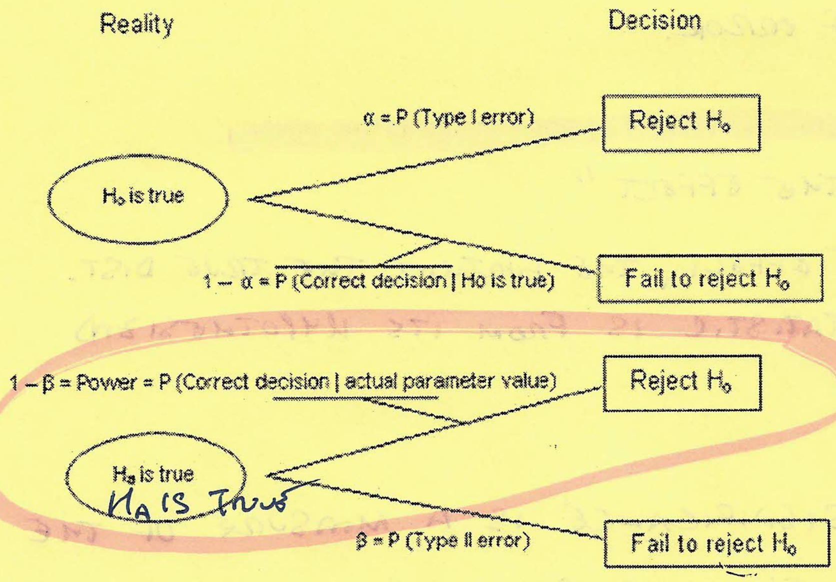
A significance test makes a Type II error when WE FAIL TO REJECT H_0 WHEN H_0 IS FALSE.

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

$$\text{Power} = 1 - \beta = P(\text{Rejecting } H_0 \mid H_A \text{ is true})$$



"GETTING IT RIGHT."
REJECT H_0 WHEN H_0 IS NOT 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. A LARGER α MEANS A LARGER REJECTION REGION FOR THE TEST AND THUS A GREATER PROB OF REJECTING H_0.

(2) The *sample size n.* A LARGER SAMPLE SIZE NARROWS THE DISTRIBUTION OF THE TEST STATISTIC.
(INCREASE TIME + RESOURCES FOR COLLECTING MORE DATA.)

(3) The inherent variability in the measured response variable.
DO A PRELIM STUDY TO DET'M SAMPLE SIZE NEEDED FOR A GIVEN MARGIN OF ERROR.

(4) The difference between the hypothesized value of a parameter and its true value.
"MAGNITUDE OF THE EFFECT"
THE LARGER THE EFFECT, THE FURTHER THE TRUE DIST. OF THE TEST STATISTIC IS FROM ITS HYPOTHESIZED DISTRIBUTION.

* STATISTICAL SIGNIFICANCE IS A MEASURE OF THE STRENGTH OF EVIDENCE OF THE PRESENCE OF AN EFFECT.
IT IS NOT A MEASURE OF THE MAGNITUDE OF THE EFFECT.

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 answer this.

① ~~CRIT VALUE = 0.53~~

SEE APPLET SHEET.

② $\alpha \downarrow$ POWER $\downarrow$
③ ALT $\downarrow$ POWER $\uparrow$
④ n $\downarrow$ POWER $\downarrow$

① $H_0: 0.63$ $H_A: 0.53$ n=250, 10000 SAMPLES $\alpha = 0.10$ $\boxed{1-\beta = 0.969}$	③ $H_0: 0.63$ $H_A: 0.50$ n=250, 10000 $\alpha = 0.01$ $\boxed{1-\beta = 0.9661}$
② $H_0: 0.63$ $H_A: 0.53$ n=250, $\alpha = 0.01$ $\boxed{1-\beta = 0.8072}$	④ $H_0: 0.63$ $H_A: 0.50$ n=100 $\boxed{1-\beta = 0.6176}$

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. 538-545; do problems 15, 19, 21, 23, 25, 31*, 32* on pp. 547-548.

WATCH CARTOON VIDEO ON CHAP 9 PAGE.