SO5032 Lecture 2

Brendan Halpin

January 29, 2023

Outline

SO5032 Lecture 2

SO5032 Lecture 2

Lecture 2

Lecture 2

Reading (for this and last week):

- Agresti, Chapter 8

Lecture 2

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

Tabular association in Stata

`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
frequency *row percentage*

co-habiting is alright	divorce better than unhappy marriage					
	strongly	agree	neithr ag	disagree	stronglyd	Total
strongly agree	2,381	1,228	304	38	19	3,970
	59.97	30.93	7.66	0.96	0.48	100.00
agree	1,462	4,159	687	103	15	6,426
	22.75	64.72	10.69	1.60	0.23	100.00
neithr agree, disagr	485	1,803	717	73	13	3,091
	15.69	58.33	23.20	2.36	0.42	100.00
disagree	156	647	252	143	15	1,213
	12.86	53.34	20.77	11.79	1.24	100.00
stronglydisagree	78	143	129	101	50	501
	15.57	28.54	25.75	20.16	9.98	100.00
Total	4,562	7,980	2,089	458	112	15,201
	30.01	52.50	13.74	3.01	0.74	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	1228	304	38	19
	1191.444	2084.113	545.578	119.614	29.251
agree	1462	4159	687	103	15
	1928.519	3373.428	883.094	193.613	47.346
neithr agree, disagr	485	1803	717	73	13
	927.646	1622.668	424.781	93.131	22.774
disagree	156	647	252	143	15
	364.036	636.783	166.697	36.547	8.937
stronglydisagree	78	143	129	101	50
	150.356	263.008	68.850	15.095	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	2084.113	545.578	119.614	29.251
	47.925	-31.654	-12.956	-8.815	-2.213
agree	1928.519	3373.428	883.094	193.613	47.346
	-16.713	25.829	-9.351	-8.703	-6.210
neithr agree, disagr	927.646	1622.668	424.781	93.131	22.774
	-19.463	7.277	17.104	-2.373	-2.303
disagree	364.036	636.783	166.697	36.547	8.937
	-13.587	0.612	7.416	18.639	2.122
stronglydisagree	150.356	263.008	68.850	15.095	3.691
	-7.173	-10.918	7.937	22.831	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
```

sociology
UNIVERSITY OF LIMERICK

Measures of ordinal association

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

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

	lopfamo	lopfaml
lopfamo	1.0000	
lopfaml	0.3831	1.0000

Non-linear correlation

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

```
. spearman lopfamo lopfaml

 Number of obs =   15201
Spearman´s rho =       0.3840

Test of H0: lopfamo and lopfaml 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 lopfaml, gamma
```

co-habiting is alright	divorce better than unhappy marriage					
	strongly	agree	neithr ag	disagree	stronglyd	Total
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

Variants

- 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)
- Good summary at
  `http://faculty.chass.ncsu.edu/garson/PA765/assocordinal.htm`

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		
	Yes	No	Total
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