Type I and Type II Error in Hypothesis Testing

There are two types of errors in hypothesis testing.

The first error is *when the null hypothesis should not have been rejected and it was*, this is known as a **type I error**. The probability of occurrence of this type of error is defined by your **alpha (α) risk** (your significance level).

The second error is *when the null hypothesis should have been rejected and it wasn't*, this is known as a **type II error.** The probability of occurrence of this type error is defined as the **Beta (β) risk**.

You can see these errors in the matrix below where across the top is truth, and we knew whether or not the null hypothesis was actually true or not. Down the horizontal axis is the result of the hypothesis test itself.

There are only 4 outcomes in this matrix, 2 are the correct decision and 2 are the incorrect decision. Errors happen when we make the wrong decision.

		The Truth	
		H_0 is True	**H_0 is False**
The Outcome of the Hypothesis Test	**Fail to Reject H_0**	**Correct Decision**	INCORRECT DECISION **(Type II Error) Beta (β) Risk**
	Reject H_0	INCORRECT DECISION **(Type I Error) Alpha (α) risk**	**Correct Decision** ***Power (1 - β)***

You can see that when the null hypothesis is actually true/accurate/correct there is a rare instance where we actually reject that null hypothesis, even though it's true. This is a type I error and is generally considered **the worse of the two error types.**

Let's go through some quick details for each of these error types.

Type I Error: The decision to reject the null hypothesis when it is actually true. (False Positive)

The **probability** of a type I error is governed by the **significance level** (**alpha risk**, α) you choose for your test. This is analogous to the "producers risk" in the world of acceptance sampling.

A type I error occurs due to random chance. This is the random chance you happen to randomly sample units from your population that happen to exist in the tails of the distribution, causing a rejection of the null hypothesis.

Type II Error: The decision to not reject the null hypothesis when it is actually false. (False Negative)

The **probability** of a type II error is governed by the **beta risk (β)**, and it is analogous to the concept of consumers risk in the world of acceptance sampling.

Oftentimes, when you lower your alpha risk, you increase your beta risk; however, both risks can be simultaneously reduced by taking larger sample sizes.

Similarly, you should know though that both errors types cannot be totally eliminated and are due to random sampling error.

The Power of a Hypothesis Test

Now that we've covered the **beta risk (β)**, let's jump into the **Power of a Hypothesis Test** as these two concepts are related.

Before I tell you the definition of the power of a hypothesis test, remember that we're only interested in the power of a hypothesis test **when the null hypothesis is in fact false**.

The power of a hypothesis test is the probability of correctly rejecting the null hypothesis (H_0) when it is actually false.

If we go back to the table above, you can see that the power of a hypothesis test is found in the bottom right-hand outcome; when the null hypothesis is actually false and we reject it.

		The Truth	
		H_0 is True	**H_0 is False**
The Outcome of the Hypothesis Test	**Fail to Reject H_0**	**Correct Decision**	INCORRECT DECISION **(Type II Error)** **Beta (β) Risk**
	Reject H_0	INCORRECT DECISION **(Type I Error)** **Alpha (α) risk**	**Correct Decision** ***Power (1 - β)***

The probability of correctly rejecting the null hypothesis is equal to 1 - β.

The power of a hypothesis test = The Probability of Rejecting a False Null Hypothesis = 1 - β

The other interpretation of Power is that is represents the probability of accepting the alternative hypothesis (H_a) when it is true.

Power can also be thought of as the probability of avoiding a Type II error.

The 5 Factors that Impact the Power of Your Hypothesis Test

There are five factors that impact the power of your hypothesis test, which include:

 # 1 - Increase Your Alpha Risk - If you hold all other factors constant, if you're willing to accept more alpha risk, you can improve the power of your hypothesis test.

When you increase alpha, you increase the probability of reject the null hypothesis which translates to a more powerful test.

The trade-off here is that you have to accept more alpha risk, which is the probability that you're rejecting the null hypothesis when it's actually true.

This method of improving the power of your hypothesis test is generally not recommended.

#2 - Increase your sample size - Increasing your sample size is the best way to improve the power of your hypothesis test!

If we take the population mean as an example, remember that the variance of the sample mean distribution is a combination of the population variance and the sample size.

$$Variance\ of\ sample\ mean\ distribution\colon V(\bar{x}) = \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$$

Increasing your sample size reduces the variability in your sample statistic distribution which improves your ability to discern between the null and alternative hypothesis when the null hypothesis is false.

 # 3 - Reduce your Process Variability - Similar to the above suggestion of increasing your sample size, the other way to reduce the variability in your sample statistic distribution is to improve your process.

When your process improves the population standard deviation goes down and the variance of your sample mean distribution goes down. This improves your ability to discern between the null and alternative hypothesis when the null is false.

 # 4 - The Difference Between the Hypothesized Value and its True Value - Remember, we only care about the power of a hypothesis test when the null hypothesis is false.

So, in the case that the alternative hypothesis is true, it's easier to reject the null hypothesis when there's a greater difference between the true population parameter and the null hypothesis.

Thus, the power of the test increases when there's a larger difference between the null hypothesis and the true value. If your actual population parameter is close to the hypothesized value it can often be difficult to detect that slight difference.

#5 - A One-Sided Hypothesis Test - The last way to improve the power of your hypothesis test is to use a one-sided hypothesis test as opposed to a two-sided hypothesis test.

When you do this, you're pooling your alpha risk in one tail of the distribution which increases the likelihood of rejecting the null hypothesis and also increases the power of your hypothesis test.

Statistical Significance versus Practical Significance

Ok, before we start wrapping up this chapter it's important to talk about the difference between statistical significance and practical significance.

In many situations, it is possible to show that a sample of data is statistically significantly different than the null hypothesis and thus result in the rejection of that, in favor of the alternative.

In those instances, it is very important for quality engineers to also **evaluate the practical significance of the test you ran.**

Let's go over a quick **example** of this. . .

Remember up above when we went through the example of the motor whose horsepower, we believe was improved through a design change.

We were able to show (with our sample data) that we should accept the alternative hypothesis that our **motor horsepower had increased**.

Ok, that was great, but the horsepower went from a historical average of **100**, and we measured a new sample average of **101.2**.

So, it's great that we can show a statistically significant improvement in horsepower, but is anyone really going to care? Is the customer really going to be excited about a 1% increase in horsepower? Are they going to pay more for that 1%. . .?

This situation might be statistically significantly different, but it's not practically significant.

Using the P-Value in Hypothesis Testing

Ok, so before we wrap this whole thing up, I wanted to talk quickly about the p-value approach to hypothesis testing.

The hypothesis testing **process** discussed above is known as the "**Critical Value Approach**" because we're comparing our test statistic against a critical value to determine if the null hypothesis should be rejected.

In today's world, most statistical software packages don't follow this method. Today most programs use an equal and alternate method called the "P-value" approach.

P-value stands for probability value and it represents the probability of observing a more extreme test statistic than the one observed.

Essentially, the p-value equals the area under the curve that is to the right, or the left of your test statistic.

With the p-value method, a statistical software page determines a p-value associated with your sample data. This p-value can then be compared against our significance level to determine if the null hypothesis can be rejected.

If the p-value is less than or equal to the level of significance (alpha risk, α), you must reject the null hypothesis.

A good tool to remember about the p-value is: ***If the p-value is low, the null hypothesis must go.***

Conclusion

Alright, let's wrap this up!!!

Ok, so **hypothesis testing** is defined as a *statistical process used to make a decision between a null hypothesis and alternative hypothesis based on information in a sample.*

This chapter started with a **refresher of inferential statistics**, specifically about **sampling distributions** to explain how hypothesis testing works as we **compare sample data against sampling distribution**.

We then move on to the **6-step process for performing a hypothesis test** which works regardless of the population parameter you're testing against.

- **Step 1.** Identify the **Null Hypothesis** (H_0) and the **Alternative Hypothesis** (H_a).
- **Step 2.** Choose the **Significance Level**, α.
- **Step 3.** Determine the **Rejection Region** for the Statistic of Interest (Mean, Variance, Proportion, etc.).
- **Step 4.** Calculate the **Test Statistic** from your Sample Data
- **Step 5.** **Compare** the Test Statistic against the Rejection Criteria and **Make a Conclusion**
- **Step 6.** State the Decision in Terms of the **Original Problem Statement**

From here we went into **the 3 specific population parameters & situations** where a hypothesis test can be conducted:

1. The Population Mean

Ok, when performing hypothesis testing for the **population mean**, there are two options.

The first option is to use the **normal distribution and the Z-transformation**. This option can only be used when the population standard deviation is known or when the sample size is greater than 30.

The second option is to use the **t-distribution**. This option must be used if your sample size is less than 30, or the population standard deviation is unknown.

2. The Population Variance & Standard Deviation

We covered two different hypotheses testing scenarios for the population variance & standard deviation.

The first scenario is when we're **comparing a sample variance against the population variance**, and in this case, we will use the **chi-squared statistic** & **chi-squared distribution**.

$$\boldsymbol{Chi\ Squared\ Test\ Statistic}: \boldsymbol{X^2 = \frac{(N-1)s^2}{\sigma^2}}$$

The second scenario is when we're **comparing two population variances against each other**, and in this case, we will use the **F-Test statistic** and **F-Distribution**.

$$\boldsymbol{F - Test\ Statistic}: \boldsymbol{F = \frac{s_1^2}{s_2^2}}$$

3. The Population Proportion

When hypothesis testing for the population proportion, we use the following z transformation to create our test statistic:

$$\boldsymbol{Z_O = \frac{\hat{p} - p_O}{\sqrt{\frac{p_O(1-p_O)}{n}}}}$$

We finalized this chapter with 4 additional concepts.

The first of these is the concept of **Type I and Type II Errors in Hypothesis Testing**.

A **Type I Error** is the decision to reject the null hypothesis when it is actually true. The probability of a type I error is governed by the significance level (alpha risk, α) you choose for your test. This is analogous to the producer's risk in the world of acceptance sampling.

A **Type II Error** is the decision to not reject the null hypothesis when it is actually false. The probability of a type II error is governed by the beta risk (β), and it is analogous to the concept of consumers risk in the world of acceptance sampling.

The second concept is the **Power of a Hypothesis Test** which is defined as the probability of correctly rejecting the null hypothesis (H0) when it is actually false.

We then discussed the 5 factors that impact the power of a hypothesis test: Your Alpha Risk, Sample size, Process Variability, The Difference Between the Hypothesized Value and its True Value, a One-Sided Hypothesis Test.

The third concept is the difference between **Statistical Significance versus Practical Significance**. The key concept to remember here is that anytime you're performing a hypothesis test, it's important to consider both the statistical significance of your result and the practical significance. For a result to truly matter, it's important to have both.

Lastly, we discussed using **the P-Value Method for Hypothesis Testing**.

P-value stands for probability value and it represents the probability of observing a more extreme test statistic than the one observed. A good tool to remember about the p-value is: *If the p-value is low, the null hypothesis must go.*