This week we will explore hypothesis testing and estimation with confidence intervals. We will explore a t-test that compares the means of two samples. If you were interested in comparing some other characteristic (proportions, standard deviations, fit to a model…) you would use a different test.

Load and attach the data in the fev.csv file. The fev column of that data set contains the forced expiratory volume of each participant, in liters. We have a problem in that this column is called fev, and so is the data frame. Let’s rename the column.

*library(dplyr)*

*fev <- fev |> rename(FEV = fev)*

Don’t forget to attach fev after the renaming.

1. Now, suppose we want to know if there is a difference in forced expiratory volume between smokers and nonsmokers. Enter the following command, and paste the output here.

*t.test(FEV ~ Smoking)*

> t.test(FEV ~ Smoking)

Welch Two Sample t-test

data: FEV by Smoking

t = -7.1496, df = 83.273, p-value = 3.074e-10

alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0

95 percent confidence interval:

-0.9084253 -0.5130126

sample estimates:

mean in group 0 mean in group 1

2.566143 3.276862

Notice the tilde (~) in the argument of t.test. We saw this in regression where we had output ~ input. The tilde is doing the same thing here, the first variable is the quantity we want to compare, and the variable after the tilde is the variable that tells us how to split into groups.

2. There are various parts to that output, so in order to talk about each of them separately, let’s assign the output a name. Then we can look at the different parts.

*fevtest <- t.test(FEV ~ Smoking)*

*fevtest$p.value*

*fevtest$statistic*

*fevtest$estimate*

*fevtest$stderr*

*fevtest$alternative*

I’m not sure what the question here is so I just copy/pasted the output of those commands below

> fevtest$p.value

[1] 3.073813e-10

> fevtest$statistic

t

-7.149608

> fevtest$estimate

mean in group 0 mean in group 1

2.566143 3.276862

> fevtest$stderr

[1] 0.0994067

> fevtest$alternative

[1] "two.sided"

A hypothesis test consists of two hypotheses (if the word ends in “sis,” it’s singular, in “ses,” it’s plural). This particular test is testing that the two means are equal, that is, that the mean forced expiratory volume is the same for smokers as for nonsmokers. The null hypothesis is that they are equal, the alternative hypothesis is that they are not equal.

When we look at fevtest$alternative, it returns “two.sided,” which means that we would reject the null hypothesis (equality) if the smokers have a higher fev and also if the smokers have a lower fev. If we want a one-sided test, we have to specify which direction leads to rejecting equality by adding an argument: alternative = “greater” or alternative = “less”.

3. Now that we know what is being tested, we can look at how the test functions. When we are testing whether two population means are equal, we look at the two sample means: fevtest$estimate. These two means are not equal, but we are interested in whether they are different enough for the difference to be “real” vs. due to random error. Let’s figure out how different they are:

*fevtest$estimate[[1]]-fevtest$estimate[[2]]*

> fevtest$estimate[[1]]-fevtest$estimate[[2]]

[1] -0.7107189

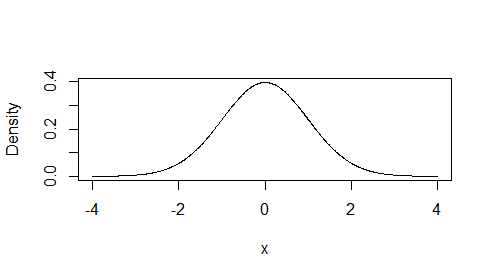
4. The units of the difference above are in liters. The question is, is this a big difference (“significant”)? To know this, we can scale the difference using the standard error (fevtest$stderr). That gets rid of the units and allows us to compare to a standard frame of reference.

*(fevtest$estimate[[1]]-fevtest$estimate[[2]])/fevtest$stderr*

Of course, this calculation has already been done

*fevtest$statistic*

We can then compare this value to the reference curve. This curve is below.



Compare our test statistic to this graph. Does it look like an unusual value in comparison to this curve?

No, a value between -1 and 1 would definitely put you in the bulk 68% of our scores.

5. When software compares the statistic value to the curve, it computes a p-value. A p-value is defined as **the probability of getting data like what was observed, or more extreme, if the null hypothesis is true.** What is the p-value for this test (fevtest$p.value)? It is reported in scientific notation. Rewrite it as a decimal value.

0.0000000003073813

As you can see, the p-value for this test is very, very, very small. The rule for hypothesis tests is **if the p-value is small, reject the null hypothesis in favor of the alternate hypothesis.** In general, unless otherwise specified, the cutoff value for what is “small” is 0.05. This cutoff is referred to as the significance level of the test.

So, for this test, since the p-value is small, we reject the null hypothesis that smokers and nonsmokers have the same fev, in favor of the hypothesis that the average fev is different for the two groups.

6. Run a t-test to see if the fev is different between men and women. Paste here your command and the output, state what the p-value was, and interpret that in terms of whether men and women have the same fev value.

fevtest2 <- t.test(FEV ~ Gender)

fevtest2

fevtest2$p.value

-------------------------------

Stated p-value: 5.604306e-08

Full t-test:

data: FEV by Gender

t = -5.5037, df = 575.75, p-value = 5.604e-08

alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0

95 percent confidence interval:

-0.4902038 -0.2323494

sample estimates:

mean in group 0 mean in group 1

2.451170 2.812446

---------------------------------

The p value suggests that we should reject the h­0 since again, the p-value is below .05

7. Knowing that the fev values are different between the two groups, we might be interested in how different they are. This is an estimation question, and requires a confidence interval to answer. To look at the confidence interval for the comparison of smokers and non-smokers, use

*fevtest$conf.int*

> fevtest$conf.int

[1] -0.9084253 -0.5130126

attr(,"conf.level")

[1] 0.95

This interval is giving reasonable values for the difference in mean fev values. The value being estimated is the mean for 0 minus the mean for 1. That is, the average fev for nonsmokers minus the average fev for smokers. The confidence level gives how sure we are that this interval contains the correct value.

The interval is (-.91, -.51), with 0.95 as the confidence level. Since both of the numbers in the interval are negative, the mean for 0 is smaller than the mean for 1 (when we subtract and get a negative answer, the first number was smaller than the second). So this is telling us that nonsmokers have a mean forced expiratory volume of between .51 and .91 liters less than the mean for smokers. The 0.95 tells us that we are 95% confident that the true value is between -.91 and -.51, which implies there is a 5% chance of error.

8. Estimate the 95% confidence interval for the difference in fev between genders.

> fevtest2$conf.int

[1] -0.4902038 -0.2323494

attr(,"conf.level")

[1] 0.95

9. Interpret the 95% confidence interval for the difference in fev between genders that you just found.

While the stat being measured is different it still seems to be a good indicator for rejecting the null. I suspect that since the mean for group one is higher group 2, men are more likely to be smokers and therefore also indicates smoking which then further indicates breathing.

There are many other tests. Which you choose depends on what kind of data you have. You could have only one sample. You could be interested in proportions rather than means. There are tests about the values of coefficients in regression, and in other types of models. But most tests yield an object like this one, that has values of the statistic, the p-value, the standard error, etc.