CHI-SQUARE TOOL: STEP-BY-STEP INSTRUCTIONS

> An overview of the relevant statistical concepts used in these instructions is provided in the **Getting Started with Statistics** guide. An overview of cross-tabulations and chi-square tool, and the differences between the tests, is provided in the **exploring the relationship between survey variables guide** and **presentation**.[1]
>
> Additional guidance is also provided in the **ANALYSIS GUIDES**, which can be accessed here. The cross-tabulations tools for both surveys can be accessed here.

INTRODUCTION

Chi-square is a form of **'hypothesis testing'** that can build on cross-tabulation calculations and be used to test whether the relationship between two categorical variables is significant (i.e. there is or isn't a relationship between the variables). The statistical test generates a **p-value** which enables you to determine whether or not to reject the null hypothesis. This p value is the probability that the null hypothesis is true. If the p value is less than a set **significance level** (usually 5% [0.05] or 1% [0.001]), it is said that the pattern is 'statistically significant'. This means that there is a low probability that the observed pattern arose by chance.

Chi-square tests can be used to explore the frequency distribution of two categorical variables in the adult social care survey (ASCS) or carers survey (PSS SACE) dataset, such as satisfaction with care and support services and self-rated health (Q1 and 13 in the ASCS). This guide provides you with a step-by-step overview of how to use the **cross-tabulations and chi-square tool** in the MAX toolkit to generate a **chi-square statistic**.

These step-by-step instructions use illustrations from the ASCS version of the cross-tabulations and chi-square tool – but are also relevant for the PSS SACE tools – and use the following question

---

[1] Please note that cross-tabulations and chi-square tests can be used to explore the differences between respondent groups and, conversely, that t-tests and ANOVAs can also be used to explore the relationship between survey variables. The format of these guides and presentations were used to introduce increasingly complex concepts to MAX toolkit users and attendees of the 2016 webinars.

Are higher ratings of satisfaction with care and support services [Q1] associated with self-rated health [Q13]?

STEP-BY-STEP INSTRUCTIONS

Before you conduct chi-square test calculations, you need to establish **what type of chi-square test** you would like to run. The analysis tool provided in the MAX toolkit can be used to conduct 2 x 2, 2 x 3 and 2 x 4 chi-square tests: that is, tests where one variable has two response categories and the other variable has two, three or four response categories. If the variables you are exploring have too many response categories, you can consider collapsing or combining these.

The example we will use in this guide will focus on the highest ratings of satisfaction and self-rated health. The highest ratings of satisfaction fall in to three response categories (extremely, very and quite satisfied) so we will collapse these into two categories (very/extremely + quite). Self-rated health, on the other hand, falls in to five response categories (very good, good, fair, bad, very bad) so we will collapse these into three categories (good/very good, fair + bad/very bad). We will therefore conduct a 2x3 chi-square test.

Overall satisfaction with services Self reported health	Extremely/ Very	Quite	Total
Good / very good	83	24	107
Fair	79	43	122
Bad / very bad	28	19	47
Grand Total	**190**	**86**	**276**

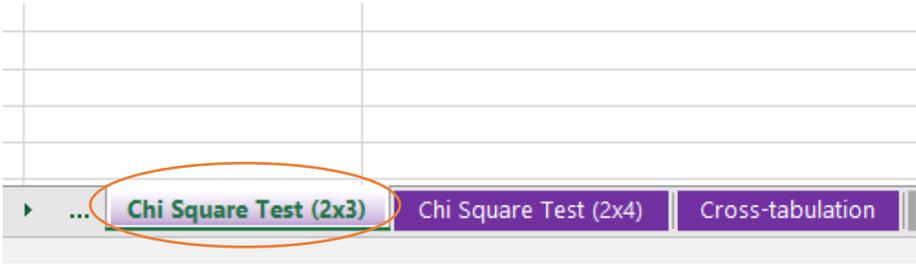**Contingency table created in Excel for illustratory purposes**

Step 1: open NHS Digital data return and relevant Cross-tabulations and chi-square tool from MAX toolkit.

Step 2: select, copy and paste data from data return into **ASCS data** tab of analysis tool, taking care to ensure columns match [note: the tool is based on 2014 data entry form. Additional amendments may therefore be required in future data collections].

	C	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU	AV	AW
	Gender	Q3a - Which of the following statements best describes how much control you have over your daily life?	Q3b - Do care and support services help you in having control over your daily life?	Q4a - Thinking about keeping clean and presentable in appearance, which of the following statements best describes your situation?	Q4b - Do care and support services help you in keeping clean and presentable in appearance?	Q5a - Thinking about the food and drink you get, which of the following statements best describes your situation?	Q5b - Do care and support services help you to get food and drink?	Q6a - Which of the following statements best describes how clean and comfortable your home is?	Q6b - Do care and support services help you in keeping your home clean and comfortable?	Q7a - Which of the following statements best describes how safe you feel?	Q7b - Do care and support services help you in feeling safe?	Q8a - Thinking about how much contact you've had with people you like, which of the following statements best describes your social situation?	Q8b - Do care and support services help you in having social contact with people?	Q9a - Which of the following statements best describes how you spend your time?	Q9b - Do care and support services help you in the way you spend your time?	Q10 - Which of these statements best describes how having help to do things makes you think and feel about yourself?	Q11 - Which of these statements best describes how the way you are helped and treated makes you think and feel about yourself
2		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
2		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2		4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1		2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
1		3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1		4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

| Introduction | Chi-Square Test (2x2) | Cross Tabulation | **ASCS Data** |

The information copied onto this tab is converted from numerical data (e.g. 1, 2 etc.) to text labels (e.g. male, female) on the hidden tab entitled 'ASCS data (2)', using the **Excel IF function**. Further information about the IF function can be found at the end of this guide.

Step 3: go to most appropriate Chi-square test tab of analysis tool. In this example, we will select the **Chi Square test (2 x 3)** tab.

| Chi Square Test (2x3) | Chi Square Test (2x4) | Cross-tabulation |

A screenshot of the chi-square tab in the MAX analysis tool is provided overleaf for reference.

Labels

Column Title	
Row Title	
Column , Category #1 Title	
Column , Category #2 Title	
Row , Category #1 Title	
Row , Category #2 Title	
Row , Category #3 Title	

Observed Values

		0		
		0		0 TOTAL
0		0		0
		0		0
		0		0
	TOTAL	0	0	0

Expected Values

Chi Square (X^2)	-
Degrees of freedom	2
p value	

If the p value is less then 0.05 (5%) then the null hypothesis (i.e. no pattern) may be rejected.

Step 4: manually add your **data labels** to the yellow section of the **labels table**. Please note that, in this example, the position of the data switches – so satisfaction with services, shown in rows of the cross-tabulation tables, is now shown in the columns; and self-rated health, formerly shown in the columns of the cross-tabulations table, is now shown in the rows.

Labels

Column Title	Satisfaction with services
Row Title	Self reported health
Column , Category #1 Title	Extremely/very satisfied
Column , Category #2 Title	Quite satisfied
Row , Category #1 Title	Very good/good
Row , Category #2 Title	Fair
Row , Category #3 Title	Bad/very bad

The labels that you input into this table will automatically copy over to the relevant fields in the observed values table (on the right of the labels table).

Observed Values

		Satisfaction with services		
		Extremely/very satisfied	Quite satisfied	TOTAL
Self reported health	Very good/good			0
	Fair			0
	Bad/very bad			0
	TOTAL	0	0	0

Expected Values

Step 5: manually add your **data** to the yellow section of the **observed values** table.

Observed Values				
		Satisfaction with services		
		Extremely/very satisfied	Quite satisfied	TOTAL
Self reported health	Very good/good	83	24	107
	Fair	79	43	122
	Bad/very bad	28	19	47
	TOTAL	190	86	276

Tip: if you are using collapsed categories (as in this example), you may find it useful to create a table in Excel first.

The chi-square statistics will be automatically calculated in the **expected values fields** below the observed values table.

Observed Values				
		Satisfaction with services		
		Extremely/very satisfied	Quite satisfied	TOTAL
Self reported health	Very good/good	83	24	107
	Fair	79	43	122
	Bad/very bad	28	19	47
	TOTAL	190	86	276

Expected Values			
		73.66	33.34
		83.99	38.01
		32.36	14.64
		1.184457181	2.616824005
		0.295947281	0.653837016
		0.586203483	1.295100718

Chi Square (X²)	6.63
Degrees of freedom	2
p value	0.036291024

If the p value is less then 0.05 (5%) then the null hypothesis (i.e. no pattern) may be rejected.

Step 6: establish whether **chi-square test assumptions** are fulfilled by your dataset.

1. **Sufficient sample size:** the sample has to be large enough to avoid **type 2 error**. See point 2.

2. **Sufficient expected observations for all categories:** depends on the type of chi-square table being generated:
    a. Expected frequency of 5 or more in all cells of a 2x2 table
    b. Expected frequency of 5 or more in 80% of cells in larger tables
    c. No cells with zero expected frequency
    If this assumption is violated, you can combine categories or conduct a Fisher's exact test. Combining categories is covered in the accompanying pre-recorded presentation. Please note that the Fisher's exact test is not covered in the MAX toolkit.

3. **Independence of observations:** each case contributes data to only one cell of the chi-square table. Chi-square analysis is therefore not suitable for data that is measured over time or data that is correlated in anyway. In these instances, a McNemar's test may be more suitable. Please note that the McNemar's test is not covered in the MAX toolkit.

4. **Missing values are excluded**: incomplete data should be excluded prior to analysis. Any missing data should also be recorded in your write up.

INTERPRETING THE CHI-SQUARE TEST STATISTICS

The chi-square test generates a **p-value** which enables you to determine whether or not to reject the null hypothesis. This p value is the probability that the null hypothesis is true. If the p value is less than a set **significance level** (usually 5% [0.05] or 1% [0.001]) it is said that the pattern is 'statistically significant'. This means that there is a low probability that the observed pattern arose by chance.

Chi Square (X²)	6.63
Degrees of freedom	2
p value	0.036291024

The p-value in this example is 0.036 and means that the **null hypothesis can be rejected** – in other words, *there is a statistically significant association between high ratings of satisfaction and self-reported health*.

A word of caution: a significant association does not imply causation but does identify potential important relationships that should be further explored using multi-variate and/or themed analysis. Further guidance is provided in the **FURTHER ANALYSIS GUIDES**, which can be downloaded from the main page of the MAX toolkit.

REPORTING THE CHI-SQUARE TEST STATISTICS

The usual format for reporting the results of a chi-square test is shown below:

$$X^2(2) = 6.63, p = 0.036$$

Test statistic

Degrees of Freedom

Actual value

P value

You may, however, want to report this in everyday terms. For example, *Chi-square analysis showed that there is a statistically significant association between higher ratings of satisfaction and self-reported health (you can include chi-square statistic in brackets). This may be further explored…..*.

ADDITIONAL INFORMATION

THE 'IF' EXCEL FUNCTION

The 'IF' function in Excel is used in the **cross-tabulations and chi-square tool** to convert the numbers in each cell of the **ASCS data tab** to the corresponding text (e.g. 1 = male, 2 = female, 9 = missing). The IF statements and corresponding text are shown in the **ASCS data (2) tab** and can be adapted, for example, if you would like to add or change categories.

Using the IF function in Excel to convert numerical data on the ASCS Data tab to text (labels) of the ASCS Data (2) tab.

For example, the formulae below from cell C3 of the 'ASCS Data (2)' tab link to cell C3 on the 'ASCS data' tab.

=IF('ASCS Data'!C3=1,"male",(IF('ASCS Data'!C3=2,"female",(IF('ASCS Data'!C3=-9,"missing","")))))

If cell C3 of the 'ASCS data' tab is 1, then the entry in the cell is the text 'male'.

=IF('ASCS Data'!C3=1,"male",(IF('ASCS Data'!C3=2,"female",(IF('ASCS Data'!C3=-9,"missing","")))))

If cell C3 of the 'ASCS data' tab is not 1, then the next statement is considered. If C3 is 2, then the entry in the cell is the text 'female'.

=IF('ASCS Data'!C3=1,"male",(IF('ASCS Data'!C3=2,"female",(IF('ASCS Data'!C3=-9,"missing","")))))

If cell C3 of the 'ASCS data' tab is not 1 or 2, then the next statement is considered. If C3 is -9, then the entry in the cell is the text 'missing'.

=IF('ASCS Data'!C3=1,"male",(IF('ASCS Data'!C3=2,"female",(IF('ASCS Data'!C3=-9,"missing","")))))

If cell C3 of the 'ASCS data' tab is not 1 or 2 or -9, then the text is left blank ("").

=IF('ASCS Data'!C3=1,"male",(IF('ASCS Data'!C3=2,"female",(IF('ASCS Data'!C3=-9,"missing","")))))

Further information about how to use IF functions and other features of Excel, please visit the **Office Training Centre** at https://support.office.com/en-gb/article/Office-Training-Center

DISCLAIMER

The MAX toolkit and website are based on independent research commissioned and funded by the NIHR Policy Research Programme (Maximising the value of survey data in adult social care (MAX) project and the MAX toolkit implementation and impact project). The views expressed on the website and in publications are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or its arm's length bodies or other government departments.