

The Statistics and Statistical Tools Used in Market Research in 2017

Ray Poynter, NewMR, September 2017

In August I ran a survey via NewMR that collected the views of 344 people involved in market research and insight, predominantly from participants in English-speaking countries. The survey looked at the sorts of statistics and statistical tools being used in 2017.

Summary

The top four things that I want to share about the use of statistics and statistical tools are:

1. Most statistical tests/approaches are not widely used. Only Correlation, Regression, z- or t-tests, and Cluster Analysis have been used by more than 50% of the participants in this research, during the first half of 2017 – and this sample probably over-represents people using statistics, and under-represents those using statistics less often.
2. SPSS is the dominant software package amongst people using statistical packages. Given SPSS is approaching 50 years old, that may not be the sign of a dynamic industry? But, there are many people using tools such as Q, Sawtooth Software, SAS – and beyond them programs such as Latent Gold, Tableau, and XLSTAT.
3. One of the growth areas is the use of tools is the use of integrated data collection / analysis solutions, for example Confront, Askia, Vision Critical, Qualtrics. The use of these tools requires the researcher to make fewer decisions. For example, survey monitoring flows into the analysis without any extra steps, the packages have a default way of looking of testing differences (for example t-tests) – making it less likely that the researcher will consider less convenient options, such as Chi-squared tests.
4. The most widely adopted complex solution is R, an open-source programming language that leverages large numbers of libraries for things like advanced analytics, data science, and data visualisation. People have been highlighting the growing role of R for a few years, and it seems to be gaining a stronger share of market research and insight analyses.

The Details

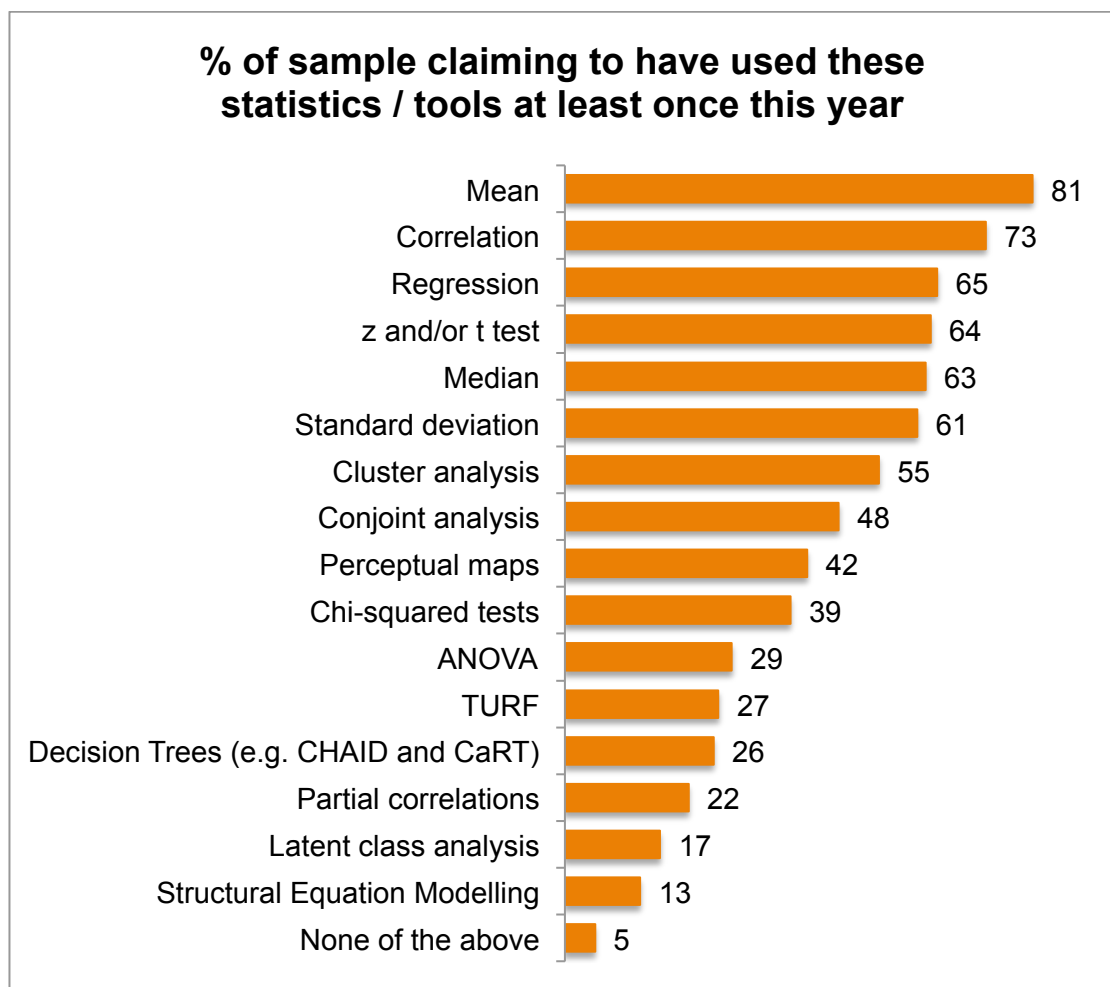
This short report looks at the sorts of statistical approaches and tools that appear to be in regular use in market research in mid-2017. The data for this report come from a survey conducted by NewMR between 16 and 31 August, 2017. The sample size was 344 individuals.

Caveat about the data

Note, this report is based on a survey created and distributed by me (Ray Poynter), it is not a representative sample, and the following biases are evident: it over-represents people who are based in English-speaking countries, it was conducted in English, and was distributed to networks (social and email) connected to me and NewMR. I believe the ordering of the data is probably reflective of the English-speaking research world, but I would have little confidence in the absolute values. Read the sample details in the Appendix to get a better sense of who took part in the survey.

A Small Group of Widely Used Statistics

Not surprisingly, the statistic that was most frequently selected was the Mean. The survey and responses included people who create statistics and those who use them.



If we take 50% as a reasonable cut value, then:

- The key statistics that people are using are the mean, median and standard deviation.

- The key techniques are correlation, regression and cluster analysis.
- The key tests are the t- or z-test.

From my own perspective, I was pleased to see Conjoint Analysis so high on the list, surprised to see Perceptual Maps so low in the list.

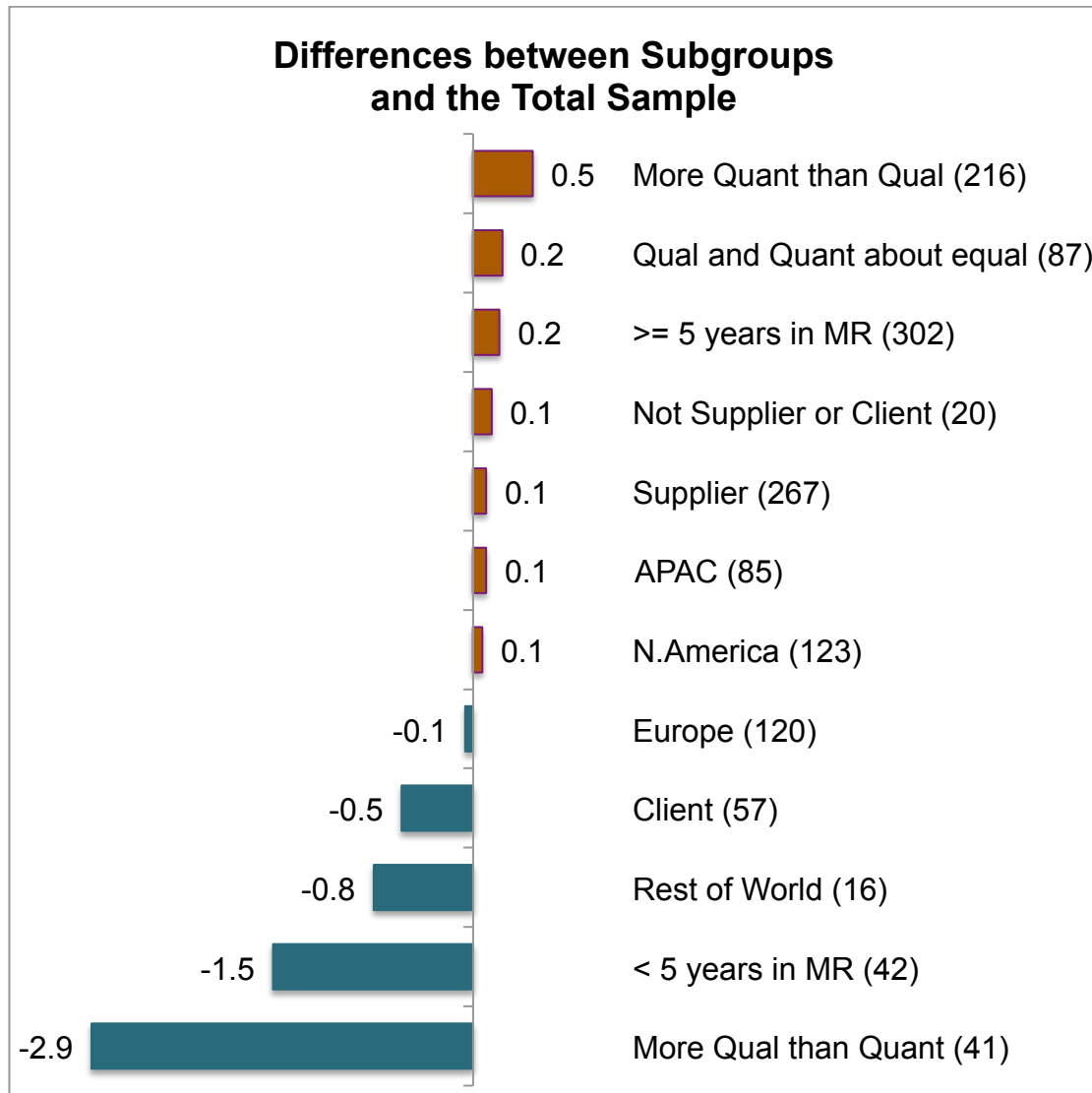
I was not surprised to see Chi-squared test and ANOVA so low on the list, the rise of automated testing via cross-tabulation reports (which tend to use t- or z-tests) have pushed the more manual tests to the periphery of commercial market research.

The t / z tests is a combined code for people who selected (in the survey) t-test, or z-test, or both. A fuller set of numbers is available in the appendix.

At the bottom of the chart are techniques such as TURF, Decisions Trees, Latent Class Analysis, and Structural Equation Modelling.

Number of Techniques Used By Different Subgroups

The chart above shows the big picture, i.e. the use of techniques and statistics by a broad cross section of researchers. The chart below looks at difference by subgroup.



Note, each row represents one of the subgroups in the data (based on client/supplier, region, how long in MR, and whether their work is Mostly Qual, Mostly Quant, or equally both). The numbers in brackets shows the sample base for that subgroup.

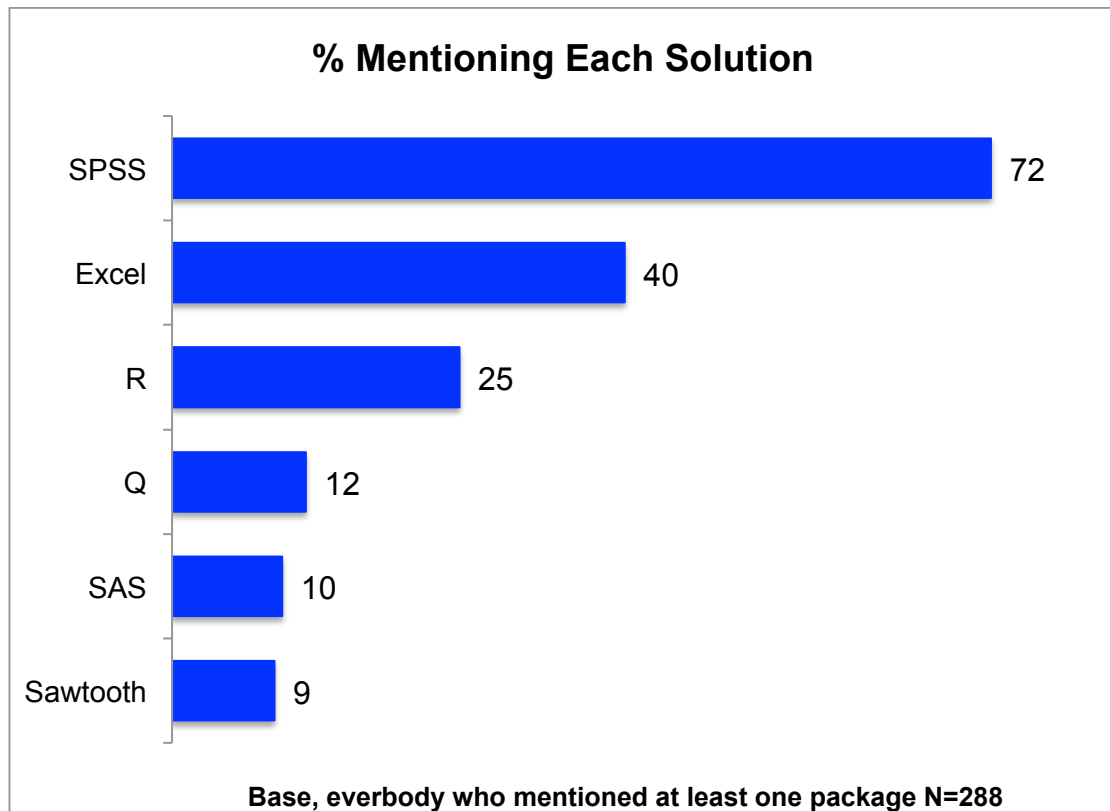
The chart above shows how subgroups differ from the overall picture. The average number of statistical tools used was 7.6 (out of a possible total of 16).

None of the subgroups had an average much higher than the total, with the people who described themselves as More Quant than Qual being 0.5 above the mean (i.e. 8.01 compared with 7.56).

There are two groups with a mean number of statistical tools that are clearly below the mean. The first group is, not surprisingly, people who do more Qual than Quant. The second group is perhaps more surprising and perhaps more worrying. People with fewer than five years in MR report a smaller average number of statistics and statistical tools used this year.

Software Packages use for Statistics

The survey participants listed the software packages they were using to find their statistics and conduct analyses, via an open-ended question (up to five responses per person). The chart below shows the coded results for the most popular options.



Amongst the people who provided at least one open-ended answer, the runaway number one option for statistical analyses was SPSS. This is interesting from several perspectives. First it speaks to the audience being relatively sophisticated, i.e. having access to an SPSS licence or to people who use SPSS. Secondly, SPSS is a very old piece of software, being first released in 1968 (as the Statistical Package for the Social Sciences), its dominance after almost 50 years perhaps indicates a lack of market innovation and competition (although SPSS has innovated over the years and continues to evolve).

Second in the list is Excel. I would assume that most of the people answering the survey use Excel for some tasks, but clearly many do not use it for advanced statistics (which is not surprising).

Third in the list is R, a programming language that is of growing interest to a variety of specialists, including statisticians, data scientists, and data visualisation experts. R is free, and there is a large and growing set of libraries for analytics, data manipulation, and charting. As well as R, there was an indication that the programming language Python is being adopted by cutting edge analysts – often as a complement to R.

The last three, in our list of the top six, represent three commercial packages with different characteristics. Q is a general-purpose analytical program that can be run as a graphical 'point and click' system or as a complex tool for advanced analysis.

Sawtooth Software is specialist software focusing on choice modelling and conjoint analysis. SAS is a complex and very powerful system, and before the rise in R it was the choice of many of the most advanced analysts (and still is for some).

In addition to these six, there were many, many other packages mentioned, such as Latent Gold, XLSTAT, Marketsight and Tableau. We have not listed all of them and we have not listed the counts because we don't want imply items not on the list are not in wide usage. Given the sample bias towards English-speaking countries some of the packages that scored less well in our survey might do much better, for example, in Germany, France or Japan.

In addition to specific packages, such as those shown in the chart, many participants referred to using the analysis options within integrated systems, such as Confront, Askia, FocusVision, Vision Critical and Qualtrics.

Quite a few participants referred to in-house systems and solutions.

Appendix

First of all I would like to thank everybody who took part in the survey and especially to those who helped promote the survey.

Here are some extra notes and details about the survey we conducted.

Suppliers and Clients

Like most surveys and indeed conferences and events, suppliers outnumbered clients.

Supplier / Client	Frequency	%
Provider of Market Research / Insight	267	78
User of Market Research / Insight	57	17
Other	20	6
Total	344	100

Years in Market Research / Insight

One of the worrying biases in the data is that most of the responses came from people with 5 or more years experience of market research / insights.

Years in/using MR and Insight	Frequency	Percent
Five or more years	302	88
Fewer than five years	42	12
Total	344	100

This bias is worrying because the data from those fewer than five years does appear to be different, they appear to be using fewer statistics/techniques. If this bias was unique to this survey, that would be worrying enough. However, most events and most of the research I see seems to under-represent people in the first part of their career. Are these people not being reached, or do they feel less engaged, or are their other factors inhibiting their involvement?

Are you More Qual or Quant?

Participants in the survey described themselves in terms of their use of quant and qual. Nearly 40% of the sample have a substantial amount of qual in their work life.

Are you more qual or quant?	Frequency	%
More quant than qual	216	63
Both about equal	87	25
More qual than quant	41	12
Total	344	100

The analysis showed that people who do more qual than quant tended to use fewer statistics and tools than others.

Countries

Self-declared (i.e. not from IP or other passive information)

Country	Frequency	%	Cumulative %
United States	101	29	29
United Kingdom	81	24	53
Australia	36	11	63
Canada	21	6	70
India	18	5	75
Netherlands	7	2	77
Singapore	7	2	79
France	4	1	80
Germany	4	1	81
Hong Kong	4	1	82
Ireland	4	1	83
Bangladesh	3	1	84
Brazil	3	1	85
Indonesia	3	1	86
Italy	3	1	87
Japan	3	1	88
New Zealand	3	1	89
South Africa	3	1	90
Spain	3	1	90
United Arab Emirates	3	1	91
Belgium	2	1	92
Egypt	2	1	92
Malaysia	2	1	93
Poland	2	1	94
Romania	2	1	94
Sweden	2	1	95
Vietnam	2	1	95
China	1	0	96
Cyprus	1	0	96
Estonia	1	0	96
Greece	1	0	97
Hungary	1	0	97
Lithuania	1	0	97
Mexico	1	0	97
Nigeria	1	0	98
Peru	1	0	98
Philippines	1	0	98
Saudi Arabia	1	0	99
Senegal	1	0	99
Slovenia	1	0	99
Sri Lanka	1	0	99
Timor-Leste	1	0	100
Trinidad and Tobago	1	0	100
Total	344	100	100

Note, 70% of the sample are from USA, UK, Australia and Canada; 53% of the sample are from just USA and UK. This means the survey should mostly be considered a study of English-speaking countries, with a dash of other countries. However, it is worth noting that in terms of market research revenues, ESOMAR estimate that USA, UK, Australia and Canada between them account for about 62% the global spend on MR – so while the 70% is high, it is not wildly wrong.