

An Introduction to tests for Categorical Outcomes

In all the exercises so far this afternoon focussed singularly on scenarios where our dependent (AKA outcome) variable is a **continuous variable**. For example, the **number of items** recalled correctly during a memory assessment, the **standardised score** on an IQ test.

We have, of course, come across **categorical variables** too. However, up until now we've only considered how to use these as independent (AKA predictor) variables in our analyses. For example: the effects of **Gender** (being *Male* or *Female*); **Group** (tested in the *Morning* versus tested in the *Afternoon*).

In this final tutorial we're going to look at the analysis options open to you when your study is interested in analysing the influence of other variables upon a **categorical dependent variable**. For example, often in medically-orientated research there is a particular interest in understanding what factors in the present predict whether someone will or won't go on to develop a certain health problem in the future (e.g. Depression, Schizophrenia, Obesity)? This type of dependent variable is what we'd call a *dichotomous (or binomial) categorical outcome*. This isn't a continuous variable because there are only two possible response values (i.e. two groups a participant can fall into): a person either will develop the specific health problem or they won't. For these sorts of analyses we need to look to use types of statistical tests, which are designed to evaluating how variables influence the distribution of response in a **categorical outcome**.

The relevant chapters in the Andy Field Textbook for this exercise is:

Chapter 19: Categorical Outcomes: Chi-square & Log-linear analysis

In these exercises first focus on a basic type of categorical analysis that will establish some key concepts within this field of analysis. Using this type of analysis we will learn how to test if a **single categorical independent variable** has an influence on single a categorical dependent variable

Exercise 1: Chi-squared testing – by hand

Let's begin with the simplest form of categorical analysis: in which we have only two variables that are both categorical, each only has two levels and we only wish to predict whether one variable has an influence on the other. In other words, we identify one of these two categorical variables as being the independent variable and the other as the dependent variable in our design

Note: the theory I'm about to discuss is covered in a more detail in the **Andy Field textbook, Chapter 19, Section 19.3**

As an example, let's imagine I'm interested in determining if elderly individuals who are identified as having 'high' blood pressure are at greater risk of being diagnosed with Vascular Dementia at some point in the 5 years following that blood pressure test, compared to those where categorised as having 'normal' blood pressure. The best way to describe such data is in the form of a **contingency table**, which displays the frequencies for each of these two variables across rows and columns (see example on next page). This should look very familiar to the **crosstabs** you were creating in SPSS, in the earlier Descriptives Statistics session.

<i>Table 1 (observed)</i>	DV: Vascular Dementia		
IV: Blood Pressure	Yes	No	Total
High	28	48	76
Normal	10	114	124
Total	38	162	200

Null Hypothesis testing using contingency tables

Considering our hypothesis in terms of the contingency table, we can think of it as taking the form of a prediction that: *the category that a participant falls into within the independent variables alters the likelihood of which category they will fall into within the dependent variable*. In other words, being a 'High' for blood pressure increases your odds of also being a 'Yes' for Vascular Dementia compared to if you were a 'Normal' for blood pressure. We test this hypothesis by testing how significantly different from **chance** the frequency distribution is across the cells highlighted in yellow in the table.

Extra tips/pointers:

These values are sometimes referred to as the 'counts' in each cell of the table (e.g. 28 participants were 'counted' as having high blood-pressure and vascular dementia)

To flip our hypothesis around, to consider it as a null hypothesis that we will actually test, we ask if the distribution of observations we see in the table are likely due to 'random chance' allocation. For example, even if there was no relationship between blood pressure and vascular dementia and we looked at the 200 elderly individuals in this example we might expect some of these people who happened to have high blood pressure would go on to develop dementia. Similarly, we'd expect some with normal blood pressure to go on to develop dementia too. If there was no systematic effect of blood pressure on your risk of dementia, we'd subsequently anticipate that the proportion of people who *overall* go on to develop Dementia would be very similar to the proportions *within each Blood pressure group* that go on to develop dementia (i.e. the risk of developing dementia is approximately the same irrespective if you're 'High' or 'Normal' for your blood pressure).

The prediction we're therefore interested in testing statistically is whether the actual **observed frequency distribution** (illustrated in the table above) is significantly different enough from an **expected frequency distribution** based on the null-hypothesis.

Can you illustrate in the table below the values that you would expect in each cell if 20% of the sample go on to develop vascular dementia and the null hypothesis is being true:

<i>Table 2 (expected)</i>	DV: Vascular Dementia		
IV: Blood Pressure	Yes	No	Total
High	15	61	76
Normal	25	99	124
Total	40	160	200

--	--

To formally test the statistical significance of the null hypothesis in this example we need to use a certain family of tests called **chi-squared tests** (χ^2).¹ The easiest of these to understand (in computational terms) is the Pearson Chi-square, which we'll apply to the example we've been working through.

Of course, in practice we are very unlikely to know this sort of information prior to collecting our sample.² Instead, all we will have is our frequency distributions in the ‘margins’ of our contingency table (i.e. the total proportion of our sample that has high/low blood pressure and the total proportion of the sample with/without vascular dementia). We can work backwards from these totals to estimate the expected frequency in each cell using the following equation though:

Below is an equation that I've used to calculate the expected frequency for having high blood pressure and vascular dementia, based on the Margins in Table 1. Can you complete the rest of the blanks in Table 3 (on the next page) using the same method?

$$model_{High,Yes} = 14.4$$

² i.e. the overall expected prevalence of our outcome (i.e. vascular dementia), irrespective of its association with the independent variable.

<i>Table 3</i>	DV: Vascular Dementia		
IV: Blood Pressure	<i>Yes</i>	<i>No</i>	Total
<i>High</i>	14.4	61.6	76
<i>Normal</i>	23.6	100.4	124
Total	38	162	200

Extra tips/pointers:

Notice how similar these are to the expected frequencies you worked out in Table 2. Once we have these two 'scenarios' for our contingency table mapped out, we can then formally contrast them, statistically testing how probable the observed distribution we have is, were the expected distribution (which reflects the null hypothesis) the true distribution of the wider population.

To quantify the degree of difference between the expected and observed frequencies we calculate the Chi-squared (χ^2) statistic³. This is difference between an Expected versus Observed frequency squared, for each 'count' cell, divided by that cell's expected value and then summed together for all count cells. This is essentially a measure of how much deviance from the cell counts expected under the null hypothesis there is in your actual observed frequencies.

To demonstrate this in action I've calculated below the χ^2 statistic based on the observed cell counts (i.e. Table 1) versus the expected cell counts (i.e. Table 3):

$$\begin{aligned}\chi^2 &= \frac{(28 - 14.4)^2}{14.4} + \frac{(48 - 64.0)^2}{64.0} + \frac{(10 - 23.6)^2}{23.6} + \frac{(114 - 100.4)^2}{100.4} \\ \chi^2 &= \frac{184.96}{14.4} + \frac{256}{64.0} + \frac{184.96}{23.6} + \frac{184.96}{100.4} \\ \chi^2 &= 12.84 + 4 + 7.84 + 1.84 \\ \chi^2 &= 26.52\end{aligned}$$

This Pearson chi-squared test has 1 degree of freedom⁴ and the value we have for χ^2 here is therefore statistically significant because this value (i.e. 26.52) is much larger than the critical value for χ^2 at one degree of freedom for rejecting the null at the 5% level (if $\chi^2 \geq 3.84$ then $p < .05$). This means we would reject the null hypothesis in this scenario and find support for our alternative hypothesis: that blood pressure level does influence the risk of going on to develop vascular dementia.

Can you now calculate the value you would get for χ^2 if, instead of using the cell counts from Table 1 you used the values you calculated and entered in Table 2 (i.e. this would represent a scenario where the observed frequencies were very similar to the expected frequencies). Can you also interpret the statistical significance of this χ^2 (i.e. based on the cut-off value mentioned for rejecting null hypothesis at the $p < .05$ level in the previous paragraph)?

³ which we can then use to compare against an expected sampling distribution to derive a p-value

⁴ Calculated as $df = (\text{Number of rows} - 1)(\text{Number of columns} - 1)$

Extra tips/pointers:

The χ^2 statistic, like all test statistics, has a known sampling distribution, which varies depending on the size of your sample (converted into “degrees of freedom [dfs]”). Therefore, if we’ve calculated a χ^2 value for our sample, and its associated dfs we can then use a look-up table (like this [one](#)) to judge the significance of our χ^2 ’s value. If it’s greater than the critical value you find in this table then we can say we have grounds to reject the null hypothesis.

Assumptions and limitations for Chi-Squared Testing

We’ll discuss the several different types of chi-squared tests SPSS can do for us in a moment but before doing this it is best to highlight some assumptions that need to be taken into account when running any of the following types of analysis:

1. Observations must be independent in your sample. This rules out using this method of analysis if you have a repeated measures experimental design (see paired-samples t-tests).

Extra tips/pointers:

An easier way of thinking about this is that each participant **can only be ‘counted’ once in the contingency table**. That is, they either get counted as a ‘High+Yes’, ‘High+No’, ‘Normal+Yes’ or a ‘Normal+No’ in the yellow cells in the previous example.

2. Within our sample, the expected frequencies for each possible response has to be greater than 5 or (in large contingency tables, no more than 20% of the expected frequencies can be below 5 but all should be above 1). Think of this as having enough data that could conceivably be consistent with the null hypothesis for it to be testable.
3. Another point Field Makes (**Chapter 19, Section 19.5.3**) is that in interpreting statistical significance we should look at row and column **percentages** rather than raw frequency scores as the latter can be misleading because their magnitudes are influenced by the actual number of observations we have in our sample. This is not an assumption as such but something to bear in mind to avoid risk of over-interpretation!

Lastly, bear in mind that the results of a chi-squared test can be interpreted much like the results of any other test for significance. That is, it is a test of whether the interaction between the two variables is statistically significant but it tells us nothing of the direction of this relationship (e.g. does High Blood Pressure increase or reduce the risk of Vascular Dementia and/or does Low Blood Pressure do the same or opposite?). To explore this further we need to look at the **standardised residuals** from our output (something I'll come back to in the following example).

Exercise 2: Chi-squared testing – application in SPSS

The example we're going to work uses the data in **session_3_dataset_Ex4.sav** file, which represents the following study:

Imagine you are a social psychologist interested in superstitious beliefs in sports people. Specifically, you are interested in what influences football managers to believe (or not) in the concept of home-field advantage¹ (i.e. believing you are more likely to win when playing at home). You ask a sample of football managers whether or not they believe in home-advantage (Believe vs. Don't Believe) and see whether their likelihood of believing in this effect depends on a number of different factors, including whether:

1. Their mentor (i.e. the first manager they worked under) believed in this advantage
2. Whether in the past season they have won more times than they have lost at home
3. Whether they consider the team to have a clear 'local rival' they play 'derby' matches against on a regular basis

¹inspired by Van Den Ven, N. (2011), Supporters are not necessary for Home Advantage: Evidence from same-stadium derbies and games without an audience. *Journal of Applied Psychology*, 41(12).

Using Crosstab as Contingency Tables in SPSS

To test the first hypothesis in the box above (i.e. Did their mentor's believed in home-advantage influence their belief) we're going to use the **Crosstabs...** contingency tables function, like so::

1. Go to **Analyze > Descriptive Statistics > Crosstabs...**
2. In the Crosstabs window, select your Independent Variable (*Mentor_Blf*) and move it to the **Row(s):** box.
3. Next select your Dependent Variable (*Man_Blf*) and move it to the **Column(s):** box.
4. Open the **Statistics...** sub-window and in here select the following tests to allow us to test our hypotheses regarding the frequency distributions within this contingency table:
 - Chi-square
 - Contingency Coefficient
 - Phi and Cramer's V

Then click **Continue** to close this sub-window.

5. We'll also ask for **Fisher's Exact Test**, which is a variation on the regular Chi-square tests. Whilst not all that necessary in this example, it is a test recommended when you have a 2x2 contingency table like in our example but you have a much smaller dataset. To select this click on **Exact...** and in the sub-window that pops up change to the **Exact** option, then click **Continue** to close this sub-window.

6. Finally, using the **Cells...** sub-window we need to request some additional output to help check assumptions and aid interpretation. Open this up and make sure the following are selected:

- **Counts:** both observed and expected
- **Percentages:** Row, Column and Total

To explore further any significant effects also ensure the following are selected:

- **Residuals:** Standardized

Then click **Continue** to close this sub-window.

7. To run the analysis click **OK**.

To interpret the output take the following steps:

Check Assumptions

In the **Crosstabulation** table (which shows *Mentors' Beliefs [Mentor_Blf]* by *Managers' Beliefs [Man_Blf]* in home advantage) first look at the Expect Count values for each cell (see figure on the next page illustrating where you can find them) and evaluate whether our assumptions for running this type of analysis have been met? If you think they have, why is this (*Hint: refer back to point 2 on page 5*)?

What assumption would we violate if we changed the research question to ask whether football players' beliefs in home advantage were influenced by their manager's belief? In such a scenario we'd likely ask teams of players from each club to participate. What problems would this present in terms of assumptions?

Investigate Chi-Squared Results: Testing our Hypothesis

In the **Chi-Square Tests** table the row we're primarily interested in is the *Pearson Chi-Square* and in particular its *Asymptotic Significance*. As with all the other kinds of significance testing we've come across, the fact that $p = .029$ here (i.e. $< .05$) tells us that we have strong grounds to reject the null hypothesis that a manager's belief in home advantage is not influence by their mentor's beliefs.

In other words, we have grounds to go on and explore further the patterns of difference between manager and mentor belief to ascertain whether (i) managers are more likely to have believed in home advantage if their mentors did or (ii) managers are less likely to believe in home advantage if their mentors did.

Did their first mentor believe in homefield advantage? * Does the manager believe in homefield advantage Crosstabulation					
		Does the manager believe in homefield advantage			
			Yes	No	Total
Did their first mentor believe in homefield advantage?	Yes	Count	87 ^a	45 ^b	132
		Expected Count	78.3	53.7	132.0
		% within Did their first mentor believe in homefield advantage?	65.9%	34.1%	100.0%
		% within Does the manager believe in homefield advantage	55.8%	42.1%	50.2%
		% of Total	33.1%	17.1%	50.2%
		Standardized Residual	1.0	-1.2	
	No	Count	69 ^a	62 ^b	131
		Expected Count	77.7	53.3	131.0
		% within Did their first mentor believe in homefield advantage?	52.7%	47.3%	100.0%
		% within Does the manager believe in homefield advantage	44.2%	57.9%	49.8%
Total		% of Total	26.2%	23.6%	49.8%
		Standardized Residual	-1.0	1.2	
		Count	156	107	263
		Expected Count	156.0	107.0	263.0
		% within Did their first mentor believe in homefield advantage?	59.3%	40.7%	100.0%
		% within Does the manager believe in homefield advantage	100.0%	100.0%	100.0%
		% of Total	59.3%	40.7%	100.0%

Each subscript letter denotes a subset of Does the manager believe in homefield advantage categories whose column proportions do not differ significantly from each other at the .05 level.

Look Along this row...

... and this row...

... and this row.

You'll notice we also get a number of other versions of the chi-square test reported in this table:

- The Yates' **Continuity Correction**, which is an adjusted version of the Pearson Chi-Square that corrects for the fact that the Pearson test can often be overly lenient in 2x2 contingency table and thus is a greater risk of making a type-1 error. However, Field suggests (see **Chapter 19, Section 19.3.5**) that Yates's is perhaps too conservative a correction (i.e. it goes too far in the opposite direction).
- The **Likelihood Ratio**.
- **Fisher's Exact test**, which is provided as both a 2-sided and 1-sided test. Although, as I've stated earlier, this variation is probably unnecessary given how large our sample size is.

Reassuringly, all these variations of the χ^2 return the same conclusion, that we should reject the null hypothesis (i.e. $p < .05$)

Exploring further the nature of the association

To explore the nature of the significant association further we could look to describe it using the observed percentages in the **Crosstabulation**. Specifically we want to summarise our results in terms of either the percentages within the rows **or** the percentages within the columns (i.e. in instances where the percentages sum to 100% in the margins of the table). Below I've provided an example of how to provide this interpretation along the rows (i.e. **% of Managers believing if mentors did**):

Of the managers whose first mentor believed in home advantage, a higher proportion believed in this advantage (65.9%) than didn't (34.1%). Meanwhile, of the managers whose first mentor didn't in home advantage there was a much more marginal bias in favour of not believing in this advantage (i.e. 47.3% did believe, whilst 52.7% did).

Can you provide an equivalent statement looking down the **columns** (i.e. % of Mentors believing if Mentee'd Manager did)

--

We can be more objective in our interpretations by also looking at the **Standardized Residuals** in this table also. These tell us the strength of relationship between each cell within the contingency table and the overall significance results of our Chi-Square test. In other words, which cell(s), out of the four cells within this 2x2 table, are contributing more to the conclusion that there is systematic variability in the frequency distribution due to the independent variable?

In terms of its absolute size, a standardised residual can be treated like a z-score, meaning that a value $\geq \pm 1.96$ indicates a cell is having a significant (at the $p < .05$ level) influence, whilst a residual value of $\geq \pm 2.58$ is significant at the $p < .01$ level. Are there any clear relationships emerging here?

--

Exploring further the strength of Associations

So far, we have some ambiguous evidence for a meaningful relationship between a manager's belief in home-advantage and their mentors. The overall significance of the χ^2 test suggests a relationship exists but the Standardised Residuals are inconclusive as to the nature of that relationship, suggesting that whilst there may be a relationship its **effect size** is likely small, and thus difficult to reliably detect. In other words, whatever the influence mentors have had it's a relatively small contributing factor in influencing whether a manager also believes in home advantage).

To properly assess effects sizes here we need to go beyond statistical significance and look at a few additional things in follow-up testing:

Firstly, we can get a quick sense of the effect size for the overall chi-square result by looking at the **Symmetric Measures** table. This table presents adjusted versions of the original chi-square test, which take into account sample size and degrees of freedom and try to restrict the range of the test statistic (i.e. the *Value* in this table) to fall between 0 to 1 (to make it similar to interpreting a correlation coefficient). Looking at these test statistics and reflecting on what you know about correlations, what conclusions would you draw about the strength of the relationship between manager and mentor belief here?

Odds Ratios

A more common method of reporting effect sizes for categorical outcomes is to use a statistic we call the **odds ratio**. To calculate the odds ratios (**ORs**) for believing in home advantage if your mentor also did, compared to if they didn't, we need to do the following calculation, using the *Count* values in the **Crosstabulations** Table:

$$\begin{aligned}\text{Odds}_{\text{Manager Yes AND Mentor Yes}} &= \text{Manger 'Yes' \& Mentor 'Yes' / Manger 'No' \& Mentor 'Yes'} \\ &= 87/45 = 1.93 \\ \text{Odds}_{\text{Manager Yes BUT Mentor No}} &= \text{Manger 'Yes' \& Mentor 'No' / Manger 'No' \& Mentor 'No'} \\ &= 69/62 = 1.11290323 \\ \text{Odds Ratio} &= \text{Odds}_{\text{Manager Yes AND Mentor Yes}} / \text{Odds}_{\text{Manager Yes BUT Mentor No}} \\ &= 1.93/1.11290323 = \underline{\underline{1.74}}\end{aligned}$$

This value of 1.74 for this Odds Ratio represents the odds of one outcome versus another (i.e. a manager believing in home advantage versus not believing) under each level of the independent variable (i.e. whether their mentor *did* or *did not* also believe in home advantage). We then look at the ratio of these two odds to see if our outcome is more likely (i.e. has higher odds) under one level of the IV compared to the other.

We now need to understand how to interpret an odds ratio of 1.74 means! To do so:

- First note that Odds ratios are expressed as a single numerical value⁵
- **An odds ratio (OR) > than 1** (with 95% CI around it that has a lower bound >1) suggests there is a **significantly elevated chance of the predicted outcome**. In other words, the odds of falling in the category of interest are greater (i.e. you're more likely believe in home advantage) if you exhibit the 'risk factor' (i.e. your mentor believed in home advantage). The size of the odds ratio is proportional to the increase in risk. For example, an OR of 2 would mean participants were twice as likely to believe in home advantage if their mentor also did, whilst an OR of 100 means they're a hundred times more likely.
- **An odds ratio (OR) = 1** (and a 95% CI that has its lower bound <1 whilst its upper bound is >1) suggests there is no systematic relationship because you are as likely to fall within the category of interest as you are within the comparator group (i.e. you're as likely to believe in home

⁵ You will often see a 95% confidence interval around this value too though because, as ever, we are estimating the odds in the wider population based on the odds we've seen in our sample.

advantage as not) irrespective of if you fall within the 'risk factor level' for your predictor (i.e. your mentor believing in home advantage).

Extra tips/pointers:

It might help to think of this as the odds being at "fifty:fifty" or 'chance' level. After all what value do you get when you divide 50 by 50 (i.e. 1!)?

- **An odds ratio (OR) < than 1** (and a 95% CI around it that has an upper bound <1) suggests there is a significantly reduced chance of the predicted outcome, whereby the odds of falling in the category of interest are less than into its comparator. In this case that would mean you're less likely to believe in home-advantage if you exhibit the 'risk factor' (i.e. your mentor believes in home advantage). The size of the odds ratio is proportional to the reduction in risk. For example, an OR of 0.5 would mean participants were half as likely to believe in home advantage if their mentor believed in home advantage, an OR of 0.01 means you are a hundred times less likely.

Extra tips/pointers:

Due to the inverse nature of this relationship you often hear 'risk factors' in such analysis instead be referred to as 'protective factors'. After all, they reduce the likelihood of the (sometimes negative) outcome you're studying when running this type of analysis. For example, regular exercise is a protective risk factor for heart disease because taking regular exercise reduces your odds of having a heart attack, compared to those who do not exercise.

Given these interpretive points, how would you interpret the odds of a manager believing in home advantage if they've worked under a mentor that also believed in home advantage?

Also, more generally how would you evaluate the evidence for claiming that mentors' beliefs in a home advantage influence the likelihood of managers' also believing in a home-advantage?

To conclude these exercises, can you repeat the same analysis but this time test whether winning more often than not at home in the last year influences beliefs about a home advantage. Summarise what you find in a format similar to the summary below, which describes the effect of whether the team has a close local rival on belief in home advantage.

There was a significant association between having a close local rival and the manager disbelieving in home advantage $\chi^2(1) = 8.68, p = .003$. The odds ratio indicated the odds of a manager believing in home advantage if you had a local rival, compared to if they didn't have such a rival, were 0.46. In other words managers with a local rival were less than half as likely as those that didn't to believe in home advantage.



Are these three results in line with your expectation and what would you infer from them about the factors that affect a manager's belief in the concept of home advantage?

