



SO5032 Lecture 2

Brendan Halpin

February 19, 2024

SO5032 Lecture 2

SO5032 Lecture 2

Lecture 2

Reading (for this and last week):

- Agresti, Chapter 8

- Expected values, residuals, adjusted residuals in Stata
- Ordinal association
- Association in multi-way tables
- Multivariate causality

`tabchi` procedure allows access to

- Percentages
- Expected values
- Residuals
- Adjusted residuals

Ordinal association

- When variables are ordinal, association may be structured
- High values on X are associated with high values on Y , low with low
- Or vice versa for negative association
- Analogous to correlation
- Examine using percentages, adjusted residuals: ordered pattern

Example: row percentages

```
. tab lopfamo lopfaml, row
```

Key
<i>frequency</i>
<i>row percentage</i>

co-habiting is alright	divorce better than unhappy marriage					Total
	strongly	agree	neithr ag	disagree	stronglyd	
strongly agree	2,381 59.97	1,228 30.93	304 7.66	38 0.96	19 0.48	3,970 100.00
agree	1,462 22.75	4,159 64.72	687 10.69	103 1.60	15 0.23	6,426 100.00
neithr agree, disagr	485 15.69	1,803 58.33	717 23.20	73 2.36	13 0.42	3,091 100.00
disagree	156 12.86	647 53.34	252 20.77	143 11.79	15 1.24	1,213 100.00
stronglydisagree	78 15.57	143 28.54	129 25.75	101 20.16	50 9.98	501 100.00
Total	4,562 30.01	7,980 52.50	2,089 13.74	458 3.01	112 0.74	15,201 100.00

Example: observed and expected values

```
. tabchi lopfamo lopfaml
      observed frequency
      expected frequency
```

co-habiting is alright	divorce better than unhappy marriage				
	strongly agree	agree	neithr agree, disagr	disagree	stronglydisagree
strongly agree	2381 1191.444	1228 2084.113	304 545.578	38 119.614	19 29.251
agree	1462 1928.519	4159 3373.428	687 883.094	103 193.613	15 47.346
neithr agree, disagr	485 927.646	1803 1622.668	717 424.781	73 93.131	13 22.774
disagree	156 364.036	647 636.783	252 166.697	143 36.547	15 8.937
stronglydisagree	78 150.356	143 263.008	129 68.850	101 15.095	50 3.691

```
1 cell with expected frequency < 5
```

```
      Pearson chi2(16) = 4.2e+03   Pr = 0.000
      likelihood-ratio chi2(16) = 3.3e+03   Pr = 0.000
```

Example: adjusted residuals

```
. tabchi lopfamo lopfaml, adj noo
      expected frequency
      adjusted residual
```

co-habiting is alright	divorce better than unhappy marriage				
	strongly agree	agree	neithr agree, disagr	disagree	stronglydisagree
strongly agree	1191.444 47.925	2084.113 -31.654	545.578 -12.956	119.614 -8.815	29.251 -2.213
agree	1928.519 -16.713	3373.428 25.829	883.094 -9.351	193.613 -8.703	47.346 -6.210
neithr agree, disagr	927.646 -19.463	1622.668 7.277	424.781 17.104	93.131 -2.373	22.774 -2.303
disagree	364.036 -13.587	636.783 0.612	166.697 7.416	36.547 18.639	8.937 2.122
stronglydisagree	150.356 -7.173	263.008 -10.918	68.850 7.937	15.095 22.831	3.691 24.601

```
1 cell with expected frequency < 5
```

```
      Pearson chi2(16) = 4.2e+03   Pr = 0.000
      likelihood-ratio chi2(16) = 3.3e+03   Pr = 0.000
```

Measures of ordinal association

- Sometimes Pearson's Correlation is used
- Equivalent to scoring the categories linearly and calculating the conventional correlation

```
. corr lopfamo lopfam1  
(obs=15,201)
```

	lopfamo	lopfam1
lopfamo	1.0000	
lopfam1	0.3831	1.0000

- Assumption of equal intervals problematic (but often reasonably OK)
- Spearman's Rank Correlation is a better solution

```
. spearman lopfamo lopfam1
Number of obs = 15201
Spearman's rho = 0.3840
Test of H0: lopfamo and lopfam1 are independent
Prob > |t| = 0.0000
```

Truly ordinal measures

- The Gamma statistic (γ) is truly ordinal
- Counts “concordant” and “discordant” pairs

$$\gamma = \frac{C - D}{C + D}$$

- Range: -1, 0, 1
- Approximately normal for large samples

Gamma in practice

```
. tab lopfamo lopfam1, gamma
```

co-habiting is alright	divorce better than unhappy marriage					Total
	strongly	agree	neithr ag	disagree	stronglyd	
strongly agree	2,381	1,228	304	38	19	3,970
agree	1,462	4,159	687	103	15	6,426
neithr agree, disagr	485	1,803	717	73	13	3,091
disagree	156	647	252	143	15	1,213
stronglydisagree	78	143	129	101	50	501
Total	4,562	7,980	2,089	458	112	15,201

```
gamma = 0.4975 ASE = 0.009
```

- Gamma is symmetrical
- Kendall's tau (τ) is also symmetrical, similar logic
- Somer's d also uses $C + D$ but is asymmetrical: one variable affecting another (takes account of ties)

Multi-way tables

- How do we think in terms of multi-way tables – more than two dimensions?
- Often, in terms of whether the $A \times B$ relationship is constant across C

Scouting example

Scout	Delinquent		Total
	Yes	No	
Yes	36	364	400
No	60	340	400
Total	96	704	800

Scouting example

Low Church Attendance			
Scout	Delinquent		
	Yes	No	Total
Yes	10	40	50
No	40	160	200

Total	50	200	250
-------	----	-----	-----

Medium Church Attendance			
Scout	Delinquent		
	Yes	No	Total
Yes	18	132	150
No	18	132	150

Total	36	264	800
-------	----	-----	-----

High Church Attendance			
Scout	Delinquent		
	Yes	No	Total
Yes	8	192	200
No	2	48	50

Total	10	240	250
-------	----	-----	-----
