

The Melody of Statistics

SPSS Guide and Statistical Tips

Developed by Prof. Dr. Francisco Tigre Moura

Introduction	3
Description of data set	4
Types of variables	5
Step 1: How to prepare the variable sheet	6
Step 2: Descriptive statistics	8
<i>2.1 Frequencies</i>	8
<i>2.2. Means</i>	9
Step 3: Chi-Square Test	12
Step 4: Comparing Two Groups of Means	16
<i>4.1 Independent t-Test</i>	17
<i>4.2 Dependent t-Test</i>	20
Step 5: Comparing Three (or more) Groups of Means	22
<i>4.3 One-Way ANOVA</i>	23

Introduction

Do you play guitar? No? (*Honestly, what is happening to humanity...*). Ok, but have you ever tried to learn? If you haven't, I'll tell you how it goes: **The first six months sucks!**

You will feel like the dumbest person on the planet because you will try to move from a simple "G major" chord to an "A major" chord, then to a "D major" chord (*because with these three chords you can play 1,267 songs from the "Ramones"*) and it will sound horrible! That's when many people simply give up...

But... If you persist you will eventually learn and develop the amazing ability to play an instrument and carry that ability for the rest of your life.

STATISTICS IS EXACTLY THE SAME!

If you hang on I can guarantee you that statistics is pretty amazing! (*maybe not as impressive with the ladies, though and it may ruin your social skills. Other than that, it's great!*). So my job here is to show you an **initial view** of statistics and how it can be **EASY** and pretty **FUN!**

BUT PLEASE KEEP IN MIND: This is only a supporting material at bachelor level. It is aimed to provide support to the topic during research courses. **So make sure to complement it with statistics and research methods books!**

So for further explanations and more in-depth understanding of the topic I would personally recommend the following amazing literature:

Additional literature

- [Field, A. *Discovering Statistics Using SPSS*. Sage. ISBN: 978-1-84787-906-6](#)
- [Field, A.; Hole, G. *How to Design and Report Experiments*. Sage. ISBN: 0857028294-9780857028297](#)
- [Field, A. *An Adventure in Statistics: The Reality Enigma*. Sage. ISBN: 1473943930-9781473943933](#)

Ps: As you can see, I only recommend books from [Andy Field](#). Simply because I think he is by far the best author on the topic (*and he is also into music!*).

Description of Data Set

For this supporting guide I will use a small data set of a study from **LiveInnovation.org**: “**General Behavior of Music Fans**”.

What did we measure? For this supporting material, we will make available **10 variables** that were measured, concerning:

Context for listening to music:

1. How often do you listen to music while doing physical exercise?
 - a. **(Type of data: Interval)**
2. How often do you listen to music while in the car?
 - a. **(Type of data: Interval)**

Music genre preference

3. What is your preferred music genre?
 - a. **(Type of data: Nominal)**

Music streaming behavior

4. How often do you stream music from Spotify?
 - a. **(Type of data: Interval)**

Music listening behavior

5. How often do you listen to music on your smartphone with headphones?
 - a. **(Type of data: Interval)**
6. How often do you listen to music on your smartphone without headphones?
 - a. **(Type of data: Interval)**
7. How often do you listen to music on your laptop?
 - a. **(Type of data: Interval)**

Demographics

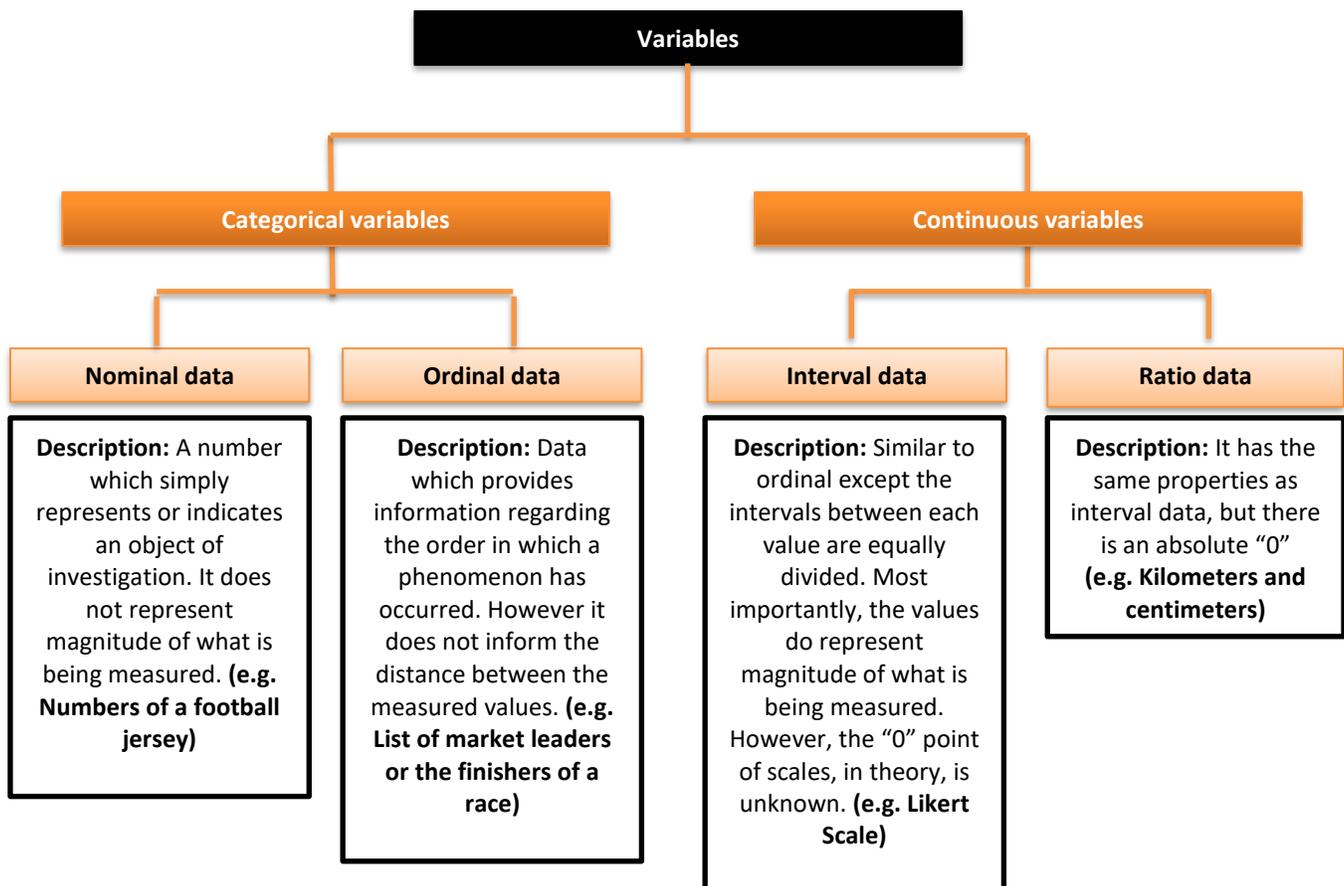
8. What is your gender?
 - a. **(Type of data: Nominal)**
9. How old are you?
 - a. **(Type of data: Nominal)**

Types of Variables

Before we start having a great time analyzing data (*please don't judge me for saying this. And yes, I do have friends*), one must understand what type of data we have in hands. For one main purpose: **the type of data influences the type of analysis that can be conducted**. And this is VERY important! **VERY important** (*It's like bass players. You may not like them, but they are VERY IMPORTANT, know what I mean?*).

Look at these numbers: **1, 2, 3, 4, 5, 6**. They are just numbers, right? **Yes**. But from now on we will call them by a different name: **DATA**. And they can represent many different things.

And how are they different? Well, the type of data depends on two main factors: **THEIR PURPOSE** and **HOW THEY WERE MEASURED**. Do not forget this, so I am going to repeat, ok? The type of data will depend on **THEIR PURPOSE** and **HOW THEY WERE MEASURED**. Ok, now I will repeat, but in bold, all caps, italics and in red hoping that the incredible visual effect will help you memorize: ***THE TYPE OF DATA WILL DEPEND ON THEIR PURPOSE AND HOW THEY WERE MEASURED***. So let's have a look at the 4 main types of DATA and the main things you need to know:



STEP 1: How to Prepare the SPSS variable sheet

What is the first thing you do when someone gives a pint of lager beer? **You toast. Then, you drink.** (*Well there must be other people with you for you to toast. If you are drinking alone, don't do it. It is sad*).

Well, the very first thing you should do when you have **DATA** and **SPSS** available to analyze it is to **PREPARE THE VARIABLE SHEET**. This is important so that your tests can run properly and you can interpret your findings later on accurately.

Here are the steps to do it!

1. Once you start the software, click in the bottom at “**variable View**”. This is where you will “prepare the variable sheet”.

The screenshot shows the SPSS Variable View dialog box. The table below represents the data shown in the dialog:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Frequency_Listen_Physical_Exercise	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
2	Frequency_Listen_Car	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
3	Favorite_Music_Genre	Numeric	8	2		{1,00, F...	None	8	Right	Nominal	Input
4	Stream_Spotify	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
5	Stream_Apple_Music	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
6	Device_Smartphone_WITH_Headphone	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
7	Device_Smartphone_WithOut_Headphone	Numeric	8	2		{1,00, Never...	None	8	Right	Scale	Input
8	Device_L...										
9	Gender										
10	Age	Numeric	8	2		{1,00, Youn...	None	8	Right	Nominal	Input
11											
12											
13											
14											
15											
16											
17											
18											
19											
20											
21											
22											
23											
24											
25											

Callouts in the image:

- 1. Variable view: Points to the 'Variable View' tab at the bottom of the dialog.
- 2. Name the variables: Points to the 'Name' column.
- 3. Label the values: Points to the 'Values' column.
- 4. Define the measure: Points to the 'Measure' column.

2. Once you are on the “variable view”, the most important elements to edit here are: “**name**”, “**values**” and “**measure**”.

a. **Name the variable:** Here you should label the variable that you have measured.

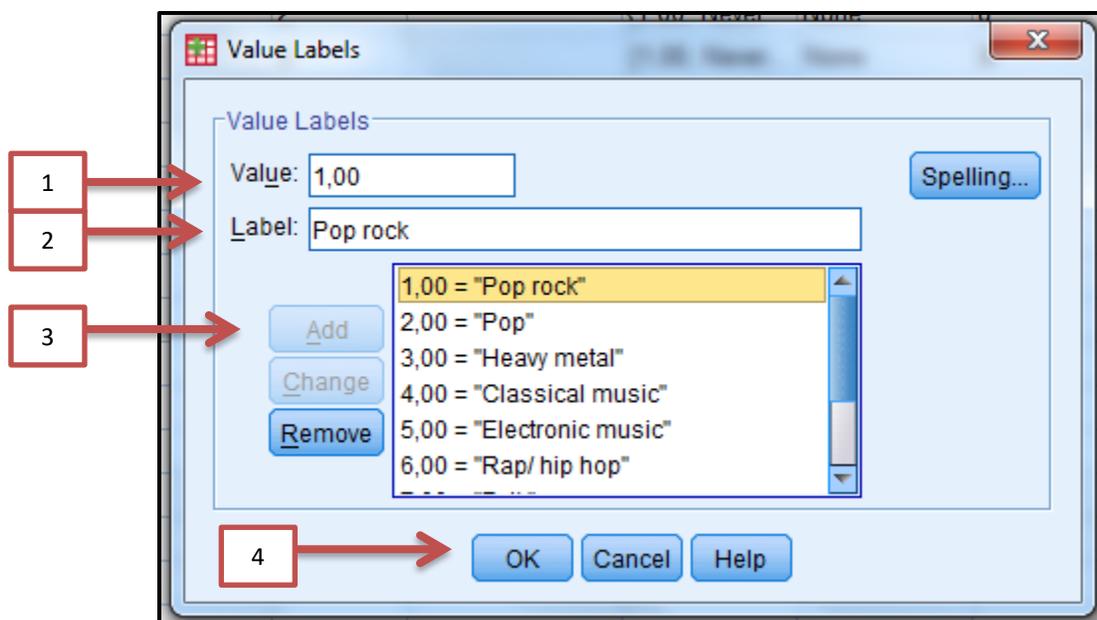
IMPORTANT: SPSS does **NOT** allow spaces between words here (*Not sure why...*). So variables should be labeled, for example, with underlines.

- **Example:** Favorite Music Genre = "Favorite_Music_Genre".

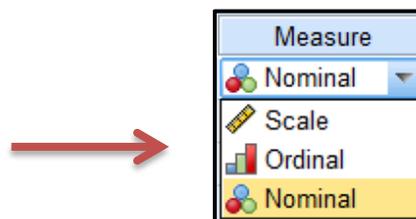
IMPORTANT 2: After you finish your analysis and you will write your paper and **report your results**, remember to **write clearly the name of the variables**. Do not use codes or underlines, otherwise your client or reader may not understand.

- **Example of how to do it:** Favorite Music Type.

- b. **Label the values:** Here you will define the value of each number. Once you click on the white box, the window below will simply pop up. Simply write the value and the desired label and "add". Once you are done, simply click on "ok".



- c. **Define measure:** For each variable, you should define what type of data it is: **Nominal, ordinal or scale (Interval or ratio)** .



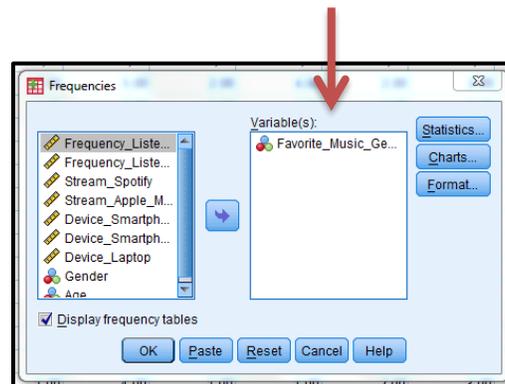
STEP 2: Descriptive Statistics

Alright, now it's time to have some fun and start analyzing some data (*don't judge me for saying this. And yes, I do have friends*). We will start by the absolute most basic analysis: frequencies and means (also called averages):

Here are the steps to do it!

2.1 Frequencies

On SPSS Menu: Analyze > Descriptive Statistics > Frequencies. Then, simply choose the variables that you wish to analyze. For this example, we will look into the variable of "Favorite_Music_Genre". Then, simply click on "ok".



Here are the results:

The symbol "N" describes your sample size (the number of respondents). So in this case, there were 301 participants in the study.

Statistics		
Favorite_Music_Genre		
N	Valid	301
	Missing	0



The second table shows the frequency of respondents (the total number of answers per category); the percent (the percentage of respondents from each category in relation to the total of respondents. **Ex: Pop Rock had 56 respondents, which represents 18,6% from the total of 301 respondents. Got it?**).

Favorite_Music_Genre

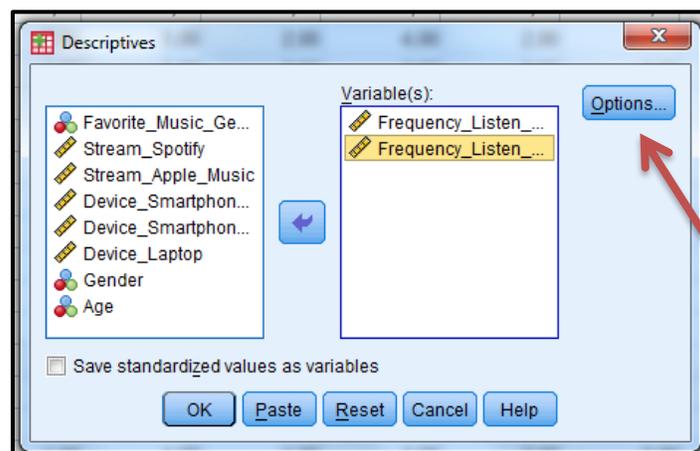
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Pop rock	56	18,6	18,6	18,6
	Pop	54	17,9	17,9	36,5
	Heavy metal	20	6,6	6,6	43,2
	Classical music	15	5,0	5,0	48,2
	Electronic music	51	16,9	16,9	65,1
	Rap/ hip hop	32	10,6	10,6	75,7
	Folk	12	4,0	4,0	79,7
	Jazz	8	2,7	2,7	82,4
	Reggae	2	,7	,7	83,1
	Other	51	16,9	16,9	100,0
	Total	301	100,0	100,0	

How to Report Frequencies

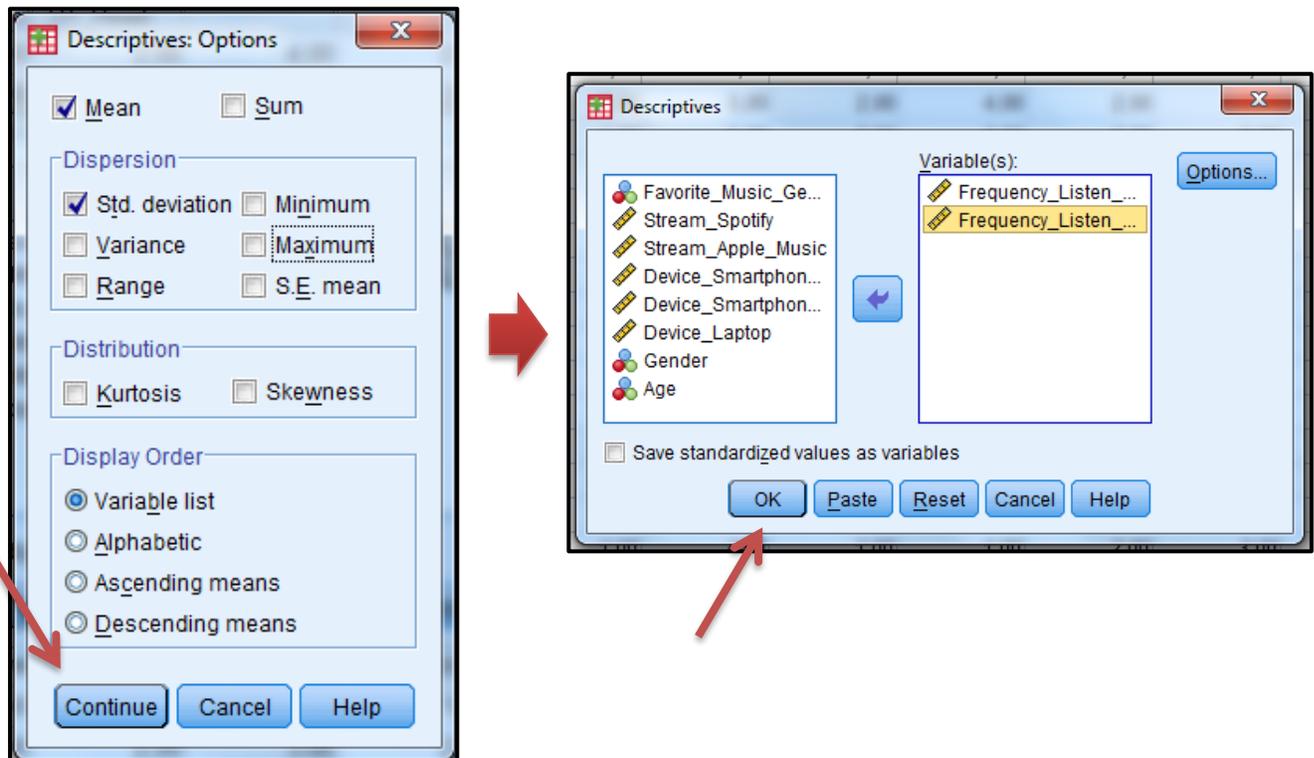
Example: The data set revealed that 56 participants (18,6%) preferred listening to Pop Rock, while 54 respondents (17,9%) indicated to prefer listening to Pop music.

2.2. Means (also referred to as “Averages”)

On SPSS Menu: Analyze > Descriptive Statistics > Descriptives. Following this, choose the variables to be analyzed. In this example, we will use “Frequency_Listen_Physical_Exercise” and “Frequency_Listen_Car”. Then, click on “options”.



Here, make sure to click on “Mean” and “Standard deviation”. These are the most basic and most commonly used tests. After click on “continue”, and afterwards on “Ok”.



Here Are The Results:

REMEMBER: Both questions were measured on a 5-point Likert scale, ranging from “1-Never” to “5-Always”. So the higher the mean, the more frequent respondents listen to music during those situations.

Descriptive Statistics

	N	Mean	Std. Deviation
Frequency_Listen_Physical_Exercise	301	4,0897	1,18121
Frequency_Listen_Car	301	4,6711	,70814
Valid N (listwise)	301		

First observation: The symbol “N” describes your sample size (the number of respondents). So in this case, there were 301 participants in the study.

Second observation: Respondents revealed an average of 4.08 towards listening to music while doing exercise and 4.67 while in the car.

How to Report Means

Means are usually reported with the letter “**M**”, followed by the variable name. Standard deviation is usually reported by the letters “**SD**”.

Example: Participants in the study revealed very high frequency while doing physical exercise (**M**_{Physical Exercise}: 4.08, **SD**: 1.18) and an even more active behavior while in cars (**M**_{Car}: 4.67, **SD**: .70).

#####

VERY IMPORTANT: In the SPSS variable view variables are usually coded or written with underlines, such as “**Frequency_Listen_Car**”, for example. During your study, you most likely will also create your own coding to the variables you measure.

HOWEVER, when reporting results on your thesis (or report) or create tables, **make sure to ALWAYS write the name of the variable correctly, NOT abbreviated.**

- **Ex:** In this case, in your results and tables, you should write “**Frequency of Listening to Music in the Car**”, **NOT** “**Frequency_Listen_Car**”.

Got it?

STEP 3: Chi-Square Test

Alright, let's make things more interesting. Let me ask you something: Who do you think is more passionate about "Pop" music, Guys or Girls? Who is more passionate about "Electronic Music", Guys or girls? Well, let's check it out!

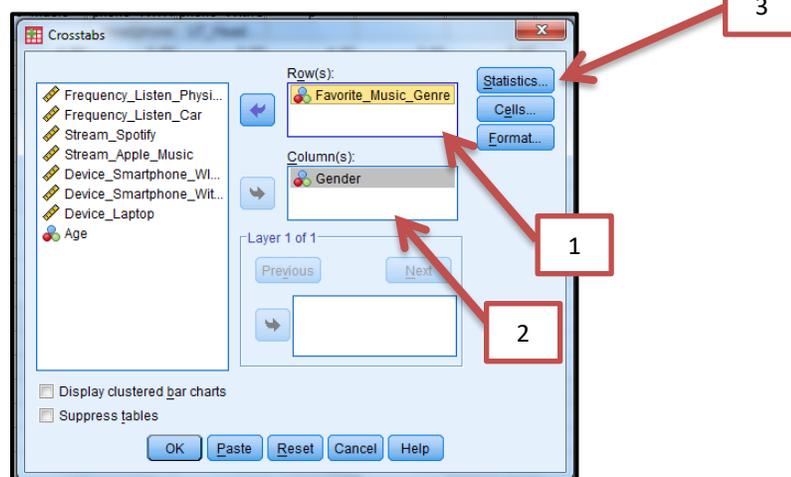
First of all, as you should know by now, gender and music types are **CATEGORICAL VARIABLES**, correct? (If you forgot why, or don't know why, go back and re-read the section on "Types of Data").

So we will introduce you to an amazing statistical test that verifies the ASSOCIATION OR RELATIONSHIP BETWEEN CATEGORICAL VARIABLES. And it has a funny name: **Chi-Square test**.

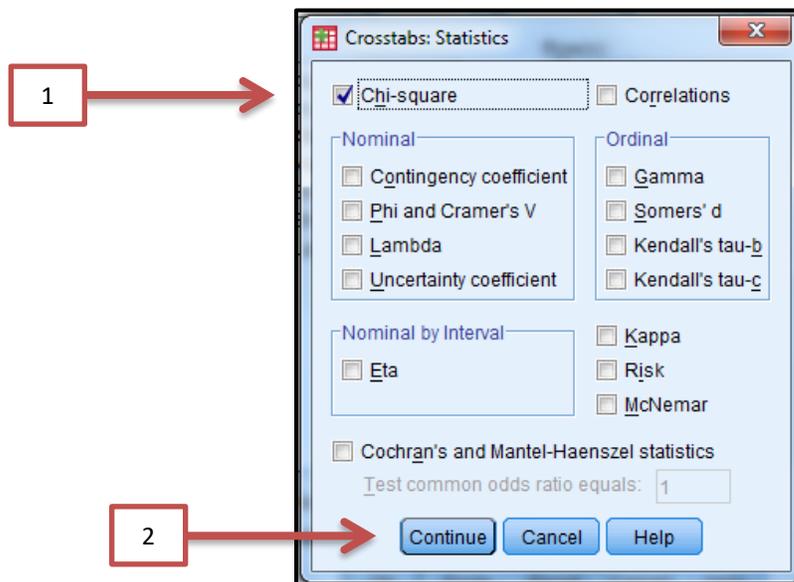
Pearson's Chi-Square test (χ^2): Is a test to check for the **association** between categorical variables (Nominal or ordinal).

Here are the steps to do it!

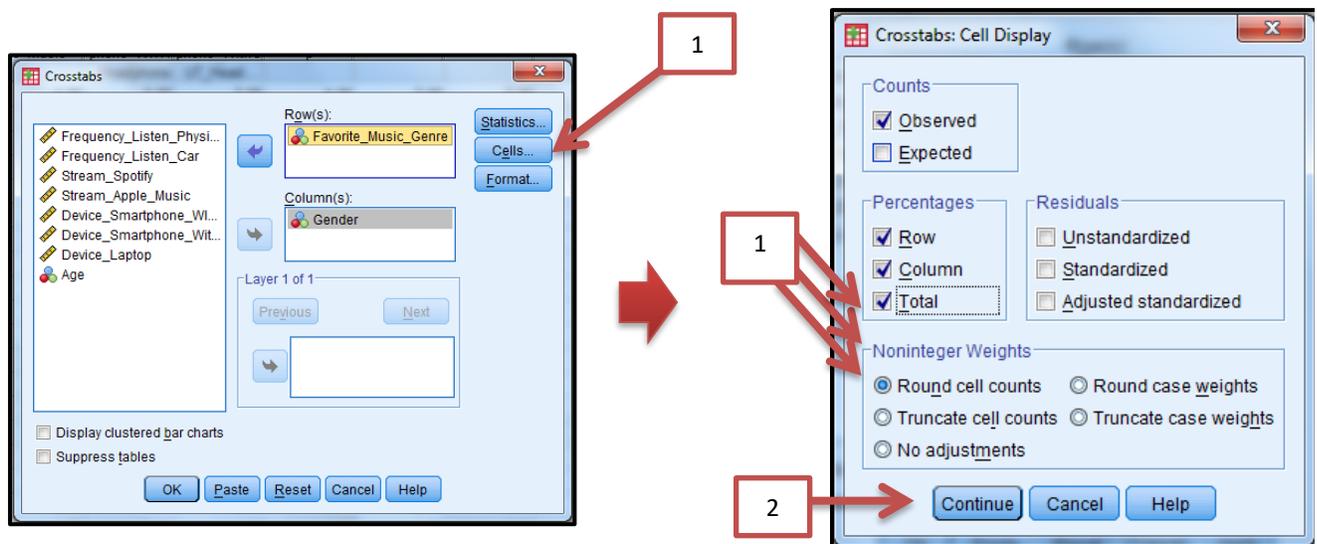
On SPSS Menu: Analyze > Descriptive > Crosstabs. Then, the following window will appear. You simply need to choose the two variables to compare: one to be a "row" and one to be "column". (It does NOT matter which variable is "row" and which is "column"). The results will be the same. For the example, we will choose "Gender" and "Favorite_Music_Genre". Afterward, click on "statistics".



Here, make sure to check for “**chi-square**”. Then, click on “**continue**”.



Afterwards, click on “**cells**”, and check for “all frequencies”: “**Row**”, “**Column**” and “**total**”. The click in “**continue**”. Finally, click on “**OK**” to have your results.



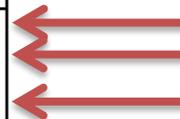
Here Are The Results:

The first table will give you a summary of the number of respondents. As you already know by now, the label “N” refers to the sample size. So, again, we have 301 participants. The table will also indicate the two variables that are being tested: In this case, “Favorite_Music_Genre” and “Gender”.

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Favorite_Music_Genre * Gender	301	100,0%	0	,0%	301	100,0%

The second table will give you a description of the distribution of frequencies. (Due to the size of the table, here is only shown the results for 5 music genres: Pop rock, Pop, Heavy Metal, Classical Music and Electronic Music).

			Gender		Total
			Male	Female	
Favorite_Music_Genre	Pop rock	Count	19	37	56
		% within Favorite_Music_Genre	33,9%	66,1%	100,0%
		% within Gender	14,4%	21,9%	18,6%
		% of Total	6,3%	12,3%	18,6%
Pop	Pop	Count	8	46	54
		% within Favorite_Music_Genre	14,8%	85,2%	100,0%
		% within Gender	6,1%	27,2%	17,9%
		% of Total	2,7%	15,3%	17,9%
Heavy metal	Heavy metal	Count	14	6	20
		% within Favorite_Music_Genre	70,0%	30,0%	100,0%
		% within Gender	10,6%	3,6%	6,6%
		% of Total	4,7%	2,0%	6,6%
Classical music	Classical music	Count	8	7	15
		% within Favorite_Music_Genre	53,3%	46,7%	100,0%
		% within Gender	6,1%	4,1%	5,0%
		% of Total	2,7%	2,3%	5,0%
Electronic music	Electronic music	Count	31	20	51
		% within Favorite_Music_Genre	60,8%	39,2%	100,0%
		% within Gender	23,5%	11,8%	16,9%
		% of Total	10,3%	6,6%	16,9%



How to Interpret?

Count: The total number of respondents per category. **Example:** from the 301 respondents, 19 were Males that liked “Pop Rock” and 37 were Females that liked “Pop Rock”.

% within Favorite Music Genre: The percentage of gender in relation to each the music style. **Example:** 56 participants equal 100% of all people that liked Pop Rock”. 19 were male (33,9%) and 37 were Females (66,1%).

% within Gender: This represents the percentage of respondents from the total within the Gender. One cannot see in the table above (the full table was too large to paste here) but the total number of Males for the study was 132 and of females was 169. The “% within Gender” reveals the percentage of respondents in relation to the total. **Example:** 19, out of 132 Males, liked “Pop Rock”. And 19 equals 14,4% out of 132 male respondents.

Example 2: Out of the 169 Females, 37 preferred “Pop Rock”, and this equals 21,9% of the total of Females.

The third table will provide you with the “Chi-Square” test results.

Chi-Square Tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	47,343 ^a	9	,000
Likelihood Ratio	51,175	9	,000
Linear-by-Linear Association	6,253	1	,012
N of Valid Cases	301		

Pearson Chi-Square value (points to 47,343^a)

Degrees of freedom (points to 9)

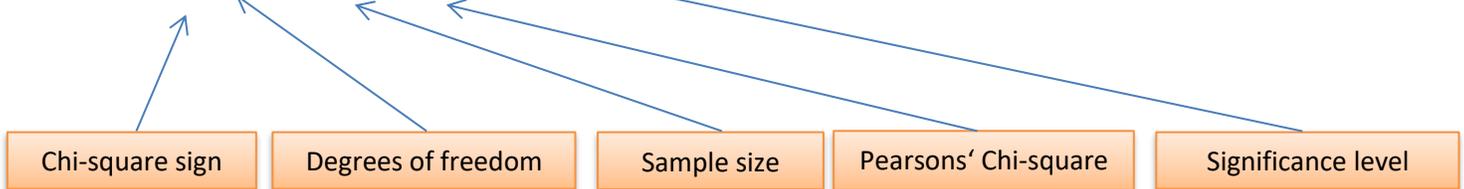
Significance level (points to ,000)

Significance level (p-value): the probability of a result from a statistical test occurring by chance. Researchers look for a probability of 5% or less that the results are caused by chance, which means a 95% chance the results are "not" due to random factors.

How to Report Chi-Square Results

Chi-Square statistics are reported with degrees of freedom and sample size in parentheses, the Pearson chi-square value (rounded to two decimal places), and the significance level.

Example: Results indicated that there was a significant association between gender and preferred music genre, $\chi^2(9, N = 301) = 47.34, p = .000$.



Reminder: Chi-Square test does **NOT DETERMINE CAUSALITY** of the association. It tests merely the significance of the association between the categorical variables!

Example: One cannot say gender “causes” someone to prefer a music genre (Example: That females like Pop because they are females). You can ONLY assume that males or females are more associated to a particular music genre.

STEP 4: Comparing Two Groups of Means

Comparing Two Groups of Means: t-Test

Who do you think was a better guitar player: **Jimi Hendrix or Jimmy Page?** (*If you don't know them, you should really go do your homework on this!*). Who was better songwriter: **Paul McCartney or John Lennon?** Who was a better singer: **Nina Simone or Billie Holiday?**

Answers: In my personal opinion: **Jimmy Page, Paul McCartney and Billie Holiday**

Ps: *A bit controversial opinion, but that is how life goes!*

Anyway, in all those examples there was something in common: **the goal of comparing two groups** (*In this case, musicians*). And the way to do it is by conducting a tested called: **t-Test**.

t-Test: Is a statistical test to compare two groups of means. A t-Test checks if two samples are significantly different from each other and is commonly used when the variances of two normal distributions are unknown.

There are two types of t-tests, and their differences are very simple: **if the two groups have the same participants (dependent) or different (independent) participants.**

The two types of t-Tests

Independent t-test: When **different** participants have been used in the each group.

Dependent (or matched pairs) t-test: When **similar** participants have been used in the two different groups.

Now let's start by showing you how to conduct the **independent sample t-Test** (when you have **different** participants in each group).

4.1 Independent Sample t-Test

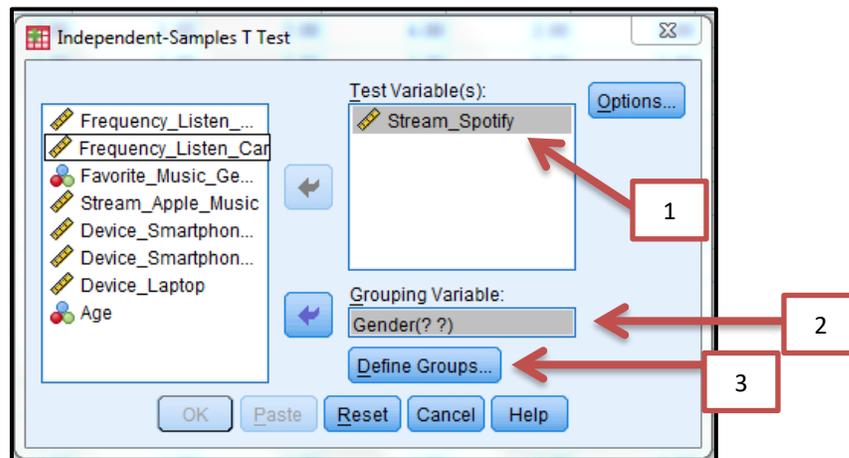
The example below will address this completely random question

Question: Who streams music more frequently on Spotify: Males or Females?

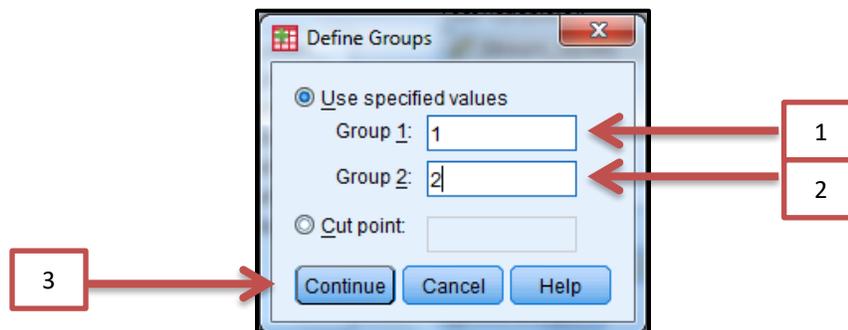
Here are the steps to do it!

On SPSS Menu: Analyze > Compare Means > Independent Samples t-Test. After this, the window below will appear. First, you will choose the variable which you want to compare (For this example, we will choose the streaming of Spotify).

Afterwards, you must choose the “**grouping variables**”: the groups you wish to compare (In this example, we will check to see if there is any difference between **Gender**).



Once you’ve chosen the “**grouping variable**”, you must now click on “**define groups**”. The following table will appear:



Basically the software is asking: **what is the label of the groups you want to compare?** For this example, on the “**variable view**” sheet the coding for **Male** is “1” and for **Female** is “2”. After this, press on “**continue**”, then “**Ok**” on the previous table.

Here Are The Results:

The first table provides a general overview of both groups: The variables compared (Streaming of Spotify), both groups being compared (Males and Females), Sample size (N), means of perceptions (in this case, “how frequently they listened to music using streaming from Spotify”), Standard deviation and standard error.

Group Statistics



Gender	N	Mean	Std. Deviation	Std. Error Mean
Stream_Spotify Male	132	2,9773	1,67818	,14607
Stream_Spotify Female	169	3,3314	1,58385	,12183

Results revealed a sample size of **132 Males and 169 Females**. Also, that Males ($M_{\text{Males}}: 2,97, SD:1,67$) listened a bit less frequently than Females ($M_{\text{Females}}: 3,33, SD: 1,58$).

By checking the means can I assume that Females listen more frequently to Spotify than Males?

NO! NO! NO! NO!! NOOOOOOOO!!!!!!!

(Ok, that was a bit dramatic. But I just wanted to emphasize it)

Remember: You must check the result from the Independent Sample t-Test to verify **IF this difference between the means is statistically significant!**
If not, you cannot assume there is a difference in behavior!

And this is why we will have a look at the second table that SPSS will produce. So here are the steps to interpret it:

1. Verify the Levene’s Test for Equality of Variances: This is a test to check for the homogeneity of variance between the two groups. You simply need to remember the following:

- If the Levene’s test is **non-significant ($p > .05$)**, then the assumption **IS MET**
 - So interpret the significance value of the t-Test from the **TOP** row
- If the Levene’s test is **significant ($p < .05$)**, then the assumption **IS NOT MET**
 - So interpret the significance value of the t-Test from the **BOTTOM** row

2. Check the t-test significance value. In the case below, the Levene's test is non-significant ($p: .128$), thus the assumption test IS MET. So interpret the independent t-Test result from the TOP row ($p: .062$).

		1		Independent Samples Test		2				
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
								Lower	Upper	
Stream_Spotify	Equal variances assumed	2,324	,128	-1,875	299	,062	-,35409	,18886	-,72575	,01757
	Equal variances not assumed			-1,862	273,473	,064	-,35409	,19021	-,72855	,02037

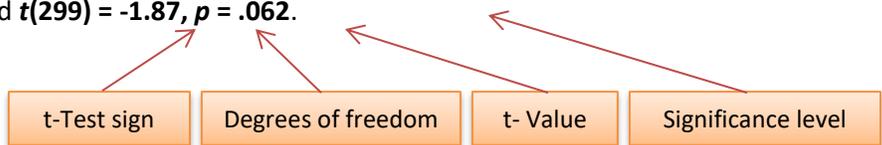
Soooooo, since the independent sample t-Test was non-significant ($p > .05$), what can we conclude???

Answer: THAT THERE IS **NO** STATISTICAL DIFFERENCE BETWEEN MALES AND FEMALES IN REGARDS TO THE FREQUENCY OF MUCSIC STREAMING ON SPOTIFY!
In other words, they stream just as much.

How to Report t-Test Results

t-Tests are reported with the t-test sign (t), degrees of freedom in parentheses, t-value and significance value.

Example: Results revealed that Males ($M_{Males}: 2,97, SD:1,67$) have similar streaming frequency to Females ($M_{Females}: 3,33, SD: 1,58$) in regards to streaming music through Spotify, as no significant difference was found $t(299) = -1.87, p = .062$.



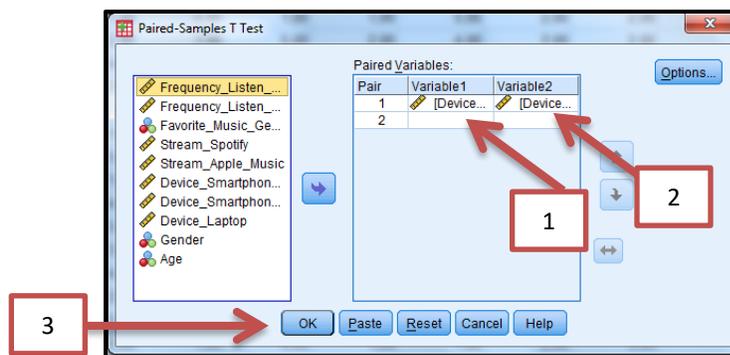
4.2 Dependent Sample t-Test (also called Paired Sample t-Test)

The example below will address the following question:

Question: When listening to music on smartphones, do people listen more frequently WITH or WITHOUT headphones?

Here are the steps to do it!

On SPSS Menu: Analyze > Compare Means > Paired Samples t-Test. After this, the following window will appear and all you will need to do is pair the two variables you wish to compare and click on “Ok”:



Here Are The Results:

Once again, the first table provides a general overview of both groups: The groups compared (Smartphones WITH and WITHOUT headphones), Sample size (N), means (in this case, “how frequently of listening to music), standard deviation and standard error.

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Pair 1 Device_Smartphone_WITH_Headphone	3,9535	301	1,17948	,06798
Device_Smartphone_WithOUT_Headphone	2,6146	301	1,21009	,06975

Results revealed a sample size of **301 participants**. Also, that respondents listened more frequently to music “with” headphones ($M_{\text{With}}: 3,95$, $SD:1,17$) than “without” headphones ($M_{\text{Without}}: 2,61$, $SD: 1,21$).

**AGAIN: Just by checking the means can I assume that people listen more to music more with headphone than without?
NOOOOOOOO!!!!!!!!!!**

Remember: You must check the result from the Independent Sample t-Test to verify **IF this difference is statistically significant! If not, you cannot assume there is a difference in behavior!**

And this is why we will have a look at the **third** table that SPSS will produce (*The **second** table shown is a paired correlation, which you do not need for mean comparison*). So here are the steps to interpret it:

Ps: In a dependent samples t-test (also called Paired sample t-test) there is no assumption of homogeneity of variance (as both groups are composed of the same participants). Thus, the table does not show a Levene’s test result and we do not need to worry about it.

What the table below does show is that the p-value is highly significant: **p=.000**.

Significance level

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Device_Smartphone_WITH_Headphone - Device_Smartphone_WithOUT_Headphone	1,33887	1,43229	,08256	1,17641	1,50133	16,218	300	,000

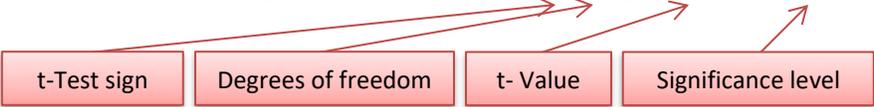
And so, what can we conclude from the results of the example above?

PARTICIPANTS STATISTICALLY LISTEN MORE FREQUENTLY TO MUSIC ON SMARTPHONES WITH HEADPHONES THAN WITHOUT!

How to Report t-Test Results

t-Tests are reported with the t-test sign (*t*), degrees of freedom in parentheses, t-value and significance value.

Example: Results revealed that participants significantly listen more frequently to music on smartphones with headphones ($M_{With}: 3,95, SD:1,17$) than without ($M_{Without}: 2,61, SD: 1,21$), $t(300) = -16.21, p = .000$.



STEP 5: Comparing Three (or more) Groups of Means

Who do you think would win a 100 meter sprint race: John Mayer (with a broken foot), Keith Richards (Completely sober) or Ed Sheeran (On a massive hangover)?

Answers: I think Keith Richards would just convince everyone that running was a bad idea. There would be no race. He would just take everyone to the pub.

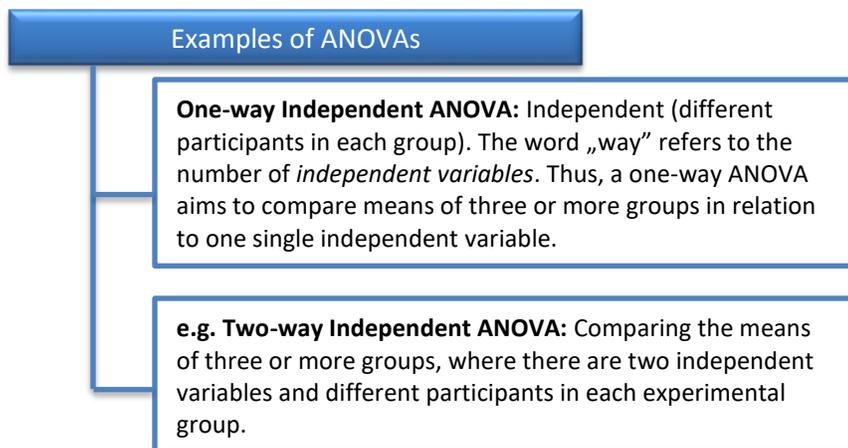
Ps: Please tell me you know who Keith Richards is...

So on the example above, the general idea is to compare three groups (**John Mayer, Keith Richards and Ed Sheeran**).

And the test to compare three (or more) groups of means to identify statistical differences among them has an absolutely beautiful name: **ANOVA**. Isn't it beautiful?

ANOVA: Analysis of variance (**AN**alysis **Of** **VA**riance) is a statistical test for comparing three (or more) groups of means from independent populations. So as you can notice, it is an extension of t-Tests.

Also, ANOVA's have usually a number before it, which refers to the number of independent variables that are being contrasted. So here are examples:



Since this is only a supporting material, we will focus on an example of a “One-way ANOVA”. Up next!

4.3 One-Way ANOVA (Analysis of Variance)

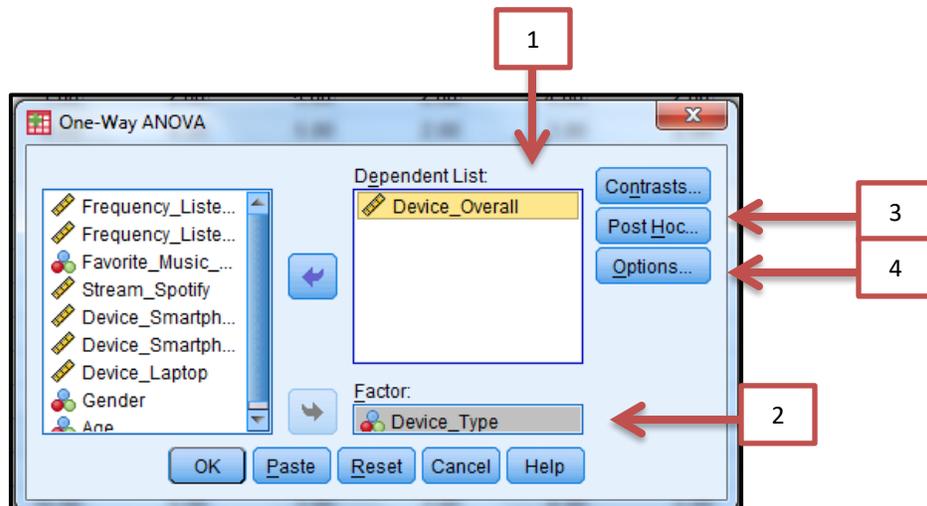
Alright, now let's take the previous example one step further. Now we will compare three groups of means. **And please keep in mind:** this is the same process if you were comparing 4, 5, 6 or more groups!

So for the example below, let's address the following research question:

Question: Do people listen to music more frequently on smartphones WITH headphones, on smartphones WITHOUT headphones or on laptops?

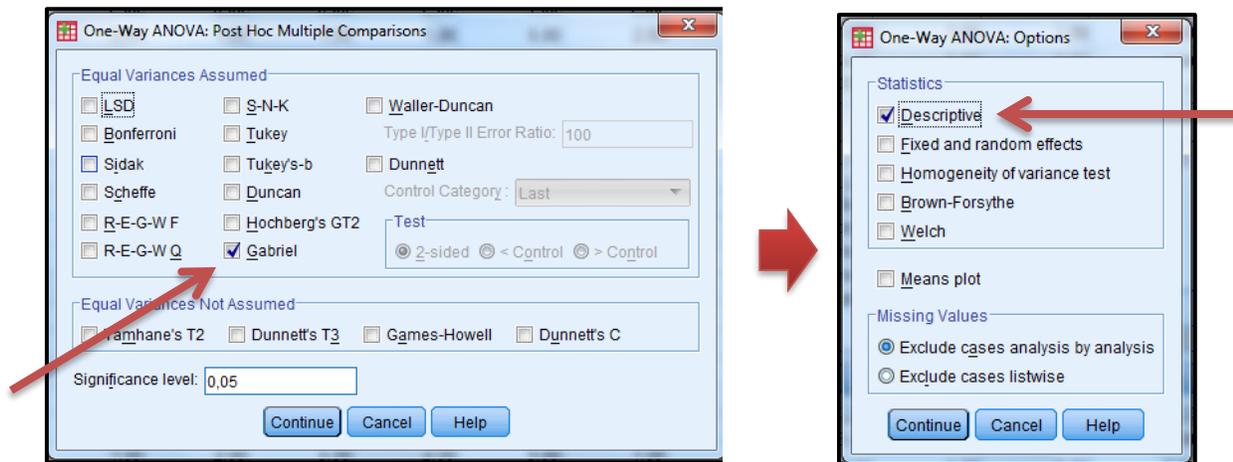
Here are the steps to do it!

On SPSS Menu: Analyze > Compare Means > One-Way ANOVA. After this, the table below will pop-up. Here, the first thing to do is to choose the "Dependent list", which represents the **variable (what was measured)** that you wish to contrast. In this case, we are comparing the "Device overall". After this, you must choose the "factor", which are the **groups** you wish to contrast.



After this, you should click on "Post-Hoc", where you will choose different types of multiple comparison tests. For the example here, since we have the same sample size in all three groups, let's choose the "Gabriel" test.

Also, click on "Options" and tick on "descriptive". This way the results will display the means, standard deviations, standard errors and more.



When you are done just click on “continue” on all windows and “Ok” in the initial window, and the results should appear.

Here Are The Results:

The first table displayed will be the “**Descriptives**”. It displays the **devices** (Smartphone with headphone, Smartphone without headphone and laptop), **sample size (N)**, **means and standard deviation** (These are the most important factors for us here!).

Remember, we measured on a scale of frequency, ranging from “**1-Never**” to “**5-Always**”. Thus, the higher the average, the more frequently people listen to music with such device.

Results shows us that listening to music on smartphone with headphone has the highest frequency (**M_{SH}**: 3,95, **SD**: 1,17), followed by laptop (**M_{SWH}**: 3,71, **SD**: 1,10), and lastly Smartphone without headphone (**M_{SWH}**: 2,61, **SD**: 1,21).

Descriptives

Device_Overall	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Smartphone With Headphone	301	3,9535	1,17948	,06798	3,8197	4,0873	1,00	5,00
Smartphone Without Headphone	301	2,6146	1,21009	,06975	2,4774	2,7519	1,00	5,00
Laptop	301	3,7176	1,10604	,06375	3,5922	3,8431	1,00	5,00
Total	903	3,4286	1,30288	,04336	3,3435	3,5137	1,00	5,00

And here we go AGAIN: Just by checking the means can I assume that people listen more to music more with headphone than without or on laptops?
NOOOOOOOO!!!!!!!!!!

Remember: You must check the result from the ANOVA test to verify **IF this difference is statistically significant! If not, you cannot assume there is a difference between the groups!**

And this is why will focus on the following table. The one with the name of the gorgeous girl: **ANOVA**.

As seen in the table below, the **ANOVA** result is **significant ($p < .05$)**. Thus, there is a statistical difference between the three groups!

ANOVA

Significance level
↓

Device_Overall

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	307,502	2	153,751	113,085	,000
Within Groups	1223,641	900	1,360		
Total	1531,143	902			

PS: If the ANOVA result was non-significant ($p > .05$) it would indicate the there was NO difference amongst the groups. Thus, we would **END** the analysis here and conclude that there was no difference in the frequency of listening on the three different devices.

IMPORTANT: Since the ANOVA result was significant ($p < .05$), the logical following question is: **Ok, but which group in relation to which group?**

- When running a t-test, differences can ONLY be one group in relation to the second.
 - Afterall, we are comparing only two groups!

But comparing three groups or more is different! When comparing three groups, for example, the differences could be found:

- (Group 1 x Group 2), (Group 2 x Group 3) or (Group 3 x Group 1).

And this is why we run a “**Post-Hoc**” test. Basically, **Post-Hoc** is a test that provides “**Multiple comparisons**” amongst the groups. In other words, if you are comparing 3 groups, it will run ALL possibilities:

- (Group 1 x Group 2), (Group 2 x Group 3) or (Group 3 x Group 1).

In our example, the Post-Hoc table displayed the results below. As you can see, it provides all possible combinations between the three groups.

In order to interpret the results, you must verify the significance column and interpret the values. In our case, **ALL comparisons are significant (p<.05)**. What does this mean?

- That there are significant differences amongst the three groups.

Multiple Comparisons **Significance level**

Device_Overall
Gabriel



(I) Device_Type	(J) Device_Type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Smartphone With Headphone	Smartphone Without Headphone	1,33887 [*]	,09505	,000	1,1115	1,5662
	Laptop	,23588 [*]	,09505	,039	,0085	,4632
Smartphone Without Headphone	Smartphone With Headphone	-1,33887 [*]	,09505	,000	-1,5662	-1,1115
	Laptop	-1,10299 [*]	,09505	,000	-1,3303	-,8756
Laptop	Smartphone With Headphone	-,23588 [*]	,09505	,039	-,4632	-,0085
	Smartphone Without Headphone	1,10299 [*]	,09505	,000	,8756	1,3303

*. The mean difference is significant at the 0.05 level.

So what can we conclude from the results?

Answer: The most common way of listening to music is through their smartphones with headphones, followed by on laptops and lastly on smartphones without headphones.

ANOVAs (both one-way and two-way) are reported like *t*-tests, but there are two degrees-of-freedom numbers to report. First report the between-groups degrees of freedom, and then report the within-groups of freedom (separated by a comma). After that report the *F* statistic (rounded off to two decimal places) and the significance level:



Example: There was a significant effect for frequency of use of devices, $F(2,900) = 113.08, p = .000$, and a significant interaction between “*Smartphone with headphones*” and “*Smartphone without headphones*”, $p = .000$, “*Smartphone without headphones*” and “*laptops*”, $p = .039$ and “*laptops*” and “*Smartphone with headphones*” $p = .000$.

The End



Is your **curiosity** alive?

Live**Innovation**.org