

# The Melody of **STATISTICS.**

**SPSS**  
**Edition**

A practical guide on how to conduct  
and interpret statistical tests like a  
rockstar using IBM SPSS.

Written and designed by  
**Francisco Tigre Moura**  
**Silke Jütte**



LIVEINNOVATION.ORG

This book is dedicated to all students with curious minds.  
Or to anyone who has banged their heads in rock concerts.  
Or while trying to learn statistics.

**Francisco and Silke.**

## **Version of the e-book**

This edition was  
published on  
**01.07.2022**

Visit [LiveInnovation.org](https://LiveInnovation.org) to download the latest version.

**LiveInnovation.org** is an independent academic initiative. The mission is to develop and share research and educational content. The main areas of interest involve marketing, creativity and technology experiences, mostly applied to contexts of live experiences.

# WHO IS IN THE BAND?



## **PROF. DR. FRANCISCO TIGRE MOURA**

Ph.D. in Marketing (University of Otago, New Zealand). Currently a Professor of Marketing at IU University of Applied Sciences (Germany). Passionate about everything that involves Music, Marketing, Technology and the Future.

**[francisco.tigre-moura@iu.org](mailto:francisco.tigre-moura@iu.org)**



## **PROF. DR. SILKE JÜTTE**

Ph.D. in Supply Chain Management (University of Cologne, Germany). Currently working as a Professor of Quantitative Methods at IU University of Applied Sciences (Germany). Passionate about singing, crafting and hanging out on football grounds (cheering for her sons).

**[silke.juette@iu.org](mailto:silke.juette@iu.org)**



Connect with us on LinkedIn!

# MAIN SETLIST

THE SOUNDTRACK TO BEAUTIFUL NUMBERS

- 1 STOP CRYING YOUR HEART OUT (Oasis)  
**Initial considerations**
- 3 ONES AND ZEROS (Jack Johnson)  
**Types of quantitative variables**
- 11 HOW OFTEN (Ben E. King)  
**How to conduct descriptive statistics**
- 18 NOTHING ELSE MATTERS (Metallica)  
**Understanding statistical significance**
- 22 NO ASSOCIATION (Silverchair)  
**How to conduct a Chi-Square test**
- 28 MUTUAL MOTION (Loure)  
**How to conduct a correlation test**
- 34 JUST THE TWO OF US (Bill Withers)  
**How to conduct a t-Test**
- 44 THREE LITTLE BIRDS (Bob Marley and The Wailers)  
**How to conduct an ANOVA**
- 55 TRUST ME (Janis Joplin)  
**How to calculate Cronbach's alpha (Reliability)**
- 63 WE ARE THE CHAMPIONS (Queen)  
**The end**

**Follow and listen to  
the full playlist  
of the book  
on Spotify.**



Stop crying your heart out (Oasis)

1. Ones and Zeros (Jack Johnson)
2. Little numbers (BOY)
3. Variables (The New Raemon)
4. Nominal (#1 Dads)
5. Thirty-three (Smashing Pumpkins)
  
1. How Often (Ben E. King)
2. What's the Frequency, Kenneth? (R.E.M.)
3. Average (Sasupt)
4. Standards (Leslie Odom Jr.)
5. Means something (Lizzy McAlpine)

Nothing else matters (Metallica)

1. No association (Silverchair)
2. Related (Kingdom of birds)
3. Free association (Gordi)
4. Square 1 (Mahalia)
5. No cause for alarm (Just Jack)
  
1. Mutual motion (Loure)
2. It takes two to tango (Louis Armstrong)
3. Up and Down (CLOVES)
4. Correlation (Ginne Marker)
5. Opposite directions (Neil Erickson)
  
1. Just the two of us (Bill Withers)
2. Two birds (Regina Spektor)
3. Two against one (Danger Mouse)
4. Nothing compares 2 U (Chris Cornell)
5. Two of us (The Beatles)

1. Three little birds (Bob Marley and The Wailers)
2. Three strikes (HONNE)
3. Thirty three (The Smashing Pumpkins)
4. 4:44 (Jay-Z)
5. 3. Stock (AnnenMayKantereit)

1. Trust me (Janis Joplin)
2. Trust in me (Scarlett Johansson)
3. Reliable (Golden Idols)
4. I'm a believer (The Monkees)
5. Believe what I say (Kanye West)

We are the champions (Queen)

**The  
Melody of  
STATISTICS.**

**SPSS**  
**Edition**

# Ah... Follow and watch the full **SPSS** tutorial video series on YouTube.



The screenshot shows a YouTube channel page for 'SPSS TUTORIAL'. The channel has 12 videos with 7,725 views and was last updated on May 6, 2020. The channel description reads: 'An easy and objective SPSS tutorial for Marketing and Management students.' A 'SUBSCRIBE' button is visible. The 12 video thumbnails are listed below:

- 1. SPSS Tutorial #1: Introduction to SPSS
- 2. SPSS Tutorial: #2: Importing a Survey Questionnaire to SPSS
- 3. SPSS Tutorial #3: Percentages and Averages (Descriptive statistics)
- 4. SPSS Tutorial #4: Dependent (Paired) Samples t-Test
- 5. SPSS Tutorial #5: Independent Samples t-Test
- 6. SPSS Tutorial #6: One-Way ANOVA
- 7. SPSS Tutorial #7: Cronbach's Alpha (Reliability test)
- 8. SPSS Tutorial #8: Correlation
- 9. SPSS Tutorial #9: Exporting SPSS Output to Excel, Word or PDF
- 10. SPSS Tutorial #10: Chi-Square
- 11. SPSS Tutorial #11: Factor Analysis/PCA

## What will you find?

1. Introduction to SPSS
2. Importing a questionnaire to SPSS
3. Percentages and averages
4. Dependent (Paired) samples t-test
5. Independent samples t-test
6. One-Way ANOVA
7. Cronbach's Alpha (reliability test)
8. Correlation
9. Exporting SPSS output to Excel
10. Chi-Square
11. Factor Analysis / PCA
12. MANOVA

# The Melody of **STATISTICS.**

# DISCLAIMER

Hello my friend, how are you? We are really glad you are here.

So if you reached this e-book we can only assume one thing: **for some weird or unfortunate reason in life, you have to conduct a few statistical tests on IBM SPSS, but you have no idea how to do them.** And this is freaking you out.

Is this a fairly accurate description?

Well, well, well... Isn't this your lucky day? Do not worry, my dear friend. We hope to help you out. We are going to explain to you how to conduct a few tests on **IBM SPSS Statistics** and also how to interpret them.

**But before we start, let us make something very clear:**

**This  
is not a  
statistics book.**

**This is a quick guide to help you conduct a few tests.  
This is a quick guide to help you conduct a few tests.  
This is a quick guide to help you conduct a few tests.**

It will surely be enough for you, depending on what your needs are and the level of the project you are involved in.

**For a more in-depth understanding,  
you will need statistics books.**

Alright, now that we are clear and on the same page, we feel like we can get the music going. Shall we rock on to the introduction stage?

# STOP CRYING YOUR HEART OUT

**Artist:** Oasis

## WELCOME!

My friend, do you play guitar? No? (Honestly, what is happening to humanity...) Ok, but have you ever tried to learn? If you haven't, let us tell you how it goes: **The first six months are terrible!**

You will feel like the dumbest person on the planet as you try to move from a simple "G major" chord to an "A major" chord, then to a "D major" chord (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, we can guarantee you that statistics is pretty amazing! (Ok, maybe not enough to impress someone during a date and it may ruin your social skills. But other than that, it's great!) So our job here is to show you an initial view of statistics and how it can be **EASY** and pretty **FUN!**

Hopefully, by the end you will have fallen in love with numbers!

**BUT PLEASE KEEP IN MIND:** This e-book is only a quick guide to support students during research courses or while conducting their own research. For a more in-depth understanding of tests, we highly recommend you to complement it with statistics and research methods books!



So for further explanations and more in-depth understanding of the topics, we would personally recommend the following amazing literature:

- Field, A. Discovering Statistics Using SPSS. Sage.
  - Field, A. & Hole, G. How to Design and Report Experiments. Sage.
  - Field, A. An Adventure in Statistics: The Reality Enigma. Sage.
- Especially this one!**

As you can see, we only recommend books from **Andy Field**.

Why? Well, simply because, in our modest opinion, he is by far the best author on the topic. And... like us, he is also a huge music fan and tries to use musical examples in almost all of his publications as well.

Sounds pretty good, right?

So yeah, welcome to the show, we hope that this publication will make your life a bit easier, facilitate your analysis and preserve your sanity! Oh, we also hope you like the music!

**And... you can try for yourself every example shown in this e-guide!**  
Download the data sets at [LiveInnovation.org](http://LiveInnovation.org) and the stage is yours.

**Are  
you ready to  
start listening to...**



# The Melody of **STATISTICS?**



# ONES AND ZEROS

Artist: Jack Johnson

## Types of quantitative variables.

And how to insert them on **IBM SPSS**.



1

Before we start having a great time analyzing data (please don't judge us for saying this- and yes, we 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 we mean?).

Look at these numbers: 1, 2, 3, 4, 5, 6... They are just **numbers**, right? Yup. Good!

But from now on we will call them by a different name: DATA. And look how interesting: **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 we are going to repeat, ok? The type of data will depend on THEIR PURPOSE and HOW THEY WERE MEASURED.

Ok, now we will repeat, but in bold, all caps, italics and in orange (again) hoping that the incredible visual effect will help you memorize: **THE TYPE OF DATA WILL DEPEND ON THEIR PURPOSE AND HOW THEY WERE MEASURED.**

And what are the different types? Let's have a look at the four main types of VARIABLES and the main things you need to know:

# What are the **types of variables?**

## Categorical Variables

### Nominal data

**Description:** A number which simply represents or indicates an object of investigation. It does NOT represent the magnitude of what is being measured (e.g. Types of music, with 1=jazz, 2=pop, 3=hip hop, etc.). You CANNOT calculate with these numbers.  
(Honestly:  $1 \text{ (jazz)} + 2 \text{ (pop)} = 3 \text{ (hip hop)}$ ?! Surely not!)

### Ordinal data

**Description:** Data which provides information regarding the order in which a phenomenon has occurred. However, it does not inform about the distance between the measured values (e.g. 1=soprano, 2=alto, 3=tenor, 4=bass, which are ordered by the pitch of the voices). You still CANNOT calculate with these numbers.

## Quantitative Variables

### Interval data

**Description:** Similar to ordinal except the intervals between each value are equally divided. Most importantly, the values do represent the magnitude of what is being measured (and hence, we are now allowed to add and subtract values). However, the 0 point of scales, in theory, is unknown (e.g. **Likert-type scales**).

### Ratio data

**Description:** It has the same properties as interval data, but there is an absolute 0 (e.g. Length of a song in minutes and seconds). As a consequence, you can now also calculate ratios (e.g. one song being 20% longer than another one).



# Likert-type Scale Alert!

## Ok, ok, ok...

We know we told you in the beginning that this is NOT A STATISTICS BOOK. And it is not. However, we feel like we need to give you a bit more information on the incredible **Likert-type scales**.

## So...

Essentially, **Likert-type scales** are rating scales which contain a series of "anchors" (displayed numerically or in words) that allow numerical measurements of an item or question.

For example:

Fully disagree - Disagree - Indifferent - Agree - Fully agree

1 Never - 2 Rarely - 3 Sometimes - 4 Often - 5 Always

**And normally we classify Likert-type scales as interval scales!**

## HOWEVER...



**Strictly speaking**, for **Likert-scale data** to be considered interval data, we have to add a numeric scale to the textual descriptions of the scale. E.g. "fully agree = 100%, agree = 80%, rather agree = 60%", etc. Otherwise, we cannot be sure the difference between e.g. "fully agree" and "agree" is the same as between "agree" and "rather agree" for all participants of a survey...

**And let me tell you: in the world of statistics, this discussion creates a massive debate!**

In case you are curious to know more about this intriguing discussion and gossip in the world of statistics, I suggest reading the following articles:

- Bishop, P. A., & Herron, R. L. (2015). Use and misuse of the Likert item responses and other ordinal measures. International journal of exercise science, 8(3), 297.
- Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. Medical education, 42(12), 1150-1152.

## HOW SATISFIED ARE YOU WITH THIS EXPLANATION ON LIKERT-TYPE SCALES?

1 Very dissatisfied - 2 Dissatisfied - 3 Indifferent - 4 Satisfied - 5 Very satisfied

# NOW THAT YOU ARE THE MASTER OF VARIABLES, LET ME ASK YOU A QUESTION:

**What is the first thing you do when you grab a guitar to play a song?** You check to see if it is in tune, right? If it isn't, you tune it. In other words, don't ever just start playing the guitar without having it tuned. It will sound horrible. So you do your preparation first, then the fun begins.

**In statistics we do exactly the same.**

**Before any analysis, the first step is always to format and organize your data.**

And what is that exactly? Well, you need to create all variables from your questionnaire on the software you are using. You tell the software all the information it needs and insert the data properly.

This is "fine tuning before playing a song" in statistics. Got it?

In the case of **IBM SPSS**, the very first thing you should do when you have all your data and conduct any type of analysis is to **PREPARE THE VARIABLE SHEET**.

This is an extremely important step, so that your tests can run properly and you can interpret your findings later on accurately. Another way to take it, is as if you are preparing a house before receiving guests (weird analogy, but we hope you get it).

**So every time you have all your data,  
play some good music in the background,  
pour yourself a nice drink and sit cozy somewhere. Then...  
prepare the variable sheet on your software with all details you need.**



**HERE IS HOW  
TO DO IT!**



**1**

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

**A) Name of variable**

**B) Label of values**

**C) Define the measure**

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	Q1_Location	Numeric	8	2	In which countr...	{1,00, Germ...	None	8	Right	Nominal	Input
2	Q2_Frequen...	Numeric	8	2	Roughly, how ...	{1,00, I did ...	None	8	Right	Nominal	Input
3	Q3_Covid19	Numeric	8	2	How concerned...	{1,00, None ...	None	8	Right	Scale	Input
4	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
5	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
6	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
7	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
8	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
9	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
10	Q4_attendin...	Numeric	8	2	Attending musi...	{1,00, Fully ...	None	8	Right	Scale	Input
11	Q5_Organiz...	Numeric	8	2	Organization of ...	{1,00, Fully ...	None	8	Right	Scale	Input
12	Q5_Organiz...	Numeric	8	2	Organization of ...	{1,00, Fully ...	None	8	Right	Scale	Input
13	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
14	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
15	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
16	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
17	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
18	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
19	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
20	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
21	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input
22	Q7_musicig...	Numeric	8	2	What are your ...	{1,00, FAVORI...	None	8	Right	Nominal	Input

**Variable view**

**2**

Once you are on the “**variable view**”, the most important elements to edit here are: “name of variable”, “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:** change back from “Favorite\_Music\_Genre” to Favorite Music Genre.

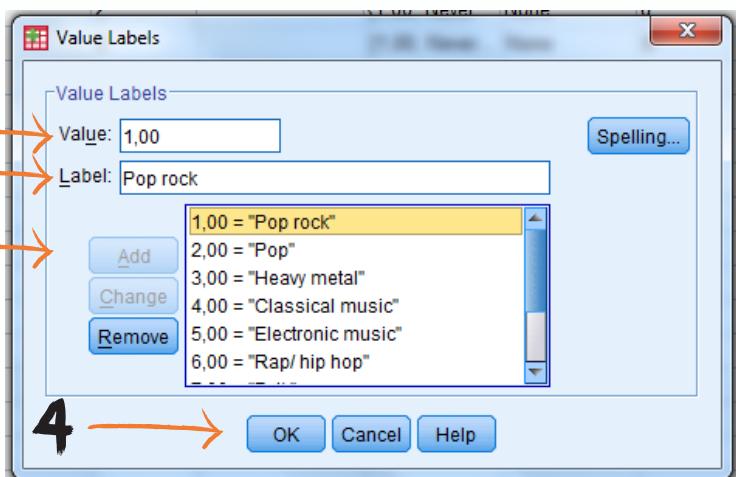
**B)**

**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, click on "ok".



1. You add the value
2. You write the label of the value
3. Click "add".
4. Once all values are inserted, click on "ok".

1  
2  
3

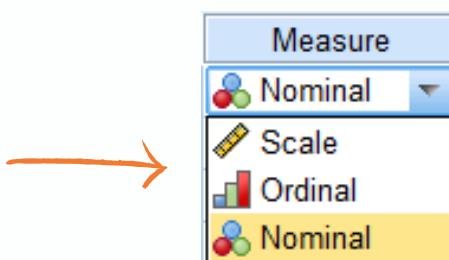
## ATTENTION!

In case your variables are **quantitative**, and therefore your measure is "**scale**", I suggest the following:

- If using **Likert-type scales**, include the label and value of each anchor (e.g. 1 = Fully disagree, 2= Disagree).
- If you have **ratio variables** (total number of something), there is no need to insert labels.

**C)**

**Define measure:** For each variable, you should define what type of data it is. So once you click on "Measure", SPSS will give you three options: Nominal, ordinal or scale (Interval or ratio) .



**Now, you might ask yourself, which measure is my variable?**

Well, generally:

- All your **multiple choice** questions will be "**Nominal**".
- All your **ranking** questions will be "**Ordinal**".
- All your **Likert-Type scales** will be "**Scale**".

**3**

Once all of your variables have been properly labelled, you have defined values and measures on the "variable view", now it is time to move to the "**Data view**" and insert all your data into SPSS!

**All variables you created on "variable view".**

	Q1_Location	Q2_Frequency_of_Attendance	Q3_Covid_19	Q4_attending_musicalconcert	Q5_Organization_concerts	Q5_Organization_concerts	Q7_music_genre_Classical	Q7_music_genre_Country							
1	1,00	2,00	4,00	2,00	2,00	2,00	3,00	3,00	4,00	3,00	4,00	4,00	4,00	2,00	2,00
2	1,00	2,00	4,00	2,00	2,00	1,00	3,00	2,00	4,00	3,00	4,00	4,00	4,00	2,00	2,00
3	1,00	2,00	4,00	2,00	2,00	2,00	3,00	2,00	4,00	3,00	2,00	3,00	2,00	2,00	2,00
4	1,00	2,00	4,00	2,00	2,00	2,00	3,00	3,00	3,00	3,00	4,00	3,00	2,00	2,00	2,00
5	1,00	2,00	4,00	2,00	2,00	2,00	3,00	3,00	3,00	2,00	4,00	3,00	2,00	2,00	2,00
6	1,00	1,00	4,00	2,00	2,00	2,00	3,00	3,00	4,00	3,00	4,00	4,00	4,00	2,00	2,00
7	1,00	2,00	4,00	2,00	2,00	1,00	2,00	1,00	3,00	2,00	3,00	3,00	3,00	2,00	2,00
8	1,00	4,00	2,00	4,00	4,00	2,00	4,00	2,00	2,00	2,00	4,00	2,00	2,00	2,00	2,00
9	1,00	3,00	4,00	2,00	1,00	1,00	2,00	3,00	4,00	2,00	3,00	5,00	2,00	2,00	2,00
10	1,00	2,00	4,00	2,00	1,00	1,00	2,00	2,00	4,00	2,00	4,00	4,00	4,00	2,00	2,00
11	1,00	2,00	3,00	2,00	2,00	2,00	1,00	2,00	5,00	3,00	3,00	4,00	2,00	2,00	2,00
12	1,00	3,00	4,00	2,00	2,00	2,00	4,00	4,00	3,00	2,00	2,00	5,00	2,00	2,00	2,00
13	1,00	2,00	3,00	2,00	2,00	3,00	2,00	4,00	4,00	4,00	3,00	3,00	2,00	2,00	2,00
14	1,00	2,00	5,00	1,00	1,00	1,00	3,00	5,00	5,00	3,00	5,00	5,00	2,00	2,00	2,00
15	1,00	3,00	1,00	4,00	4,00	5,00	3,00	4,00	2,00	3,00	5,00	5,00	2,00	2,00	2,00
16	1,00	2,00	4,00	3,00	3,00	4,00	3,00	3,00	4,00	4,00	4,00	4,00	4,00	2,00	2,00
17	1,00	2,00	4,00	2,00	2,00	2,00	2,00	2,00	5,00	2,00	4,00	4,00	4,00	2,00	2,00
18	1,00	2,00	4,00	2,00	2,00	1,00	3,00	3,00	3,00	2,00	4,00	4,00	4,00	2,00	2,00
19	1,00	3,00	4,00	2,00	2,00	2,00	2,00	3,00	2,00	1,00	4,00	5,00	2,00	2,00	2,00

Data View      Variable View

**A) Data view**

**B) INSERT DATA!**

**A)**

**Data view:** Once you click on "data view", you will see that all variables you have created on "variable view" are displayed on the top.

**B)**

**Insert data:** Alright, now that you are on the data view and all variables are created, simply insert all the data. You can do this simply by doing the classic "**Ctrl+C** **Ctrl+V**" (copy and paste), from Excel or any other place where your data is.

**4**

You are done with the preparation of your data set! You have tuned the guitar and now you can play. Oh, wait... there is one important step missing: **SAVE YOUR FILE! SAVE YOUR FILE!**



1. Click on "File"
2. Click on "Save as..."
3. DONE.



# Alright, great stuff!

So by now you should have understood:

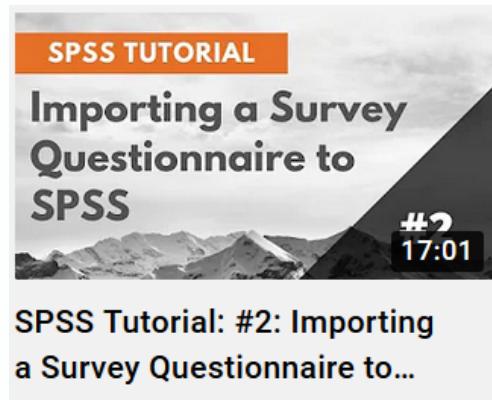


- 1 The different types of variables.**
- 2 The importance of always preparing the software prior to an analysis.**
- 3 How to insert variables and all your data into SPSS.**



## SPSS Tutorial playlist series.

In case you feel like you still need some further support on **how to insert your data into SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of [LiveInnovation.org](https://www.youtube.com/user/LiveInnovation).



## Full soundtrack for Types of variables

1. Ones and Zeros (Jack Johnson)
2. Little numbers (BOY)
3. Variables (The New Raemon)
4. Nominal (#1 Dads)
5. Thirty-three (Smashing Pumpkins)



# HOW OFTEN

Artist: Ben E. King

# Descriptive statistics.

And how to conduct it on **IBM SPSS**.



# Thinking Out Loud, during music festivals.

I am not going to lie: I love "people watching". My friend, don't judge. It's just that observing the weird behaviors of others can be much more entertaining than watching a movie, traveling or doing underwater camping.

I don't know, it's just that sometimes people are simply so interestingly strange. And hey, there is no better place to go people watching than music festivals. I think it is the mixture of a sense of freedom, combined with music, alcohol and other questionable substances. It all simply makes people go wild!

I remember the last time I visited a festival, there was so much to see: people fighting, couples breaking up, others vomiting, some falling on the dirt, others clearly having sex in tents (we could all tell by the noises), some being taken by the medical team...

And I haven't even mentioned the music. Clearly the fans of the different genres had opposite tastes. They dressed differently and the songs contained very different words. I had never heard so often the word F\*\*\*\* being sung on songs. Is this a new trend? Crazy.

I mean, I will tell you: for someone like me that enjoys "people watching", attending music festivals is a treat!

But all those crazy scenes made me wonder if I could actually calculate those behaviors, and understand how frequently they happened on that festival. For example:

- 
- Q1. Which percentage of visitors vomited during the festival?**
  - Q2. Which percentage of couples broke up during the festival?**
  - Q3. What was the average time of sex in the tents?**
  - Q4. On average, how often did bands say the word F\*\*\* on their songs?**

These questions were way too interesting to be left alone. I had to find a way to answer them. So I started wondering if there are any descriptive statistics tests that could provide me with a general overview of the audience...



# Wait... What is Descriptive Statistics?

**Another disclaimer:** Dear friend, we made it quite clear at the start: [this is not a statistics book](#). So, again, we will not provide a detailed explanation of the mathematical rationale of the tests. Here, we will only provide a brief explanation, enough for you to remember what the test is all about. And importantly, how to conduct and interpret it on SPSS.

## Hope we are still friends?

Great! Glad to know we are clear and still friends. So then let us us get back to it.

So... In general terms, descriptive statistics are analyses, which (as the name "surprisingly" suggests) allow us to [describe](#), or [summarize](#), patterns from a data set we have collected. It is pretty simple tests, nothing too fancy.

Some examples of descriptive statistics tests are: [frequency distributions](#) (percentages), [averages](#) (means), [mode](#) and [median](#).

**Perfect!** Now that we got that clear, shall we look at how to answer those four questions about crazy behaviors during music festivals?



## Frequency distribution (percentage)

The first two questions regarding the festival that we would like to answer are:

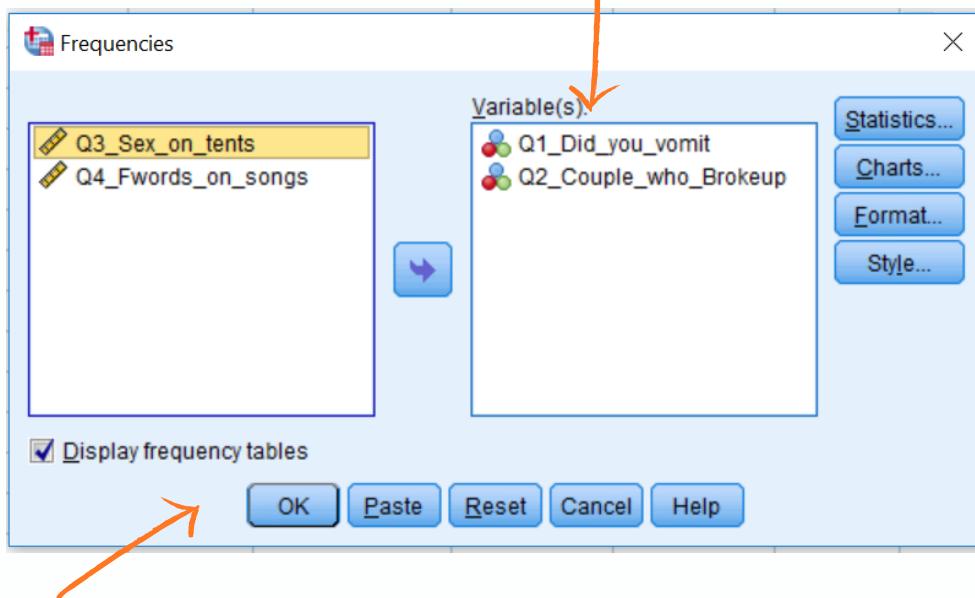
- Q1. Which percentage of visitors vomited during the festival?**
- Q2. Which percentage of couples broke up during the festival?**

And how will we answer them? by running frequencies (percentages).

And here is how to do it:

1

On the top menu of SPSS, you must click on "**Analyze**", then "**Descriptive Statistics**", then "**Frequencies**". The window shown below will appear, then, simply choose the variables that you wish to analyze ("Did you vomit" and "Couple who broke up"), drag them to the box on the right and click on "ok".



## Results

So the first window we see gives an overview of all responses. **N** always refers to the **sample size**. In this case, for both questions, there were 75 respondents (75 people that were at the festival).

Statistics		
	Q1_Did_you_vomit	Q2_Couple_who_Brokeup
N	75	75
Missing	0	0

The second table shows the **frequency** of respondents (the total number of answers per category) and the **percent** (the percentage of respondents from each category in relation to the total of respondents). Ex: out of the 75 respondents, 48 (or 64%) had vomited during the festival. On the positive note, only 2 respondents (or 2.7%) broke up after the festival.

Q1_Did_you_vomit			Q2_Couple_who_Brokeup		
	Frequency	Percent	Valid	Frequency	Percent
Valid	Yes, vomited	48	64,0	73	97,3
	Nope, no vomit	27	36,0	We are still together	2,7
Total		75	100,0	Total	100,0

So here are the answers to our questions:

**Q1. Which percentage of visitors vomited during the festival?**

- o **Answer: 64%.**

**Q2. Which percentage of couples broke up during the festival?**

- o **Answer: 2.7%.**

## Averages (means)

Alright, now that we know that people are throwing up frequently, but still staying together with their loved ones, let us answer the remaining two questions:

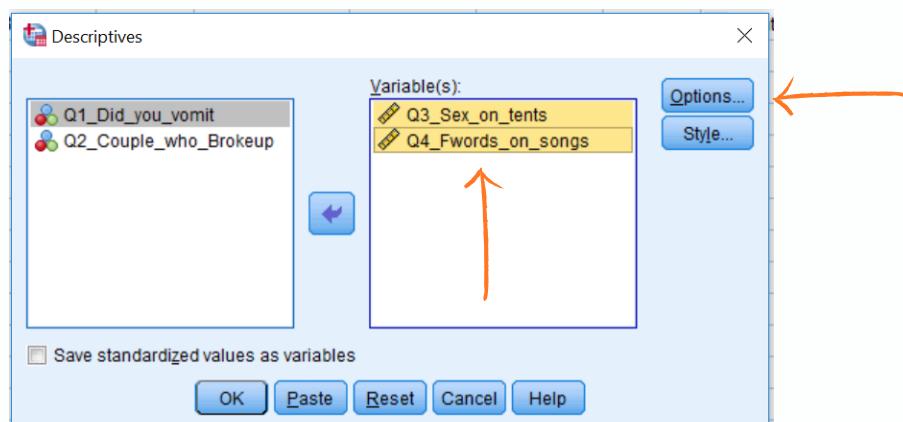
**Q3. What was the average time of sex in the tents?**

**Q4. On average, how often did bands say the word F\*\*\* on their songs?**

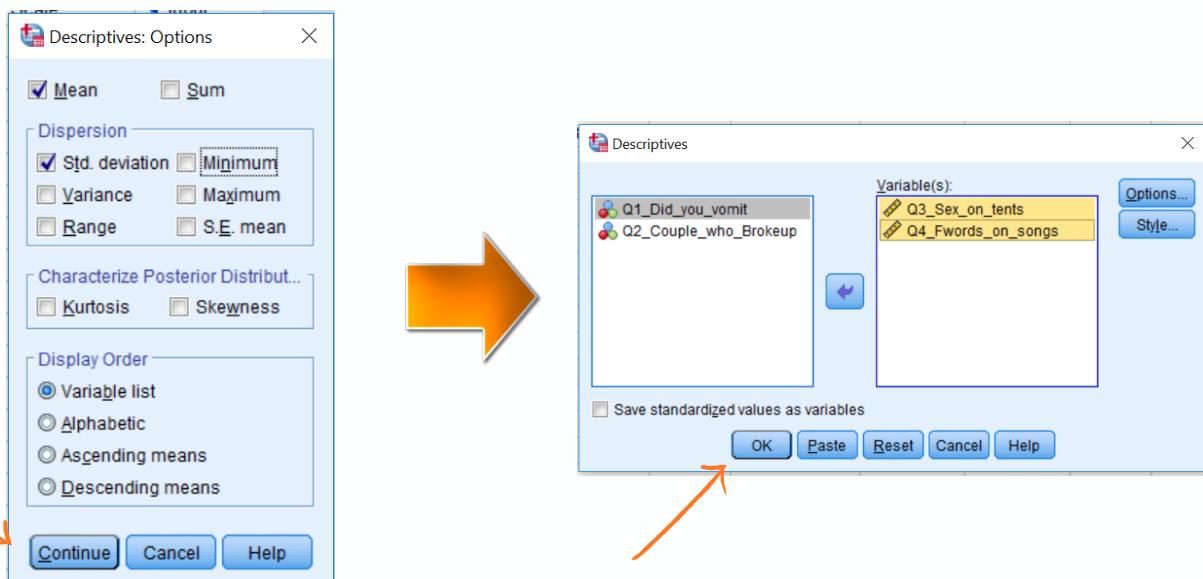
The **mean** is the average of the data. The **standard deviation** measures the spread of the data. For example, imagine you asked five people about how many months they have been playing the guitar. If their answers are: 2, 1, 2, 2, 2, this is a rather small spread of the data, so the standard deviation will be close to zero here. If their answers are: 5, 0, 17, 3, 15, this is a much larger spread of the data, so the standard deviation will be a lot larger than zero.

2

On the top menu of SPSS, you must click on "**Analyze**", then "**Descriptive Statistics**", then "**Descriptives**". The window shown below will appear, then, simply choose the variables that you wish to analyze ("Sex on tents" and "Fwords on songs"), drag them to the box on the right and click on "**Options**".



After clicking on options, a new window will appear. In it, 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**".



## Results

So, again, the first important result we see is an overview of all responses. **N** always refers to the **sample size**. In this case, for both questions, there were 75 valid respondents.

**And look how interesting:** on average people had sex on tents for 11.9 minutes. But importantly, the standard deviation was of an incredible 10.3 minutes. Which indicates a very heterogeneous data. This means that many were having sex for only 1 minute (!!), while others for longer than 20 minutes (now that is more acceptable, right?).

Descriptive Statistics			
	N	Mean	Std. Deviation
Q3_Sex_on_tents	75	11,9067	10,39123
Q4_Fwords_on_songs	75	14,5067	21,50196
Valid N (listwise)	75		

**Shockingly**, there was an average of 14.5 times the use of the F\*\*\* word during songs! Wow, no wonder I thought they could be more creative. If you are using the "F" word so often, you almost have no time to say anything else (I wonder what Shakespeare would think of their vocabulary).

Also, to my relief, the standard deviation is pretty high again. So, while some bands apparently use hardly anything else than F\*\*\* in their lyrics, lucky enough there are still others who use this word only rarely!



# Important reminder!

In the SPSS variable view, variables are usually coded or written with underlines, such as "Sex\_on\_tents", for example. Again, this is because SPSS does not allow for spaces between words when creating variables.

**HOWEVER**, when reporting results on your paper, thesis, report or any other format, make sure to **ALWAYS** write the name of the variable correctly. **NOT** the abbreviated versions with underlines.

So in the example before, instead of writing on your report "Sex\_on\_tents", for example. you would write the accurate name of the variable. So something like: "Sexual intercourse on tents during the music festival".

**Got it? Got it? Got it?**



**SPSS Tutorial  
playlist series.**

In case you feel like you still need some further support on how to run **frequencies or means on SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of **LiveInnovation.org**.



## Full soundtrack for Descriptive statistics

1. How Often (Ben E. King)
2. What's the Frequency, Kenneth? (R.E.M.)
3. Average (Sasupt)
4. Standards (Leslie Odom Jr.)
5. Means something (Lizzy McAlpine)



# NOTHING ELSE MATTERS

Artist: Metallica

# Understanding Statistical Significance!



LIVEINNOVATION.ORG





# Statistical Significance Alert!

**Read this section carefully!**

If needed, read again. And again. And again... Until the end of time.

My dear friend, up to now you only saw tests which simply **described the data**: averages, frequencies, standard deviations. But from now on we will discuss more sophisticated tests. The tests you will see next will try to use the sample to deduce conclusions in regard to the population of the study. We call this:

## Inferential statistics!

**In short: with inferential statistics we analyze a data set which represents a sample of the population, and based on it, derive conclusions for the entire population.**

**But if we aim to derive conclusions for the entire population, we might be wrong (just because we only know what happened in the sample, so what is "outside the sample" we cannot control). So we can never be 100% sure that our conclusion is correct. But still, we want to be "nearly 100%" sure.**

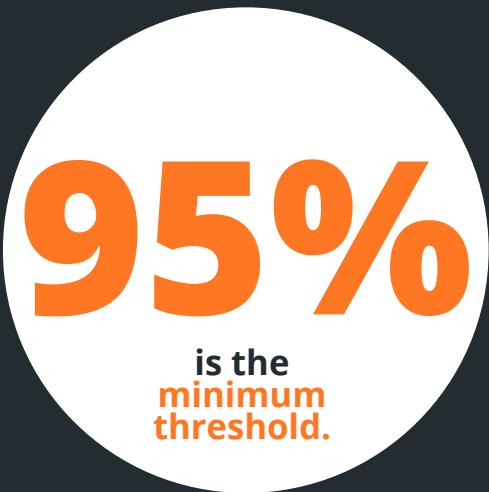
## And how can we try to be sure ?

Well, in inferential statistics tests, there is always a very important value which is reported: **the significance value! And it is represented by the value "p".**

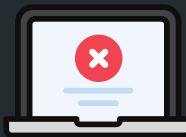
**In short, the significance value tells us the extent to which we can be sure of what we are testing. So it is a very important value. Without it, we cannot assume any inferential conclusion!**

# And how sure do you need to be to assume inferential conclusions?

In science or the industry, we work with a **MINIMUM threshold** of:



If we are **95% OR MORE** sure, we **DO** assume inferential conclusions.



If we have **LESS than 95%** assurance, we **DO NOT** assume inferential conclusions.

## And how do you know if you have reached the threshold or not?



**VERY IMPORTANT:** All significance values are reported on a scale ranging from 0-1, and they tell us the amount of randomness (and thus, less assurance) we have on the results. **Since our threshold of assurance is 95%, the maximum amount of randomness we allow on results is 5%.**

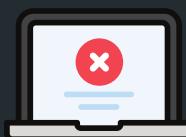
Therefore... if our significance value (or "p-value") is:

**Lower  
than 0.05**



**We have less than 5% randomness.** Thus, we are **95% OR MORE** sure. Consequently, we **DO** assume inferential conclusions.

**Higher  
than 0.05**



**We have more than 5% randomness.** Thus, we are **LESS than 95% sure**. Consequently, we **DO NOT** assume inferential conclusions.

# MY FRIEND, IF YOU HAVEN'T FULLY UNDERSTOOD, WHY ARE YOU MOVING FORWARD?

Come on! Go back and read it again until  
you fully understand.

I am waiting...



# NO ASSOCIATION

Artist: Silverchair

# Chi-Square test.

And how to conduct it on **IBM SPSS**.



# The death of you and me.

(This is the title of a wonderful song from Noel Gallagher)

My friend, have you ever heard of the class of 27? Kurt Cobain, Janis Joplin, Amy Winehouse, Jimi Hendrix... All of them died at the age of 27, unfortunately due to pretty sad reasons.

So I started researching on the [fictitious sources of my own imagination](#), and my friend, I found some pretty interesting data. You will not believe: musicians are currently dying for the weirdest reasons!

**Here are the Top 4 main reasons for deaths of musicians in 2021:**

1. Instagram overdose
2. Excessive Tik-Toking
3. Uncontrolled Call of Duty use
4. Missing texts from their mothers.

Yes, this is a strange generation.

But anyway, I wanted to go further. [I wanted to understand if musicians from certain genres were more likely to die from some of these causes](#). So I focused on four music genres: rap, heavy metal, jazz and electronic music.

So with this incredible scenario in mind, I became obsessed in knowing one thing:



**Q5. Is the music genre associated with the cause of death?**

This question was haunting me, day and night. All I could think of was the possible answer to it. I would walk past cemeteries and crave the answer to it. But the issue was: I did not know how to calculate it.

If only there was a test that would show me if there was any association between the genre played by musicians and the cause of death...



# Wait... What is a Chi-Square test?

So... in the random example I showed before, it contained one very important characteristic: I wanted to check the association between two variables (music genre and type of death).

And what do these two variables have in common? They are both categorical variables (both nominal)!

In case you have forgotten what categorical variables are, or nominal variables, please go back to the first section: "Ones and Zeros" (Jack Johnson), where we discuss types of variables!



So... in a nutshell... **Pearson's Chi-Square test:** is a test that checks for the association or relationship between categorical variables (nominal or ordinal).



**IMPORTANT:** You could also use it for quantitative variables, but there are more powerful tests in this case, see the next chapters.



## Chi-Square test

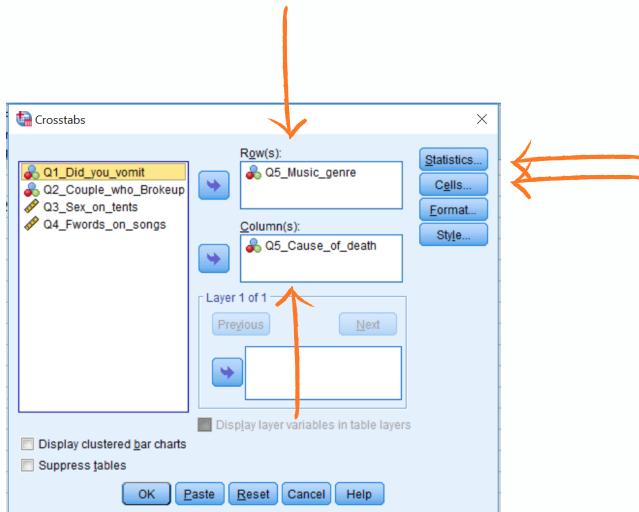
In the example at the start, we had one main burning question to answer, remember:

- **Q5. Is the music genre associated with the cause of death?**

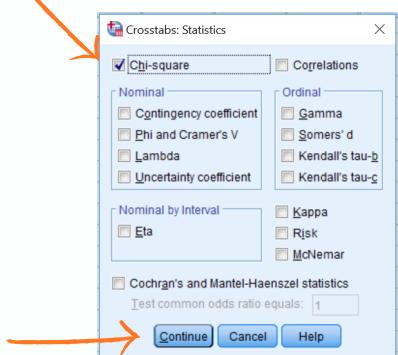
So, here is how to do it on IBM SPSS:

1

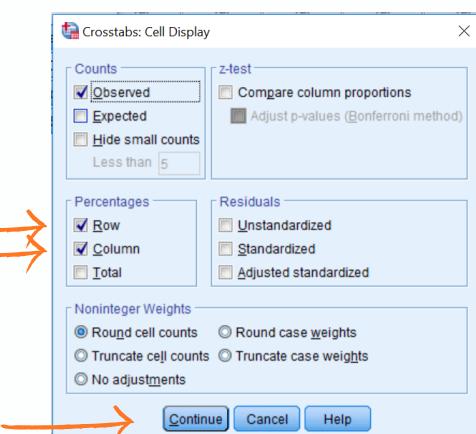
On the top menu of SPSS, you must click on "**Analyze**", then "**Descriptive Statistics**", then "**Crosstabs**". The window shown below will appear, then, simply choose the variables that you wish to analyze ("Music genre" and "Cause of death"), drag them to the box on the right and click on "**Statistics**".



On "statistics", make sure to check the box for "**Chi-square**". Then click on "**Continue**".



On "Cells", make sure to check the box for "**percentages**", on row and column. Then click on "**Continue**" and "ok".



## 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 **75 participants**. The table will also indicate the two variables that are being tested: In this case, "**Music Genre**" and "**Cause of death**".

	Case Processing Summary					
	Cases		Missing		Total	
	Valid N	Percent	N	Percent	N	Percent
Q5_Music_genre * Q5_Cause_of_death	75	100,0%	0	0,0%	75	100,0%

## Chi-Square: No Association

The second table will give you a description of the distribution of frequencies. It will provide the total combination between both variables: Music genre and cause of death.

			Q5_Cause_of_death Crosstabulation			
		Instagram overdose	Q5_Cause_of_death			Total
Q5_Music_genre	Rap		Count	Excessive Tik-Toking	Uncontrolled Call of Duty use	
	Count	10	5	5	23	
	% within Q5_Music_genre	43,5%	21,7%	21,7%	100,0%	
Q5_Music_genre	Heavy Metal	% within Q5_Cause_of_death	58,8%	26,3%	26,3%	100,0%
		Count	4	7	11	30
		% within Q5_Music_genre	13,3%	23,3%	36,7%	100,0%
Q5_Music_genre	Jazz	% within Q5_Cause_of_death	23,5%	36,8%	57,9%	40,0%
		Count	3	7	3	22
		% within Q5_Music_genre	13,6%	31,8%	13,6%	100,0%
Total		% within Q5_Cause_of_death	17,6%	36,8%	15,8%	29,3%
		Count	17	19	19	75
		% within Q5_Music_genre	22,7%	25,3%	25,3%	100,0%
		% within Q5_Cause_of_death	100,0%	100,0%	100,0%	100,0%

Got it? Good! But look, so far all we had was a frequency distribution. In other words, we only know how many musicians from each genre died from each cause. But remember: in a Chi-square test, we wish to know the extent to which there is an association or relationship between both categorical variables.

For this, you must interpret the following table, which shows the Chi-square result:

Chi-Square Tests			
	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	12,970 <sup>a</sup>	6	,044
Likelihood Ratio	12,543	6	,051
Linear-by-Linear Association	5,485	1	,019
N of Valid Cases	75		

a. 1 cells (8,3%) have expected count less than 5. The minimum expected count is 4,99.

## How to interpret?

**Count:** The total number of respondents per category. Example: from the 75 respondents, 10 were Rappers that died from "Instagram overdose", 5 from "excessive Tik-Toking", 5 from "uncontrolled CoD" and 3 from "Missing their moms' texts".

**% within Music Genre:** The percentage of each music genre in relation to each cause of death. Example: from all 23 rappers, 10 (43,5%) died from Instagram overdose, 5 (21,7%) from Excessive Tik Toking, 5 (26,3%) from uncontrolled CoD and 3 (13%) from missing their moms' texts.

**% within Cause of death:** The percentages of each causes of death. Example: from the 17 people that died from Instagram overdose, 10 were Rappers (58,8%), 4 were Heavy Metal (23,5%) and 3 were Jazz players (17,6%).

## How to interpret?

**The table provides three main values:** Pearson Chi-Square value, degrees of freedom (df) and the significance value (p value).

**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.

**RESULT:** as the p value result is 0.044 (below the threshold of 0.05), we can assume that there IS a statistical association between the genre of music and type of death!

Pretty interesting, right? But there is one more thing I would like to draw your attention to: under the table it says... "**1 cells (8,3%) have expected counts less than 5**". **This is important!** Because on Chi-Square tests, there is an assumption for the test which must be met: for all cells to have a minimum count of 5.



So in this fictitious example the assumption has been violated. And when reporting results, it is important to make this clear. But if you want to know why that is and what to do about it, we strongly suggest having a further read of the statistics books we recommended at the start of this guide, ok?



## Attention! Attention! Attention!

My friend, here is one very important detail when interpreting Chi-Square tests:

### Association is not Causality!

Chi-Square tests **do NOT DETERMINE CAUSALITY** of the association. It tests merely the significance of the association between the categorical variables!

For this reason, we **cannot** say that **being a rapper causes** you to die from Instagram overdose. We can only assume that those variables are related. Got it???



### SPSS Tutorial playlist series.

In case you feel like you still need some further support on how to run a **chi-square test on SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of **LiveInnovation.org**.



### Full soundtrack for Chi-square test

1. No association (Silverchair)
2. Related (Kingdom of birds)
3. Free association (Gordi)
4. Square 1 (Mahalia)
5. No cause for alarm (Just Jack)



27

# MUTUAL MOTION

Artist: Loure

## Correlation test.

And how to conduct it on **IBM SPSS**.



LIVEINNOVATION.ORG

# Oops!... I did it again. (Got pregnant!)

Look, I do not like talking about people's lives. But since no one is listening here, I will share a secret with you. Following music festivals, there is usually a baby boom! I am not kidding.

I mean, people go wild, so what would you expect, right? The mixture of alcohol, music, dancing, socializing and forgetting about "normal" life makes people do the craziest things. And guess what? After having visited a summer rock festival, both my imaginary friends, Janis and Joplin, are expecting another baby!

Being the curious person I am, I had to ask them: was it planned? They said...

But this situation is not exactly new. So it really got me wondering... Do longer music festivals lead to more pregnancies?

**"Absolutely not!  
But hey, the festival took so long...  
We spent so much time in the tent..."**

Fascinating! So the number of days of the festival had impacted their life plans. This got me wondering...



**Q6. Do longer festivals lead to more pregnancies?**

This question really puzzled me. I spent long nights thinking about it. I even checked online to see the average amount of days of festivals and where they would happen. All I could think was:

**Will more babies be born because of this festival?**

A bit weird to be obsessed with this question, I know. But still an interesting question with genuine scientific curiosity. So I have to find a way to answer it...

If there was only a test which could tell me if there was an association between these two types of quantitative variables (number of days of a festival and number of babies born)...

# Wait... What is a correlation test?

So... before we discuss what a correlation test is, I would like to emphasize an important detail in the weird example I introduced before: I wanted to check the association between two variables ([number of days](#) and [number of babies](#)).

And what do these two variables have in common? [They are both quantitative variables.](#)

**In case you have forgotten what quantitative variables are, please go back to the first section: "Ones and Zeros" (Jack Johnson), where we discuss types of variables!**



So... in a nutshell... **Correlation test:** is a test which checks for the association or relationship between quantitative variables (interval or ratio).



**HERE IS HOW  
TO DO IT!**

## Correlation test

So... Want to find out a bit about the consequence of people's intimate shows in tents during festivals? Great, so let's answer this fascinating question:

- **Q6. Do longer festivals lead to more pregnancies?**

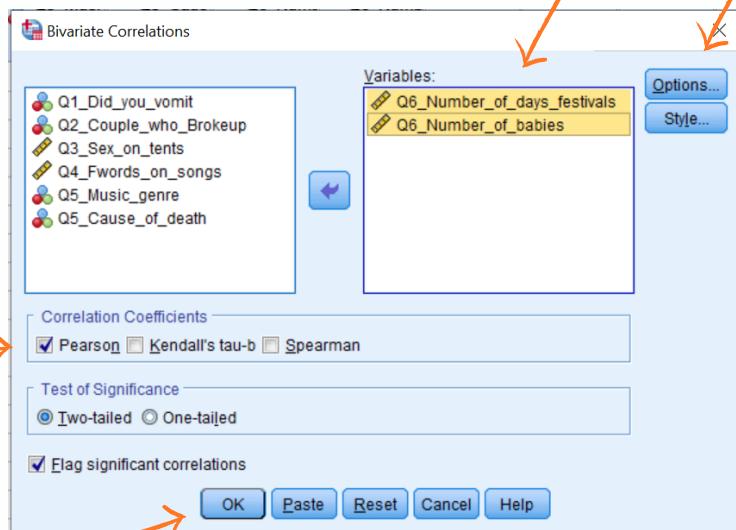
So, here is how to do it on IBM SPSS:

**1**

On the top menu of SPSS, you must click on "**Analyze**", then "**Correlate**", then "**Bivariate**". The window shown below will appear, then, simply choose the variables that you wish to analyze ("Number of days of the festival" and "Number of babies"), drag them to the box on the right.

**2**

Next, define the "**Correlation coefficient**".



Apply **Pearson's coefficient** when having **parametric data**. Apply **Kendall's** or **Spearman's coefficient** when having **non-parametric data**.

**3**

In case you would like to have the means (averages) and standard deviations of each factor, simply click on "**Options**" then tick the box for "**means and standard deviations**".

**Done? Great!**  
**Click on "OK" to see the results!**

## Results

In case you chose before to have the means and standard deviations, these you will find in the first table. Exactly as seen below.

So you will see, for example, that the average number of days of festivals was **3.04** and standard deviation of **1.89**. And also, that the average number of babies was an incredible **6.68**! And a standard deviation of **6.54**. **Wow. Simply... wow.**

Descriptive Statistics			
	Mean	Std. Deviation	N
Q6_Number_of_days_festivals	3.0400	1.89908	75
Q6_Number_of_babies	6.6800	6.54333	75

Ah, and as you already know by now, the label "N" refers to the sample size. So we have investigated 75 festivals.

# Here is where the magic is revealed!

Correlations			
		Q6_Number_of_days_festivals	Q6_Number_of_babies
Q6_Number_of_days_festivals	Pearson Correlation	1	.888**
	Sig. (2-tailed)		.000
	N	75	75
Q6_Number_of_babies	Pearson Correlation	.888**	1
	Sig. (2-tailed)	.000	
	N	75	75

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Alright, now how do we interpret the table above?



**Significance level:** our p-value here is .000. As it is **below** our famous significance threshold of 0.05, this indicates to us that there is a **significant correlation** between both variables.

**PS:** although the value is .000, this is a rounded value.  
The p-value is never "zero flat". So in this case, it is just very close to zero.



**Correlation coefficient:** here there are two important things to keep in mind:



**Positive or negative?** Positive values will indicate a positive correlation. In other words, if one variable increases, so does the other. If there is a "negative" sign, it will indicate a negative correlation. In this case, if one variable goes up, the other tends to go down. **In this case, since there is no "negative" sign before the value, it indicates a positive correlation between number of days of a festival and the number of babies.**



**Pearson's correlation coefficient value:** The value here is displayed on a scale from 0-1, and represents the "magnitude", "strength" or "size" of the correlation. Broadly speaking, there is general benchmark to interpret it:



**Small:** .1-.3

**Medium:** .3-.5

**Large:** .5-1.0

Given that the Pearson's correlation coefficient value was of **.888**, we can assume that there is a **large positive correlation between the number of days of a music festival and the number of babies that are born afterwards**.



# Attention! Attention! Attention!

My friend, once again, I need to draw your attention to one very important detail when interpreting correlation tests:

## Correlation is not Causality!

Correlation tests do NOT DETERMINE CAUSALITY of the association. They test solely the significance of the association between the quantitative variables!

## Exactly like with Chi-Square tests!

For this reason, we **cannot** say that people had more babies **because** the festivals were longer. We can only say that they were correlated. Got it???



## SPSS Tutorial playlist series.

In case you feel like you still need some further support on how to run a **correlation test on SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of **LiveInnovation.org**.



## Full soundtrack for Correlation test

1. Mutual motion (Loure)
2. It takes two to tango (Louis Armstrong)
3. Up and Down (CLOVES)
4. Correlation (Ginne Marker)
5. Opposite directions (Neil Erickson)



# JUST THE TWO OF US

Artist: Bill Withers

## Independent and dependent (paired) samples t-tests.

And how to conduct it on **IBM SPSS**.



# Milk a Cow. Or Die Trying.

Cows are beautiful animals, aren't they? And it is such a shame that they are often so exploited. Poor things... And one of the main ways they are exploited is to produce **milk**.

**Ps: don't blame me. I actually only drink almond milk! It's delicious. Really.**

Anyway... But let us assume that you wanted to boost the milk production in a farm. How would you do it? Well, I hear from different sources that there is a great hack: **play music!**

But which style of music works best? Tough question. So a fictitious friend of mine decided to do a bizarre experiment. Since he is a major "**50 Cent**" fan, he decided to play some rap songs from 50's first album, "Get rich or die trying", to the cows. Apparently it is a classic album.

And since his other favorite genre is heavy metal, he decided to play some **Iron Maiden** tunes, like "Run to the hills". And so he did.

**No I will give you a minute to imagine a bunch of cows in a farm, listening to 50 Cent and Iron Maiden. Enjoy this thought.**

Good, now that you have imagined this crazy scene, we are left with an important question. What would have happened?

This intriguing situation left us with an important question to answer:



**Q7. Do cows produce more milk when listening to rap when compared to heavy metal?**

Wow, what an interesting question to answer. Finding the answer could potentially revolutionize farming, the entire milk production industry and even the music industry itself. So yeah, it was a pretty big deal.

If there was only a test which could tell me if there were statistical differences between two groups I wanted to compare...

That would be so nice and useful...

# Wait... What is a t-test?

So... before we discuss what a t-test is, it is important to remember that we are comparing two groups (rap music and heavy metal), and that these groups are **categorical variables** (nominal). And that the outcome, amount of milk, is a **quantitative variable** (ratio).

Thus, we have two categorical groups being compared through quantitative variable measures.

**In case you have forgotten what categorical or quantitative variables are, please go back to the first section: "Ones and Zeros" (Jack Johnson), where we discuss types of variables!**



So... in a nutshell... **t-test:** compares the means (averages) of two groups. It checks if the means of the two samples are significantly different from each other and is commonly used when the variances of two distributions are unknown.

# But wait a minute...

Perhaps you might be wondering here: There has got to be a difference if the cows heard **both** types of music or **simply one**, right? **YES.**

**That is why there are 2 different types of t-tests:**



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

e.g. **each cow only heard one music genre (either rap or heavy metal).**



**Dependent (paired) samples t-test:** When **the same** participants have been used in both groups.

e.g. **each cow heard both music genres.**



**Tutorial Alert!**

# HERE IS HOW TO DO IT!

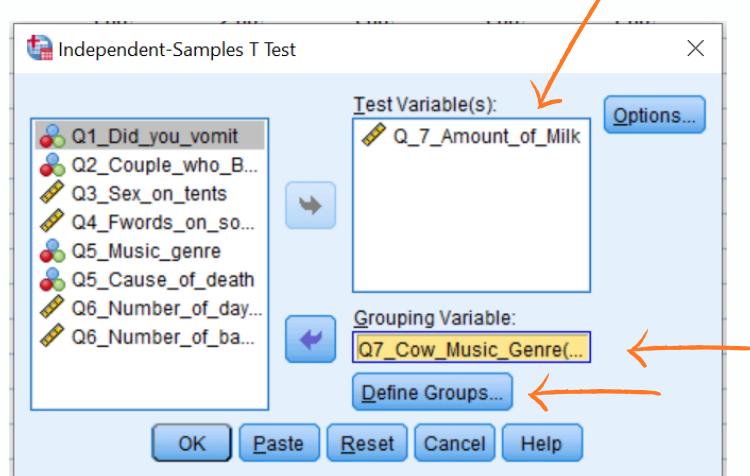
## Independent samples t-test

Alright... Are you ready to find out whether rap from 50 Cent or heavy metal from Iron Maiden makes cows produce more milk? Personally, I cannot wait to see the results!

### Here is how to do it on IBM SPSS:

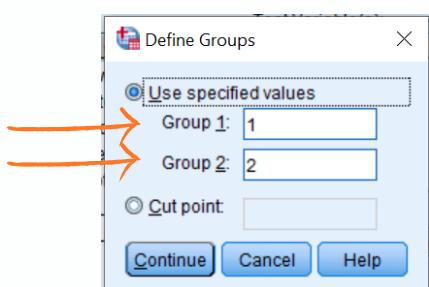
1

On the top menu of SPSS, you must click on "**Analyze**", then "**Compare means**", then "**Independent samples t-test**". The window shown will appear. Then, choose the variables that you wish to compare ("amount of milk"), and drag them to the "test variables" box on the right.



2

You must also drag the variable which describes the groups you wish to compare to the "grouping variable" space ("Cow music genre"). After you do this, a button "**define groups**" will light up. Essentially SPSS wants to know what is the coding on the variable view for each value. In this case, we had previously set "1 as the value for Rap - 50 Cent and 2 for heavy metal - Iron Maiden".



After setting the values, click on "continue" to close the window and then on "ok" on the previous window to run the results!

# Results

Alright... Let's see what we have got! The first table you will find is the one you see below. It gives a few important information: N (sample size), so 39 cows listened to rap, while 36 cows listened to heavy metal.

Group Statistics					
		N	Mean	Std. Deviation	Std. Error Mean
Q_7_Amount_of_Milk	Rap - 50 Cent	39	9.5641	3.11021	.49803
	Heavy Metal - Iron Maiden	36	4.3611	1.88457	.31409

The table also tells us that on **average (mean)**, cows listening to rap produced **9.56 liters** of milk. Meanwhile, cows listening to heavy metal produced **4.36 liters**.



## Attention! Attention! Attention!

**Perhaps you are wondering:** "If one average is higher than the other, I can assume that listening to rap makes cows produce more milk than listening to heavy metal". **Correct?**

**NO! YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

**NO! YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

**NO! YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

**NO! YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

**NO! YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

Ok, I understand that all this repetition was a bit dramatic from my side. I admit it. But my friend, I am just trying to make a point here because this is something many students, professionals and even journalists get wrong.

**So, always remember....**

**IN ORDER TO CHECK IF THE DIFFERENCES BETWEEN THE MEANS ACTUALLY EXIST AND ARE NOT DUE TO RANDOMNESS, WE MUST ANALYZE THE SIGNIFICANCE VALUE OF THE T-TEST!**

The table below is the second one you will find on your results. And here are the key values to look for when interpreting it:

1

**Verify the Levene's Test for Equality of Variances:** In a nutshell... it is a test to check for the homogeneity of variance between the two groups. If you wish to understand further what that means, I suggest reading Andy Field's book "Discovering statistics".

But if you simply wish to interpret the result, all you need to remember is:



If the Levene's test is **non-significant ( $p>.05$ )**, then the assumption of "homogeneity of variances" 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.

Independent Samples Test										
	Levene's Test for Equality of Variances			t-test for Equality of Means					95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper	
Q_7_Amount_of_Milk	Equal variances assumed	5.551	.021	8.672	73	.000	5.20299	.59996	4.00727	6.39871
	Equal variances not assumed			8.837	63.358	.000	5.20299	.58881	4.02649	6.37949

2

**Check the independent t-test significance value:** In the case above, the Levene's test is significant ( $p=.021$ ), thus the assumption test of homogeneity of variance between both groups **IS NOT MET**. So we need to interpret the independent t-Test result from the **BOTTOM** row ( $p=.000$ ).

So... Since the independent samples t-test result is significant ( $p=.000$ )...

# What can we conclude?

**Answer:** That there is a statistical significant difference between the amount of milk produced when cows are exposed to rap music or heavy metal.



And since the average of milk produced by cows when listening to rap (9.56 liters) is higher than when listening to heavy metal (4.36 liters), **we can say that cows listening to rap produce more milk than cows listening to heavy metal!**

# Dependent samples t-test

On the example we discussed, there was one important detail: **each cow had heard only one genre of music (either rap or heavy metal). This is why we applied an independent samples t-test.**

**And what if each cow had listened to both music genres?**

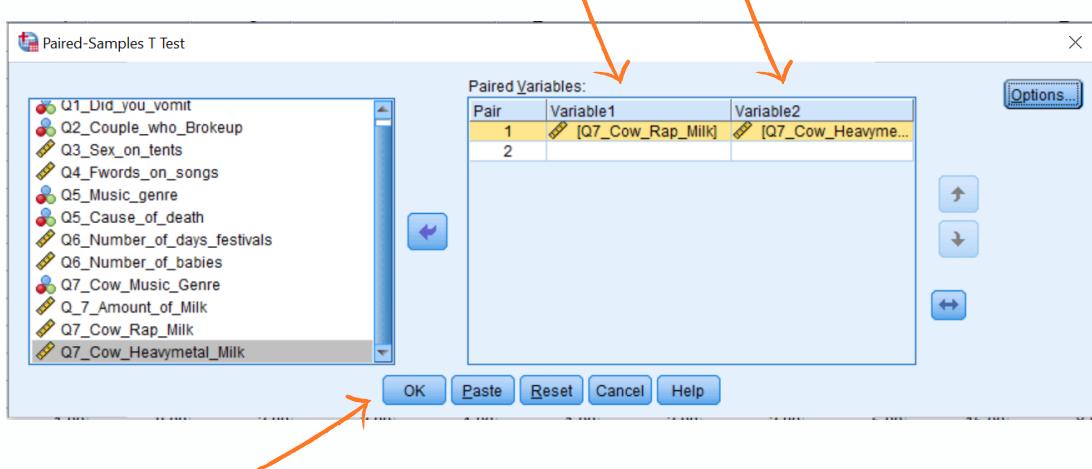


Well, as mentioned before, when we are comparing **two groups** and we have the same participants in both groups we must use a **dependent (paired) samples t-test**.

**Here is how to do it on IBM SPSS:**

**1**

On the top menu of SPSS, you must click on "**Analyze**", then "**Compare means**", then "**Paired samples t-test**". The window shown will appear. Then, choose the two variables that you wish to compare ("Cow rap milk" and "Cow heavy metal milk"), and drag them to the "paired variables" box on the right.



**2**

Click on "**OK**" and you will see the results.

# Results

So... The first table you will see on the results is the one you see below. It will give you some interesting information: N (sample size). So as we can see, there were 75 cows in each group. Also, it gives us the average. While listening to rap, a cow produced an average of 10.56 liters, while listening to heavy metal it produced an average of 4.36 liters.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Q7_Cow_Rap_Milk	10.5467	75	3.05929	.35326
	Q7_Cow_Heavymetal_Milk	4.3600	75	1.49449	.17257



So... Let me try asking you something: if the average amount of milk produced by cows while listening to rap is higher than while listening to heavy metal, can we assume that listening to rap makes cows produce more milk than listening to heavy metal?

# Nooooooooooooo!

## And why? Because...

**YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**  
**YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**  
**YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**  
**YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**  
**YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.**

**PERFECT!** In order to check for significant statistical differences between the groups, we must check the result of the paired sample t-test table. As you can see below the p-value is .000, which means it is significant! **Therefore, we conclude that cows had a significantly higher production of milk when listening to rap as compared to when listening to heavy metal!**

Paired Samples Test								
	Mean	Std. Deviation	Std. Error Mean	Paired Differences		t	df	Sig. (2-tailed)
				Lower	Upper			
Pair 1	Q7_Cow_Rap_Milk - Q7_Cow_Heavymetal_Milk	6.18667	3.29493	.38047	5.42857	6.94476	16.261	.000



# Fantastic, my friend!

What should you remember from this section?



1

**You must conduct different types of t-tests depending if you have the same sample on both groups or different samples.**

2

**Only looking at the averages you cannot assume statistical differences.**

3

**How to conduct and interpret dependent (paired) and independent samples t-tests!**



## SPSS Tutorial playlist series.

In case you feel like you still need some further support on how to conduct **t-tests in SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of [LiveInnovation.org](http://LiveInnovation.org).



### Full soundtrack for t-tests

1. Just the two of us (Bill Withers)
2. Two birds (Regina Spektor)
3. Two against one (Danger Mouse)
4. Nothing compares 2 U (Chris Cornell)
5. Two of us (The Beatles)



# DISCLAIMER

No cows were harmed in  
the writing of this  
statistics guide.



# THREE LITTLE BIRDS

Artist: Bob Marley and The Wailers

# One-Way ANOVA.

And how to conduct it on **IBM SPSS**.



LIVEINNOVATION.ORG

6

# Highway to hell!

Are you also one of those people that as soon as they get in a car turn up the radio and start singing loud as if there was no tomorrow? **Me toooo!**

But my friend, have you ever seen the scary fictitious statistics we just made up on how many people have car accidents while driving? Wow. I mean wow....

One of the main causes of distraction, which we believe has been getting too little attention in official statistics, is the **type of music you are listening to while driving.**

It is not so hard to imagine why, right? For example:

1. **If you listen to punk-rock you will be more likely to step on the gas and speed.**
2. **If you listen to classical music, you will probably fall asleep from being over relaxed.**
3. **If you listen to trash metal you probably want to crash on purpose!**
4. **And if you listen to reggae maybe you will feel so relaxed that you will forgive all other drivers and not get involved in any accident.**

This intriguing scenario led me to wonder one important thing: if those types of music lead to such behaviors, and given that certain radio stations only play certain genres, which one would lead to greatest amount of car accidents?

**In other words....**



**Q8: Are drivers more likely to have accidents when listening to certain radio stations?**

Oh my God... This question haunted me for weeks! I was obsessed about it. Every time I saw a driver singing, all I could think was: what radio station is he/she listening to? Is he/she safe?

If there was only a statistical test that would allow me to compare the number of accidents of all of those stations to see if there was any statistical significant difference among them....

# Wait... What is an ANOVA?

(Analysis of Variance)

First of all, let us stop for a minute and contemplate how beautiful this name is: **ANOVA**. Honestly, if I ever have a daughter I will seriously consider naming her after it. I can already imagine these moments:

"ANOVA, you look beautiful today"... "ANOVA, come here. We need to have a conversation"... "I cannot believe ANOVA still has not called in weeks". You never know, this might be my future!

But anyway, let us stop daydreaming about future daughters and get back to reality. So before we discuss what an ANOVA is, it is important to remember that in the previous example we aim to compare four groups, or radio stations (punk-rock, classical, trash metal and reggae), and that these groups are **categorical variables** (nominal). And that the outcome, amount of crashes, is a **quantitative variable** (ratio).

Thus, we have four categorical groups being compared through one quantitative variable measure.

In case you have forgotten what categorical or quantitative variables are, please go back to the first section: "Ones and Zeros" (Jack Johnson), where we discuss types of variables!



So... in a nutshell... **ANOVA (Analysis of variance)**: compares three (or more) groups of means (averages). It checks if three or more samples are significantly different from each other.

## So, essentially, an ANOVA is an extension of t-tests.

### But wait a minute... what's that "One-way"?

If you paid attention to the cover of this chapter, it said: **One-way ANOVA**. And what is this "**one-way**" for? It refers to the **number of independent variables** that you are comparing. In this case, there was only one: "the radio stations": punk-rock, classical, trash metal and reggae...

## So it is **not** exactly rocket science:



If you have **one** independent variable, you apply a **One-way ANOVA**.  
If you have **two** independent variables, you apply a **Two-way ANOVA**.

**And so on... And importantly, each one is conducted slightly different.**

What I will show you here is how to conduct the one-way ANOVA. Shall we have a look?



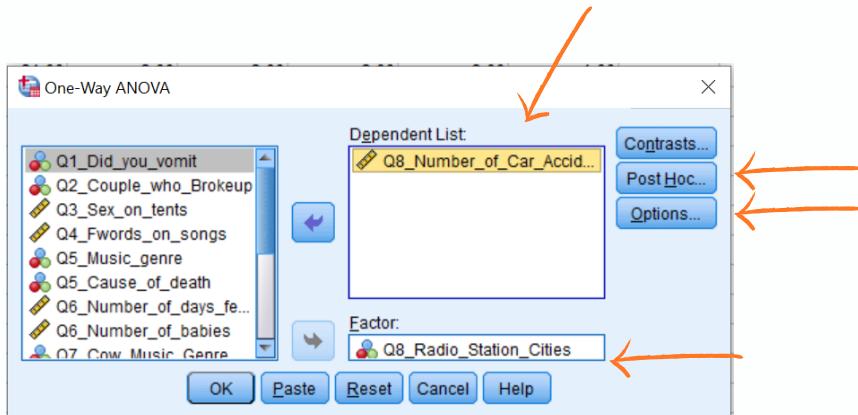
## One-way ANOVA

Honestly, I cannot wait to know which radio station led to the greatest number of accidents! So let us dive straight into the data set and figure out what happened.

**Here is how to do it on IBM SPSS:**

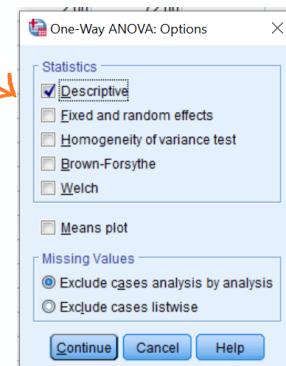
1

On the top menu of SPSS, you must click on "**Analyze**", then "**Compare means**", then "**One-Way ANOVA**". The window shown will appear. Then, choose the "factor" as the variable of the groups ("radio stations") and the variables that you wish to compare to the "dependent list" ("number of accidents").



2

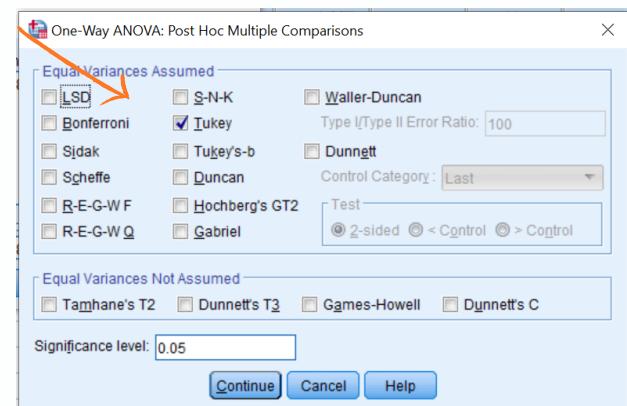
Next, you will click on "options", then select "descriptive". This way you will have all the averages, sample size, standard deviations and more.



3

Next, you will click on "Post-Hoc", then select "Tukey". This way you will have all between group comparisons.

Finished? Click on "Continue" and "Ok" for the results!



# Wait... What is Post-Hoc?

As I told you many times: **this is not a statistics book**. And it continues not to be. However, I feel like I need to give you a bit more on this topic. Soooo... Remember on the last chapter how we discussed **t-tests**? And how it allows us to compare **two groups**? So whatever significance result we have, it is always when comparing two groups, let us say, group A against group B.

**t-test compares two groups!**

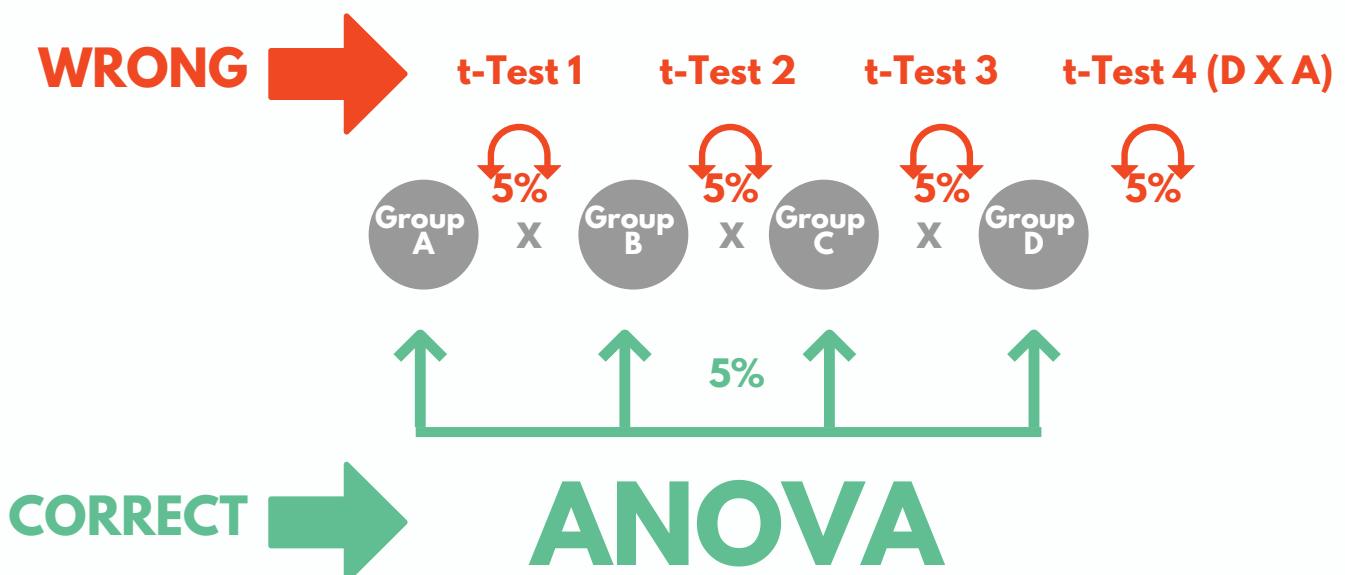


And when comparing these two groups, remember: **the maximum amount of randomness allowed is 5%! ( $p<0.05$ )**.

# And what happens when we compare more than two groups?

Well... For example, you **could** think: "if I have four groups to compare, can't I just run four t-tests and contrasts them all simply by running multiple t-tests?" **Noooooooooooooo**.

**Why?** Because since each t-test allows a maximum amount of randomness of 5% (p-value threshold), after only four t-tests, you could end have an aggregate total amount of 20% randomness. And there are even six combinations when comparing four groups. So you could end up with 30% of randomness!



The ANOVA is a test that will contrast three or more groups, but still remaining with **the maximum amount of randomness allowed of 5%! ( $p<0.05$ ) for all comparisons**. So it is a much more effective and accurate way of contrasting three or more groups than conducting multiple t-tests. Got it?

**So you might be wondering:**  
**Ok... But what does that have to do with Post-Hoc?**

**Post-Hoc test will provide a multiple comparison between all groups!**

In other words, it will give us the significance results between all possible comparisons. In this case: **AxB, AxC, AxD, BxC, BxD, CxD**.

# And why did I ask you to click on "Tukey" for the Post-Hoc test?

As you can see in Figure 3, there are many different types of Post-Hoc tests. And explaining them is beyond the scope of this statistics guide. And hey, do not complain, I told you this was not a statistics book!

But the reason for choosing **Tukey** is because it is a recommended Post-Hoc test **when you have fairly similar sample sizes on all groups.** In case you have very uneven sample sizes, please check for support on actual statistics books!

## Results

Alright! Enough of explaining Post-Hoc tests. Let's see the results and find out which radio station leads to the greatest number of car accidents. Yeah!!!!

Alright, so the first table will show us the "**descriptive statistics**". As usual, it will include things like N (sample size). In this case, the sample size refers to the number of cities where the radio stations play. So the Punk Rock radio is played in 19 cities and led to an average of 49.1 crashes per year. The Classical radio is played in 19 cities and an average of 21.0 drivers had accidents. The radio playing trash metal music was found in 19 cities, with an average of 48.3 accidents annually, and the reggae station in 18 cities with 19.1 accidents per year.

Q8_Number_of_Car_Accidents	Descriptives							
	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
Punk Rock	19	49.1579	13.93741	3.19746	42.4403	55.8755	36.00	86.00
Classical music	19	21.0526	5.08236	1.16597	18.6030	23.5023	7.00	29.00
Trash metal	19	48.3684	7.31976	1.67927	44.8404	51.8964	38.00	61.00
Reggae	18	19.1667	3.29438	.77649	17.5284	20.8049	13.00	25.00
Total	75	34.6400	16.64952	1.92252	30.8093	38.4707	7.00	86.00

**And now, the million dollar question:** can I assume their differences only by looking at the means? In other words, can I derive from the table above that people listening to punk rock and trash metal radios get more involved in accidents than the other two stations simply because their averages are higher?

# Noooooooooooooo! And why? Because...

YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.  
 YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.  
 YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.  
 YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.  
 YOU CANNOT ASSUME STATISTICAL DIFFERENCES ONLY BY LOOKING AT THE MEANS.



Very good, my friend! I am proud of you. See, sometimes being a bit dramatic when trying to explain something really helps. I am glad you are not assuming such differences only by looking at the means.

And how do you find the statistical significance on an ANOVA?

## First, have a look at the second table!

ANOVA					
Q8_Number_of_Car_Accidents	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	15402.885	3	5134.295	71.332	.000
Within Groups	5110.395	71	71.977		
Total	20513.280	74			

The second table shows us the ANOVA result. **Essentially, it tells us if there are any statistical differences between the groups.** Here is the deal:



If the the ANOVA result is **NON-SIGNIFICANT ( $p>.05$ )**, we assume that there are **no differences between the groups**. **Therefore, the analysis ends here with this conclusion:** there are no differences among groups.

## However...



If the the ANOVA here is **SIGNIFICANT ( $p<.05$ )**, we assume that there **are differences between the groups**. **Therefore, the analysis does not end here.**

In this case, the ANOVA result was highly significant ( $p=.000$ ). So there is one important question missing:



# Which groups are statistically significantly different from which groups?

# To answer that, we need to look at the third table: The Post-Hoc!

The third table shows us the Post-Hoc test, which refers to the **multiple comparisons**.

**Why does it have this name?** Exactly because it provides us with all the significance results among all groups and possible combinations.

Here are all the significance values of all comparisons!

→ Here it shows each group compare with all other three!

Multiple Comparisons						
		Dependent Variable: Q8_Number_of_Car_Accidents		Tukey HSD		
(I)	(J)	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
Punk Rock	Classical music	28.10526*	2.75256	.000	20.8635	35.3471
	Trash metal	.78947	2.75256	.992	-6.4523	8.0313
	Reggae	29.99123*	2.79052	.000	22.6495	37.3329
Classical music	Punk Rock	-28.10526*	2.75256	.000	-35.3471	-20.8635
	Trash metal	-27.31579*	2.75256	.000	-34.5576	-20.0740
	Reggae	1.88596	2.79052	.906	-5.4557	9.2277
Trash metal	Punk Rock	-.78947	2.75256	.992	-8.0313	6.4523
	Classical music	27.31579*	2.75256	.000	20.0740	34.5576
	Reggae	29.20175*	2.79052	.000	21.8601	36.5434
Reggae	Punk Rock	-29.99123*	2.79052	.000	-37.3329	-22.6495
	Classical music	-1.88596	2.79052	.906	-9.2277	5.4557
	Trash metal	-29.20175*	2.79052	.000	-36.5434	-21.8601

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

I will interpret the Punk rock stations to illustrate how to interpret it all, ok?

For example: as you see above, the **Punk Rock radio stations** (average 49.1 crashes), is **significantly different** from the **Classical radio stations** (average 21.0 crashes), and **significantly different** than the **Reggae radio stations** (average 19.1 crashes). In both comparisons, p-value was .000 (highly significant). **But NOT from Trash metal radio stations** (48.3 crashes per year). Here the p-value was .992 (>0.05).

## Conclusion?

We can confidently say that drivers have more car crashes when listening to Punk Rock and Trash metal radio stations, than when listening to Classical and Reggae radio stations while driving.

# And to conclude the analysis, you need to apply the same logic to analyze all other multiple comparisons shown in the Post-Hoc table!

So, if you do it, here is a summary of how your results would look like:

- **Classical radio stations:** drivers have significantly been involved in less accidents than punk rock and trash metal. But not in comparison to reggae stations. I guess drivers feel so relaxed they make no mistakes!
- **Trash metal stations:** drivers have significantly been involved in more accidents than classical and reggae stations. But not in comparison to punk rock stations. No wonder, with all the screaming in trash metal drivers will maybe crash on purpose!
- **Reggae radio station:** Drivers listening to reggae have significantly had less accidents than punk rock and trash metal stations. But no significant differences were found in comparison to classical music stations.

## So... That was a lot, right?

Since I am a nice guy, here is a quick summary on how to interpret the One-Way ANOVA:



**Descriptive table:** First, analyze the descriptive table to interpret the averages of each group you are comparing.



**ANOVA table:** Analyze the ANOVA table. If it is non-significant, it is game over because it means that there is no significant difference among the groups. If it is significant, you must then interpret the Post-Hoc table (multiple comparison).



**Post-Hoc table:** Investigate all group comparisons to identify which interaction (between the groups) is statistically significant.

# Congrats, my friend!

What should you remember from this section?



1

**One-Way ANOVA will compare three or more groups, when you have only one categorical variable.**

2

**Only looking at the averages you cannot assume statistical differences.**

3

**How to interpret Post-Hoc tests and the multiple comparison it provides!**



## SPSS Tutorial playlist series.

In case you feel like you still need some further support on how to conduct a **One-Way ANOVA in SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of [LiveInnovation.org](https://www.youtube.com/user/LiveInnovation).



## Full soundtrack for a One-Way ANOVA

1. Three little birds (Bob Marley and The Wailers)
2. Three strikes (HONNE)
3. Thirty three (The Smashing Pumpkins)
4. 4:44 (Jay-Z)
5. Stock (AnnenMayKantereit)



# TRUST ME

Artist: Janis Joplin

# Cronbach Alpha test.

(Scale reliability test)

And how to conduct it on **IBM SPSS**.



# The Lost Beatle.

My friend, I will ask you tricky question: do you consider **The Beatles** as one of the best bands of all times? If your answer to this question was "No", please close this e-guide. Like, right now. We are not friends anymore and I believe we should stop talking. Forever.

Ok, a bit dramatic. But I guess you understand what The Beatles mean to me. They are true legends: John Lennon, Paul McCartney, Ringo Star and George Harrison. Four legends.

But let me tell you a secret: the band had a fifth member that, believe it or not, chose to leave the group before they became famous. His name? **Stuart Sutcliffe**. At that time they were still doing tours in Hamburg, Germany. And Stu, who used to play bass, decided to leave the band to concentrate on his art career. And the band decided not to replace him.

So McCartney took over the bass, they became a four-piece band... **AND THE REST IS HISTORY**.

## So what can we take from this story?

**Sometimes things get better when someone leaves.**

(Keep this thought. I will get back to it. This introduction will make sense. I promise).

## Anyway... changing topics: Have you heard of multi-item marketing scales?

Explaining marketing scales in detail is, again, beyond the scope of this e-guide. But briefly, they are a group of **ITEMS** (**statements about the concept, which are measured through Likert-type scales**), that together, are applied to measure factors such as: risk, quality, attitude, satisfaction and many more. Such scales are extremely common in marketing and psychology, for example.

Also, when using them, to measure consumer perception, it is very important to measure the **reliability of the scale**. And the main test to do it is called **Cronbach's Alpha**.

# Marketing Scale Handbooks

**Are you interested (or do you simply need) to measure the perception of consumers towards factors such as:**

Quality  
Satisfaction  
Attitude  
Risk  
Purchase intention  
Affective response  
Visual design  
**AND MANY OTHERS?**

Then the safest way is to use previously validated scales. On **marketing scale handbooks**, you can find decades of scale development, all complied for your use.

Here are three recommendations:

- Bearden, W. O., & Netemeyer, R. G. (1999). **Handbook of marketing scales: Multi-item measures for marketing and consumer behavior research.** Sage publications.
- Bruner, I. I., & Gordon, C. (2012). **Marketing scales handbook.**
- Bruner, G. C., Hensel, P. J., & James, K. E. (2005). **Marketing scales handbook.** Chicago, IL: American Marketing Association.

In case you do not find the scale you need, check other editions of the handbook.  
Different editions have different scales!

# Wait... What is Cronbach's alpha?

Cronbach's alpha is a reliability test for scales. **It measures the internal consistency of a scale.** So, essentially, it checks how related the items of scale are to each other, when measuring the factor it is intended to measure. Also, it suggests how reliable the scale can be, if we remove one or more items to enhance the reliability (don't worry, I'll show you later).

The Cronbach alpha score ranges from 0-1. And the score represents how reliable the scale is. Here is how to interpret it:

- **Scores under .6** are considered "unacceptable".
- **Scores under .7** are considered "questionable" (some authors accept it, some do not).
- **Scores above .7** are considered "acceptable" in most social sciences.
- **Scores above .8** represent that your scale has achieved a "good" reliability level.
- **Scores above .9** represent that your scale has achieved an "excellent" reliability level.



**Do not worry, trying to memorize this.  
If you forget, come back here to remember it!**

## Also important!

Cronbach's alpha can only be calculated with a **minimum of three items on a scale**. It is not the appropriate test when having two items, and with only one item you do not even have a multi-item scale!

Alright... Now that you are a pro, shall we have a look on how to run the reliability test?



**HERE IS HOW  
TO DO IT!**

# Cronbach's Alpha

My friend, I am pretty sure that when you woke up this morning you had no idea you would end up doing some scale reliability test, right? But hey, look how exciting! It will brighten your day, I am sure. So... For this example, I will use an example of a **5 item scale** that measures "**involvement**", which was used to measure how involved a young generation of fans are with "The Beatles".

For clarification purposes, I have made up a few items to illustrate how the times could look like (but please note these are fictitious!):



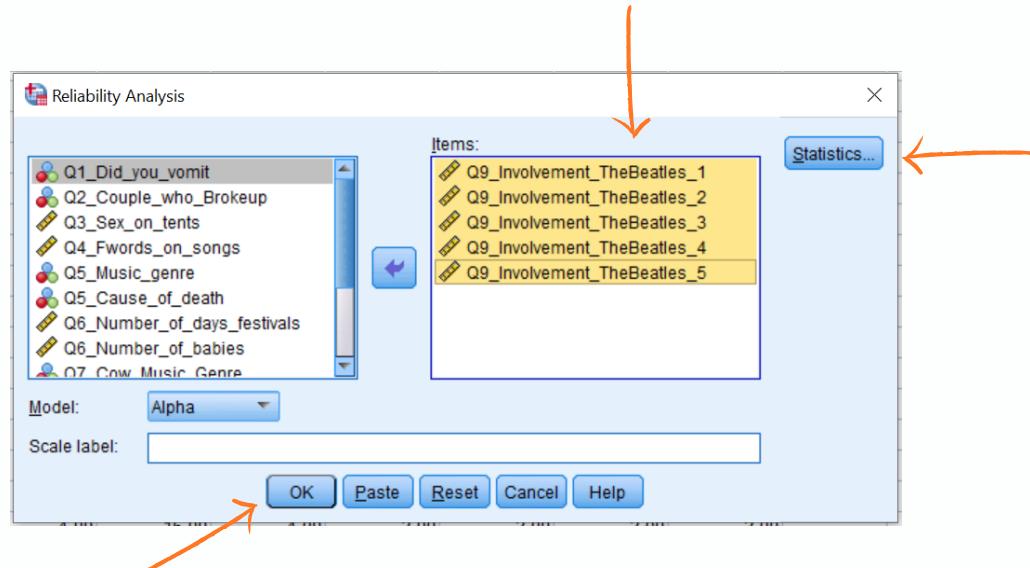
1. I care deeply about The Beatles
2. The Beatles are my favorite band
3. I cannot think of another band I enjoy more than The Beatles
4. I feel close to The Beatles
5. The Beatles represent who I am

**Here is how to do it on IBM SPSS:**

1

On the top menu of SPSS, you must click on "**Analyze**", then "**Scale**", then "**Reliability analysis**". The window shown will appear. Then, choose "only the items of the scale" to the box labelled as "items".

**You must always analyze one scale at a time! Do not mix items from different scales!**

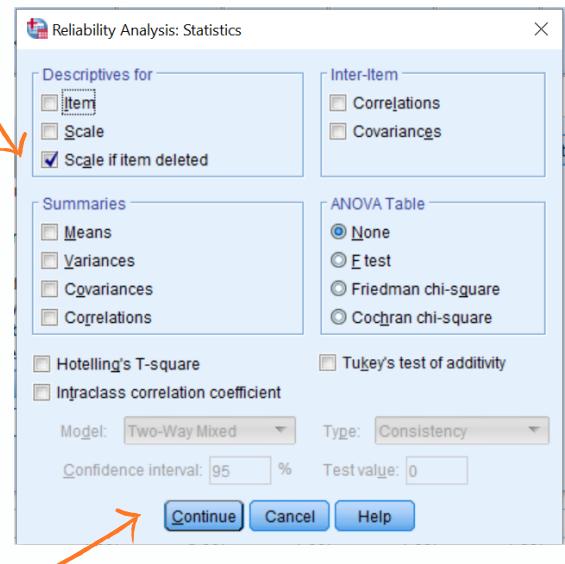


2

Next, you will click on "**Statistics**", then select "**scale if item deleted**". This way you will know if the reliability can be improved if one or more items are removed.

3

Next, you will click on "**Continue**", then "**Ok**" to see the results!



## Results

The interpretation of results could not be easier, my friend. The first table shows you two things: the actual **Cronbach's alpha score** and the **number of items of the scale**.

The number of items is 5 (we knew it already), and, importantly, our reliability score is **.59**. This is not good. If you remember, I showed you the benchmark before and below .6 is often considered **unacceptable**.

But wait a minute... The last column of the second table shows me something important: **Cronbach's Alpha if item deleted**. And it is telling me that if I remove the first item of the scale, and run the test again, the reliability value will increase to **.72**! (considered an **acceptable** reliability score). There was probably something weird with that item...

Reliability Statistics	
Cronbach's Alpha	N of Items
.591	5

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q9_Involvement_TheBeat_les_1	10.9730	9.397	.010	.721
Q9_Involvement_TheBeat_les_2	10.8514	7.115	.511	.448
Q9_Involvement_TheBeat_les_3	11.4730	7.376	.422	.495
Q9_Involvement_TheBeat_les_4	10.9189	7.665	.459	.484
Q9_Involvement_TheBeat_les_5	11.1351	6.913	.446	.477

# Shall we remove the first item and run the test again?

After going back to the menu, removing the first item and running the test again, **LOOK HOW BEAUTIFUL!** I now have a reliability score of .72 (acceptable) with only four items. And, importantly, the second table shows me that my reliability score would not improve (higher than .72) if I removed any of the other items.

**So these 4 items will provide me with the best reliability score for the scale. Thus, I would delete the other item from my data set and proceed with all other analysis!**

Reliability Statistics	
Cronbach's Alpha	N of Items
<b>.721</b>	<b>4</b>

Item-Total Statistics				
	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Q9_Involvement_TheBeatles_2	7.9865	6.150	.441	.698
Q9_Involvement_TheBeatles_3	8.6081	5.776	.495	.668
Q9_Involvement_TheBeatles_4	8.0541	6.024	.546	.642
Q9_Involvement_TheBeatles_5	8.2703	5.159	.564	.625

## So what can we take from this Cronbach's Alpha test?

### Sometimes things get better when someone leaves!

**Exactly like the story of Stuart Sutcliffe and The Beatles!** Not that Stu was the problem, but the band only got famous after he left. And do you want to know something sad about this story? Stu left the band in July 1961 and died less than a year later, in April 1962. Sad... right?



I am pretty sure you never imagined feeling emotional while reading about reliability tests for scales. I guess life sometimes simply surprises us. Please dry your tears before reading any further.

# Well done, my friend!

What should you remember from this section?



1

**The Beatles are the best band of all times. And that Stu left before they became famous.**

2

**Cronbach's Alpha is a reliability measure of multi-items scales and requires a minimum of three items to be calculated.**

3

**How to run and interpret Cronbach's Alpha and how to improve the score by removing possible "bad" items!**



## SPSS Tutorial playlist series.

In case you feel like you still need some further support on how to conduct a **Cronbach's Alpha in SPSS**, do not forget to watch the tutorial video with Francisco Tigre Moura on the SPSS playlist on the YouTube channel of **LiveInnovation.org**.



## Full soundtrack for Cronbach's Alpha

1. Trust me (Janis Joplin)
2. Trust in me (Scarlett Johansson)
3. Reliable (Golden Idols)
4. I'm a believer (The Monkees)
5. Believe what I say (Kanye West)



I've paid my dues  
Time after time  
I've done my sentence  
But committed no crime  
And bad mistakes  
I've made a few  
I've had my share of sand  
Kicked in my face  
But I've come through  
And we mean to go on and on and on  
and on

**We are the champions, my friends!**

(Freddy Mercury)

# WE ARE THE CHAMPIONS

Artist: Queen



So this is the end, my friend.

We are so proud of you for coming all the way here and for sticking through our questionable jokes along the different statistical tests. We honestly hope that this e-guide has helped you somehow. And who knows, that it may have even made you like statistics!

Obviously, there is much more to learn on statistics and SPSS. We have only covered a few tests. Who knows, maybe in the future we'll cover more.

So from now on we hope you will start enjoying the moments when you have a new exciting data set to analyze, either for a research project, a consulting deal or simply for your curiosity.

When these moments come, make yourself a nice drink, put some good music on and enjoy running the tests. Finding insights in data is fascinating. And beautiful. Sometimes all it takes is for someone to show us it from a beautiful angle.

And this is what we have tried to do in this e-guide. So if you feel it helped you, we would love to hear from you. Just reach out on any of the social media channels of [LiveInnovation.org](#).

Hope you also enjoyed the music. Life is always more beautiful with music.

And wherever you are, we hope you are having a wonderful day.

Cheers,

**Francisco and Silke.**



# Other Publications

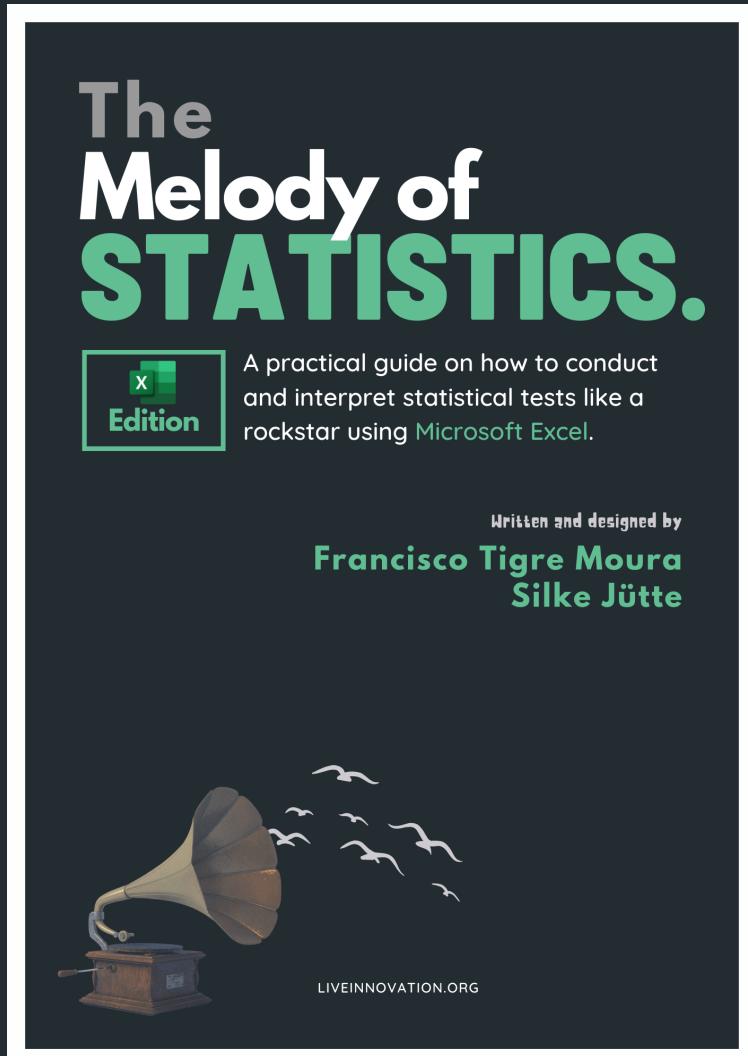
Trying to find an easier way of conducting many of these tests? Then you will like our version of the e-book for **Microsoft Excel**.

## DOWNLOAD THE E-BOOK THE MELODY OF STATISTICS - EXCEL EDITION

@ LIVEINNOVATION.ORG

The e-guide will walk you through most of the statistical analyses of this e-guide, with **Microsoft Excel**.

Chapters and sections are also labelled after songs and you can listen to the **full playlist on Spotify!**



# Other Publications

Are you planning to start writing your Bachelor or Masters thesis soon?

**DOWNLOAD THE E-BOOK **SOUNDS LIKE A THESIS****

@ LIVEINNOVATION.ORG

Published in 2021,  
the e-book will guide  
you through all  
chapters and stages  
of the thesis writing  
process!

Chapters and  
sections are also  
labelled after songs  
and you can listen to  
the **full playlist on**  
**Spotify!**



## **Sounds LIKE A THESIS!**



A quick guide on how to write and present your Bachelor or Masters thesis like a **Rockstar**.

Written and designed by  
**Francisco Tigre Moura**  
Academic collaborator  
**Clare Hindley**

LIVEINNOVATION.ORG



**Research to collaborate.  
Educate to inspire.**

LIVEINNOVATION.ORG

## **CONTACT**

**Prof. Dr. Francisco Tigre Moura**  
[francisco.tigre-moura@iu.org](mailto:francisco.tigre-moura@iu.org)