The Melody of STATISTICS.

Edition

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

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

04.07.2022

Visit 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

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

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

10 HOW OFTEN (Ben E. King)
How to conduct descriptive statistics

17 NOTHING ELSE MATTERS (Metallica)
Understanding statistical significance

21 NO ASSOCIATION (Silverchair)
How to conduct a Chi-Square test

28 MUTUAL MOTION (Loure)
How to conduct a correlation test

34 PREDICTABLE (Good Charlotte)
How to conduct a regression test

42 JUST THE TWO OF US (Bill Withers)
How to conduct a t-Test

52 THREE LITTLE BIRDS (Bob Marley and The Wailers)
How to conduct an ANOVA

64 WE ARE THE CHAMPIONS (Queen)
The end

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

The **Melody of STATISTICS.**

Edition

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. Predictable (Good Charlotte)
2. Hold the Line (Toto)
3. I Predict a Riot (Kaiser Chiefs)
4. All Down the Line (The Rolling Stones)
5. I Predict (Sparks)

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)

We are the champions (Queen)

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 Excel, 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 **Microsoft Excel** 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:

Especially this one!

- 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.

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** 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 **Microsoft Excel.**

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 green (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 (jazz) + 2 (pop) = 3 (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 **Microsoft Excel**, however, things are a bit difficult. **There is no concrete way of telling Excel upfront which variable is nominal, ordinal, interval or ratio scale**. A number in Excel is still just a number. So, if you tell Excel to calculate the average of the values 1 (=jazz) and 3 (=hip hop) in our example above, Excel will tell you the average is 2 (=pop) without batting an eye. However, of course:

THIS IS COMPLETE NONSENSE!

It is up to you to decide which measures you are dealing with (and hence, which analyses make sense or do not make sense). 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 "**Interval**".

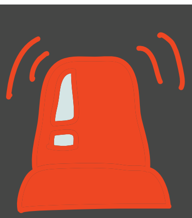
**So every time you have all your data,
play some good music in the background,
pour yourself a nice drink and sit cozy somewhere. Then...
become clear on the types of variables you are using and keep them in mind (or even note them down somewhere).**

Alright, great stuff!

So by now you should have understood:

1. **The different types of variables.**
2. **How the type of variable influences which calculations make sense.**
3. **The importance of being clear on the types prior to an analysis.**

Alright! Now you are almost ready to learn some statistical tests.



But before you must activate the Analysis ToolPak in Microsoft Excel.

We'll show you next!

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)

Activating the Analysis ToolPak

Excel users, pay attention...

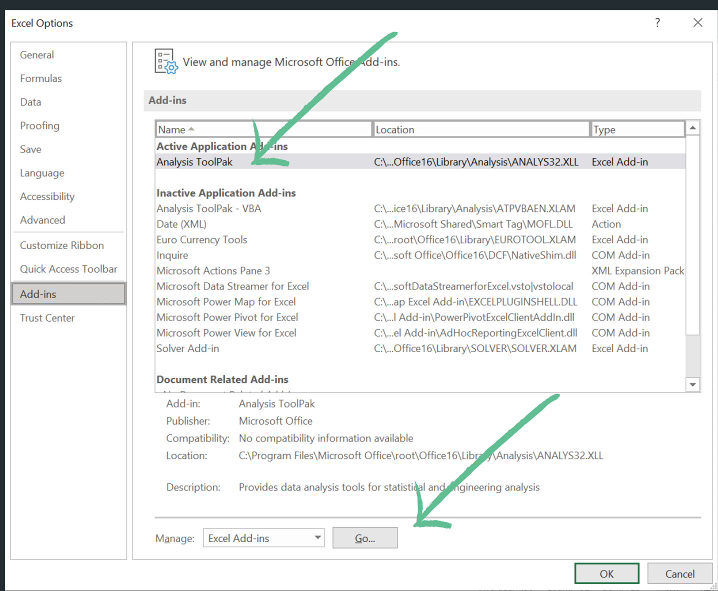
In the following chapters, we will walk you through running statistical analyses with Microsoft Excel. Now, if you have worked with it before, you might know that "using Excel" sometimes means applying pretty freaky formulas in Excel. And this can be very intimidating and exhausting...

But...

Lucky enough, most statistical tests in Excel can also be executed in a "select and click" manner, without the use of formulas. To do so, you have to activate a specific add-in in your Excel installation: the so-called **Analysis ToolPak...**

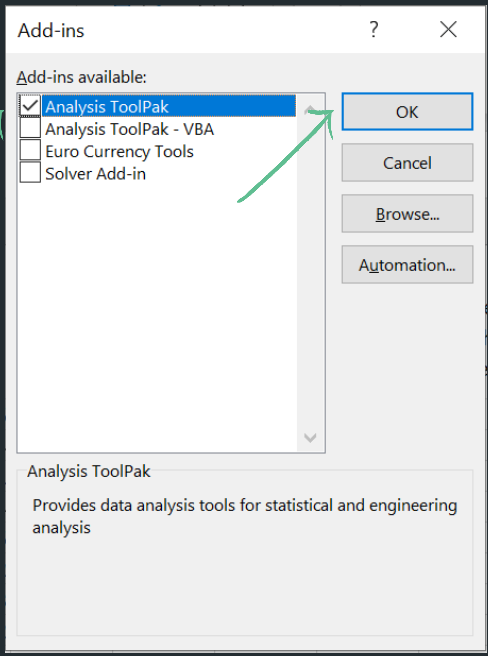
So here is how the magic is revealed:

1. With a **Windows operating system**, click on "**File**", then "**Options**". In the window that appears, choose "**Add-ins**" on the left menu bar, then select "**Go...**" next to "Manage: Excel Add-ins" on the bottom of the window.



2

In the next window, select "**Analysis ToolPak**" and click on "**ok**".



If you have a **Mac operating system**, activate the Analysis ToolPak by choosing **Extras > Excel Add-ins > Analysis Functions.**

The **Analysis ToolPak only has to be activated once**. If it is activated, you will see a separate option "Data Analysis" on the very right of the menu "Data".

Got it?

Then let's start with some **funky data analysis...**

HOW OFTEN

Artist: Ben E. King

Descriptive statistics.

And how to conduct it on **Microsoft Excel**.

2

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 work 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 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?

Tutorial Alert!

HERE IS HOW TO DO IT!

Frequency distribution (percentage)

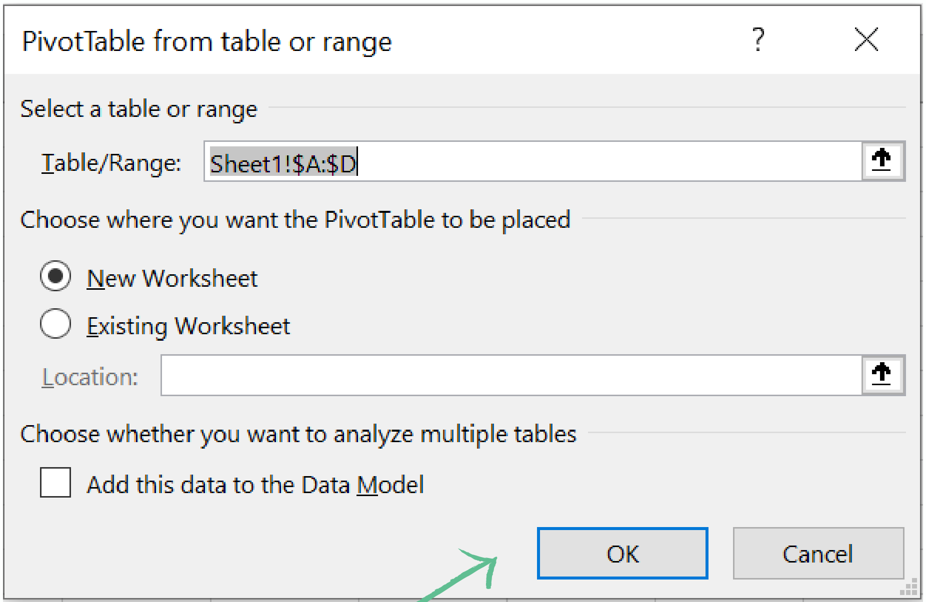
The first two questions regarding the festival that we would like to answer are:

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:

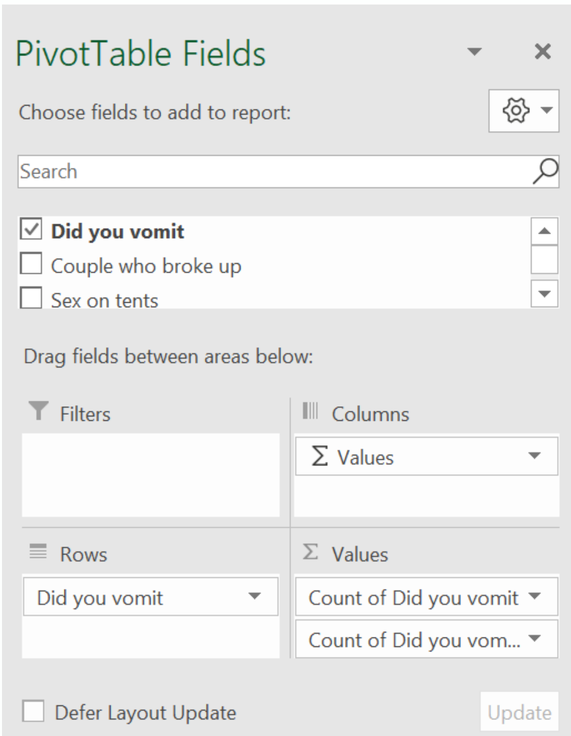
1

On the top menu of Excel, you must click on **"Insert"**, then **"Pivot Table"**. The window shown below will appear. At the field "Table/Range", simply choose the range of the variables that you wish to analyze (e.g. all variables in columns A to D), then click on "OK".



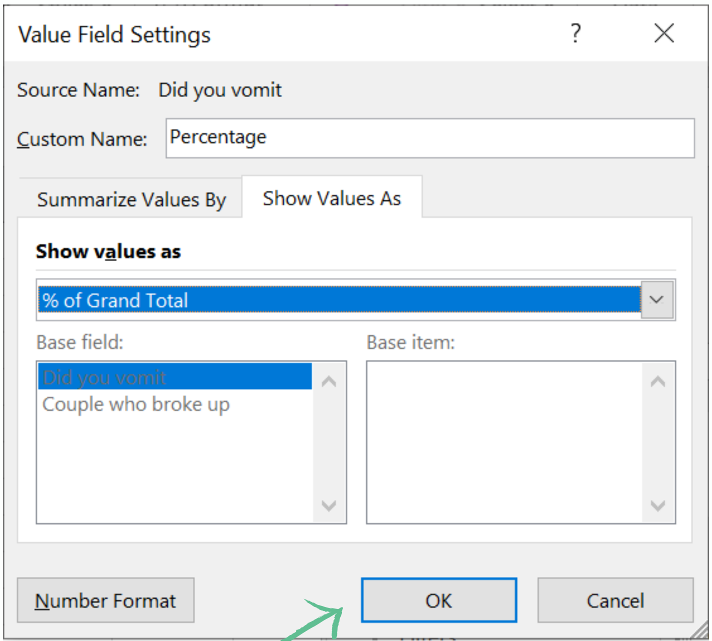
2

In the next window, tick the variable you would like to analyze first (e.g. "Did you vomit"). Then, drag and drop the variable to the "Values" field. Do this two times (since we want to display both the absolute and percentage frequencies of the variable values).



3

Click on the dropdown arrow next to the second entry at the "Values" list and choose "Value Field Settings...". The window shown below will appear. Choose the "Show Values As" tab and select "% of Grand Total". Change the "Custom Name" of the field to "Percentage". Then click on "ok":



Results

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

Row Labels	Count of Did you vomit	Percentage
Nope, no vomit	27	36,00%
Yes, vomited	48	64,00%
(blank)		0,00%
Grand Total	**75**	**100,00%**

The last two columns show the frequency of respondents ("count of..." means 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.

To obtain the frequency and percent values for the second variable ("Couple who broke up"), re-do steps 1-3 above for this variable. However, make sure to "unselect" the previous variable in step 2. Your result table will look like the table on the right. On the positive note, only 2 respondents (or 2.7%) broke up after the festival.

Row Labels	Count of Couple who broke up	Percentage
We are still together	73	97,33%
We broke up	2	2,67%
(blank)		0,00%
Grand Total	**75**	**100,00%**

So here are the answers to our questions:

Q1. Which percentage of visitors vomited during the festival?
- **Answer: 64%.**

Q2. Which percentage of couples broke up during the festival?
- **Answer: 2.7%.**

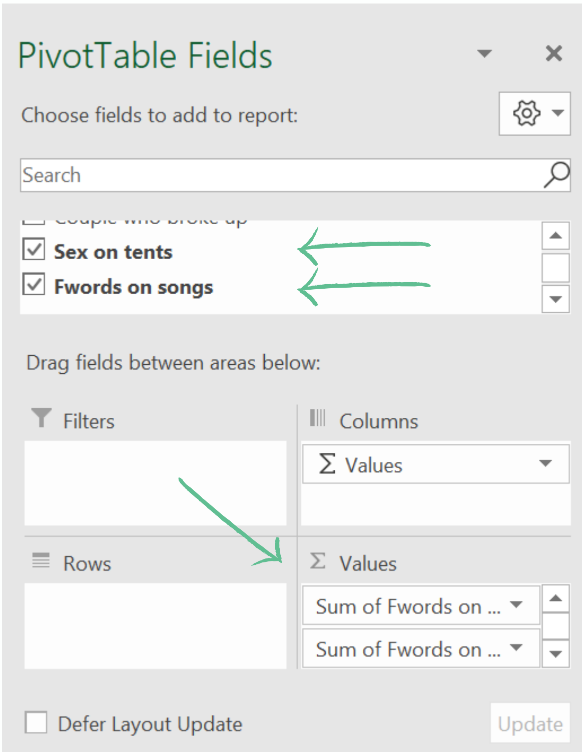
Averages (means)

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:

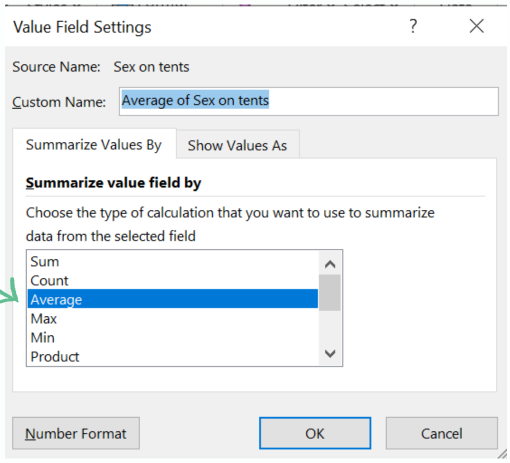
Q3. What was the average time of sex in the tents?
Q4. On average, how often did bands say the word F* on their songs?**

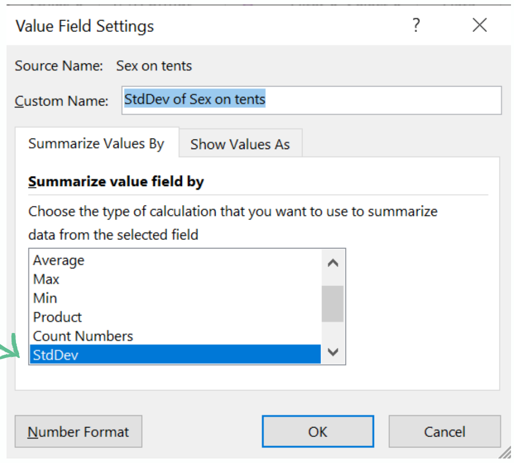
The **mean** is the average of the data. The **standard deviation** measures the spread of the data. For example, imagine you asked five people about how many months they have been playing the guitar. If their answers are: 2, 1, 2, 2, 2, this is a rather small spread of the data, so the standard deviation 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.

1 Get back to the "Pivot Table" you generated in the "Frequencies" section above. Unselect the previous variables and select both "Sex on tents" and "Fwords on songs". Doing this will automatically create an entry "Sum of..." for each of the variables in the "Values" section. Drag and drop each variable to the "Values" section, such that you end up with two "Values" entries for each variable:



2 Click on the dropdown arrow next to each entry in the "Values" list and select "Value Field Settings...". At "summarize value field by", once choose "Average" and once "StDev" for each of the two variables:





Results

So, again, the results table provides an overview of all responses.

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?).

Average of Sex on tents	StdDev of Sex on tents	Average of Fwords on songs	StdDev of Fwords on songs
11,90666667	10,39122984	14,50666667	21,50196304

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!

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!

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:

95% is the **minimum threshold.**

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 **Microsoft Excel**.

3

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.

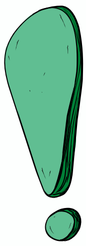
If only there was a test that would show me if there was any association between the genre played by musicians and the cause of death...

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.

Tutorial Alert!

HERE IS HOW TO DO IT!

Chi-Square test

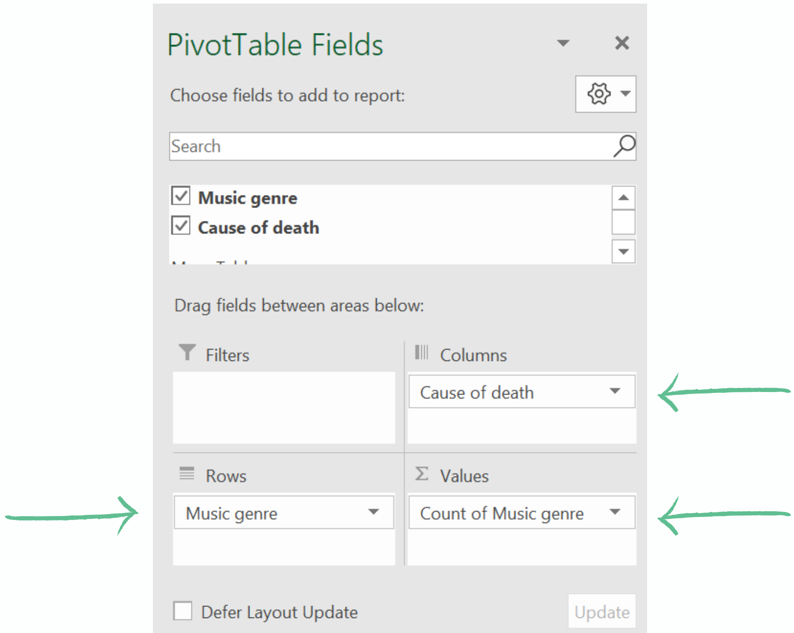
In the example at the start, we had one main burning question to answer, remember:

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

So, here is how to do it on Microsoft Excel:

1

On the top menu of Excel, click on **"Insert"**, then **"Pivot Table"**, and select the values of the variables "Music genre" and "Cause of death" as "Table/Range".
In the next window, drag and drop "Cause of death" to the "Columns" section and "Music genre" to both the "Rows" and the "Values" section:



Observed Frequencies

The resulting table will give you a summary of the so-called **observed frequencies**: For each combination of "Music genre" and "Cause of death", the number of observations in our sample have been calculated. For example, there are 3 Jazz musicians who died from Instagram overdose:

Count of Music genre	Column Labels					
Row Labels	Excessive Tik-Toking	Instagram overdose	Missing texts from their mothers	Uncontrolled Call-of-Duty use	(blank)	Grand Total
Heavy Metal	7	4	8	11		30
Jazz	7	3	9	3		22
Rap	5	10	3	5		23
(blank)						
Grand Total	19	17	20	19		75

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 wish there is an association or relationship between both categorical variables.

To do so, we next have to find the so-called **expected frequencies**. These are the frequencies for each combination of "Music genre" and "Cause of death" we would expect in case the two variables are completely independent.

Unfortunately, there is no Excel routine to automatically calculate these expected frequencies- so we will have to **assist Excel a bit**:

Side Note: Excel Handicaps

Microsoft Excel is NOT a dedicated statistics software. It was originally designed for spreadsheet processing, and the statistics features we use in this e-book have only been added in later versions of the software.

Though Excel's Analysis ToolPak already provides a reasonable scope of statistical analyses, there are still some handicaps. It is a bit like playing a song on your guitar, and in the middle of the song you realize that one of the strings is missing.

So, just like when playing a song, you sometimes have to improvise. We will give you a hint whenever this happens and will guide you through.

2

Select the cells of our Pivot table of "observed frequencies", right-click and select "copy". Then, select an empty cell below the Pivot table, right-click and select "Paste Special...", then "Values".
This way, you have copied the values of the original Pivot table, but not the formatting. **Insert headlines** saying **"observed"** and **"expected**" for the two tables. For the "expected" table, remove all values of observed frequencies, leaving only the row and column totals.

3

Next, we have to manually **calculate the expected frequencies**, which can be calculated as follows:

$$\text{EXPECTED FREQUENCY} = \frac{\text{COLUMN TOTAL} * \text{ROW TOTAL}}{\text{GRAND TOTAL}}$$

For example, the expected frequency of jazz musicians dying from Instagram overdose is (17*22)/75 = 4.99.
You can perform these calculations using Excel formulas as follows:

	A	B	C	D	E	F	G
1							
2	Observed:						
3	**Count of Music genre**	**Column Labels**					
4	**Row Labels**	**Excessive Tik-Toking**	**Instagram overdose**	**Missing texts from their mothers**	**Uncontrolled Call-of-Duty use**	**(blank)**	**Grand Total**
5	Heavy Metal	7	4	8	11		30
6	Jazz	7	3	9	3		22
7	Rap	5	10	3	5		23
8	(blank)						
9	**Grand Total**	**19**	**17**	**20**	**19**		**75**
10							
11	Expected:						
12	**Count of Music genre**	**Column Labels**					
13	**Row Labels**	**Excessive Tik-Toking**	**Instagram overdose**	**Missing texts from their mothers**	**Uncontrolled Call-of-Duty use**	**(blank)**	**Grand Total**
14	Heavy Metal	7,60	6,80	8,00	7,60		30
15	Jazz	5,57	=C$18*$G15/G18	5,87	5,57		22
16	Rap	5,83	5,21	6,13	5,83		23
17	(blank)						
18	**Grand Total**	**19**	**17**	**20**	**19**		**75**

Column total (blue value) is 17, row total (red value) is 22, grand total (purple value) is 75

4

Finally, we have to perform a Chi-Square test and calculate its corresponding p-value. This is done using the formula **CHISQ.TEST** in Excel:

The first parameter (blue range) specifies the observed frequencies, the second parameter (red range) the expected frequencies.

	A	B	C	D	E	F	G
1							
2	Observed:						
3	Count of Music genre	Column Labels					
4	Row Labels	Excessive Tik-Toking	Instagram overdose	Missing texts from their mothers	Uncontrolled Call-of-Duty use	(blank)	Grand Total
5	Heavy Metal	7	4	8	11		30
6	Jazz	7	3	9	3		22
7	Rap	5	10	3	5		23
8	(blank)						
9	Grand Total	19	17	20	19		75
10							
11	Expected:						
12	Count of Music genre	Column Labels					
13	Row Labels	Excessive Tik-Toking	Instagram overdose	Missing texts from their mothers	Uncontrolled Call-of-Duty use	(blank)	Grand Total
14	Heavy Metal	7,60	6,80	8,00	7,60		30
15	Jazz	5,57	4,99	5,87	5,57		22
16	Rap	5,83	5,21	6,13	5,83		23
17	(blank)						
18	Grand Total	19	17	20	19		75
19							
20	Chi-Square p-value:	=CHISQ.TEST(B5:E7;B14:E16)					

Expected Frequencies and Chi-Square

Having performed all these calculations, we arrive at the following tables and values:

	A	B	C	D	E	F	G
1							
2	Observed:						
3	Count of Music genre	Column Labels					
4	Row Labels	Excessive Tik-Toking	Instagram overdose	Missing texts from their mothers	Uncontrolled Call-of-Duty use	(blank)	Grand Total
5	Heavy Metal	7	4	8	11		30
6	Jazz	7	3	9	3		22
7	Rap	5	10	3	5		23
8	(blank)						
9	Grand Total	19	17	20	19		75
10							
11	Expected:						
12	Count of Music genre	Column Labels					
13	Row Labels	Excessive Tik-Toking	Instagram overdose	Missing texts from their mothers	Uncontrolled Call-of-Duty use	(blank)	Grand Total
14	Heavy Metal	7,60	6,80	8,00	7,60		30
15	Jazz	5,57	4,99	5,87	5,57		22
16	Rap	5,83	5,21	6,13	5,83		23
17	(blank)						
18	Grand Total	19	17	20	19		75
19							
20	Chi-Square p-value:	0,043516363					

First, recall the interpretation of "observed" and "expected" frequencies:

How to interpret: Observed and Expected Frequencies

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

Expected: The expected number of respondents per category. Example: if there is no association between music genre and cause of death, we would expect that, from the 23 rappers in our sample, 5.83 died from "excessive Tik-Toking", 5.21 from "Instagram overdose", 6.13 from "missing texts from their moms" and 5.83 from "uncontrolled CoD"

Aaaand finally (**please add a drum roll in the background**):

Here is the interpretation of the Chi-Square test result:

How to interpret: Chi-Square Test

Chi-Square 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!

Attention! Attention! Attention! 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???

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)

MUTUAL MOTION

Artist: Loure

Correlation test.

And how to conduct it on **Microsoft Excel**.

LIVEINNOVATION.ORG

4

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).

Tutorial Alert!

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:

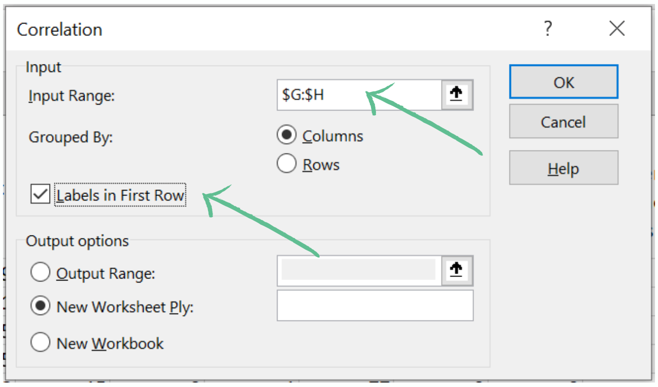
- **Q6. Do longer festivals lead to more pregnancies?**

So, here is how to do it on Microsoft Excel:

1 In Excel's main menu, select "**Data**", then "**Data Analysis**". In the window popping up, select "**Correlation**".

This only works if you have activated Excel's **Analysis ToolPak** (see Chapter 1 for details...)

2 In the next window, select the columns of the variables ("number of days of festivals" and "number of babies") as "Input Range", and tick "Labels in First Row".



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

Results

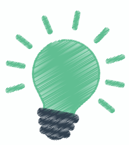
Excel will display a correlation matrix. The value 0.888 displayed below is the so-called **Pearson correlation coefficient (or "R-value")** between "Number of babies" and "Number of days of festivals". Here is how to interpret this value:

	Q6 Number of days of festivals	*Q6 Number of babies*
Q6 Number of days of festivals	1	
Q6 Number of babies	0,888440319	1

Interpreting Pearson's correlation coefficient:

1 **Positive or negative?** Positive values will indicate a positive correlation. In other words, if one variables 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.

2 **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, for the festivals we investigated, there is a **large positive correlation between the number of days of a music festival and the number of babies that are born afterwards.**

My friend, I trust that, by now, you are an attentive reader. And being attentive, you have probably recognized that I only said "for the festivals we investigated" when interpreting the correlation coefficient value. But was that just by random for the festivals we investigated, or **does that mean this positive association also holds for ALL music festivals?**

This is a matter of significance.

So here is the important second step of our analysis that we MUST NOT FORGET:

2

Significance test:
Now that I have (hopefully) made you enthusiastic about a significance test in correlation, I am afraid to tell you that **Excel's Analysis ToolPak misses such a test (don't know why).**

BUT DON'T PANIC: THERE IS A WAY AROUND IT!

"Remember what we told you about **Excel handicaps**? Here is the next one. This one feels a bit like someone provides you with the sheet music to a song, but did not add the final chord...

"Regression analysis" is a statistical method very much related to correlation analysis. (You will get to know more about this method in the next chapter.) For correlation analysis, we can simply **"borrow" the significance test of regression analysis.**

So for now, keep on reading the next chapter on regression analysis, then get back here to perform a regression significance test for our "length of festival" and "number of babies" question.

Spoiler alert:

The p-value Excel will show is **2.14E-26**, which is a VERY small value: namely 0.0000000000000000000000000214. As it is **below** our famous significance threshold of 0.05, this indicates that there is a significant correlation between both variables.

PS: An E within a number in Excel stands for "times 10 to the power of", so 2.14E-26 means "2.14 times 10 to the power of -26", which moves the decimal point 26 places to the left...

Attention! Attention! Attention!

My friend, once again, I need to draw you 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???

Are you really still thinking about what happens inside tents during music festivals?

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)

PREDICTABLE

Artist: Good Charlotte

Regression test.

And how to conduct it on **Microsoft Excel**.

5

Can you predict the success of a rap song based on the number of f*** words used?

My friend, I must admit: I am a very impulsive and impatient person. While waiting at the supermarket checkout, I frequently find myself murmuring curse words towards the people at the top of the queue if they dilly-dally while putting their shopping back to their cart. If you are the same, you should be aware that children accompanying you are very attentive listeners. So next time you find yourself at the playground, don't be surprised if you hear your innocent little son shouting:

"C'MON, MOVE YOUR ASS!"

to the little guy standing on the slide...

Ever since, I have become more careful with what I say when my kids are around. But there is one thing I we cannot control, no matter how hard we try: the songs our kids listen to, especially when they learn how to stream music. And if they come to love rap music, you'll be in a for a treat!

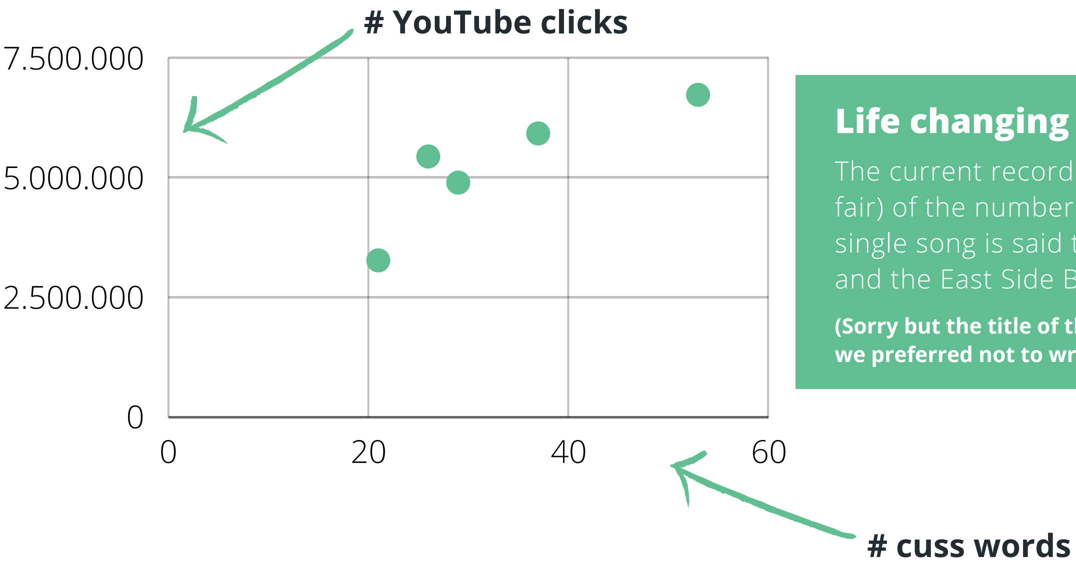
Have you ever noticed how many cuss words there are on average in each rap song? And trust me: your kids will soak these up like a sponge! So this got me wondering why there are so many f*** words in the lyrics, especially in rap music. And made me come up with a pretty interesting research question:

Q7. Can we predict a song's success by the number of cuss words used?

For example, here's a **fictitious** list of songs of our all-time favorite made-up rap artist: the one and only... **Dr. Stats!**

TITLE	# CUSS WORDS	# YOUTUBE CLICKS
"Dr Stats is in da house"	26	5,437,212
"That f***in' numbr"	53	6,722,085
"Ain't no good analysis"	21	3,270,589
"Got da significance"	29	4,892,661
"Nuthin' to correlate"	37	5,917,332

To visualize this data, we can put it in a so-called **scatter plot**. Here's how the scatter plot would look like for our data:



Life changing fact

The current record (questionable, to be fair) of the number of swear words in a single song is said to be held by Lil Jon and the East Side Boyz with **329!**

(Sorry but the title of the song is quite... strong. So we preferred not to write it here).

And my friend, look how interesting! by looking at the graph, apparently, there is a positive relationship between the variables: **The more cusses, the more clicks**. We can confirm this with correlation analysis, which might even tell us that this is a strong relationship.

But what if we want to know more?

If there is such a strong relationship, would there be a way to predict the number of youtube clicks just by knowing the number of cuss words used in the lyrics?

What we need is then a straight line approximating the relationship between our variables. And now dig-deep in your memory, we are sure your high-school math teacher tortured you with this: **What do you need to describe a straight line?**

Think about it.
We'll wait.

My friend, you need a **slope** and a **y-intercept.**

And what is that? Well...

THE SLOPE The slope will describe the steepness of the line. In our case, it will indicate how many clicks we get per additional cuss word used

THE Y-INTERCEPT The y-intercept will indicate where the line intersects the y-axis. In other words: What is the number of clicks for a rap song without any single cuss word used in the lyrics? (Well, admittedly, that's a rather theoretic concept for our example...)

Once you know the slope and y-intercept, you can describe any point on our straight line with the following equation, the so-called regression equation:

ESTIMATED #CLICKS = SLOPE * #CUSS WORDS + Y-INTERCEPT

Now let us show you how Excel can help finding the slope and y-intercept:

Tutorial Alert!

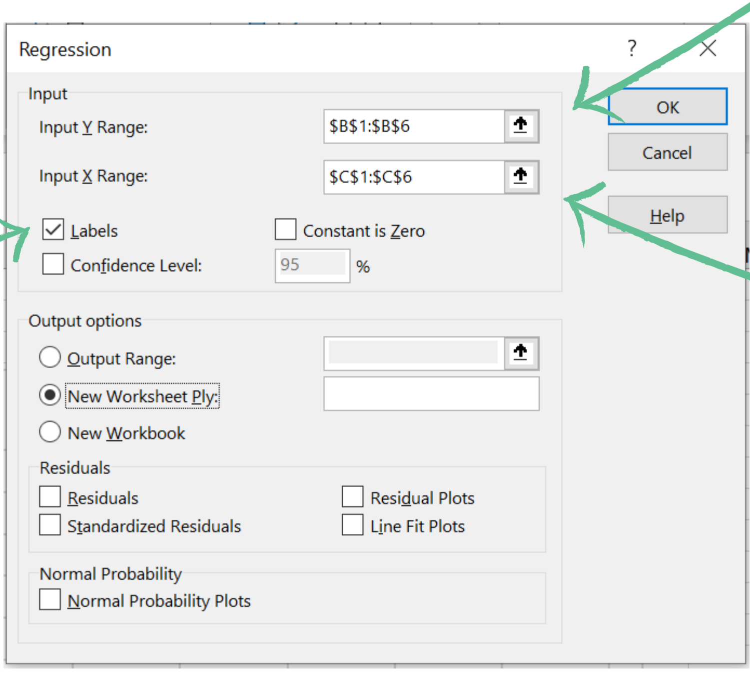
HERE IS HOW TO DO IT!

1 To start a regression analysis in Excel, choose the **"Data"** menu. Choose **"Data Analysis"** on the very right and select **"Regression"** in the following pop-up window.

If the "Data Analysis" option does not show, repeat step 1 of the correlation chapter to **activate the Analysis ToolPak**.

2 Tell Excel which data you want to run the regression analysis on:

cells of "dependent variable" (this is the variable you want to predict, so **#clicks** in our example)



tick this checkbox if your selection of cells contains headlines

cells of "independent variable" (this is the variable you base your prediction on, so **#cuss words** in our example)

IMPORTANT: Excel can be very picky (just like an annoying band colleague who keeps telling you he cannot continue like this, because you are not exactly following the beat). Here, **Excel does not allow any empty cells in the Input X Range or Input Y Range**. You can save lots of discussion with this annoying band colleague by sticking to his rules right away...

3 Once you have managed your way across potential error pop-ups, Excel will display the following **results of the regression analysis:**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.865735629
R Square	0.749498179
Adjusted R Square	0.665997572
Standard Error	747697.3726
Observations	5

B) model fit

A) coefficients

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	5.01802E+12	5.01802E+12	8.975960828	0.057853455
Residual	3	1.67715E+12	5.59051E+11		
Total	4	6.69518E+12			

C) significance

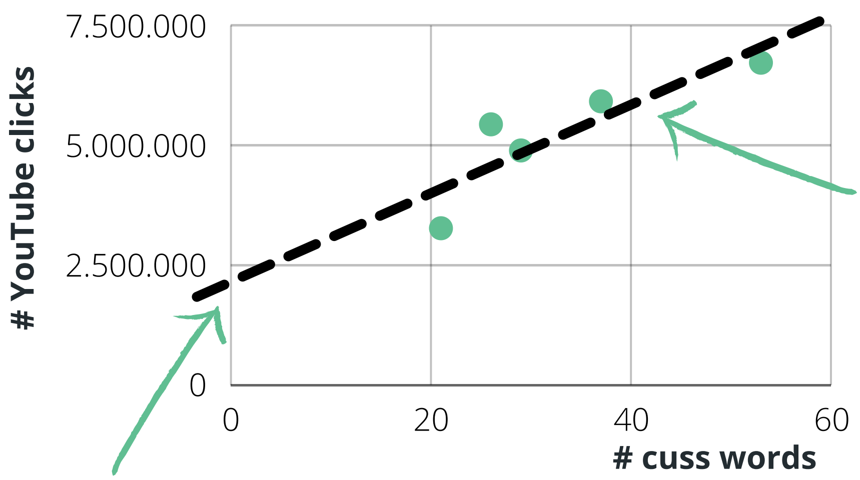
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2272278.489	1047883.537	2.168445642	0.118629287	-1062554.602	5607111.58	-1062554.602	5607111.5
Q7: number of cuss words	89618.1178	29912.68132	2.995990792	0.057853455	-5577.384335	184813.6199	-5577.384335	184813.619

OMG, THAT'S AN AWFUL LOT OF NUMBERS!

But don't worry, I will walk you through... Follow me!

A) **Coefficients:** Remember we wanted to find the slope and y-intercept of the straight line approximating our relationship. These can be found in the "coefficients" column. Hence, our regression equation looks like follows:

ESTIMATED #CLICKS = 89,618* #CUSS WORDS + 2,272,278



slope: Each additional cuss word that Dr Stats adds to his lyrics would yield 89,618 clicks extra

y-intercept: a Dr Stats song without any cuss words would yield 2,272,278 clicks

So now we have a mathematical description of the straight line in the picture above. (From now on, we will call this line the **regression line**.) But, as we can clearly see, this straight line is simplifying reality: the points of the scatter plot are *not exactly* on the line. Hence, how good does the line represent our data? This is a question of the model fit:

B) **Model fit:** Excel calculates a measure called "**R-Square**". (If you want to impress people, you can also call it the **coefficient of determination**.) This value shows how good your model (so, the straight line approximation) represents your data (so, the points of the scatter plot). R-Square can be any value between 0 and 1. The closer the points are to the line, the higher the R-Square value. Hence, the best model fit is obviously one with an R-Square of 1 (and this happens when all points of the scatter plot are *exactly* on the regression line).

SMART-ASS KNOWLEDGE

There is a reason why R-Square is called R-Square: Remember the correlation coefficient is called R? Well, using some freaky math, you can show that, in fact, the coefficient of determination IS the square of the correlation coefficient...

In our case, Excel indicates that the R-Square value is around 0.7495. In other words:

Round about 75% of the variation in the number of youtube clicks can be explained by our model (and about 25% cannot).

(VERY ROUGH) RULE OF THUMB:

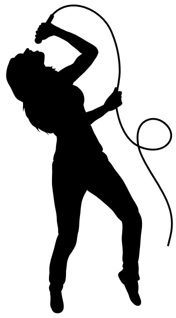
If R-Square is at least 70%, the model fit is acceptable.

(If not, you typically continue by collecting more data and/or adding additional variables to your model...)

So this means we are dealing with a model of acceptable quality. So can we go ahead using this model for predicting the number of clicks for Dr Stats' new record?

NO!!!!!

By now, there is hopefully a bell ringing somewhere in your mind, and you can join us singing the chorus:



"It's all about significance..."

C) **Significance:** Remember that we still don't know whether our model holds for the whole population- so, all (past, present and future) rap songs of Dr Stats- or whether the relationship we modeled in the regression line is born out of pure coincidence.

To judge on this, we have to evaluate the **p-value of the underlying significance test** which Excel was so kind to execute for us. You can find this p-value in the column "P-value" and the row "number of cuss words" of the last table: it is about 0.0579.

Unfortunately, our p-value is *greater* than 5%, which means that, unfortunately, the relationship is **NOT significant**.

So sorry Dr Stats, but we are afraid this model is not a money-making machine: it could still be that the relationship we observed in the sample was pure coincidence.

NOTE: The reason for OUR model being non-significant is the small sample size. In any case, never trust an analysis that builds on only five observations...

Full soundtrack for
Regression test

1. Predictable (Good Charlotte)
2. Hold the Line (Toto)
3. I Predict a Riot (Kaiser Chiefs)
4. All Down the Line (The Rolling Stones)
5. I Predict (Sparks)

JUST THE TWO OF US

Artist: Bill Withers

Independent and dependent (paired) samples t-tests.

And how to conduct them on **Microsoft Excel**.

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:

Q8. Do cows produce more milk when listening to rap as 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:

1. **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).

2. **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!

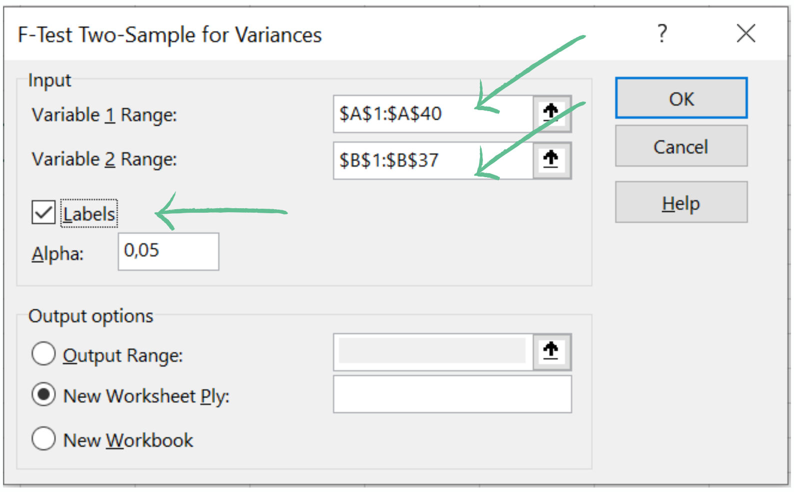
Independent samples t-test

Alright... Are you ready to find out whether rap from 50 Cent of 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 Microsoft Excel:

1. First, we have to **separate the values of the "amount of milk" variable for the two music genres**. You can do so by first filtering by "music genre = rap" and copy-paste the corresponding "amount of milk" values to one separate column, then do the same with "music genre = heavy metal". Add appropriate column headings.

2. Because there are several "independent samples t-tests" in Excel, we first have to decide which one to use. To do so, you must click on **"Data"**, then **"Data Analysis"**, then **"F-Test Two-Sample for Variances"**. The window shown will appear.



Select the "amount of milk" values for rap music as "Variable 1 Range", and the values for heavy metal music as "Variable 2 Range".

Don't forget to tick "Labels", and confirm with "ok".

This only works if you have activated Excel's **Analysis ToolPak** (see Chapter 1 for details...)

Results of F-Test

Alright... Let's see what we have got! The table you will find is the one you see below. It gives a few important information: First, reading from the "observations" row, we we can see that 39 cows listened to rap, while 36 cows listened to heavy metal.

F-Test Two-Sample for Variances		
	Q7 Amount of milk RAP only	*Q7 Amount of milk HEAVYMETAL only*
Mean	9,564102564	4,361111111
Variance	9,673414305	3,551587302
Observations	39	36
df	38	35
F	2,723687603	
P(F<=f) one-tail	0,001744263	
F Critical one-tail	1,74330402	

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!

Now let us concentrate on the second part of the F-test results table. Remember that we wanted to find out **which independent t-test is appropriate for our setting**. This is what we exectued the F-test for:

In a nutshell... The F-test 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 F-test is non-significant (p>.05), then the assumption of "homogeneity of variances" IS **MET**. We will hence execute an **independent t-test assuming EQUAL variances**.

If the F-test is significant (p<.05), then the assumption **IS NOT MET**. We will hence execute an **independent t-test assuming UNEQUAL variances.**

The p-value of the F-test can be found in the "P(F<=f) one-tail" row. In our case, it is about 0.0017, hence smaller than 0.05, so we can assume unequal variances.

3

In Excel's top menu, click on **"Data"**, then **"Data Analysis"**, then **"t-Test: Two-Sample Assuming Unequal Variances"**. In the window shown next, specify the cells of "milk with rap" and "milk with heavy metal" as input ranges, tick "Labels", and click on "ok".

Results of Independent t-Test

t-Test: Two-Sample Assuming Unequal Variances		
	Q7 Amount of milk RAP only	*Q7 Amount of milk HEAVYMETAL only*
Mean	9,564102564	4,361111111
Variance	9,673414305	3,551587302
Observations	39	36
Hypothesized Mean Difference	0	
df	63	
t Stat	8,836525522	
P(T<=t) one-tail	6,21435E-13	
t Critical one-tail	1,669402222	
P(T<=t) two-tail	1,24287E-12	
t Critical two-tail	1,998340543	

Here's the result table. The p-value of the t-test can be found in the **"P(T<=t) one-tail"** row.

One-tail test means we test whether one mean is **greater than** another one. If we merely wanted to test whether one mean is **different to** another one, this would have been a **two-tail test**.

So... Since the independent samples t-test result is significant (p=6.21E-13)...

We conclude that the **amount of milk** produced **when** cows are **exposed to rap music is statistically greater** than when exposed to heavy metal.

Paired 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.**

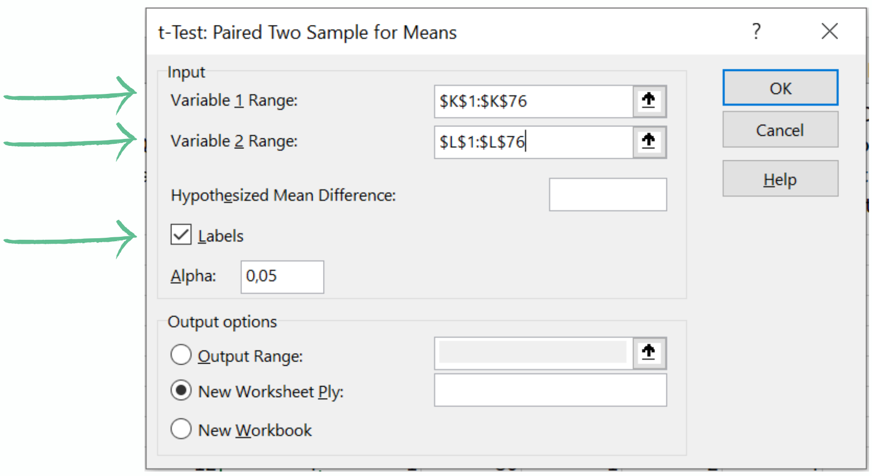
And what if each cow had listened to both music genres?

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**.

Recall that you need to have activated Excel's **Analysis ToolPak** for this to work (see Chapter 1)...

Here is how to do it on Microsoft Excel:

1. On the top menu of Excel, you must click on **"Data"**, then **"Data Analysis"**, then **"t-Test: Paired Two Samples for Means"**. The window shown will appear.
Then, choose the cells of the two variables' values ("Cow rap milk" and "Cow heavy metal milk") and their headings as "Variable 1 Range" and "Variable 2 Range". Also, make sure to tick "Labels".



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

Results of paired samples t-test

So... The result table is the one you see below. the first three lines will give you some interesting information: In line 3, you see the number of observations, so there were 75 cows in each group. Line 1 it gives us the averages. While listening to rap, a cow produced an average of 10.55 liters, while listening to heavy metal it produced an average of 4.36 liters.

t-Test: Paired Two Sample for Means		
	Q7 Cow rap milk	*Q7 Cow heavymetal milk*
Mean	10,54666667	4,36
Variance	9,359279279	2,233513514
Observations	75	75
Pearson Correlation	0,080511865	
Hypothesized Mean Difference	0	
df	74	
t Stat	16,26076221	
P(T<=t) one-tail	1,94485E-26	
t Critical one-tail	1,665706893	
P(T<=t) two-tail	3,8897E-26	
t Critical two-tail	1,992543495	

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**. The p-value is given in the **P(T<=t) one-tail** row. As you can see above the p-value is 1.94485E-26, so veeeery close to zero, 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!**

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!**

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 **Microsoft Excel**.

7

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 the greatest amount of car accidents?

In other words....

Q9: 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:

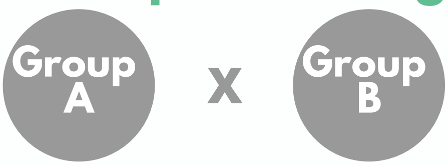
If you have **one** independent variable, you apply a **One-way ANOVA**.
If you have **two** independent variables, you apply a **Two-way ANOVA.**

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

But why don't we just do several t-tests?

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!

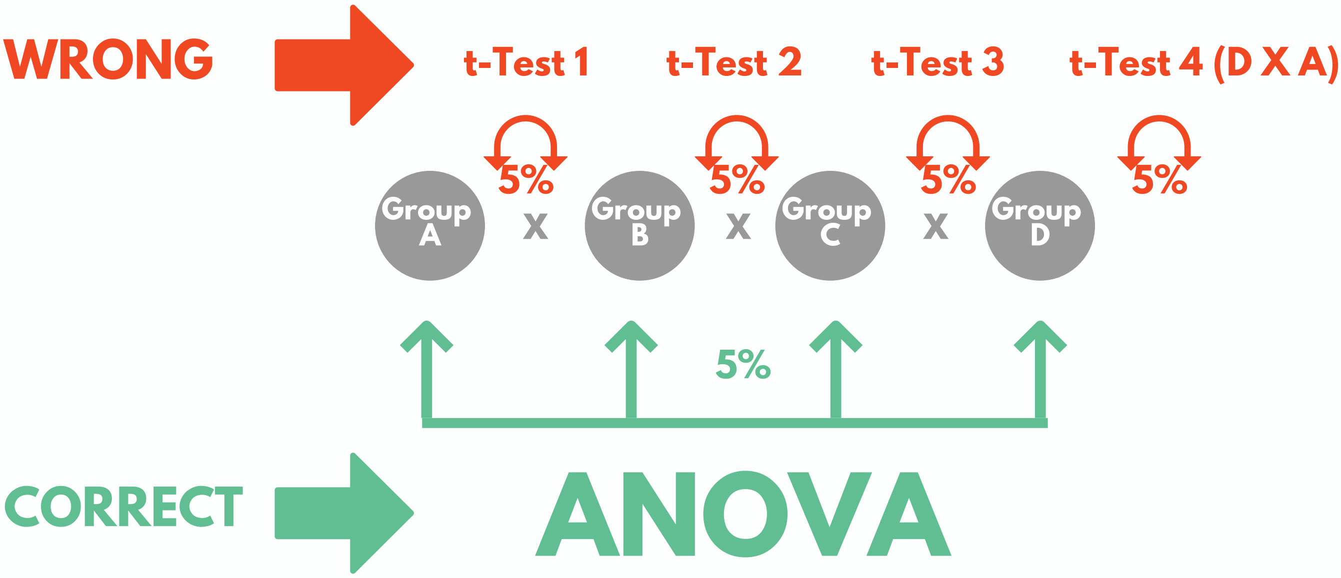


And when comparing these two groups, remember: **the maximum amount of randomness allowed is 5%! ($p<0.05$).**

And what happens when we compare more than two groups?

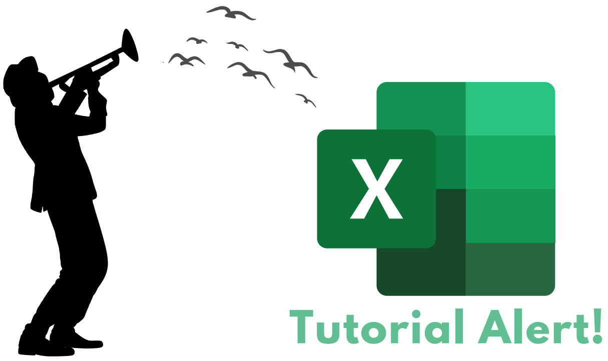
Well... For example, you **could** think: "if I have four groups to compare, can't I just run four t-tests and contrasts them all simply by running multiple t-tests?" **UNFORTUNATELY, THINGS ARE A BIT MORE COMPLICATED...**

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?

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



HERE IS HOW TO DO IT!

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 Microsoft Excel:

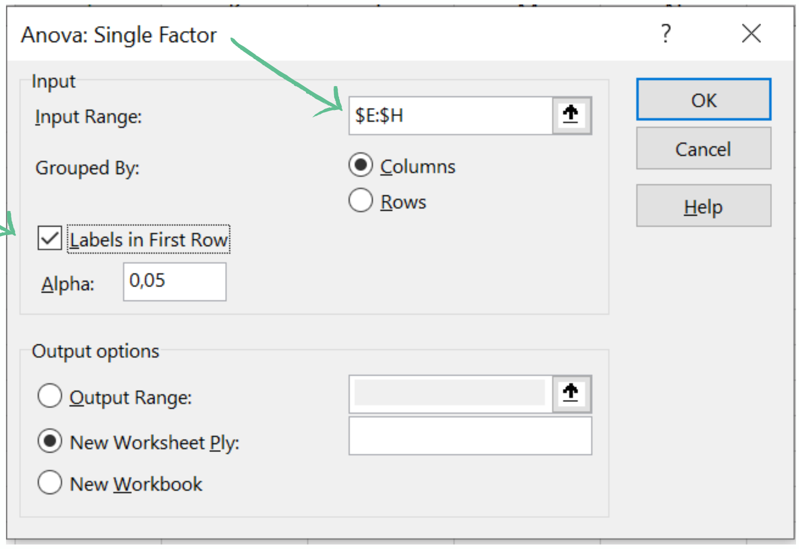
1. First, we have to **separate the values of the "number of car accidents" variable for the four radio stations**. You can do so by first filtering by "Radio stations cities = Punk rock" and copy-paste the corresponding "number of car accidents" values to one separate column, then do the same with the other radio stations. Add appropriate column headings.

2 Next, go to the "**Data**" menu and select "**Data Analysis**" on the very right, then choose "**Anova: Single Factor**".

This only works if you have activated Excel's **Analysis ToolPak** (see Chapter 1 for details...)

3 Next, specify the columns you have created in step 1 as the Input Range, and make sure to tick "Labels in first Row".

Finished? Then click on "OK" for the results!



Results

Alright! 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 Count (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.

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Punk rock	19	934	49,15789474	194,251462
Classical music	19	400	21,05263158	25,83040936
Trash metal	19	919	48,36842105	53,57894737
Reggae	18	345	19,16666667	10,85294118

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?

Noooooooooooo!

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?

Well, have a look at the second table!

ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	15402,88526	3	5134,295088	71,33205359	2,2118E-21	2,73364716
Within Groups	5110,394737	71	71,97739066			
Total	20513,28	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:

1. If the the ANOVA result is **NON-SIGNIFICANT (p>.05)**, we assume that there are **no differences between the groups**.

2. If the the ANOVA here is **SIGNIFICANT (p<.05)**, we assume that there **are differences between the groups**.

In this case, the ANOVA result was highly significant (p=2.2118E-21, so veeeery close to zero).

Conclusion?

We can confidently say that the type of music you are listening to (punk rock, classical music, trash metal or reggae) makes a difference when considering the number of accidents you have.

However, this is only one part of the story. There is still a very important open question:

Which groups are statistically significantly different from which groups?

To find the answer, we will now have to compare our groups in pairs. If you are working with a dedicated statistics software, this so-called **post-hoc test** is included automatically in an ANOVA analysis. But I think by now you already have an unpleasant suspicion here: **Excel's Analysis ToolPak does not provide an in-built post-hoc test...**

But don't worry:

Armed with all the statistical knowledge you have gained by now, you have all the skills it takes to find a way around...

Yet another example of **Excel handicaps.** This time, it even feels like someone stops a song in the middle of the chorus. So again, we have to run the last few meters by foot...

Essentially, we can still **compare our groups in pairs using several t-tests**. If you have many groups to compare, this will unfortunately take some time... If you have four groups to form pairs of (as in our case), you will have to conduct six t-tests.

Yes, this is pretty annoying- which is one of the reasons we recommended you NOT to do this in the first step, you might remember. You only conduct these t-tests if the ANOVA analysis tells you that there IS a difference between at least two groups. **(And only because the Excel Analysis ToolPak does not have an implemented post-hoc test, which would have been way more comfortable...)**

But haven't we just learned that, when conducting multiple t-tests, the single percentages of randomness could add up?

Yup, indeed.

I am proud of you, my friend, for being such an attentive reader.

And to overcome this problem, we will adjust the significance level, using the so-called Bonferroni correction:

Essentially, all this Bonferroni correction does is **dividing our former significance level by the number of comparisons we are making:**

$$\text{Corrected significance level} = \frac{\text{original significance level}}{\text{number of t-tests executed}}$$

In our case, the original significance level was 5% = 0.05, and we will have to execute six separate t-tests now, **so the Bonferroni-corrected significance level will be 0.05/6 = 0.0083.**

So, now let's go ahead executing six separate t-tests, just as we learned in Chapter 6:

- **t-test #1** will compare punk rock and classical music
- **t-test #2** will compare punk rock and trash metal
- **t-test #3** will compare punk rock and reggae
- **t-test #4** will compare classical music and trash metal
- **t-test #5** will compare classical music and reggae
- **t-test #6** will compare trash metal and reggae.

The **Bonferroni correction** is the simplest of several ways to compensate for a potential accumulation of randomness, and it is comparably strict. Check out a statistics book if you want to know more....

We know, we know, this is tedious work... But clench your teeth and stay with us, we are close to the finishing line!

If you do not remember t-tests, go back to Chapter 6 now. What we will need here is **two-tailed independent t-tests assuming unequal variances**:

- **two-tailed**, because we are merely interested in whether there is a difference between the groups considered
- **independent**, because our groups do not overlap
- **unequal variances**, because our sample variances (as calculated above) are pretty different already, so we want to be on the safe side (and skip executing additional F-tests)

Results

Here is the result of our six t-tests (we copied and pasted the Excel result tables to one single sheet to get a better overview on the results):

t-Test: Two-Sample Assuming Unequal Variances						
	Punk rock	*Classical music*	*Punk rock*	*Trash metal*	*Punk rock*	*Reggae*
Mean	49.15789474	21.05263158	49.15789474	48.36842105	49.15789474	19.16666667
Variance	194.251462	25.83040936	194.251462	53.57894737	194.251462	10.85294118
Observations	19	19	19	19	19	18
Hypothesized Mean Difference	0		0		0	
df	23		27		20	
t Stat	8.257951378		0.218593459		9.11477917	
P(T<=t) one-tail	1.23885E-08		0.41430532		7.34104E-09	
t Critical one-tail	1.713871528		1.703288446		1.724718243	
P(T<=t) two-tail	2.47771E-08		0.82861064		1.46821E-08	
t Critical two-tail	2.06865761		2.051830516		2.085963447	
	Classical music	*Trash metal*	*Classical music*	*Reggae*	*Trash metal*	*Reggae*
Mean	21.05263158	48.36842105	21.05263158	19.16666667	48.36842105	19.16666667
Variance	25.83040936	53.57894737	25.83040936	10.85294118	53.57894737	10.85294118
Observations	19	19	19	18	19	18
Hypothesized Mean Difference	0		0		0	
df	32		31		25	
t Stat	-13.36148469		1.346281297		15.78384257	
P(T<=t) one-tail	6.13595E-15		0.093987129		8.25206E-15	
t Critical one-tail	1.693888748		1.695518783		1.708140761	
P(T<=t) two-tail	1.22719E-14		0.187974257		1.65041E-14	
t Critical two-tail	2.036933343		2.039513446		2.059538553	

The relevant p-values are highlighted already. Recall that we are conducting two-tailed tests, which means we have to look up the p-value in the **P(T<=t) two tail** row

We will interpret the Punk rock stations (p-values in top row) to illustrate how to interpret it all, ok?

For example: as you see above, the **Punk Rock radio stations** (average 49.2 crashes), is significantly different from the **Classical radio stations** (average 21.1 crashes), and significantly different than the **Reggae radio stations** (average 19.2 crashes). In both comparisons, p-value was very small (having an E-08 in it), so definitely smaller than our corrected (!) significance level of 0.0083. **But Punk Rock is NOT significantly different from Trash metal radio stations** (48.4 crashes per year). Here the p-value was 0.8286 (>0.0083).

Conclusion?

We can confidently say that drivers have more car crashes when listening to Punk Rock 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 we're nice and we like you, here is a quick summary on how to interpret the One-Way ANOVA:

1. **Descriptive table:** First, analyze the descriptive table to interpret the averages of each group you are comparing.
2. **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 conduct a Post-Hoc analysis (multiple comparison).
3. **Post-Hoc analysis:** Use independent t-tests for a pairwise comparison of groups. Make sure to use the two-tailed p-values and compare them against the Bonferroni-corrected significance level.

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!**

So this is it.
You have reached the end of the guide!

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. 3. Stock (AnnenMayKantereit)

WE ARE THE CHAMPIONS

Artist: Queen

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)

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 Excel. 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

Annoyed by all these tiny Excel handicaps when running statistical analyses? Then you might consider switching to a dedicated stats software like IBM SPSS.

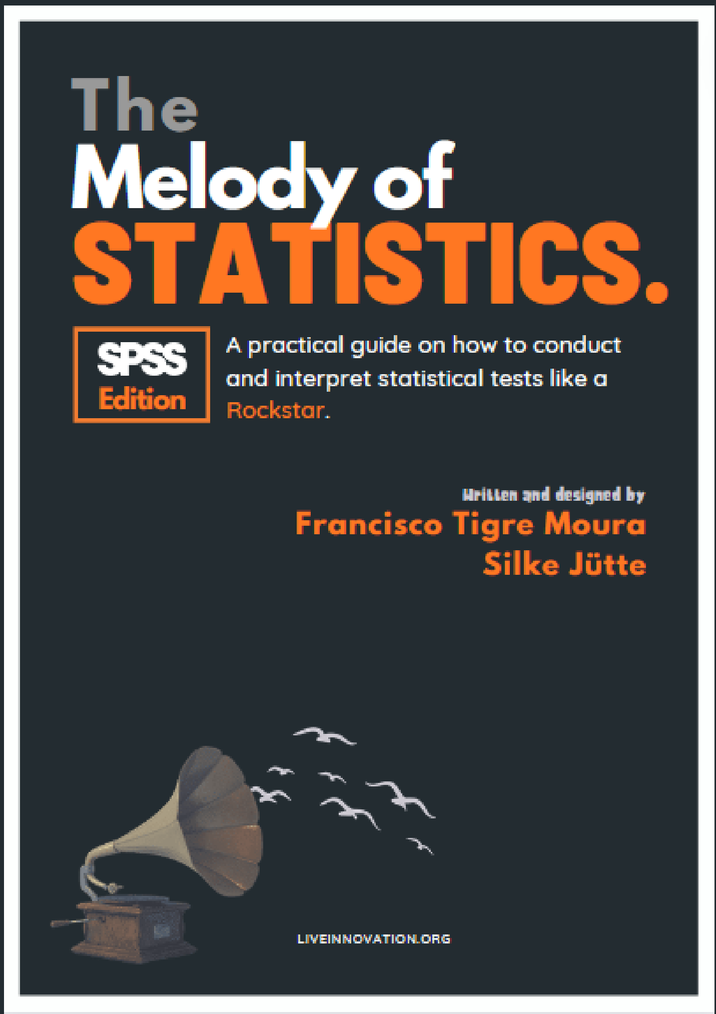
DOWNLOAD THE E-BOOK

THE MELODY OF STATISTICS - SPSS EDITION

@ LIVEINNOVATION.ORG

The e-book will walk you through essentially the same statistical analyses as this one, though with the software **IBM SPSS**, which is a very powerful statistics engine.

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



...And even more Publications

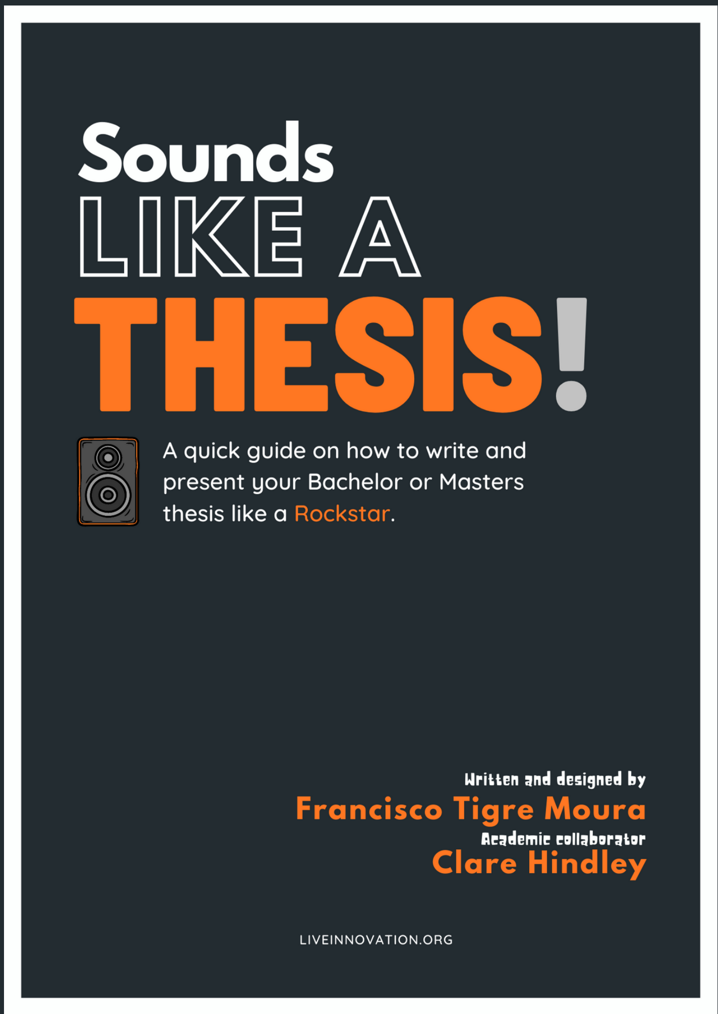
Are you planning to start writing your Bachelor or Masters thesis soon?

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!**



Is your curiosity alive?

LiveInnovation.org

Research to collaborate.
Educate to inspire.

LIVEINNOVATION.ORG

CONTACT
Prof. Dr. Francisco Tigre Moura
francisco.tigre-moura@iu.org