2 Sample Hypothesis Tests

Lesson Goals

In this lesson we will continue to explore hypothesis tests and learn about performing tests with two samples of data.

Introduction

It is common for researchers to perform studies to compare two groups and check if they behave differently. One example is applying a treatment to one group while another is left untreated. The results are then compared to see if the two groups differ. Hypothesis tests shine the most in such problems where we need to use statistics to inform decision making. In this lesson we will learn about the different ways we can compare two samples to see if they differ significantly.

Matched Pairs

When we approach a two sample hypothesis test, we must first determine what type data we have. The first type of 2 sample hypothesis test is performed on matched pairs. This means that the data in the two samples is dependent. For example, in a clinical drug trial, we may give a blood pressure medication to a group of people and look at their blood pressure before and after the treatment. We will then treat the before and after as two samples and compare them. However, since the two groups both contain the same people, we are able to match each entry in the before data with its corresponding entry of the after data.

A Bit of Theory

Since the we can match the data between the samples, we take the difference between the two samples in each row and then revert to using a one sample hypothesis test. Our hypothesis test will check whether the mean of the differences is significantly different from zero (we could also test that the mean is greater than or less than zero using a one-sided hypothesis test). For a two sided hypothesis test, this is our hypothesis:

$$H_0: \mu_d = 0$$
$$H_1: \mu_d \neq 0$$

Matched Pairs in Python

In our example we will look at a blood pressure study with 100 participants. Our participants all had their blood pressure measured before the beginning of the study and a month into the study. We will compare the systolic blood pressure for the participants before and after.

Let's start by looking at our dataset. The file blood_pressure.csv can be downloaded here.

```
import pandas as pd
```

```
blood_pressure = pd.read_csv('blood_pressure.csv')
blood_pressure.head()
           before       after
0  136.713072   92.432965
1  134.735618  105.022643
2  127.529115   82.242766
3  144.527126   93.607172
4  124.214720  103.212223
```

We will be using the scipy function ttest_rel. This function is used for hypothesis testing of dependent data.

```
from scipy.stats import ttest_rel


ttest_rel(blood_pressure.after, blood_pressure.before)
Ttest_relResult(statistic=-27.29184176756024, pvalue=7.303035069607835e-48)
```

Our result is a very small p-value. This means that we will reject the null hypothesis.

Since a matched pairs test is equivalent to a one sample test of the difference, we can also perform a one sample test and get the exact same result.

```
from scipy.stats import ttest_1samp
ttest_1samp(blood_pressure.after-blood_pressure.before, 0)
Ttest_1sampResult(statistic=-27.29184176756024, pvalue=7.303035069607835e-48)
```

We can see that the p-value is identical since the tests are equivalent.

Independent Samples

The second type of two sample hypothesis tests are independent samples. In this case, we have two groups where we cannot match the rows to one another. For example, we compare the effect of a certain medication on a sample of men and a sample of women. We then perform a hypothesis test to see whether there is a significant difference in the way the medication affects the groups. Another example is an A/B test on a website. We can implement a number of changes in the UI of an e-commerce website. We will release version A to a sample of customers and version B to another sample. We will then test if there is a difference in revenue between the different samples.

A Bit of Theory

When looking at two independent samples, we need to check that a few assumptions hold. The first assumption is obviously independence. An example of what could cause a dependence between two groups is if we had a study on the impact of nutrition on health and we had a husband in one group and a wife in the other. While they are not the same person, they most likely live in the same household. Therefore, there are some things that they do that might be similar like sleep habits or commuting habits. As researchers, when this happens, we cannot be sure whether the intervention in our study was the main cause of the difference (or similarity) between the subjects.

With a 2 sample test, our hypothesis test (for a 2 sided test) is a comparison of the two means.

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

We must also assume that the samples were drawn at random from a normally distributed population.

Equal Variances

If we make an additional assumption that the variances of the two populations are equal, we may use a pooled standard deviation in our hypothesis test.

In scipy, this means that we will be setting equal_var=True in our function.
The following is an example of a 2 sample hypothesis test with equal variance. We will load a sample dataset of transaction amounts from an e-commerce website. The dataset ab_test.csv can be downloaded here.

```
ab_test = pd.read_csv('ab_test.csv')
ab_test.head()
        a     b
0  0.27  13.61
1  6.08  21.53
2  13.74  9.23
3  9.70   5.36
4  7.00  12.90
```

The rows are not matched and the data is not stored in any order.

We make the assumption that the variances of both populations are equal based on prior knowledge of the data. Now we will test that there is a significant difference between the website layouts with a 95% degree of confidence.

```
from scipy.stats import ttest_ind


ttest_ind(ab_test.a, ab_test.b, equal_var=True)
Ttest_indResult(statistic=-2.637533181209767, pvalue=0.009713140852447347)
```

Our p-value is very small. This means that there is a significant difference between the two sample means.

Unequal Variances

When we don't feel comfortable that we can make the equal variance assumption with great certainty, we can use a more robust test instead. Instead of using a test with pooled variance, we use a test called Welch's t-test. This test is considered robust since it does not need to make as many assumptions about the data.

Let's use our A/B test data to perform a t-test that does not require the equal variance assumption:

```
ttest_ind(ab_test.a, ab_test.b, equal_var=False)
Ttest_indResult(statistic=-2.637533181209767, pvalue=0.009776243024828825)
```

In this case the p-value slightly differs from the one we get with equal variances. However, since it is very small in this case as well, we will still reject the null hypothesis and conclude that there is a significant difference between the two sample means.

Summary

In this lesson, we have looked at how to compare 3 different kinds of two sample tests. We first looked at matched pairs where our data was not independent. We then looked at the two different options for independent data. Hypothesis tests are an important tool in many areas of business and research. Therefore, it is important to master them and know to distinguish between the different types of tests.