An Introduction to Market Research Tables

Ray Poynter, The Future Place

1. Statistical Tables versus Market Research Tables

The single most common analysis tool in market research is the use of cross-tabulation reports. However, these tables are quite different from the sorts of cross-tabulations that statistical and most market research textbooks talk about.

Classical cross-tabulations combine two (or occasionally three) categorical variables (for example, gender by employment category), and is typically combined with totals, percentages, and Chi-squared tests of association. For example, the table in Table 1 is a classical cross-tabulation, which is also called a contingency table.

Table 1	**Male**	**Female**	**Total**
Rented accommodation	22 (49%)	8 (19%)	30 (34%)
Owner occupier	18 (40%)	26 (62%)	44 (51%)
Other	5 (11%)	8 (19%)	13 (15%)
Total	45 (100%)	42 (100%)	87 (100%

Chi-square statistic = 8.587, p=0.014

However, market research cross-tabulation reports do not confine themselves to just two variables, nor to just categorical variables, nor do they tend to use chi-square as their main statistical test. Indeed, one reason that they are so rarely written about in textbooks is that they tend to break many of the statistical rules for good practice! Another reason is that once the discussion goes beyond a fairly basic level, the nature of the tables is quite strongly linked to the procedures created by the authors of the particular 'tab' package being used.

2. A Simple Table

Tables 2 and 3 show two versions of a highly simplified Market Research Cross-Tab.

Table 2
Q1 Do you want to learn about cross-tabs?
Base: All who were asked Q1

	Total	Men	Women	18- 29 years	30 - 49 years	50 and older
Base	200	105	95	60	70	70
Yes	160	70	90	50	60	50
	80%	67%	95%	83%	86%	71%
No	35	32	3	7	9	19
	18%	30%	3%	12%	13%	27%
No response	5	3	2	3	1	1
	3%	3%	2%	5%	1%	1%

Table 2 shows a simple table where the two variables (Gender and Age) are arranged as columns. Down the side, we have the answers to Question 1, "Do you want to learn about cross-tabs?". The columns, and the headings of the columns are known by a variety of names, including cross-breaks, banners, breaks, and column variables.

The rows are also referred to by a variety of names, including stubs, rows, downbreak, and answers.

Table 2 comprises the following elements:

- A table number.
- A title (often the question that identifies the rows).
- The base (at the top of the page), which people were used.
- The base (in the table), how many people are in that column.
- The integers are the counts, the number of people who are described by the column heading, and who match the row answer. For example, in Table 2 there are 70 men who said *Yes* to the question about learning about cross-tabs.
- The percentage values are the result of dividing the count, by the correct base. For example, the 7 18-29 year olds who said *No* to the question, represented 12% of all the 18-29 year olds.

3. Percentages

In market research tables the most common use of tables is to apply them to columns, as in Table 2. However, percentages can be applied to rows, or to an entire table.

Because percentages are rounded, often of no decimal places, the visual total of a column does not always equal 100%. This is well illustrated in Table 2.

In Table 2 the three percentages in the Total column are shown as 80%, 18%, and 3%, which would appear to give a total of 101%. However, if the numbers are shown with one decimal place they are 80%, 17.5%, and 2.5%, which sums to 100%.

In Table 2 the three percentages in the 50 and older column appear as 71%, 27%, and 1%, which would appear to total to 99%. However, looking at these numbers to two decimal place shows them to be 71.43%, 27.14%, and 1.43%, which adds to 100%.

The main consequence of this rounding effect is that researchers should not simply add together percentages from the tables to create combined totals. Researchers should either ask the spec writer to create nets in the tables (see section later) or should calculate their new percentage using the counts and the base. For example to find out what percentage of the 50 and over sample said *Yes* or No, the researcher should add 50 to 19 to get 69, and then divide it by the base, i.e. by 70, to get 99% (adding 71% to 27% would have given the wrong total, i.e. 98%).

4. Bases

Researchers need to pay close attention to the base for the tables. There are various ways that the base can be altered, for example filters and weights, which are covered later. Table 3 shows a case where the researcher has chosen a different base from Table 2.

Table 3
Q1 Do you want to learn about cross-tabs?
Base: All who were answered Q1

	Total	Men	Women	18- 29 years	30 - 49 years
Base	195	102	93	57	69
Yes	160	70	90	50	60
	82%	69%	97%	88%	87%
No	35	32	3	7	9
	18%	31%	3%	12%	13%
No response	5	3	2	3	1

Table 3 shows the same data, but shows how the preferences of one researcher, compared with another, can produce different tables. In Table 2 the base was all of the people who had a chance to answer the question, in Table 3 the base is those who entered an answer for Q1. In this example the differences in the percentages are small, but in a study with a large amount of missing data the differences can be large.

Another difference, between Table 3 and Table 2, is that the age group 50 and over has been omitted. There are no rules that say each group of column variables has to include all of the data.

5. Significance Testing

Most market research tables packages have a facility to show significance testing. One way of doing this is shown in Table 4.

Table 4
Q1 Do you want to learn about cross-tabs?
Base: All who were answered Q1

	Total	Men (A)	Women (B)	18- 29 years (C)	30 - 49 years (D)
Base	195	102	93	57	69
Yes	160 82%	70 69% B	90 97% A	50 88%	60 87%
No	35 18%	32 31% B	3 3% A	7 12%	9 13%
No response	5	3	2	3	1

Lower case, sig=95%, Caps sig=95%

Software packages differ in whether they conduct their testing based on a z-test or a t-test. Most research organisations or departments will have a view about whether they prefer to use z-tests or t-tests. Some software packages are capable of using both t-tests and z-tests, but some packages will only support one of these.

Statisticians tend to say that if sample sizes are greater than about 120 there is no practical difference between the z-test and the t-test. Similarly, most statisticians would recommend the t-test for any sample size less than about 30.

Different software packages vary in how they display significance, they may use numbers, or * and ** symbols, or, as in Table 4, upper and lower case letters. In Table 4 each break (i.e. each column) has a letter below the heading. In the body of a table, a letter indicates that the cell is significantly different from the value in the same row that belongs to the column indicated. For example, the 70 Men who said *Yes*, are marked with an upper case B, this means the 70 is statistically significantly different from the *Yes* value in the B column (i.e. from the 90 Women who said *Yes*). If a difference was significantly different, at the 95% level but not at the 99% level, it would be shown as a lower case 'b', by this particular software package.

Researchers should note that standard significance testing should only be applied to non-overlapping samples. In Table 4 the researcher can test Men versus Women (columns A versus B), and they could test 18-29 versus 30-49 (columns C versus D). However, it would not be appropriate

to test Women versus 18-29 (B versus C) because some of the Women are likely to be aged 18-29 and some of the 18-29 years olds are likely to be Women.

6. Interval Scales

Market researchers often show interval scales (or a scales they are willing to assume are interval scales) as rows in a table, such as the one in Table 5.

Table 5
Q5 Agreement with statement "Tables are really cool"
Base: All respondents

	Total	Men	Women
Base	400	200	200
Agree Strongly (+2)	150	20	130
	38%	10%	65%
Agree (+1)	40	30	10
	10%	15%	5%
Neither (0)	50	40	10
	13%	20%	5%
Disagree (-1)	100	80	20
	25%	40%	10%
Disagree Strongly (-2)	20	20	-
	5%	10%	
Don't Know (-)	40	10	30
	10%	5%	15%
Mean	0.56	-0.26	1.47
Std Dev	1.405	1.166	1.039
Std Err	0.074	0.085	0.080

In Table 5 the spec writer has assigned numerical values ranging from +2 to -2 to the agree/disagree scale, and has said that Don't Knows will be treated as missing data. Another researcher may have assigned values ranging from 5 to 1 for this scale. Other researchers might feel that this agree/disagree scale was an ordinal scale, and should not be treated as an interval scale.

The table shows the mean, standard deviation, and the standard error. This sort of table can be configured to show statistical differences in both the counts and the mean.

7. Other Options

Most research agencies, organisations, and departments have a range of options that they will use with tables. These are highlighted below, but the student is advised to find out what is available in their organisation.

7.1. Sorting Rows

Sorting is often used with a scale such as "Which of the following brands have you heard of?". The rows could be arranged in questionnaire order, but the researcher can request that the rows be ordered such that the brand which is the most familiar will be at the top of the list, and the least familiar at the bottom (usually followed by Other and Don't Know).

7.2. Nets

In tables such as Table 5, the researcher may ask for a total showing all those who agree or agree strongly. This total is referred to as a 'net', as in 'net agree'.

7.3. Summary Tables

When a researcher has a number of agree/disagree scales, such as the one shown in Table 5, the information is spread over many pages. The researcher can request a summary table which shows, for each of these scales, just the mean scores, allowing all the scales to be shown on a single page.

7.4. Nested Breaks

Table 6 shows a banner where age has been nested with gender.

Table 6
Q1 Do you want to learn about cross-tabs?
Base: All who were asked Q1

		Men			Women		
	Total	18- 29 years	30 - 49 years	50 and older	18- 29 years	30 - 49 years	50 and older
Base	200	30	35	40	30	35	30

Researchers should note that nesting variables reduces the sample size for each column dramatically.

7.5. Filters

A filter can be used to produce table based on a sub-set of respondents. For example the researcher can specify that a table be based on, for example, users, or non-users, young or old.

7.6. Weights

The researcher can request that the respondents who comprise the data can be weighted so that they better represent some target population. In Table 6, the data could have been weighted so that each age/gender combination represented 33 people. This would be achieved by assigning a weight of less than one to Men aged 30-49 or 50 and over, a similarly down-weighting women aged 30-49. The weighting would up-weight (i.e. give them a value greater than one) Men aged 18-29 and Women aged 18-20 years and over 50 years. An example of this weight in shown in Table 7.

Table 7
Q1 Do you want to learn about cross-tabs?
Base: All who were asked Q1

		Men			Women		
	Total	18- 29 years	30 - 49 years	50 and older	18- 29 years	30 - 49 years	50 and older
Unweighted Total	200	30	35	40	30	35	30
Weighted Total	200	33	33	34	33	33	34
Yes	100	17	20	14	15	18	16
	50%	52%	61%	41%	45%	55%	47%
No	100	16	13	20	18	15	18
	50%	48%	39%	59%	55%	45%	53%

In Table 7 the cell values are based on the weighted data. The heading will often show both the unweighted and weighted totals, to help the researcher appreciate the degree of difference the weighting has made.

7.6.1. Issues with Weighting

There are a number of issues and concerns about weighting that the researcher should keep in mind.

Some clients are nervous about weighting and will reject its use. The researcher should, ideally, be aware of this before designing the data collection, so that the sample is collected so that it matches the target population without weighting.

Researchers should note that weighting the data reduces the effective sample size and makes it harder to calculate whether a difference is statistically significant.

One concern with weighting is the range of weights that are applied to respondents. Most researchers expect that, in most cases, the average weight that is applied will be 1, which means the total sample size is not increased or decreased by the weighting process.

One key statistic is the ratio of the largest respondent weight to the smallest. Some, conservative, research organisations will tend the limit this ratio to 2:1 (which would tend to mean the largest weight was 1.333 and the smallest was 1.667). A more common limit is a ration of 10:1 (implying a

range of something like 0.5 to 5 or 0.2 to 2). Some researchers take a more relaxed view and rations as high as 100:1 do occur.

7.6.2. How Weighting is Performed

The researcher will not often be involved in the details of how the weighting is performed. The researcher's role is normally limited to specifying the characteristics that should be matched to a target population, and the values for these characteristics. For example, a researcher might supply the spec writer with a table such as Table 8.

Table 8	**Sample**	**Target**
Younger Male	31%	25%
Older Male	22%	25%
Younger Female	26%	25%
Older Female	21%	25%

In table 8 each respondent will fall into only one row, and this makes the weighting process straightforward. The cells in Table 8, such as Younger Males, are known as interlocking cells.

However, the researcher can only specify interlocking cell target percentages if they are known. In many cases the target interlocking cells are not known and the researcher can only specify the targets as in Table 9.

Table 9	**Sample**	**Target**
Male	48%	50%
Female	52%	50%
Cappuccino Drinkers	50%	33%
Cappuccino Non-Drinkers	50%	67%

In these sorts of cases, the spec writer will tend to use a technique called Rim Weighting. Rim Weighting uses an iterative approach to modify the respondent weights to bring the sample percentages into alignment with the target percentages.

It should be noted that the sample profile in Table 9 is not unusual. If a research project is looking at Cappuccino Drinkers it is often a good idea to over sample them, so they can be examined with greater reliability. The weighting is only relevant to examination of the data at the total level. Examination of the data amongst Drinkers will be largely unaffected by the weighting.

8. Producing Tables

In most organisations the Table Report is not produced by market researchers, it is more normally produced by either a DP department or via a third-party provider. However, the researcher has to specify the table they want. The researcher will need to specify the breaks, the questions they want to form tables from, bases, treatment of non-responses, whether to use significance testing, filters, weighting, summary tables, and the use of derived information such as nets and means.

Most organisations will have an existing process for specifying tables, often including forms or a computerised system.

The people who produce the tables are often called spec writers.

9. Dynamic Tables

Since the 1990s, there has been a growing trend away from producing large books of printed tables towards more dynamic, electronic solutions. Broadly these dynamic solutions fall into two categories: electronic format tables, and simplified, on demand tables.

The term primary analysis is sometimes used to describe the creation of traditional tables, in these cases the use of on demand table is referred to as secondary analysis.

9.1. Electronic Format Tables

These tables are specified by the researcher in the traditional way, and are run in the traditional way. However, instead of being printed out these tables are used to create an electronic, searchable file or document. A leading example of this approach is eTabs. The user does not have a paper printout, they have a piece of software that allows them to search for information, to pull up the crosstab they want (provided it was run in the first place), view it, chart it, and paste it into another package, such as PowerPoint.

9.2. Simplified on Demand Tables

These systems can sometimes, initially, look like the electronic format tables, but they are based on the data, not on pre-run tables. Users can specify any cross-tab and run it dynamically. Examples of this type of product include Quanvert from SPSS and Reflect from QPSMR. The researcher should note that the more powerful the software, the steeper the learning curve. The most powerful tools are almost as powerful as the tools that the professional script-writers use, but they are almost as difficult to master.

10. Software

This document cannot list all the software available, if you are aware of a software package that ought to be listed, email ray.poynter@thefutureplace.com, and it will hopefully be added to a future update.

10.1. Stand Alone Packages

These packages have as their main function the creation of market research tables, however, they may have some other functions as well.

Merlin	Merlin is a tabs package produced by Merlinco. http://www.merlinco.co.uk/
QPSMR	QPSMAR is tabs package produced by QPSMR. Options include a script based system, a menu based system, and an end user system (called Reflect). http://www.qpsmr.ltd.uk/

10.2. Integrated with larger systems

Some data collection systems have and integrated tables package, similarly some statistical packages have a market research tables option.

Confirmit: Confirmit is a leading CAWI provider with a range of data processing options. http://www.confirmit.com/

10.3. Online directories

Tim Macer of Meaning [http://www.meaning.uk.com/] maintains a large online databases of market research software and reviews which includes crosstabulation packages.
http://www.meaning.uk.com/rscentral/rscentral.html

11. Acknowledgements

Many thanks to the people who have helped me produce this note, including:

Matt Raistrick, Virtual Surveys, for reviewing some of my text and suggesting improvements.

Peter Jackling, Merlinco, whose chapter on Analyzing Data in "A Handbook of Market Research Techniques", Kogan Page, 1990, was one of the very few existing references to market research tables I was able to find.

Richard Collins, Confirmit, for reviewing some of my text, suggesting improvements, and supplying reference material.

Sue York, The Future Place, for reviewing all the material and suggesting improvements.