Research for Marketing Decisions

Vaibhav Chawla

Department of Management Studies

Indian Institute of Technology Madras

Week - 08

Lecture - 37

**Data Analysis using SPSS: Hypothesis Testing for Differences in Means
(Independent Sample T test)**

So this was the one sample t-test. The next one is do a if you look at the document do a t-test on the compare means procedure for internet usage against the gender, whether there is a difference among the gender regarding internet usage, this will you remember, we did with the box plot also just to see right now if a statistical test is done has to be done let's say you have two sec you have two segments you want to see whether they differ in the consumption regarding your product or you are going to launch a face cream and you have noted down the behavior of women. Let's say you want to see whether 18 to 25 behavior is different from 25 to 35. So there are two groups and their consumption behavior is there or whichever metric variable you have taken their readings, you want to see whether there is a difference.

Or let's say you have two groups, 18 to 25, 25 to 35. You want to know whether they differ in the importance they give to a particular attribute regarding face cream. Let's say somebody wants beauty more, other wants confidence more. You want to see whether there is a difference among the two age groups in the woman regarding the beauty, regarding the importance they place on beauty. So two groups differ.

Difference on a metric variable, you want to see whether there is significant difference, you have to use this. You want to look at people who are doing work from home versus those who are coming to office daily, whether their satisfaction differs. You will have to do this test. You have to do male and female, whether they scored differently in market research exam. In the past, you have to do this test.

In the B2B, let's say, again this can be used let's say, you have MSMEs and large customers two groups you want to look at whether they differ in terms of their sales in the last year how much they have given you business? you have to use this two groups so we will use this, so whenever there are two groups for which you want to know the difference on a metric variable. On a variable which is interval scale or ratio scale, you will have to use independent sample t-test,

and this is very useful again. These are very very simple things. If you just know these things you could actually do. This is the starting of the data science. So what to do now?

Analyze, so we are doing independent sample t-test. Analyze, compare means independent sample t-test. For what we have to do? Internet usage against the gender.

So internet usage will be the test variable. Gender will be the grouping variable. And gender is asking, define the groups. So group number 1 is male or female? I think group number 1 was male.

So let's give group number 1 value of 1, group number 2 value of 2. No, you have to type here. Group number 1, 1, group number 2, 2. So remember that group number 1 you have given the value of 1 and group number 1 is male. Group number 2 is female,

click continue and click ok, and you are getting this t-test. Independent sample t-test. Now there are two significance tests happening, one here,

one here, other one here. The first one is the test for assumption. So there is some assumption behind this independent sample t-test. The assumption is that the two groups in the population, the two groups have equal variances.

If the equal, if, so first test is about the assumption. It is just checking whether the assumption is met. If the assumption is met, then you have to look at which means, if the variances of the two groups in the population regarding that particular metric score whatever variable which means, for male and female if you look at their, if you infer about the variance in the population regarding this internet usage, if that variance is equal, then you will have to, then you will get in the first significance, you will get P value greater than 0.05.

So, first null hypothesis, which is, if you look at this particular row, one null hypothesis for the assumption, assumption testing is going on. That null hypothesis is sigma 1 square or sigma square for female is equal to sigma square for male. On that particular variable

that is the null hypothesis, and alternative is not equal to, so hypothesis testing is happening and you are getting a p value, p value here is less than 0.05, which means it is not same when it is not same then you will have to for you will have to go to the next test and look at the lower row value, and if the assumption is met then you will have to look at the upper row value. So in this sec in this if this there the independent sample t test is happening, here is the test for assumption, here the test for equality of means, whether the test for equality of mean, whether the two groups there is difference in their means regarding the variable or there is no difference. So null hypothesis in the population, there is no difference in the mean internet usage between male and female.

That is the null hypothesis. Alternate is, there is difference. And again, p-value test is given in the row number 5th, the significance two-tailed bracket. Whether we will look at row 1 or row 2, it will be decided by our assumption. Assumption is not met.

Equal variance is 0. Assumed is not correct because p-value is less than 0.000, less than 0.05, it is less than 0.05, which means, null hypothesis rejected equal variances is not there so, equal variance is not assumed that row we will have to look for test for equality of means, and here also we are getting 0.000, which means, in the population the mean internet usage between male and female different, different, is significantly different mean internet you will differ among the two genders in the population, and that you will have to do here, it is just an exercise, but your map your likely to be doing something related to this when you are in in your jobs, then nobody will tell you, then you will have to figure out what test to use, whether one sample t-test, whether independent sample t-test, then you will have to figure out if you read, if you understand, and know that independent sample t-test is when there are two independent groups and for a particular variable you are seeing whether their mean on that particular variable is different, significantly different, or not, and you will have to then take business decisions based upon your text which would be worth crores and multi crores and 100 crores, then what you will do if you do not know this?. Here there is nothing at stake, if with this somebody business matters you know let's say somebody comes to you and ask for conducting some survey related to a problem and you will have to do this and his entire business depends upon this, then the stakes would come

and at that time, if you don't pay attention, you do not know what to do. Then you might give an approximate, maybe conduct some interviews and based upon that approximations, you might give them. Which may not work because as I said, interviews,

you can do with very small sample that the sample of the interview may not be representative. You might end up ruining somebody's business. But this is easier.

See, you can pick up any value and see whether it is different from there, from that. You need not be picking the midpoint, neutral point. You can pick up 5 also, out of scale of 5, you can pick up 4. Scale of 7, you can pick up 5, 6. Depends upon your business problem.

That is null hypothesis, yes. Whether both the groups differ, whether the population of both the groups differ with respect to mean on a particular variable, whether their mean is different on particular variable for two of the groups in the population. That is what you are doing. You have to infer and recommend businesses some actions, point of actions based upon what your test would say.

The next one is fourth one. Do a Mann-Whitney test. Now what is this animal? Mann-Whitney. This must be a person.

I am sorry. I am just saying this test. What is this new test? So Mann-Whitney test to examine the difference. So now it is same.

It is saying, do the same but is non-parametric. When the data is non-parametric, which means, if the internet usage data right now it is in ratio. Let's say, it would have been ordinal. If the data on your variable is ordinal in nature and you want to see whether there is a difference among the two groups with respect to a particular variable.

Now this will be a difference in medians. Because ordinal data whether you know you have entire internet usage data right now in ratio.

Let us say it would have been in rankings from 1 to 30. If that data would have been rankings, you cannot use independent sample t test,

which is parametric. Parametric for metric variable, interval and ratio. You have to use non parametric. Which is for ordinal.

So what to do then, the same test, but you have to do it non-parametric. Analyze non-parametric test, independent sample.

Now here it is the test between the medians because data is ordinal. Assume, now in this case the data is ratio, still we are using it. Now what SPSS would do is based upon the ratio data, it would assign them ranks and then conduct the test.

Click on run. So what is the data we want? Internet usage right? And groups is gender. And if you click run.

Now in this case, it will give you statement. It will make your life easy. It will state everything. Reject the null hypothesis. What is the null?

What is the alternate? See null hypothesis statement is given itself. So, this will make life more easy because we need not identity, we need not state what is null, what is alternate. So, in non-parametric, they have made life more easy for students. For parametric, they did not.

I do not know in newer versions, it might be. No, no, it is similar only. In non-parametric, they are giving so much description, but parametric is not. How do you interpret this? You can say the median difference, the median difference, the difference of the medians between the two categories is different, significantly different because we are talking about ranks here.

Sorry? Yeah, go to analyze and non-parametric test, independent sample. Let me reset it and test with internet usage in hours per week, gender is groups, and click run, you will get this. Okay. Let's move to the next one, which is somebody's favorite, parity test. Parametric is when your data test variable test variable is in interval or ratio scale. Non-parametric is when your test variable is in ordinal scale, then you will have to do non test variable is in ordinal scale then you will have, which means in ranks then you will have to do the non-parametric test, but when your test variable is in ratio or interval you have, you will have to do the parametric test.

What it is doing is you can do it for ratio, interval and ordinal. The algorithm is even if you have ratio or interval it is converting it into ranks and doing it. If the data would have been ranks it is easier, but it need not do the back rankings, but with ratio or interval data also it is converting first them into ranks, and then doing this test non-parametric, it is same as independent sample but it is non-parametric.