**Math 247: Invisible Errors: Type I and Type II Errors in Hypothesis Testing** (Section 8.3)

Suppose we've done everything correctly in gathering our data, doing our analysis via hypothesis testing, then forming a conclusion based on the P-value.

There is still the possibility, due to sampling variability (chance), that the evidence led us to a conclusion that is incorrect!

Remember BUZZ and Doris! If you flip a coin 16 times you won't get exactly 8 heads everytime! Results will VARY!

**Type I Error (the book calls this "the first type of error")** when the evidence tells us to reject the null hypothesis when, unbeknownst to us, the null hypothesis is actually true.

in the sample

The significance level, alpha, sets the maximum risk of making this error. The P-value further refines this and tells us precisely what the chance is we've made this error, if we do reject $H_0$.

$$\alpha = P(Type\ I\ error)$$

(*) TRIAL
$H_0$: innocent
$H_a$: guilty

$\alpha$ = reasonable doubt
Type I error: Reject innocence when person really IS innocent!

**Type II Error (the book calls this "the second type of error"):**

We make this error (and will never know we did!) when the evidence tells us to NOT reject the null hypothesis when, unbeknownst to us, the null hypothesis is actually false.

We make this error if we decide to not reject the null hypothesis, when the null hypothesis actually was false.

The likelihood of this type of error (failing to reject the null hypothesis when it really wasn't true) goes up as the significance level gets smaller.

Type II Error: $\beta = P(Type\ II\ error)$. Calculating this value is much more complex than finding a P-value, so we won't do it in this course.

(*) We dont reject innocence when we should! The person is actually guilty!

**Type I and Type II Error Grid**

| | | **Reality** (what's actually true of the **population**) | |
|---|---|---|---|
| | | ***$H_0$* is True**<br>***Ha* is False** | ***$H_0$* is False**<br>***Ha* is True** |
| **Conclusion of Hypothesis Test**<br>(based on the evidence given by the **sample data**) | **Reject *$H_0$* and accept *$H_a$***<br>**POSITIVE result** | **Type I error**<br>**False Positive**<br>$\alpha = P(Type\ I\ error)$ | Correct |
| | **Don't reject *$H_0$***<br>**NEGATIVE result** | Correct | **Type II error**<br>**False Negative**<br>$\beta = P(Type\ II\ error)$ |

**Example 1:** Right now, there is a lot of research going on to determine whether the hydroxychloroquine is helpful in treating COVID-19 (the disease caused by the coronavirus). Researchers in Wuhan did a controlled, randomized experiment with 62 COVID-19 patients to determine whether the proportion of improved* patients in the treatment group was higher than the proportion of improved* patients in the control group.
(*Specifically, they were looking at pneumonia outcomes in the patients.)

Hypotheses: p = proportion of patients who would improve in POPULATION

$H_0$: $p_{drug} = p_{control}$ proportion of improved patients would be the same IN POPULATION regardless of whether they take the drug.

$H_a$: $p_{drug} > p_{control}$ A higher proportion of ~~drug~~ patients would improve IN THE POPULATION if they took the drug as compared to those who don't.

Describe what a Type I error would be in this context:

If the evidence in the sample in the study led us to conclude a higher proportion of patients would improve (reject Ho, accept Ha) with the drug when really the drug doesn't help when the entire population of COVID-19 patients takes it.

Describe what a Type II error would be in this context:

If the evidence in the sample led us to conclude the drug does not increase the proportion of people who would improve, when in reality the drug WOULD increase improvement if it were given to the entire COVID-19 population.

The P-value from the analysis of pneumonia outcomes was .0176. What is our conclusion, based on this P-value?

$P\text{-value} = .0176 < .05 = \alpha$ (default)

Reject Ho, accept Ha

We have convincing evidence that there is a statistically significant ~~improvement~~ (increase) in the percent of patients who improve if they take the drug.

What error could we be making here? $\boxed{\text{False positive = Type I Error}}$ False negative = Type II Error

What harm could this error cause?

People taking the drug wouldn't improve (technically, a higher percent of people wouldn't improve)

- Delay in using other treatments
- SIDE EFFECTS of drug!

**Example 2:** Another drug under study right now, lopinavir–ritonavir, is one that has been used successfully for HIV treatment. The researchers in Wuhan used a randomized, controlled trial to see whether this drug, which looked promising in vitro (meaning in a lab setting), would actually be effective in vivo (meaning with actual patients). They were measuring, among other things, the percentage of patients who had improved on each day in the treatment vs. control groups to see whether there was a significant difference.

Hypotheses: $p$ = percent (proportion) of patients who would improve IN POPULATION!

$H_0$: $p_{drug} = p_{control}$ There is no difference in the percentage (proportion) of people who improve whether they take the drug or not.

$H_a$: $p_{drug} \neq p_{control}$ There is a difference

Describe what a Type I error would be in this context: We conclude based on the evidence there IS a difference in the percent (proportion) of patients who improve, when REALLY there would be no difference if we gave this drug to the entire population of COVID-19 patients

Describe what a Type II error would be in this context:
We conclude based on the evidence that there's NOT a difference in percent improved between drug and no drug when actually there REALLY would be a difference.

If the P-value was .12, what conclusion would the researchers make?

P-value $= .12 > .05 = \alpha$ (default)
We fail to reject Ho
We do not have sufficient evidence to say that taking the drug causes a significant change to the outcome (of the percentage of patients improving).

Suppose in reality, if the drug were given to the entire population of patients with severe COVID-19, there WOULD be an improvement by day 10. By failing to reject the null, what type of (invisible) error would we have made?

False positive = Type I Error

False negative = Type II Error

What harm could this error cause?
Not using a drug that, in reality, actually could help patients (technically, could change the percentage who improve)