Significance level and type I and II errors

Whenever we're using hypothesis testing, we always run the risk that the sample we chose isn't representative of the population. Even if the sample was random, it might not be representative.

For instance, if we've been told that 15% of American females have blue eyes, and we've set up null and alternative hypotheses to test this claim,

H_0: 15% of American females have blue eyes

H_a: the percentage of American females with blue eyes is not 15%

then when we take a sample to investigate our null hypothesis, we still run the risk of committing two types of errors.

Type I and Type II errors

Let's assume that the null hypothesis is true and that the percentage of American females with blue eyes is 15%. But let's say we take a sample of 100 women, find that 40 of them have blue eyes, and therefore calculate a sample proportion of $\hat{p}=40\%$.

If, based on the large difference between the sample proportion and the hypothesize proportion (40% versus 15%), we reject the null hypothesis, we've just made a **Type I error**. In other words, we make a Type I error when we mistakenly reject a null hypothesis that's actually true. The probability of making a Type I error is alpha, α, also called the **level of significance**.

Now let's consider the opposite situation and assume that the null hypothesis is false, such that the percentage of American females with blue eyes is not 15%. But imagine that we take a sample of 100 women and find a sample proportion of $\hat{p} = 15\%$.

If, based on the equality of the sample proportion and the hypothesized portion, we accept the null hypothesis, we've just made a **Type II error**. In other words, we make a Type II error when we mistakenly accept the null hypothesis when it's actually false. The probability of making a Type II error is beta, β.

	H_0 is true	H_0 is false
Reject H_0	Type I error P(Type I error)=alpha	CORRECT
Accept H_0	CORRECT	Type II error P(Type II error)=beta

There are lots of other ways to describe Type I and Type II errors, including

Type I error: Supporting the alternative hypothesis when the null hypothesis is true.

Type II error: Not supporting the alternative hypothesis when the null hypothesis is false.

Thinking about Type I and Type II errors can get people a little twisted around sometimes, so if we find that there's one description of them that makes more sense to us than the others, we can stick with that one.

Because α is literally the probability of making a Type I error, and β is literally the probability of making a Type II error, we can say that the **alpha level** is

- the probability of making the wrong decision when the null hypothesis is true, or
- the probability of rejecting the null hypothesis when it's true, or
- the probability of making a Type I error

and the **beta level** is

- the probability of making the wrong decision when the null hypothesis is false, or
- the probability of accepting the null hypothesis when it's false, or
- the probability of making a Type II error

Since the probability of committing a Type I error is α, the probability of making a correct decision when H_0 is true is $1 - \alpha$. And since the probability of committing a Type II error is β, the probability of making a correct decision when H_0 is false is $1 - \beta$.

Let's do an example where we look at the α and β levels in a hypothesis test.

Example

Lynnie is testing the hypothesis that people in her town spend more money on coffee on Monday than they do on Tuesday. She doesn't know it, but her hypothesis is false: people *don't* spend more money on coffee on Monday. She picks a random sample of people in her town and asks them how much money they spent on coffee each day. Say whether Lynnie will make a Type I or Type II error.

	Monday	Tuesday
Average spend	$6.75	$5.45

Lynnie's null and alternative hypotheses are

H_0: People spend no more on coffee on Monday than they do on Tuesday

$$C_M \leq C_T$$

H_a: People spend more on coffee on Monday than they do on Tuesday

$$C_M > C_T$$

In reality, her alternative hypothesis is false. But her sample data is showing that people spend more on Monday than they do on Tuesday. Which means she's in danger of rejecting the null hypothesis when she shouldn't, since the null hypothesis is true.

	H_0 is true	H_0 is false
Reject H_0	**Type I error P(Type I error)=alpha**	CORRECT
Accept H_0	CORRECT	Type II error P(Type II error)=beta

From the table we looked at earlier, the intersection of "reject the null" and "the null is true" is a Type I error. Lynnie is in danger of committing a Type I error.

Power

Sometimes we say that the **power** of a hypothesis test is the probability that we'll reject the null hypothesis when it's false, which is a correct decision. Rejecting the null hypothesis when it's false is exactly what we want to do.

	H_0 is true	H_0 is false
Reject H_0	Type I error P(Type I error)=alpha	**CORRECT Power**
Accept H_0	CORRECT	Type II error P(Type II error)=beta

So, the higher the power of our test, the better off we are. Power is also equal to $1 - \beta$.

Confidence levels and the α value

This α value, or level of significance, is the same α value we talked about when we looked at confidence levels and confidence intervals.

Remember that we usually pick a confidence level of 90%, 95%, or 99%, and these correspond to α values of

At 90% confidence, the alpha value is $\alpha = 1 - 90\% = 10\%$

At 95% confidence, the alpha value is $\alpha = 1 - 95\% = 5\%$

At 99% confidence, the alpha value is $\alpha = 1 - 99\% = 1\%$

So, in the same way that we said we normally pick a confidence level of 90%, 95%, or 99%, we could equivalently say that we normally pick an α value of 10%, 5%, or 1%.

When we decide on α, we're actually deciding how much we want to risk committing a Type I error. In other words, if we choose $\alpha = 0.05$, we're saying that, 5% of the time, or 1 out of 20 times, we'll reject the null hypothesis when the null hypothesis is actually true.

Choosing a significance level of 1% means we want to be more confident about the result than if we'd picked $\alpha = 10\%$. While it's true that we always want to be as confident as possible about our result, remember that picking a higher confidence level (and therefore lower alpha value) comes at a cost. The lower the alpha value, the wider the confidence interval and the larger the margin of error. This means that it'll be less likely that we detect a true difference between our sample statistic and the hypothesized value if that difference actually exists.

Similarly, since α is the probability of making a Type I error, and β is the probability of making a Type II error, we'd obviously like to minimize α and β as much as possible, because of course we always want to minimize the possibility that we'll make an error.

However, keep in mind that if we decrease α, it becomes more difficult to reject the null hypothesis, because the region of acceptance grows while the region of rejection shrinks. And if the null hypothesis is false while we decrease α, the risk of committing a Type II error increases because β gets larger and the power of the test decreases.

Alternatively, if we increase α, it becomes easier to reject the null hypothesis, because the region of rejection grows while the region of acceptance shrinks; the power of the test increases. We're at risk of committing a Type I error whenever we reduce the risk of committing a Type II error.

In other words, reducing the α value increases the β value, and vice versa. The only way to reduce them both simultaneously is to increase the sample size. If we could increase the sample size until it's as big as the population, the values of α and β would be 0.

Because of the inverse relationships between α and β, we're always trying to decide which type of error is more dangerous, and the answer to that depends on the situation. The question we need to ask ourselves is "What's the worst-case scenario?" For example, let's say a factory produces car parts and has a strict quality-control process in place. They assume that their production meets the minimum quality requirements, so that's their null hypothesis. The factory wants a low α, because that means they have to reject fewer parts as defective, which saves them money. But this lower α value might mean that more defective parts make it into cars, which could lead to cars that are less safe for consumers.

On the other hand, if we're a consumer who purchases a car made with these parts, we might prefer that the factory uses a higher α, rejects more defective car parts, thereby making sure our car is as safe as possible. However, if the factory uses a higher α value to keep the car safer, we may have to pay more for the car to account for the increased number of wasted defective parts.

So increasing the α level will decrease the Type II error risk for the consumer, but increase the Type I error risk for the producer. In other words, there are competing interests that are affected by changing the α value, and we have to decide exactly what α value gives us the balance we want.