2 X 2 Contingency Chi-square

The 2 X 2 contingency chi-square is used for the comparison of two groups with a dichotomous dependent variable. We might compare males and females on a yes/no response scale, for instance.

The contingency chi-square is based on the same principles as the simple chi-square analysis in which we examine the expected vs. the observed frequencies. The computation is quite similar, except that the estimate of the expected frequency is a little harder to determine.

Let's use the Quinnipiac University poll data to examine the extent to which independents (non-party affiliated voters) support Biden and Trump.[1] Here are the frequencies:

	Trump	Biden	
Party affiliated	338	363	701
Independent	125	156	281
	463	519	982

To answer the question whether Biden or Trump have a higher proportion of independent voters, we are making a comparison of the proportion of Biden supporters who are independents, $156/519 = .30$, or 30.0%, to the proportion of Trump supporters who are independents, $125/463 = .27$, or 27.0%. So, the table appears to suggest that Biden's supporters are more likely to be independents then Trump's supporters. Notice that this is a comparison of the conditional proportions, which correspond to column percentages in cross-tabulation output.[2]

First, we need to compute the expected frequencies for each cell. R_1 is the frequency for row 1, C_1 is the frequency for row 2, and N is the total sample size. The first cell is:

$$E_{11} = \frac{R_1 C_1}{N} = \frac{(701)(463)}{982} = 330.51$$

Filling in the rest of the cells in the same way for each expected value, E_{ij}, using the same equation but by using frequencies from the corresponding row R_i and column C_j for each cell, I obtained the following expected values:

$E_{11} = \frac{(701)(463)}{982} = 330.51$	$E_{12} = \frac{(701)(519)}{982} = 370.49$
$E_{21} = \frac{(281)(463)}{982} = 132.49$	$E_{22} = \frac{(281)(519)}{982} = 148.51$

$$\chi^2 = \sum \frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} = \frac{(338 - 330.51)^2}{330.51} + \frac{(363 - 370.49)^2}{370.49} + \frac{(125 - 132.49)^2}{132.49} + \frac{(156 - 148.51)^2}{148.51}$$
$$= 0.170 + 0.151 + 0.423 + 0.378$$
$$= 1.12$$

---

[1] These results are based on a recent national Quinnipiac University poll from Oct 4-7, 2019, https://poll.qu.edu/national/release-detail?ReleaseID=3643. Methodological details are here https://poll.qu.edu/images/polling/us/us10082019_demos_uljv62.pdf/.

[2] There are other questions we might ask, of course. Asking whether the proportion of independents who support Biden is higher than the proportion of non-independents (affiliated) who support Biden is an equivalent question to the one above (comparison of conditional row proportions rather than conditional column proportions). We also might ask whether independents are more likely to support Biden than Trump, which is a simple two-cell comparison among independents, which would be made by simply selecting out independent voters and using the z-proportions or chi-square test previously discussed.

The result of the chi-square is compared to the tabled critical value based on $df = (R - 1)(C - 1)$, where R and C represent the number of rows and the number of columns, respectively.[3] So, with $df = (R - 1)(C - 1) = 1$, the critical value is 3.84, and the computed value is not significant.

Minimum Expected Frequencies and Fisher's Exact Test
Fisher's exact test, proposed by R.A. Fisher (Fisher, 1935) and sometimes called the "Fisher-Irwin" test, is often printed along with the Pearson χ^2. It is not so much a modification of the chi-square test as an alternative approach to testing the association between two binary variables for significance. The test has been suggested for use with small samples in which the expected frequencies in some cells are low. The concept is to use the hypergeometric distribution to compute the exact probability of the particular configuration of obtained frequencies. The problem with Fisher' exact test is that it can be overly conservative and its use is often recommended when not necessary. Some software packages print a warning when 20% of the cells have an expected frequency below 5 (known as Cochran's rule). First thing to notice, however, is that it is the expected frequency that is of concern and not the observed frequency. Secondly, simulation studies (e.g., Camilli & Hopkins, 1978) suggest that Pearson's χ^2 has nominal alpha values with expected values as low as 1 as long as the total sample size is 20 or larger. So, the upshot is that Fisher's exact test is not needed in very many circumstances.

Yates' Continuity Correction
Yates suggested a correction to the Pearson's χ^2 based on the notion that a test of discrete variables should follow a discrete distribution are tested using a normal approximation, the chi-squared distribution. The Yates' correction for continuity is a simple modification of the chi-squared test formula by subtracting ½ or .5 from the frequency difference.

$$\chi^2 = \sum \frac{\left(\left|O_i - E_{ij}\right| - .5\right)^2}{E_{ij}}$$

There is good evidence and fairly wide consensus that the results with the Yates correction are too conservative (e.g., Grizzle, 1967; Camilli & Hopkins, 1978).

Magnitude of Effect
The most commonly used effect size measure associated with the 2 × 2 chi-square test is phi, ϕ (the Greek lower case "fee", as pronounced by statisticians). Phi is a simple computation, based on chi-square.

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

According to Cohen's (1992) guidelines, .1 is a small effect, .3 is a medium effect, and .5 is a large effect. Cramer's V is used for more than a 2 × 2 chi-square, and it is equivalent to phi for the 2 × 2 design. It is also the case the Cohen's w is equivalent to phi in this circumstance, and if you look back at the two-group chi-square, you will see that the computations are the same. Cohen's w can be used for any chi-square test, whether for a one-or two-dimensional table or other. These are all what Howell (2010) refers to as r-type effect size measures, because, as we will soon see, phi is the same as the Pearson correlation coefficient. Howell also discusses what he calls d-type effect size measures, odds ratios and relative risk, and we will discuss those next term when we discuss logistic regression.

Three or More Dimensions
Although 2 × 2 contingency table looks like a 2 × 2 factorial table (to be discussed later in the term), they are not analogous. The homogeneity conceptualization of chi-squared tests involves a two-group comparison of a binary outcome, which is analogous to a t-test in the continuous case. Because one of the columns (or rows)

---

[3] I use Howell's notation, which is understandably confusing in this case, because above R_i and C_j refer to the frequencies (i.e., number of cases in a row or column) and here the R and C without subscript refer to the number of rows and columns.

is for the dependent variable, it is really the three-way table that is analogous to the factorial design in ANOVA, which requires an analysis of a three-way contingency table (2 × 2 × 2) in the binary outcome case (to be discussed later this term and next term in greater depth).

Partitioning

The chi-squared values for the set of all possible *orthogonal* (or independent) chi-squares add up to the chi-square for the whole design. The likelihood ratio test (discussed next term), G^2, however, cannot be partitioned in the same way. Planned follow-up analyses to a significant Pearson χ^2 for contingency tables are simply chi-square analyses based on chi-squared tests for two or more cell comparisons, including smaller contingency tables (e.g., a 2 × 2 from a 5 × 3 design; Delucchi, 1993). Such tests may involve marginal proportions or individual cell proportions as well.

Chi-square Software Examples

SPSS

Syntax

```
crosstabs /tables=ind by response
  /cells=count row column expected
  /statistics=chisq phi.
```

Menus

1. Analyze→Descriptive statistics→ crosstabs
2. Move variables over
3. Click on “statistics”
4. Check Chi-square box and the Phi and Cramer's V box, click “continue,” click “ok”
5. Click on “cells” and choose row and column percentages.

ind ind vs affil * response intended vote Crosstabulation

			response intended vote		
			0 Trump	1 Biden	Total
ind ind vs affil	.00 affiliate	Count	338	363	701
		Expected Count	330.5	370.5	701.0
		% within ind ind vs affil	48.2%	51.8%	100.0%
		% within response intended vote	73.0%	69.9%	71.4%
	1.00 independent	Count	125	156	281
		Expected Count	132.5	148.5	281.0
		% within ind ind vs affil	44.5%	55.5%	100.0%
		% within response intended vote	27.0%	30.1%	28.6%
Total		Count	463	519	982
		Expected Count	463.0	519.0	982.0
		% within ind ind vs affil	47.1%	52.9%	100.0%
		% within response intended vote	100.0%	100.0%	100.0%

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	1.122[a]	1	.290		
Continuity Correction[b]	.977	1	.323		
Likelihood Ratio	1.123	1	.289		
Fisher's Exact Test				.322	.161
Linear-by-Linear Association	1.121	1	.290		
N of Valid Cases	982				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 132.49.
b. Computed only for a 2x2 table

Symmetric Measures

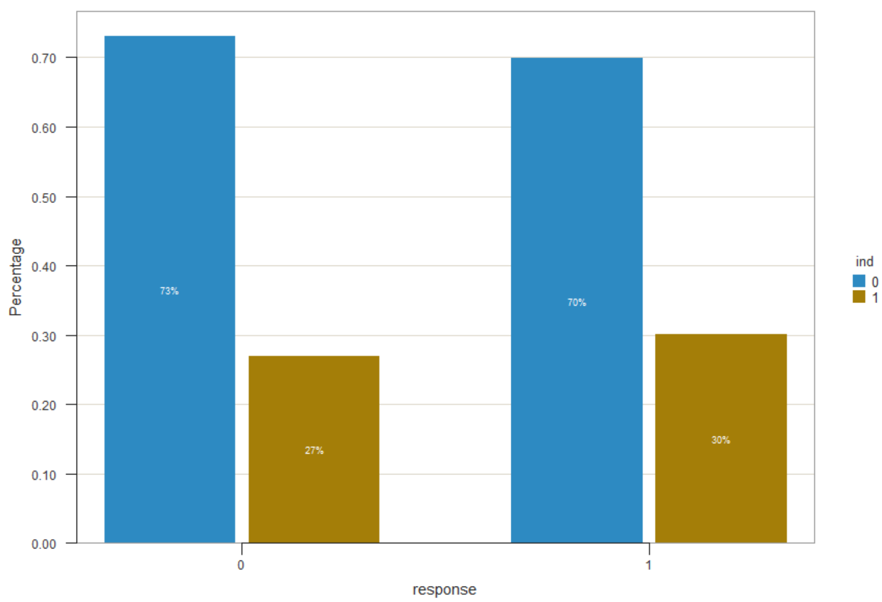
		Value	Approximate Significance
Nominal by Nominal	Phi	.034	.290
	Cramer's V	.034	.290
N of Valid Cases		982	

R

```
> #this lessR BarChart function produces a chi-square test by default⁴
> BarChart (response, by=ind, horiz = FALSE, stat = "proportion", beside = TRUE)
```

[4] This simpler syntax, `BarChart (response, by=ind)`, provides the same chi-square test, but the extra statements I give above produces a better figure with the conditional proportions.

```
Joint and Marginal Frequencies
------------------------------

response
ind      0   1 Sum
  0    338 363 701
  1    125 156 281
  Sum 463 519 982


Cramer's V (phi): 0.034

Chi-square Test:  Chisq = 1.122, df = 1, p-value = 0.290


Cell Proportions within Each Column
-----------------------------------

response
ind        0     1
  0    0.730 0.699
  1    0.270 0.301
  Sum 1.000 1.000
```



To get marginal frequencies and proportions in R

```
> tbl = table(mydata$ind, mydata$response)
> tbl

      0   1
  0 338 363
  1 125 156
> #and get marginal and cell proportions
> #margin.table(tbl, 1) # Frequencies summed over response
> margin.table(tbl, 2) # Frequencies summed over ind

  0   1
463 519
>
> #prop.table(tbl) # cell proportions
> #prop.table(tbl, 1) # row proportions (within each level of ind)
> prop.table(tbl, 2) # column proportions (within each level of response)

            0         1
  0 0.7300216 0.6994220
  1 0.2699784 0.3005780
```

Example write-up. A chi-square test was used to determine whether there was a significant difference between the proportion of Biden and Trump's supporters who are independent. Results indicated that 30.1% of Biden's supporters were independents, whereas 27.0% of Trumps supporters were independents. This difference was not significant, $\chi^2(1) = 1.12$, p = .29 The phi coefficient, $\phi = .03$, suggested a small effect.

References

Cohen, J. (1988). Statistical power analysis for the behavioral sciences Lawrence Earlbaum Associates. *Hillsdale, NJ*, 20-26.
Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.
Camilli, G., & Hopkins, K. D. (1978). Applicability of chi-square to 2× 2 contingency tables with small expected cell frequencies. *Psychological Bulletin, 85*(1), 163-167.
D'Agostino, R. B.: & Rosman, B. (1971) A normal approximation for testing the equality of two independent chi-square values. *Psychometrika, 36*, 251-252.
Fisher, R. A. (1935). T*he design of experiments*. Edinburgh, UK: Oliver and Boyd.
Grizzle, J. E. (1967). The Teacher's Corner: Continuity Correction in the χ2-test for 2× 2 Tables. *The American Statistician, 21*(4), 28-32.
Howell, D. C. (2010). *Statistical methods for psychology, seventh edition*. Cengage Learning.