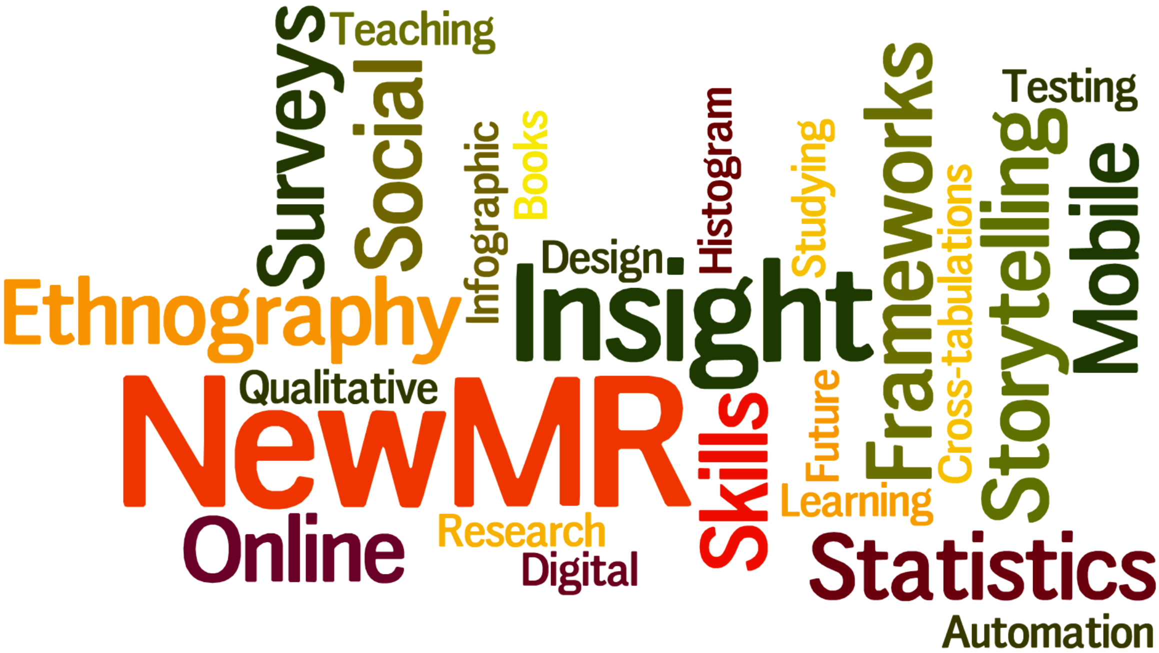
An Introduction to Market Research Tables

Ray Poynter, NewMR & The Future Place – updated 2016

Beta Consultative Release



Note, this version of the document has been released as part of the drafting/creation process. You are welcome to use this material but please quote the source (Poynter, Tables, 2016), and watch out for errors. If you would like to suggest corrections and raise queries, please email me via ray.poynter@thefutureplace.com

Contents

1. Statistical Tables versus Market Research Tables

Cross-tabulation tables and reports are the most common analysis tools used in market research. However, these tables are quite different from the sorts of cross-tabulations that statistical (and most market research) textbooks talk about.

Classical (statistical) cross-tabulations display two (or occasionally three) categorical variables (for example, gender by employment category). The data tends to be shown in combination with totals, percentages, and Chi-squared tests of association. For example, Table 1 shows a typical cross-tabulation. This type of table 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. This eBook has been created to look at these alternative tables.

One reason that market research cross-tabulation reports 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 a specific set of tables is quite strongly linked to the specifics of the particular software package or platform being used. In this eBook we show you all of the key principles that you need to adopt when working with tables, regardless of the software being used.

2. A Simple Table

Tables 2 shows a simplified page from a market research cross-tabulation report.

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 two variables (Gender and Age) arranged across the page, as columns. Down the side, are the answers to Question 1, "Do you want to learn about cross-tabs?" – note the wording of the question in the tables is often a shorter version of the question asked in the research. 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, downbreaks, and answers.

Table 2 comprises the following elements:

- A table number.
- A title (often the question that identifies the rows).
- The base (a label at the top of the page in this example) stating which people were are being reported.
- The base (numbers in the table), how many people are being reported in each 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 answered *Yes* to the question.
- The percentage values are the result of dividing the count, by the correct base. For example, in the 18-29 year old column, the 7 people who said *No* to the question, represent 12% of all the 18-29 year olds reported on this page (the value 7, divided by the base of 60).

3. Percentages

In market research tables the most common use of percentages 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, in most cases, rounded, often with no decimal places, the column does not always appear to sum to 100%, even when common sense suggests it should. 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 99%. However, looking at these numbers to two decimal places 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 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 this by the base, i.e. by 70, showing the result to be 99% (adding 71% to 27% would have given the wrong total, i.e. 98%).

3.1. Columns do not always sum to 100%

In the example in Table 2, the question represents a single response question. Each participant said Yes, or they said No, or they did not answer the question, so the total will be 100%.

However, if a multi-response question is being reported, then there is no need for the column to necessarily sum to 100%, as is illustrated in Table 3.

Table 3
Q5 Which consumed in last 24 hours?
Base: All who were asked Q5

	Total	Men	Women	18- 29 years	30 - 49 years	50 and older
Base	200	105	95	60	70	70
Coffee	110	60	50	30	40	40
	55%	57%	53%	50%	57%	57%
Tea	80	50	30	25	25	30
	40%	48%	32%	42%	36%	43%
None of these	50	20	30	15	20	15
	25%	19%	32%	25%	29%	21%

Table 3 uses the same banner as Table 2 (i.e. Total, Gender, and Age as the columns headings, and the responses to a question about hot beverages drunk in the last 24 hours.

The Total column shows that 55% said that they had drunk coffee, 40% said they had drunk tea, and 25% said they had drunk neither coffee nor tea – which adds to 120%. This is because 70 people said they drink coffee (but not tea), 40 said they drink tea (but not coffee), and 40 said they drink coffee and tea.

Table 3 is a fairly basic table and there are a number of ways that the table can be modified, depending on what the aims and needs of the researcher are. For example:

1. Recoding the data so that the rows were:
    a. Coffee Only
    b. Tea Only
    c. Coffee and Tea
    d. Neither
2. Adding a row that showed 'Coffee and Tea' to the existing rows.
3. Adding an 'Average number of drinks mentioned' row. In this case we would have noted that 50 people said 0 drinks and 0 said 2 drinks (which means that the remaining 110 participants drank 1 drink). The average number of drinks mentioned would be 0.95, i.e. (50 x 0 + 40 x 2 + 110 x 1) ÷ 200, or (0 + 80 + 110) ÷ 200.
4. In some cases it might make sense to express the percentage values as a percentage of all mentions. This is often done when looking at tables created from a coding exercise and the meaning of the numbers is taken to be 'as a percentage of all mentions' – as distinct from what percentage of the sample mentioned each term reported.

4. Bases

Researchers need to pay close attention to the table base. There are various ways that the base can be adjusted, for example filters and weights. Table 4 shows a case where the researcher has chosen a different base from Table 2, even though the same question and data are being analysed.

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

	Total	Men	Women	18- 29 years	30 - 49 years	50 and older
Base	195	102	93	57	69	69
Yes	160	70	90	50	60	50
	82%	69%	97%	88%	87%	72%
No	35	32	3	7	9	19
	18%	31%	3%	12%	13%	28%
No response	5	3	2	3	1	1

Table 3 shows the same data as Table 2, 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 a valid answer for Q1. The people who did not respond to the question are no longer part of the base, and the values for *Yes* and *No* now add to 100% for each column.

In this example the differences in the percentages are small, but in a study with a large amount of missing data the differences could be large.

5. Significance Testing

Market researchers tend to use significance testing to identify those differences in the data that are big enough to deserve further consideration. The underlying theory (which is widely disputed) is that significance testing highlights differences that are large enough that they are unlikely to have been caused by random sampling error.

Most market research packages that produce cross-tabulations have some ability to show significance testing. There should be no need for researchers to conduct significance testing using manual calculations.

One of the many ways that tables highlight the results of significance testing is illustrated in Table 5.

Table 5
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	70	90	50	60
	82%	69%	97%	88%	87%
		B	A		
No	35	32	3	7	9
	18%	31%	3%	12%	13%
		B	A		
No response	5	3	2	3	1

Lower case, sig=95%, Caps sig=99%

Different software packages vary in how they display significance, they may use numbers, or * and ** symbols, or (as in Table 5) upper and lower case letters. In Table 5 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 by the letter. 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.

Software packages differ in whether they default their testing is 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, other packages only support one of these. If the cell sizes are over 50 then the t-test and z-tests give very similar results, if the cells sizes are smaller than 50 then the researcher should consider whether they should even be looking at the data, never mind significance testing it.

Note, it is unusual for a market research cross-tabulation package to use chi-squared tests. In academia and in the social sciences, cross-tabulations are usually looking at categorical variables, which means the correct test is the chi-squared test. Market researchers are often looking at a mixture of categorical, ordinal, and metric data. Market researchers tend to take a leap of faith that they can treat all of their data as proportions or means, and test them with the z-test or t-test.

Researchers should note that standard significance testing should only be applied to non-overlapping samples. In Table 5 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 (i.e. columns B and 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. There are tests that deal with overlapping samples, but researchers are advised not to go there, given the concerns about using even simple statistical significance testing.

5.1. What is Significance Testing?

Whenever a sample is taken from a population there is a chance that the sample is not representative. Significance testing looks at differences in the data and indicates whether they are so large that there is only a very small chance they were caused by random sampling error.

When a researcher says that the difference is significant at the 95% level, they are saying that they think there is a 95% chance that the difference is not caused by sampling error, and a 5% chance that it is.

Another way that significance is used is to put a 'plus or minus error' around an estimate. For example if a pollster says that 50% of people intend to vote for Jane Doe in the upcoming election +/-3% at the 95% confidence level, they are saying that sampling error could mean that 95% of the time the real number was between 47% and 53% - and 5% of the time it would be outside that range, because of sampling error.

However, significance testing does not check for every type of error. If a large randomly selected sample were researched, but they were the wrong people, who were asked the wrong questions, in the wrong ways – the errors in the results could be enormous, even though the statistical significance seemed fine.

Market research data rarely conforms to the conditions required for statistical significance testing. The samples are not random probability samples, the response rate is too low, the hypotheses were not established before viewing the data, and the data are often of the wrong type for either the t-test of z-test.

5.2. Should Significance Testing be Used With Market Research Cross-Tabulations?

Many researchers have made the point that since significance testing with market research tables does not really do what the novice might assume it does, researchers should not conduct statistical significance testing.

Against this concern there are two reasons to continue to conduct significance testing, albeit with caveats:

1. Many users of the data (which can include the person paying for the study) request them, even after being told about their shortcomings.

2. Whilst statistical significance does not ensure the data are meaningful, the absence of a difference big enough to be statistically significant suggests the data should not be relied upon, since:

    a. There is no big difference.

    b. Or, the sample was too small to show the difference.

    c. Or, the research tool used was not sensitive enough to show the difference.

6. Interval Scales

Market researchers often show interval scales (or scales they are willing to assume are interval scales) as rows in a table, such as the one in Table 6.

Table 6
Q7 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

**Standard Deviation, can also be shown as SD*

In Table 6 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 created by this +2 to -2 coding. This sort of table can be configured to show statistical differences in both the counts and the means.

The use of cross-tabulations to create, report and investigate the differences between means is widespread in market research. The preferred process in academic circles, when compare differences between the means of different groups, is ANOVA (Analysis of Variance); but ANOVA is very rare in market research.

7. Other Options

Most research agencies, organisations, and departments have a range of options that they will use when working 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 nominal variables, 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 etc).

For example, in Table 7 the data shows responses for 'drink most often', in questionnaire order, showing just percentage values.

Table 7
Q8 Which of these do you drink most often?
Base: All who were answered Q8

	Total	Men	Women	18- 29 years	30 - 49 years	50 and older
Base	200	105	95	60	70	70
Coke	21%	23%	19%	25%	22%	17%
Dr Pepper	5%	6%	4%	6%	5%	4%
Fanta	11%	9%	13%	14%	11%	8%
Pepsi	20%	21%	19%	22%	21%	17%
Sprite	8%	7%	9%	9%	8%	7%
Other	18%	21%	15%	22%	18%	15%
None of these	17%	13%	21%	8%	16%	26%

By contrast, Table 7b shows the same data as Table 7, but the responses are sorted in descending order, in terms of the Total Column, leaving Other and None of these in questionnaire order.

Table 7b
Q8 Which of these do you drink most often?
Base: All who were answered Q8

	Total	Men	Women	18- 29 years	30 - 49 years	50 and older
Base	200	105	95	60	70	70
Coke	21%	23%	19%	25%	22%	17%
Pepsi	20%	21%	19%	22%	21%	17%
Fanta	11%	9%	13%	14%	11%	8%
Sprite	8%	7%	9%	9%	8%	7%
Dr Pepper	5%	6%	4%	6%	5%	4%
Other	18%	21%	15%	22%	18%	15%
None of these	17%	13%	21%	8%	16%	26%

7.2. Nets

In tables such as Table 8, 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'.

Table 8
Q7 Agreement with statement "Tables are really cool"
Base: All respondents

	Total	Men	Women
Base	400	200	200
Agree Strongly (+2)	150	20	140
	38%	10%	70%
Agree (+1)	60	35	25
	15%	18%	13%
Neither (0)	50	40	15
	13%	20%	8%
Disagree (-1)	100	80	15
	25%	40%	8%
Disagree Strongly (-2)	40	25	5
	10%	13%	3%
Net Agree	210	55	165
	53%	28%	83%
Net Disagree	140	105	20
	35%	53%	10%

Nets, such as Net Agree in Table 8 are often referred to as Top Boxes, and Net Disagree as Bottom Boxes. Many researchers find that reporting the percentages for Top Two Boxes and Bottom Two Boxes is more helpful than the (methodologically less sound) practice of assigning values and calculating means.

7.3. Summary Tables

When a researcher has a number of agree/disagree scales, such as the one shown in Table 8, the information is spread over many pages. The tables can (in most cases) also be shown in a summary table. A summary table shows, for each of these scales, a summary statistic such as mean scores or top box percentages, allowing all the scales to be reviewed on a single page, as in Table 9.

Table 9 Q7-11* Net Agree % Base: All respondents	Total	Men	Women
Base	*400*	*200*	*200*
Tables are really cool	53%	28%	83%
I use tables regularly	68%	74%	62%
The future is interactive	44%	52%	36%
Tables are old fashioned	54%	44%	64%
Visualisation is the next step	62%	68%	56%

**Q7-11, all of the questions from 7 to 11*

7.4. Nested Breaks

Table 10 shows a banner where age has been nested under gender.

Table 10 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

Nesting, for example showing gender split into three age for both men and women, allows the researcher to focus on specific interactions and sub-groups. Researchers should note that nesting variables reduces the sample size for each column dramatically.

7.5. Filters

A filter can be used to produce tables based on a sub-set of respondents. For example the researcher can specify that a table be based on: users, non-users, gender, age etc.

7.6. Creating New Variables

Researchers sometimes find it useful to create additional variables, which are then reported in the tables.

A simple example of an additional variable would be the creation of nets, such as Net Agree, as mentioned in the Nets section earlier. As well as showing nets as rows, they can also be used as banner variable. For example, one column each for Net Agree and Net Disagree.

A more complex example could be a case where a segmentation study has been conducted on the data. Each participant in the research would have been assigned a cluster/segment number in the segmentation and this can be added to the data set and used in the cross-tabulation tables – typically as a banner variable.

Table 11 Q5 Which consumed in last 24 hours? Base: All who were asked Q5	Total	Men	Women	Segment 1	Segment 2	Segment 3
Base	200	105	95	60	70	70
Coffee	110	60	50	30	40	40
	55%	57%	53%	50%	57%	57%
Tea	80	50	30	25	25	30
	40%	48%	32%	42%	36%	43%
None of these	50	20	30	15	20	15
	25%	19%	32%	25%	29%	21%

7.7. Weights

The researcher can request that the participants who comprise the data can be weighted so that they better represent some target population.

In Table 12 the unweighted total shows the data as it was collected. However, the researcher may know that (in the relevant population) the number of men and women in each of the three age groups should have been 33, 33, and 34. This can be achieved by using weighting.

Table 12 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%

Looking at the data for men, we can see that the initial total for men aged 18-29 years is 30. In order for this to be 33, each male in this age group needs to have a weight slightly greater than 1. In the men aged 30-49 years group, the initial total is 35 and the target is 33, so each male in this age group would have a weight slightly below 1.

The average weight that is applied, in most cases, will be 1, which means the total sample size is not increased or decreased by the weighting process (but the effective sample size is always reduced by weighting).

7.7.1. Issues with Weighting

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

Some clients are nervous about weighting and will reject its use. Researchers 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 key concern with weighting is the range of weights that are applied to research participants.

One key statistic is the ratio of the largest participant weight to the smallest. Some, conservative, research organisations will tend the limit this ratio to 2:1 (which would mean the largest weight was something like 1.3 and the smallest was 0.7). A more common limit is a ratio 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 ratios as high as 100:1 do occur.

7.7.2. How Weighting is Performed

In many cases the researcher will not 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 13.

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

Table 13

In Table 13, each participant will fall into only one row, and this makes the weighting process straightforward. The cells in Table 13, such as Younger Males, are known as interlocking cells.

However, the researcher can only specify interlocking cell targets if they are known. In many cases the values for the interlocking cells are not known and the researcher can only specify the targets in a non-interlocking way, as in Table 14.

Table 14	**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 an iterative method to create the weights. Iterative weighting seeks to repeatedly make small adjustments to weights until the sample percentages are in alignment with the target percentages.

8. Left-to-Right Tables

In most cases, market research tables are created using a banner as the heading for the tables, with one or more pages per question, with the primary focus being to read down the page, comparing columns. This focus on reading down means that the base (the number of people) refers to the people in that column, and the percentage values refer to the people in that column. Significance testing, when used, refers to differences between columns (usually with an assumption that no individual is in both columns – i.e. the samples are independent). All of the examples shown so far have followed this pattern.

However, this is not always the case. In Japan, for example, tables are often produced that run across the page, with the banners on the left, and the percentages summing along the rows. Table 15 shows an example of a table laid out as a left to right table.

Table 15
Q6 Want to study tables?

Base: all answering Q5	**Base**	**Yes**		**No**		**Not Sure**		**Total**
Total	200	160	80%	35	18%	5	3%	100%
Men	105	70	67%	32	30%	3	3%	100%
Women	95	90	95%	3	3%	2	2%	100%
18- 29 years	60	50	83%	7	12%	3	5%	100%
30 - 49 years	70	60	86%	9	13%	1	1%	100%
50 and older	70	50	71%	19	27%	1	1%	100%

9. Producing Tables

In many organisations the cross-tabulation reports have tended not to be produced by either a DP department (data processing) or via a third-party provider, rather than by market researchers themselves. However, even when the tables are produced by somebody else, the researcher (usually) has to specify the tables they want. The researcher will need to specify the breaks, the questions they want to create 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.

However, many platforms are now more DIY (Do It Yourself) and many researchers are preparing their own tables. In some cases, researchers will prepare straightforward tables. From the insight generated, the researcher can request more complex tables, to be created by a spec writer. This shift from spec writers producing tables, to DIY tables, is carried even further by the rise in interactive or dynamic tables.

10. 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 tables, and on-demand/interactive tables.

10.1. Electronic Tables

These tables are specified by the researcher in the traditional way, and are run in the traditional way. However, instead of being printed, these tables are used to create an electronic, searchable file or document. The user does not have a paper printout, instead they have a piece of software that allows them to search for information, to view 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.

A version of these electronic tables can be produced by exporting traditional tables into Excel, sometimes with a variety of hyperlinks and dynamic options included.

10.2. On-Demand/Interactive Tables

On-demand or interactive tables are based on the data, not on pre-run tables. Users can specify any cross-tab and run it dynamically. Researchers should note that the more powerful the software, the steeper the learning curve tends to be. 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.

Some researchers argue that using interactive tools is a better way to analyse data, since it brings the researcher closer to the data. With traditional approaches, the researcher tends to make some decisions about what is likely to be interesting early in the process. If these assumptions do not yield useful findings, then the researcher might try other ideas. However, if they do produce useful results the tendency is to stop, even if other (potentially more useful) results remain undiscovered. With an interactive approach, the researcher makes fewer initial decisions and is potentially freer to follow leads discovered during the analysis process.

11. Using Tables to Find the Story in the Data

Finding the story in the data is a major topic in its own right, and I have shared other material on that topic – for example a webinar series titled 'Find and Communicate the Story', produced in 2016. However, there are a few core approaches that can be highlighted in terms of tables:

- Framing the problem
- Finding the big picture
- Finding relevant exceptions
- Re-coding to refine the answer

11.1. Framing the Problem

Market research is usually conducted to help solve a business problem. However, the problem is frequently not adequately defined. The process of defining the problem and the sort of outcomes that would help solve the business question is referred to as framing. The framing of the problem should start before the research has been designed, before the survey was written, and before the data is collected. However, the framing should be re-examined prior to working with the tables. The process includes:

- Understanding the business problem that is being tackled.
- Knowing what success looks like to the end-user of the research.
- Knowing what business actions the business want to take as a consequence of the research.
- Listing what people think the results will be. The reason for collecting these hypotheses is so that the analysis can attempt to confirm and rebut each one.

11.2. Finding the Big Picture

Before looking at differences between sub-groups, for example differences between users and non-users, researchers should find the 'big picture'. For example, the big picture in a concept test might be that two of the five concepts were well received, one concept scored moderately well, and two concepts scored badly. The detailed picture will add information to this big picture, but the big picture creates a reference point for everything else.

The initial step when working with tables is to check that the data seems to be right, and that the coding has been done correctly. This checking process will normally provide the researcher with a sense of what is in the data – especially if the researcher also checks any open-ended comments.

After the checking process, the next step is to look at the main messages arising from the data. These main messages are usually found in the Total column. For example, what do most people like? What do most people do? Which brands, concepts etc are they most and least drawn to?

Once the big picture has been created, the researcher can move on to explore and describe this picture, and find relevant exceptions.

11.3. Finding the Relevant Exceptions

Once the big picture has been established, the next step is to find the relevant exceptions. The key in this process is the term 'relevant'. A relevant exception is one that creates a different business insight – which means the differences need to be big enough to matter, and in an area that matters.

When using tables, the key way to explore differences is to create banners that show relevant differences.

Many researchers start by comparing the classic headings such as gender, age, region etc. There is nothing wrong with looking at these, but the interesting differences are frequently not driven by demographic characteristics. One way to move beyond simple demographics is to re-code some of the data to make clearer/bigger differences.

11.4. Re-coding to Refine the Answer

One way to move beyond banners based on simple demographics is to construct new variables in the data. For example, if we see that young people seem to like the advertising and that men seem to like it, let's look at what young-men think – they might be driving both results.

For example, assume that the data from a project looking at a service suggest that there are four groups of users:

1. People who like the service, and use it.
2. People who like the service, and do not use it.
3. People who do not like the service and use, it.
4. People who do not like the service, and do not use it.

Allocating people to one of these four groups and using that variable as the banner, may help the researcher find out what elements are most associated with these differences.

12. Glossary

This glossary explains some of the key terms used in this eBook. If there are other terms that you think need to be explained (or if you disagree with my definitions), please email ray.poynter@thefutureplace.com with your suggestions.

Categorical variable Another term for nominal variables, variables that have no numerical sequence, such as gender (male versus female) or favourite drink (e.g. Coke, Fanta, Pepsi, etc.)

Effective sample size (ESS) The effective sample size describes the reduction in the sample (in terms of its ability to be representative) that happens when weighting is used. Consider a case where 4 interviews are conducted, 3 male and 1 female. We can weight the data so that it looks like 2 male and 2 female, by applying a weight of 2 to the female participant and a weight ⅔ to each male. But, there is still only one female in the data, so the sample will not truly reflect 2 males and 2 females, meaning the effective sample size will be smaller than 4, in this case 3. If you are interested in the formula for calculating the ESS, then you can Google 'Effective Sample Size'.

Integer An integer is a number with no fractional part. For example 1, 2, 3 and 4 are all integers. The following are not integers 1.4, 3.1416 and ⅔.

Integer scale An integer scale is a metric scale (a numerical scale) that consists of integers. Examples of integer scales include: number of children, number of drinks consumed per day or number of televisions owned.

Interlocking cells
Interlocking quotas In market research the term interlocking is used to describe a sample specification or weighting specification where the specifications combine different elements. For example, a simple sample specification might comprise Male 50%, Female 50% and Under 40 years 50%, Over 40 years 50%. An interlocking specification for this study might be:

- Male under 40 years, 25%
- Male over 40 years, 25%
- Female under 40 years, 25%
- Male over 40 years, 25%

Interval scale A metric scale where the intervals are equally spaced, for example the gap between 2 and 4 is the same as the gap between 6 and 8. With an interval scale there is not a meaningful 0. Both Centigrade and Fahrenheit are both interval scales as 0 on both scales do not mean zero heat. When there is a meaningful 0, the scale is a ratio scale. One the limitations of interval scales is that one number cannot be said to be X% bigger or smaller than another. For example, 30C is not 50% hotter than 20C (if we were to convert them to the [Kelvin scale](), where 0 is absolute 0, then the two values would be 283K and 293K, and we could say that 30C is about 4% hotter than 20C).

Metric variable	A numerical scale that is at least an interval scale, but includes ratio scales too.
Nominal variable	Another term for categorical variables, variables that have no numerical sequence, such as gender (male versus female) or favourite drink (e.g. Coke, Fanta, Pepsi, etc.)
Ordinal variable	Ordinal variables are ranked variables, such as most preferred, second most preferred, etc. Some scales might be categorised as nominal by some researchers and as ordinal by others. For example, highest level of education achieved. Some scales might be categorised by some researchers as interval by some researchers and as ordinal by others. For example, an Agree / Disagree scale might be seen as an interval scale running from 1 to 5 by some researchers, and might be treated as ordinal by others.
Random Probability Sampling	The gold standard, for many researchers, is random probability sampling. When random probability sampling is used every member of the target population has a known and non-zero chance of being interviewed. Examples of random probability sampling include: door-to-door research with a proper method for randomly picking houses and a protocol for dealing with people who are out and random digit dialling in situations where everybody has a phone. Random probability samples are very rare in commercial market research because of: • The high cost. • The absence of high quality sample frames. • The low rate of people agreeing to do a survey when asked.
Random sampling error	When a sample is interviewed the sample will only occasionally be exactly the same as the population. If you were to toss a coin 10 times, you would sometimes get 5 Heads and 5 Tails, but often you would get a different result such as 6 Heads and 4 Tails, for 3 Head and 7 Tails. The difference between what you get and what you might want/expect is sampling error.
Ratio scale	A ratio scale is a metric scale where there is a real 0, for example age and income. With a ratio scale we can describe one value as being a percentage bigger or smaller than another. For example, somebody who has owned a product for 5 years has owned it for 50% as long as somebody who has owned one for 10 years.

Significance testing Significance testing is used to give a researcher an indication about whether a finding in the data is likely to be big enough to be ‘real’. Because of sampling error, most results have a margin or error to them. Significance testing does not tell a researcher whether a difference is real or not, but it will indicate whether it is likely to be real.

Spec writer A spec writer is a specialist in producing market research cross-tabulation reports. The spec writer is normally given a description of what tables are needed and what treatment of the data is required (for example re-coding an NPS scale to show Promoters, Detractors, and a net score of Promoters minus Detractors). Spec writers may use a graphical interface to produce the tables, but many spec writers prefer to use a coding language to specify the tables. Most of the powerful tables packages have their own programming language.

t-test A statistic often used to test statistical significance in market research tables. The t-test is based on the t distribution and is one of two most commonly used statistics for this purpose, the other being the z-test. The t-test is more reliable than the z-test for smaller samples. If the sample (i.e. the column base) is less than 30, then the t-test should be preferred.

z-test A statistic often used to test statistical significance in market research tables. The z-test is based on the normal distribution and is one of two most commonly used statistics for this purpose, the other being the t-test. The z-test is less reliable than the t-test for smaller samples. If the sample (i.e. the column base) is less than 30, then the t-test should be preferred.

14. Thanks and Acknowledgements

Several people have helped me produce this document by reviewing the text, making suggestions, contributing ideas and I would like to thank them.

If you make a suggestion or contribution during this review stage, I will gladly add you to the list.

- Betty Adamou, Research Through Gaming, UK
- Sue York, NewMR, Australia

15. About the Author

This eBook is written by Ray Poynter.

Ray is the author of The Handbook of Online and Social Media Research, The Handbook of Mobile Market Research, editor of ESOMAR's book Answers to Contemporary Market Research Questions, a content author for the University of Georgia's Principles of Marketing Research online course, Manager Director of The Future Place and the founder of NewMR.org.

Ray has spent the last 35 years at the interface of research, technology and insight and is in wide demand as a consultant, trainer, workshop leader and keynote speaker.

Ray Tweets @RayPoynter and you can find him on LinkedIn.