# Descriptive data and inferential statistic

## Transcript

Video: <https://www.youtube.com/watch?v=5nTDFUPtICM>

Full resource: <https://www.ncrm.ac.uk/resources/online/all/?id=20846>

Chiara Dall’Ora: Hello, and welcome to this tutorial on reading tables in quantitative papers.

In this tutorial, I'm going to walk you through how to read tables reporting descriptive statistics and tables of significance from scientific papers. I assume that you already have prior knowledge of study design elements, for example, what a randomised controlled trial or a case control study is, and my focus will be only on reading and interpreting tables.

Try to imagine you're a researcher who has finally collected all the data they needed for their study. You will have plenty of sheets full of numbers. Your first job as a researcher is to find a meaningful way to sort and summarise all those numbers and to simply describe what they look like. You might wonder why do we need to summarise data? Why can't we just jump straight into the part of the results to evaluate whether the aim has been met or what the answer to the research question was? For example, in a randomised controlled trial, we might want to check whether a treatment was effective or in a cohort study we might want to know if there was an association between one predictor and the outcome variable. Nonetheless, the data have a story to tell, and while it's tempting to skip to the end of the book to find how the story ends, it often does not make much sense if we don't follow the whole plot without focusing on the descriptive statistics first, it is difficult to spot trends and patterns and to draw robust conclusions from the available data.

Let's make an example. You are reading a paper where the research question is, “What is the effect of a group dance exercise intervention on care workers’ intention to leave their jobs?” First, it's important to go and check what the sample looks like at baseline. It is worth looking at what do our participants look like according to their allocation group. This is often Table 1 in a quantitative paper.

So, consider the following table. A quick aside note, tables tend to be longer in a paper with more baseline characteristics reported, but for the sake of brevity I have only reported three variables. All three are useful because they summarise three different types of data.

Let's have a look at the first characteristic, age. What do you notice? The mean age of the intervention group is 29.7, which is lower than that of the control group, 36.8. Should we be worried about this difference in mean age? Not necessarily, but it's important to keep in mind when reading on about the effectiveness of the dance exercise intervention. Could the fact that participants in the intervention group tend to be younger influence, for example, their willingness to take up the intervention? You know, younger people might be more inclined to try out a group dance exercise package and sustain it over time, but our assumptions might be biased.

Between brackets you see the standard deviation. It is a measure of dispersion. It tells us about the average amount by which our data deviate from the mean. We can see that these values are similar.

We now move on to the next characteristic, gender. This is a categorical type of variable, one where the categories have been predetermined and tend to be fixed. It is useful to focus on the percentages of the distribution rather than the raw numbers, although these are important too. We can see that all cells in the column intervention group and control group for the variable gender when summed up equal 100%. This means that we can easily compare distributions between the two groups. What do we see here? There is a higher proportion of females in the intervention group compared to the control group, 63.1% versus 54.9%. It is important. You might think that females are more likely to be open to and to embrace a group dance exercise programme, but it's always important to check that your assumptions are based on evidence. Do a quick literature search to verify your assumptions.

[0:04:44]

Last, we have a variable that has been categorised. It was a numerical variable using profession, but researchers have decided to report it in categories. You can see that these categories are not equal. The first two span over three years, while the third one spans over five years and the last one does not have a maximum boundary. Researchers could have reported these categories as a mean and certain deviation, but they probably chose to report these descriptive statistics in categories, because these are somehow meaningful. For example, there might be research that shows that the first two years are crucial for retention of care workers, and if care workers stay for the first three years, they're likely to stay for longer. Five years might be another important turning point and after ten years, care workers’ intention to stay might change again.

As with previous variables, always ask yourself what the implication of summarising variables in different ways are. Is it more valuable to know if a variable is mean or median or mode? Or is it worthwhile to see how a variable is distributed in categories? When you note a difference in distribution, how do you know if the distribution might affect the quality of the research, for example, is it problematic if age is different between the intervention and control group, and if more females are in an intervention group than males? Always check your assumptions with the literature.

It's also important to remember that with randomised controlled trials, if the sample size calculation is appropriate and the allocation to intervention and control group is correct, then we should be reassured that any differences we see in the outcomes between groups are due to the intervention rather than to differences between group characteristics or other factors.

Another way of reducing the imbalance between groups is to use stratified sampling, which is a separate randomisation of subgroups of participants. For example, subgroups by age or gender.

We now take a step further. We look at tests of significance reported in quantitative papers. Let's think about our research question and let's imagine that researchers, before digging into testing whether the intervention was effective, start to worry about the differences in percentages they have noticed in age and gender between groups. They decide to run a series of tests to check if the differences between groups are statistically significant. In a nutshell, researchers are asking whether the differences between groups are big enough to represent the difference also in the population.

They will formulate a null hypothesis in this case, age is equal between intervention and control group in the population, and they will run a test to determine whether the null hypothesis can be rejected. In this case they will run as two sample t-test. And this is what they report. What does it mean? The t value itself does not say much on its own. There are tables you can consult to look at all the different t values and their corresponding level of statistical significance. But let's focus on the p value.

The p value is the frequency probability of the distribution meeting the null hypothesis. The null hypothesis in this case is that the mean edges of the two populations are equal. This becomes problematic actually because we do not have two populations. We have a sample drawn from a single population, so this t test is not particularly helpful. The p value is not 0.0001. This means that the frequency probability of meeting the null hypothesis is incredibly low, so we have to reject it. But we should not attach any meaning to these tests because the assumptions of this test, checking if two population means are equal, are not valid in a study drawing on a sample from a single population.

Let's look at gender now. Because it is a categorical variable, researchers performed a X2 test. In a nutshell, the X2 measures the discrepancy between the observed and the expected counts where the null hypothesis failed to be rejected. In this case, the null hypothesis is that gender distribution is equal between intervention and control group in the population. This is what researchers will report. Similarly to above, let's focus on the p value. It is 0.324, so it's higher than 0.05, which is standard threshold for statistical significance. This means that we cannot reject the null hypothesis.

[0:09:53]

I just wanted to remark that in a well conducted RCT where sampling and randomisation have occurred correctly, there should not be any concerns around demographic differences between intervention and control groups, and we know that the differences are solely due to chance. I use these examples to show how to interpret p values when reading tables that report significance testing.

My last remark is always treat these tests with caution and don't assume that a statistical significant difference means that there is an association or correlation or let alone an effect of a variable on an outcome. There are more sophisticated techniques for those, and we will explore them in our next tutorial.

National Centre for Research Methods (NCRM)  
Social Sciences  
Murray Building (Bldg 58)  
University of Southampton  
Southampton SO17 1BJ  
United Kingdom

**Web** www.ncrm.ac.uk   
**Email** info@ncrm.ac.uk  
**Tel** +44 23 8059 4539  
**Twitter** @NCRMUK