

6.5: Errors and Statistical Significance

It is important to understand what statistical significance does and does not tell a statistician and how it is determined. Therefore, we will review some important concepts connected to statistical significance in this chapter before learning about the kinds of statistical tests which can have significant results in the subsequent chapters.

There are a few different ways that statistical significance can be understood, all of which are connected to and complement each other. So far we have defined significance in two ways: 1. It refers to the determination that a hypothesis is likely true in the population because there is sufficient evidence in the sample to support the hypothesis and 2. The hypothesized result was observed in the sample with enough power to conclude that it was unlikely to be due to random chance. These are two ways of stating the same overarching concept.

Significance can also be described as the determination that the risk of a Type I Error is sufficiently low. Type I Error is one of two types of error in conclusions that are possible anytime a hypothesis is tested. These two types of conclusion error are tied to something known as an alpha level. Therefore, we must review their connection to get a complete understanding of what it means when we determine that a result is statistically significant.

No Conclusion Error

When a hypothesis is tested, a determination is made as to whether the hypothesis has been supported or not supported. When a result is in favor of a hypothesis and is significant, it is concluded that the hypothesis is likely true. When a result is not significantly in favor of a hypothesis, it is concluded that the hypothesis is too likely to be untrue to be supported. Another way to say this is that when a result is not significantly in favor of a hypothesis, the null hypothesis is retained. When the results from samples and their corresponding conclusions match what is true of the population, there is no conclusion error and the process of hypothesis testing is functioning as hoped.

The Ideal Situation

Ideally, population truths are reflected in sample data allowing consistent, accurate conclusion to be drawn about the population. This can happen one of two ways:

1. The hypothesis is true in the population and is reflected in the sample with sufficient evidence to support the hypothesis and reject the null or,
2. The hypothesis is false in the population and this is reflected in the sample such that the hypothesis is not supported and the null is retained.

In each of these versions of the ideal situation, there is no conclusion error.

Unfortunately, a statistician cannot know if their data from any one sample accurately reflects truths in the population. This is because what is true in a population is unknown. If we could test and know what was true of the population, there would be no need for statistics, but (for better or worse) statistics are necessitated by the fact that we simply cannot always know what is true of populations most of the time. This leaves the possibility of two kinds of conclusion error: Type I Error and Type II Error.

Type I Error

Type I Error is an erroneous conclusion wherein the null hypothesis is rejected when, in fact, it is true of the population. These types of errors are the same as “false positives.” Another way to say this is that a Type I Error occurs when the evidence suggests the hypothesis is true when it is not. This can happen because, though sample data are expected to approximate what is true of populations, samples vary in their sampling error (i.e. how accurately they reflect the population). Thus, false positives can occur. Critical values are set at levels that minimize this risk by requiring that the preponderance of the evidence be in support of the hypothesis before it is concluded that the hypothesis is likely true.

Type II Error

Type II Error is an erroneous conclusion wherein the null hypothesis is retained when, in fact, it is not true of the population. These types of errors are the same as “false negatives.” Another way of saying this is that a Type II error occurs when the evidence suggested the hypothesis was not true when it actually was true. This, like Type I Errors, is possible because of the existence of sampling error. Type II Errors are usually considered preferable to Type I Errors. Therefore, the critical values are set to minimize Type I Error. However, setting critical values to minimize Type I Error causes the risk of a Type II Error to be higher.

Reducing the Risk of Conclusion Errors

The conclusion errors are important because research findings are used to understand the world and to guide actions; it is, therefore, important to consider and reduce the risk of conclusion errors. Statisticians cannot guarantee that neither a Type I Error nor a Type II Error will occur. However, there are a few things that can be done to reduce the risk of one of these two forms of conclusion error. First, researchers should ensure they are using the best measures and methods available to collect data about each variable. Doing so can increase the accuracy of the data being used to test the hypothesis. Second, replication can and should be used. **Replication** is the act of repeating a study under the same or similar conditions and methods as were used in a prior study to see if the same results are obtained. If you have ever searched a research database for articles, you may have noticed that there are often many articles on the same topic, some of which used the same procedures to test the same hypotheses. This is done because each study has the risks of Type I and Type II Error so any one study is often considered insufficient support on its own. Therefore, studies can be repeated to see what the pattern of results is. If the same result keep occurring in the vast majority of studies, it increases the confidence of the scientific community that the conclusions in those studies are correct. A third way to reduce the risk of a conclusion error is to use a stringent threshold for determining significance. The threshold is set using something called an *alpha* level.

Meta-Analyses

Meta-analyses are a unique and very useful form of research. These are used to analyze and summarize the results of many replication studies together. Doing this is favorable over relying on any single study to determine whether a hypothesis is likely true. Thus, when a meta-analysis exists relevant to a topic of interest, it can be a great resource. This is a good thing to keep in mind whenever you are researching a topic for a class paper or even developing your own hypotheses to test. When reviewing empirical research, always check whether a relevant meta-analysis exists.

Alpha Levels

The acceptable probability of a Type I Error is referred to as an alpha level which is represented with the symbol α . The probability of a Type II Error is referred to as a beta level which is represented with the symbol β . Alpha levels are used to set thresholds for significance because the main conclusion error that a statistician is trying to avoid is a Type I Error. Recall that the null hypothesis is presumed to be true by default until sufficient evidence supports rejecting it in favor of the alternative hypothesis. This process prioritizes reducing the risk of a false positive (Type I Error) over reducing the risk of a false negative (Type II Error). In fact, the naming of the errors makes clear which error is the primary one to avoid: Type I.

Statisticians must identify an acceptable alpha level as part of step 3 of hypothesis testing. When a statistician sets an alpha level, they are identifying the risk they are willing to take of a Type I Error if they reject the null and conclude that the alternative hypothesis is supported. Unfortunately, the alpha level cannot be set to 0, which is part of why there are no guarantees in statistics. Setting alpha to 0 would essentially mean that there was no chance of a Type I Error and that a significant result would mean the researcher was 100% sure of this. Though the alpha level cannot be set to 0, it can be set pretty low. If an alpha level is set at .05 (meaning 5%), it would mean that the statistician was taking up to a 5% risk of a Type I Error. This is generally considered an appropriate level of risk. However, if the alpha level was set to .50 (meaning 50%), for example, it would mean that the statistician was taking a 50% risk of a Type I Error; this would mean it could be just as likely that the hypothesis was wrong as that it was right. At that point the researcher might as well save themselves the trouble of collecting and analyzing data and flip a coin to decide if a hypothesis is supported. Thus, a useful and acceptable alpha level is one that is set fairly low.

Most areas of the behavioral sciences consider a 5% risk of a Type I Error to be the maximum acceptable level. Alpha levels are typically written in decimal form so this risk can be summarized as follows: $\alpha = .05$. This means that a result will be considered non-significant when it is equal to or greater than .05 (i.e. 5%). When this alpha level is used, a researcher must be more than 95% sure that they are not making a Type I Error, based on the power of the evidence, to conclude a hypothesis is supported. This would mean that there is less than a 5% risk of a Type I Error. Sometimes this is worded by saying that the researcher is at least 95% confident they are not making a Type I Error. Another commonly used alpha level is .01. This translates to being at least 99% confident that there is no Type I Error (and a less than 1% risk of a Type I Error).

Though an alpha level cannot be set at 0, it is reasonable to wonder why it isn't set as close to 0 as possible by using alpha levels such as .01 all the time or even .000001 so a statistician could be 99% confident or even 99.9999% confident, respectively. Reducing the alpha level can be advisable in situations where making a Type I Error could have especially problematic

consequences. However, the lower a Type I Error is, the higher the risk of a Type II Error becomes. Therefore, decreasing the alpha rate doesn't necessary reduce the overall risk or an erroneous conclusion; instead, it may just be shifting it between the two risks.

Bonferroni correction

An additional consideration for setting alpha levels should be considered when multiple hypotheses are being tested together. Specifically, each time a hypothesis is tested with an alpha level of .05, there is a 5% risk of a Type I Error. That would mean that if two hypotheses were tested in a study, that there would be a 10% chance that at least one of them had a Type I Error, without being able to identify which, if either, had the error. When there are three hypotheses, the risk increases to 15% and so on. Therefore, when multiple tests are used, statisticians are advised to consider reducing the alpha level for each test to keep the overall risk of a Type I Error for the tests together below 5%.

The commonly used correction for this is known as a Bonferroni correction. The application of this method for addressing increased risks of Type I Error is attributed to a biostatistician named Olive Jean Dunn (Dunn, 1961; though it is named for the mathematician Carlo Emilio Bonferroni who focused on foundational issues of probability). To adjust risk of a Type I Error using a **Bonferroni correction**, the alpha level (α) is divided by the number of tests used (m) to get the reduced alpha level to be used for each test. For example, if the researcher wants a total risk of Type I Error to be less than 5% ($\alpha = .05$) when two hypotheses are tested ($m = 2$), the alpha level that would be used for each test is .025 or 2.5% (because $\frac{\alpha}{m} = \frac{.05}{2} = .025$).

Reading Review 6.4

1. In statistics, a false positive is referred to as which type of error?
2. In statistics, a false negative is referred to as which type of error?
3. Which alpha level is most commonly used?
4. When and why would a Bonferroni correction be recommended?

This page titled [6.5: Errors and Statistical Significance](#) is shared under a [CC BY-NC-SA 4.0](#) license and was authored, remixed, and/or curated by .