

The Melody of STATISTICS.

2nd
Edition

 **sas**
Edition

A practical guide on how to conduct and interpret statistical tests like a rockstar using **SAS Studio**.

Written 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
and while trying to learn statistics.

Francisco and Silke.

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.

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
- 8** **HOW OFTEN** (Ben E. King)
How to conduct descriptive statistics
- 15** **NOTHING ELSE MATTERS** (Metallica)
Understanding statistical significance
- 19** **NO ASSOCIATION** (Silverchair)
How to conduct a Chi-Square test
- 26** **MUTUAL MOTION** (Loure)
How to conduct a correlation test
- 32** **PREDICTABLE** (Good Charlotte)
How to conduct a regression test
- 40** **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
- 63** **TRUST ME** (Janis Joplin)
How to conduct a Cronbach Alpha test
- 71** **WE ARE THE CHAMPIONS** (Queen)
The end

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 SAS, 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 **SAS Studio** and also how to interpret them.

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

**This
is not a
statistics book.**

**This e-book is a quick guide to help you conduct a few tests.
This e-book is a quick guide to help you conduct a few tests.
This e-book 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. & Miles, J. *Discovering Statistics Using SAS*. 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!

Finally: Download the data sets!

You can try for yourself every example shown in this e-book! Download the data sets and the stage is yours.



ONES AND ZEROS

Artist: Jack Johnson

Types of quantitative variables.

And how to insert them on **SAS Studio**.



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: $1 (\text{jazz}) + 2 (\text{pop}) = 3 (\text{hip hop})$?! 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 SAS Studio, however, things are a bit difficult. **There is no concrete way of telling SAS upfront which variable is nominal, ordinal, interval or ratio scale.** SAS does recognize the difference between a number and text, but that's about it: A number in SAS is still just a number. So, if you tell SAS to calculate the average of the values 1 (=jazz) and 3 (=hip hop) in our example above, SAS 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 let's start with some funky data analysis...



HOW OFTEN

Artist: Ben E. King

Descriptive statistics.

And how to conduct it on **SAS Studio**.



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 Microsoft Excel.

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?



HERE IS HOW
TO DO IT!

Frequency distribution (percentage)

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

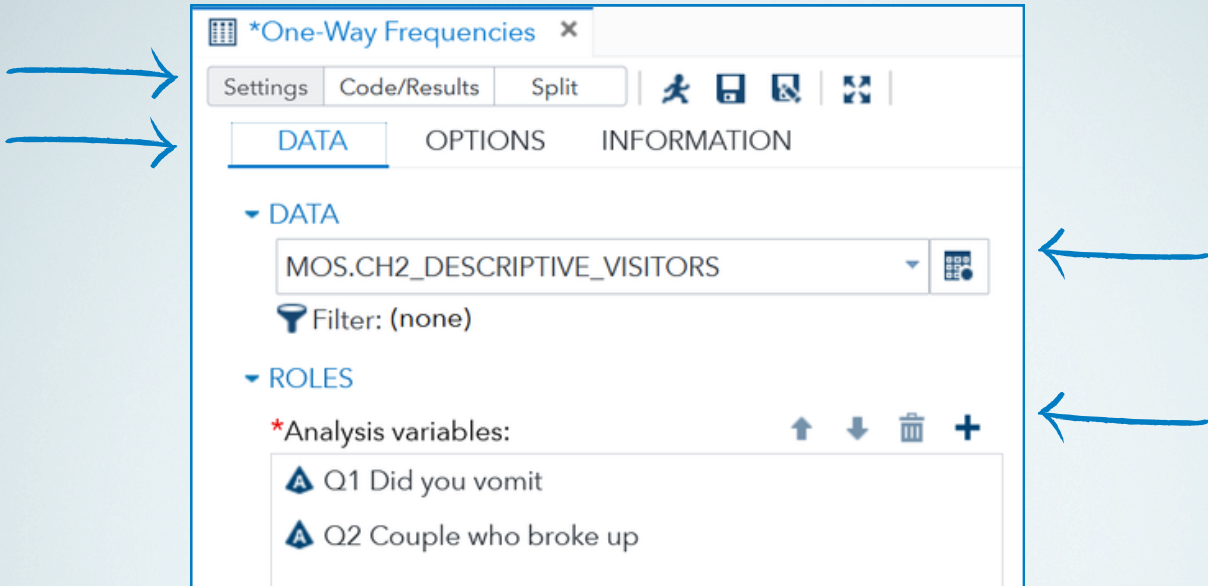
- Q1. Which percentage of visitors vomited during the festival?**
- Q2. Which percentage of couples broke up during the festival?**

And how will we answer them? by running frequencies (percentages).
And here is how to do it:



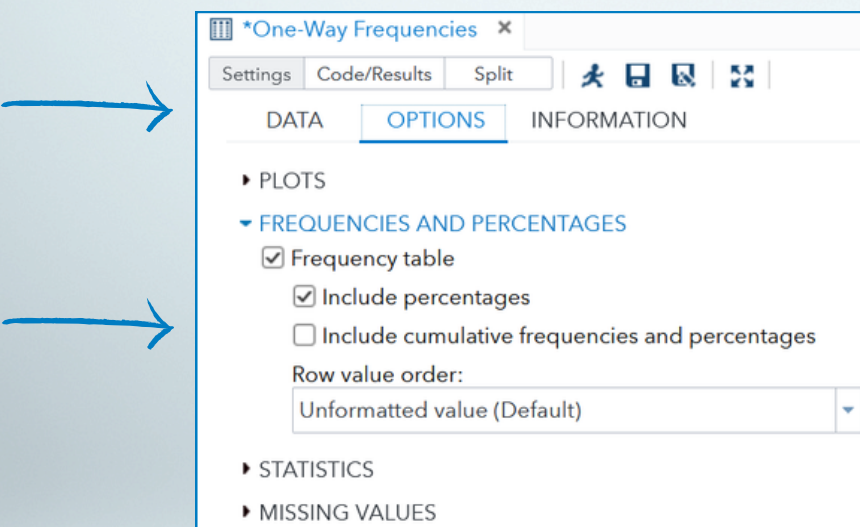
1

In the “Tasks and Utilities” menu of SAS, you must select the folder **“Tasks”**, then **“Statistics”**, then double-click on **“One-Way Frequencies”**. The window shown below will appear. In the **“Settings”** view, select the tab **“DATA”** and choose the data table that you want to use. At **“Analysis variables”**, click on the **+** sign to select the variables that you wish to analyze, then click on “OK”.



2

Next, select the **“OPTIONS”** tab. Here, you can specify details on the analysis you want to conduct. “Include percentages” is fine, but you have to unselect the option “Include cumulative frequencies and percentages” for our analysis.



“Cumulative frequencies”
 means that we add up all frequencies “up to the current value”. Now that you are an expert of data types, you know of course that this requires at least ordinal type data (but our data are nominal type only, so cumulative calculations do not make sense here).



THIS IS JUST THE START.

The full e-book walks you through every major statistical test you'll need for your thesis, step by step, with real data, screenshots, and zero formulas to memorize: *Descriptive Statistics, Chi-Square, Correlation, Independent Samples t-Test, Paired Samples t-Test, ANOVA, and Cronbach's Alpha.*

Each one explained through music analogies and plain English. No prior SAS experience needed.

€4,99. Less than a coffee.

The Melody of **STATISTICS.**



GET THE FULL E-BOOK

