Section 3E – Type 1 and Type 2 Errors

Vocabulary

Population: The collection of all people or objects to be studied.

Sample: Collecting data from a small subgroup of the population.

Random Sample: Sample data collected in such a way that everyone in the population has an equal chance to be included.

Bias: When sample data does not represent the population.

Statistic: A number calculated from sample data in order to understand the characteristics of the data. For example, a sample mean average, a sample standard deviation, or a sample percentage.

Parameter: A number that describes the characteristics of a population like a population mean or a population percentage. Can be calculated from an unbiased census, but is often just a guess about the population.

Hypothesis Test: A procedure for testing a claim about a population.

Null Hypothesis (H_0): A statement about the population that involves equality. It is often a statement about "no change", "no relationship" or "no effect".

Alternative Hypothesis (H_A or H_1): A statement about the population that does not involve equality. It is often a statement about a "significant difference", "significant change", "relationship" or "effect".

Population Claim: What someone thinks is true about a population.

Sampling Variability: Also called "random chance". The principle that random samples from the same population will usually be different and give very different statistics. The random samples will usually be different than the population parameter.

P-value: The probability of getting the sample data or more extreme because of sampling variability (by random chance) if the null hypothesis is true.

Significance Level (α): Also called the Alpha Level. This is the probability of making a type 1 error. The P-value is compared to this number to determine significance and sampling variability. If the P-value is lower than the significance level, then the sample data significantly disagrees with the null hypothesis and is unlikely to have happened because of sampling variability.

Type 1 Error: When biased sample data leads you to support the alternative hypothesis when the alternative hypothesis is actually wrong in the population.

Type 2 Error: When biased sample data leads you fail to reject the null hypothesis when the null hypothesis is actually wrong in the population.

Beta Level (β): The probability of making a type 2 error.

Introduction

It is very difficult to understand large populations. Statisticians and data scientists spend years studying data, sampling variability, sample statistics and population parameters. Yet they can sometimes come to the wrong conclusion about the population. It has nothing to do with their knowledge or ability. The problem is sampling variability. In a sense, data can lead us to a wrong conclusion sometimes. Think about it this way. We are asked to try to understand millions of people in a population, but all we have is one random sample of 250 people. What if that one random sample is biased or is not reflective of what is going on in the population? That random sample can lead us to conclude something is true about the population that really is not true. When this happens, it is not the fault of the statistician or data scientist. They analyzed the data correctly and came to the correct conclusion about the population. The data has lead them astray. These wrong conclusions are called Type 1 or Type 2 errors.

Type 1 Error (False Positive)

A type 1 or type 2 error occurs because the random sample data does not reflect what is really going in the population. A type 1 error is when random sample data gives us a low P-value that is not reflective of the population. Maybe if we collected other random samples they would all have high P-values, but this one biased random sample has lead us to the wrong conclusion.

So biased sample data has given us a low P-value less than the significance level. From this, we rightly make the conclusion that we have evidence to reject the null hypothesis or support the alternative hypothesis. The problem is this the exact opposite conclusion from what is really going in the population. Type 1 errors can look bad on the statistician or data scientist. Remember a type 1 error involves a low P-value and having evidence. Later studies about the population may find that there is not significant evidence. Because the data scientist thinks they have evidence, a type 1 error is sometimes called a "false positive".

So how can we limit the possibility of making a type 1 error? The probability of making a type 1 error is the significance level (or alpha level "α"). So if you want to decrease the chances of making a type 1 error, lower the significance level (lower the alpha level). This is the reason that significance levels are always set low (1%, 5%, or 10%). If the significance level is 5%, then we have only a 5% chance of making a type 1 error. The most common significance level is 5%, but when a statistician is really worried about making a type 1 error, they may lower the significance level to 1% (or a 99% confidence level). Now they have only a 1% probability of making a type 1 error. Making the wrong conclusion about a population can have serious consequences. You may see a statistician decrease the significance level to 0.5% even (99.5% confidence level). In a sense they are making sure the P-value is really close to zero before considering it evidence.

Type 1 Error Summary

- Biased random sample data leads to a low P-value.
- We support the alternative hypothesis when in the population the alternative hypothesis is wrong.
- The probability of a type 1 error is the significance level or alpha level (α)
- To decrease the probability of a type 1 error, decrease the significance level (decrease α).

So why don't statisticians always set the significance level at 1% or 0.5%? Why is the most common significance level 5%? To understand this, we need to take a look at type 2 errors.

Type 2 Error (False Negative)

Type 1 and type 2 errors are inversely related. As the probability of one decreases, the probability of the other tends to increase. So if we decrease the significance level to 1%, the probability of type 1 error will decrease, but the probability of type 2 error will increase. Let's take a look at type 2 errors so we can better understand this.

A type 2 error occurs when random sample data gives a high P-value. The statistician fails to reject the null hypothesis, but later it turns out that this is the wrong conclusion about the population. In the population, the null hypothesis is wrong. This one biased random sample indicated that they did not have evidence, but later studies about the population find that there is evidence. The null hypothesis is actually wrong and the alternative hypothesis is actually correct in the population. Since the data scientist did not have evidence, a type 2 error is sometimes referred to as a “false negative”.

So how do we decrease the probability of making a type 2 error. The probability of a type 2 error is called a “Beta Level” (β). Beta levels are highly impacted by sample size. Remember the principle that more data = less error. This is particularly true for type 2 errors. A small sample size usually has a larger probability of making a type 2 error. A small sample size will have a larger beta level. So the main technique for decreasing the probability of making a type 2 error, is to simply collect more data. Increasing the sample size (n) will decrease the probability of making a type 2 error.

Notice that the best way to limit the chances of making a type 1 or type 2 error is to have a large random sample and a small significance level. This is the standard for collecting data.

Type 2 Error Summary

- Biased random sample data leads to a high P-value.
- We fail to reject the null hypothesis when in the population the null hypothesis is wrong.
- The probability of a type 2 error is called the beta level (β).
- To decrease the probability of a type 2 error, increase the sample size. Collect more data.

Setting your Significance and Confidence Levels

Before ever collecting data, a statistician thinks about the consequences of making a type 1 or type 2 error. They set the significance level for the test based on that assessment. Let’s look at some of the situations that may lead to using a 1%, 5% or 10% significance level in a hypothesis test.

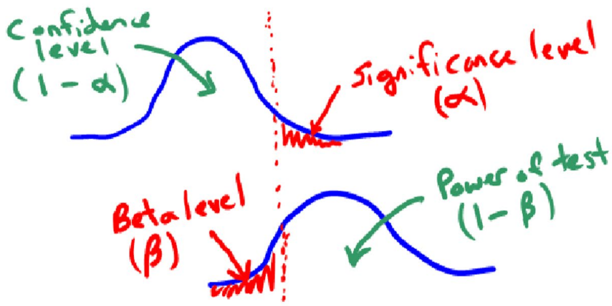
1% Significance Level ($\alpha = 0.01$): This corresponds to a 99% confidence level. Setting the significance level this low is trying to decrease the probability of type 1 error. At a 1% significance level, the probability of type 2 error will be higher. So this is usually a situation, where someone is trying to avoid a type 1 error, at the expense of allowing a higher probability of type 2 error. A good rule is that if you set the significance level at 1%, collect more random data so that the probability of type 2 error stays relatively low. Collecting more data is not always possible though.

10% Significance Level ($\alpha = 0.10$): This corresponds to a 90% confidence level. Setting the significance level this high will increase the probability of type 1 error. Remember, setting the alpha level high will decrease the beta level. So the probability of type 2 error will be lower. 10% significance levels are sometimes used to decrease the probability of type 2 error when you are not able to collect more data. The scientist is willing to risk the type 1 error, but does not want a type 2 error.

5% Significance level ($\alpha = 0.10$): This corresponds to a 95% confidence level. This is the most common significance and confidence level for a good reason. At a 5% significance level, the alpha and beta levels are both relatively low. So setting the significance level at 5%, keeps the probabilities of type 1 and type 2 errors relatively low.

So why are alpha and beta levels inversely related? Think of the probability distributions associated with a type 1 error is the “alpha curve”. The probability of a type 1 error is the significance level and the complement of the significance level is the confidence level. The probability distribution associated with a type 2 error is the “beta curve”. The complement of the beta level is called the “power” of the hypothesis test. So high confidence will correspond to a low alpha level or a low probability of type 1 error. A hypothesis test with a high power has a low beta level or a low probability of type two error. Remember that larger sample sizes result in a lower probability of type 2 error. As the sample size increases, your hypothesis test power also increases.

Diagram of Alpha levels, Beta levels, confidence levels, and power.



A 1% significance level is like taking the line above and moving it to the right. In the alpha curve on top, the probability in the right tail (significance level) decreases, but look what happens to the beta curve on the bottom. As we pull the line to the right, the beta level increases. Similarly, if we set the significance level at 10% we are pulling the line to the left. The significance level is increasing, but the beta level is now getting smaller.

Example

A pharmaceutical company wants to sell a new medicine in the U.S. To get approval they need to convince the FDA that the medicine is safe and has few side effects. If side effects happen in 3% or more of the people taking the medicine, then the FDA may not approve sale of the medicine in the U.S. If side effects happen in less than 3% of people taking the medicine, then the FDA may approve sale of the medicine in the U.S.

What is the null and alternative hypothesis?

Ho: $p \geq 3\%$ (FDA does not allow medicine to be sold in U.S.)

Ha: $p < 3\%$ (FDA does allow medicine to be sold in U.S.)

Describe the consequences of a type 1 error and what we could do to limit the probability of a type 1 error.

Because of some biased sample data, we got a low P-value and think that the alternative hypothesis is correct when it is not. (In reality, the null hypothesis is correct.) That would mean that the FDA approved sale of the medicine by mistake (false positive). The medicine causes serious side effects in a lot of people. People could die or become very sick. They may sue the pharmaceutical company or the FDA.

This is about as bad of a type 1 error that we can possibly have. To make sure this doesn't happen, we will lower the significance level to 1% (or even lower).

Describe the consequences of a type 2 error and what we could do to limit the probability of a type 2 error.

Because of some biased sample data, we got a high P-value and failed to reject the null hypothesis when in the population, the null hypothesis is really wrong and the alternative hypothesis is correct. So we think that the null hypothesis may be correct when it is not. That would mean that the FDA blocked the sale of a good medicine that rarely causes any side effects. Patients will be deprived of a good medicine and the pharmaceutical company will lose a lot of money in potential profits.

To make sure a type 2 error does not happen, we can increase the sample size. We can collect more data before calculating the P-value and making a decision. We should not increase the significance level just to avoid this type 2 error. We should keep the significance level low because a type 1 error of deaths and side effects is much worse than the pharmaceutical company losing money is.
