Ordinal Variables in 2–way Tables

Edps/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS

Fall 2018

Outline

Inference for ordinal variables.

- Linear trend instead of independence.
- Greater power with ordinal test.
- Choosing scores for categories.
- Trend tests for $2 \times J$ and $I \times 2$ tables

Testing Linear Trend instead of Independence

Consider the example from the GSS where we had 2 items both with ordinal response options:

- Item 1: A working mother can establish just as warm and secure a relationship with her children as a mother who does not work.
- Item 2: Working women should have paid maternity leave.

	Item2					
Item 2	Strongly Agree	Agree	Neither	Disagree	Strongly Disagree	
Strongly Agree	97	96	22	17	2	234
Agree	102	199	48	38	5	392
Disagree	42	102	25	36	7	212
Strongly Disagree	9	18	7	10	2	46
	250	415	102	101	16	884

GSS Example

Statistic		*df*	Value	*p*-value
Pearson Chi-square	X^2	12	47.576	$< .001$
Likelihood Ratio Chi-square	G^2	12	44.961	$< .001$

There is a "**linear trend**" in these data, so we may be able to describe this relationship using a single statistic:

(Pearson Product Moment) **Correlation**

$$r = \frac{\text{cov}(X, Y)}{s_X s_Y}$$

To compute r, we need **scores** for both the row (item 1) categories and the column (item 2) categories.

Category Scores and r

- For the categories of the row variable X:

$$u_1 \leq u_2 \leq \ldots \leq u_I$$

- For the categories of the column variable Y:

$$v_1 \leq v_2 \leq \ldots \leq v_J$$

When the scores have the same order as the categories, they are "monotone".

Assume for now that we have scores. (we'll discuss possible choices and their effect later).

Given scores $\{u_i\}$ and $\{v_j\}$, the correlation equals. . .

The Correlation for an $(I \times J)$ Table

$$r = \frac{\text{cov}(X,Y)}{s_X s_Y} = \frac{\sum_i \sum_j (u_i - \bar{u})(v_j - \bar{v}) n_{ij}}{\sqrt{\left[\sum_i \sum_j (u_i - \bar{u})^2 n_{ij}\right]\left[\sum_i \sum_j (v_j - \bar{v})^2 n_{ij}\right]}}$$

where

- Row mean

$$\bar{u} = \sum_i \sum_j u_i n_{ij}/n = \sum_i u_i n_{i+}/n$$

- Column mean

$$\bar{v} = \sum_i \sum_j v_j n_{ij}/n = \sum_j v_j n_{+j}/n$$

Properties of r for Contingency Table Data

- $-1 \leq r \leq 1$
- $r = 0$ corresponds to no (linear) relationship.
- The further r is from 0, the greater the strength of the relationship.
- Perfect association implies that $r = \pm 1$.
- $r = 1$ if all observations fall into cells on the "diagonal" that runs from the top left to bottom right of the table.
  item $r = -1$ if all observations fall into cells on the "diagonal" that runs from the top right to bottom left of the table.

Testing Null Hypothesis of Independence

(i.e., no linear trend or $H_o : \rho = 0$)

Test statistic $$M^2 = (n-1)r^2$$

- "Mantel–Haenszel" or "Cochran–Mantel–Haenszel" statistic.
- As n increase, M^2 gets larger.
- As r^2 increases, M^2 gets larger.
- Under independence, $\rho = 0$, $M^2 = 0$.
- For perfect association, $M^2 = (n-1)$.
- Larger values of M^2 provide more evidence against H_O.
- If H_O of independence is true, then M^2 is approximately chi-square distributed with $df = 1$.
- $\sqrt{M^2} = \sqrt{(n-1)}r$ is approximately distributed at $\mathcal{N}(0,1)$, which can be used to test one-sided alternative hypotheses that the correlation is > 0 or < 0.

Example: Testing $H_o : \rho = 0$

Try integer (Likert) scores for our categories:

Rows	Response	Columns
$u_1 = 1$	Strongly Agree	$v_1 = 1$
$u_2 = 2$	Agree	$v_2 = 2$
	Neither	$v_3 = 3$
$u_3 = 3$	Disagree	$v_4 = 4$
$u_4 = 4$	Strongly Disagree	$v_5 = 5$

$$r = .203 \text{ and } M^2 = (884 - 1)(.203)^2 = 36.26$$

With $df = 1$, p–value for observed M^2 is $< .001$.

SAS INPUT to Compute M^2

- You must have two numeric variables, one for the rows ("row") and one for the columns ("col"), whose values are the scores.

```
DATA gss;
INPUT item1 $ item2 $ row col count;
DATALINES;
 strongagree   strongagree   1   1   97
 strongagree   agree         1   2   96
 ⋮              ⋮
 strongdis     strongdis     4   5    2
```

- For the TABLES command, use the numeric variables that contain the row and column scores.

```
PROC FREQ;
TABLES row*col / chisq measures;
```

SAS (continued)

In the output:

- "Mantel-Haenszel Chi-Square" is M^2.
- "Pearson correlation" is r.

R to Compute M^2 (and r)

Need the package vcdExtra...I think

```
# The GSS data in case form
 gss ← read.table("gss_data.txt",header=TRUE)

gss.tab ← xtabs(count ∼ fechld + mapaid, data=gss)

# Cochran-Mantel-Haenszel test of association
CMHtest(gss.tab, strata=NULL, rscores=1:4,
cscores=1:5, types="cor" )

# To get r, use the fact that M = (n − 1)r^2
n ← sum(gss.tab)
( r ← sqrt( 36.26132 /(n-1)) )
```

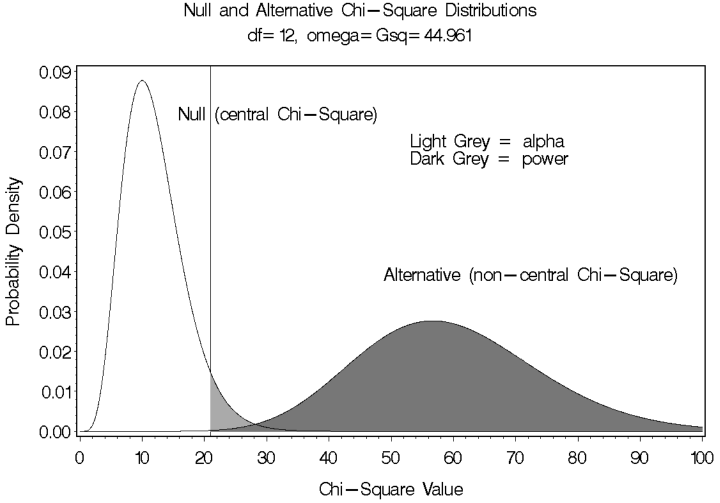
Extra Power with Ordinal Test

Statistic		*df*	Value	*p*-value
Pearson Chi-square	X^2	12	47.576	$< .001$
Likelihood Ratio Chi-square	G^2	12	44.961	$< .001$
Mantel-Haenszel Chi-square	M^2	1	36.261	$< .001$

- X^2 and G^2 are designed to detect any type association.
- M^2 is designed to detect a specific type of association.
- With ordinal data, we can summarize the association in terms of 1 parameter (i.e., r) rather than $(I-1)(J-1)$ of them (i.e., a set of $(I-1)(J-1)$ odds ratios).
- Advantages of M^2 over X^2 and G^2 when there is a positive or negative association between variables;
  - M^2 is more powerful.
  - M^2 tends to be about the same size as G^2 and X^2, but only has $df = 1$ rather than $df = (I-1)(J-1)$.
  - For small to moderate sample sizes, the true sampling distribution of the test statistics are better approximated for those with smaller df.

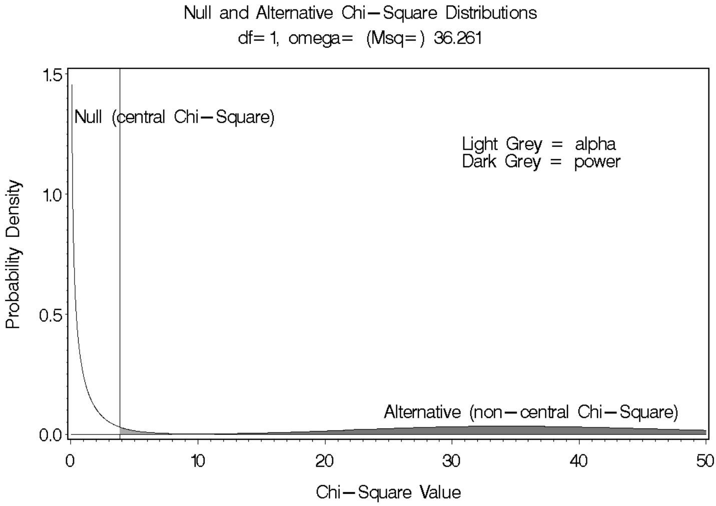
Power for Chi-square Tests: G^2

GSS data: For $G^2 = 44.961$, $df = 12 \longrightarrow$ power $= .99907$.



Power for M^2

For $M^2 = 36.261$, $df = 1 \longrightarrow$ power $= .99998$.



Computing Power

- π_{ij} = probabilities under the alternative model (which we'll take as the "saturated" model).
- π^*_{ij} = probabilities under the null hypothesis.
- N = total sample size.
- Note: $\mu_{ij}(= n_{ij}) = N\pi_{ij}$ and $m_{ij} = N\pi^*_{ij}$.
- "omega" (non-centrality parameter) for G^2

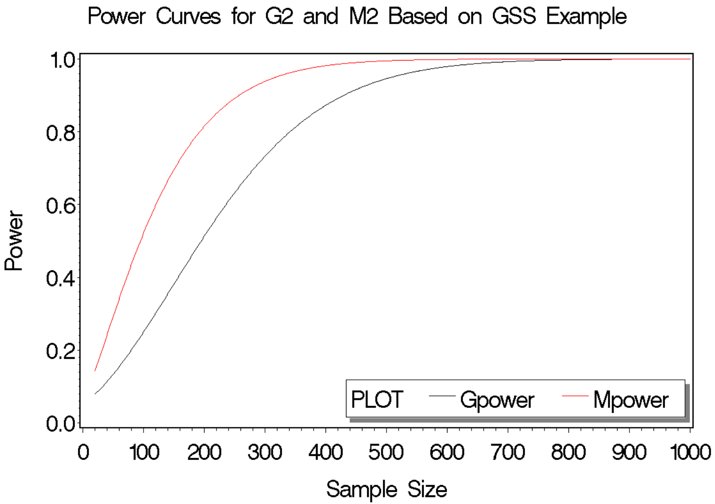
$$G^2 = 2N\sum_i\sum_j \pi_{ij}\log\frac{\pi_{ij}}{\pi^*_{ij}} = \omega$$

- "omega" for M^2

$$M^2 = (N-1)r^2 = \omega$$

- Sample Size and Power: $\uparrow N \Longrightarrow \uparrow \omega \Longrightarrow \uparrow$ Power

Power and Sample Size



Choice of Scores

- The choice of scores often does not make much difference with respect to the value of r and thus test results.
- For the GSS example, an alternative scoring that changed the relative spacing between the scores leads to an increase of r from .203 (from equal spacing) to .207 (from one possible choice for unequal spacing).
- The "best" scores for GSS table that lead to the largest possible correlation, yields $r = .210$. (Score from correspondence analysis).
- Different scoring tends to have a larger difference when the margins of the tables are unbalanced; that is, when there are some vary large margins and some relatively small ones.

Choice of Scores: Example 2

- Data from Farmer, Rotella, Anderson & Wardrop (1996) on gender differences in science careers. The data consist of a cross-classification of individuals by their gender and the prestige level of their occupation. (All subjects/individuals in this study had chosen a career in a science related field).

	Prestige Level of Occupation						
Gender	40–49	50–59	60–69	70–79	80–89	90–99	
Women	22	2	12	11	10	4	61
Men	3	0	11	6	25	7	52
	25	2	23	17	35	11	113

Statistic	DF	Value	Prob
Chi-Square	5	24.640	0.001
Likelihood Ratio Chi-Square	5	27.372	0.001
Mantel-Haenszel Chi-Square	1	19.840	0.001
Pearson Correlation		.421	

Different Possible Choices of Scores

- Equal Spacing. This is the SAS default.
- Midranks are a "no thought" approach to selecting scores.
  - Rank all observations on each variable and then use the ranks to compute the correlation — "Spearman's Rho" or the rank order correlation.
  - All individuals in the same category get the same rank, which equals the "midrank" for them.
  - e.g., Farmer et al data:

Category	Midrank/Score
40–49	$(1+25)/2 = 13.0$
50–59	$(26+27)/2 = 26.5$
60–69	$(28+50)/2 = 39.0$
70–79	$(51+67)/2 = 59.0$
80–89	$(68+102)/2 = 85.0$
90–99	$(103+113)/2 = 108.0$

  - In SAS to mid-ranks: PROC FREQ;
    TABLES row*col / cmh1 scores=ridits;

Different Possible Choices of Scores

- Midranks (continued)
  - In our example, different scores don't change our conclusion, if margins are really extreme (see example in Agresti), it can change results.
- Midpoints. When a categorical variable is a discretized numerical one, a good choice of scores often the midpoint.
  In our example, this leads to equal spacing.
- Use what you know about the data and your best guess as to what the relative spacing should be between the categories.
- Analytical method. Use row-column or "*RC*" association model or correspondence analysis.
- Try a few different ones to see if it makes a difference — a "sensitivity analysis".
- My preference: model the association.

Example and Results with Different Scores

Summary of Results for Farmer et al. using different scoring methods

Scoring	M^2	p	Pearson r	ASE
Midranks (Ridits)	19.142	$< .01$	.413	0.081
Equally spaced	19.840	$< .01$	.421	0.077
Unequal spacing*	18.281	$< .01$	.404	0.078
Unequal spacing†	21.664	$< .01$	.440	.076

$*$ Column scores were -4, -2, -1, 1, 2, and 4

† Column scores were -4, -3, -0.5, 0.5, 3, 4

Didn't really make much of a difference... now for one where scores do matter.

School of Psychiatric Thought

Wrong ordering of scores:

	SCHOOL	ORIGIN		
Scores		1	2	3
	Frequency	bio	env	comb
1	eclectic	90	12	78
2	medical	13	1	6
3	psychan	19	13	50

Statistic	DF	Value	Prob
Chi-Square	4	22.378	0.001
Likelihood Ratio Chi-Square	4	23.036	0.001
Mantel-Haenszel Chi-Square	1	10.736	0.001
Pearson Correlation		0.195 (ASE=0.056)	

A Better Ordering of Categories

Uniform Scores for row and column with good ordering:

		bio	env	comb	
	Frequency	1	3	2	Total
eclectic	2	90	12	78	180
medical	1	13	1	6	20
psychan	3	19	13	50	82

Statistic	DF	Value	Prob
Chi-Square	4	22.378	0.001
Likelihood Ratio Chi-Square	4	23.036	0.001
Mantel-Haenszel Chi-Square	1	20.260	0.001
Pearson Correlation		0.269 (ASE=0.056)	

A Better Ordering and Scores: *RC* Model

Scale values from RC association model (scores are estimated from the data):

Statistic	DF	Value	Prob
Chi-Square	4	22.378	0.001
Likelihood Ratio Chi-Square	4	23.036	0.001
Mantel-Haenszel Chi-Square	1	22.042	0.001

Statistic		Value	ASE
Pearson Correlation	0.280	0.055	

Trend Tests

Situation: the row variable X is an explanatory variable and the column variable Y is a response/outcome variable.

- When one variable just has two levels (e.g., Farmer et al), we can assign the categories any two distinct values, e.g., 0 and 1, -1 and 1, 0 and 5000 — the choice does not effect r.
- Binary X: (i.e, $u_1 = 0$ and $u_2 = 1$) and polytomous ordinal Y with scores $v_1, \ldots, v_J$.
- The term in the covariance $\sum_i \sum_j u_i v_j n_{ij}$ between X and Y simplifies to
$$\sum_i \sum_j u_i v_j n_{ij} = \sum_j v_j n_{2j}$$
- When this is divided by the number of individuals in the 2nd row, we get
$$\bar{v}(i = 2) = \sum_j v_j n_{2j}/n_{2+}$$
- So, testing a linear trend in this case is the same as testing

Trend Test for $2 \times J$ Tables

- Testing a linear trend in this case is the same as testing whether the mean on Y is the same or different for the two rows.
- When midranks are used, the test for linear trend using M^2 is the same as the *Wilcoxon* and *Mann-Whitney* non-parametric tests for mean differences.
- Now for the other case... $I \times 2$ Tables.

Trend Test for $I \times 2$ Tables

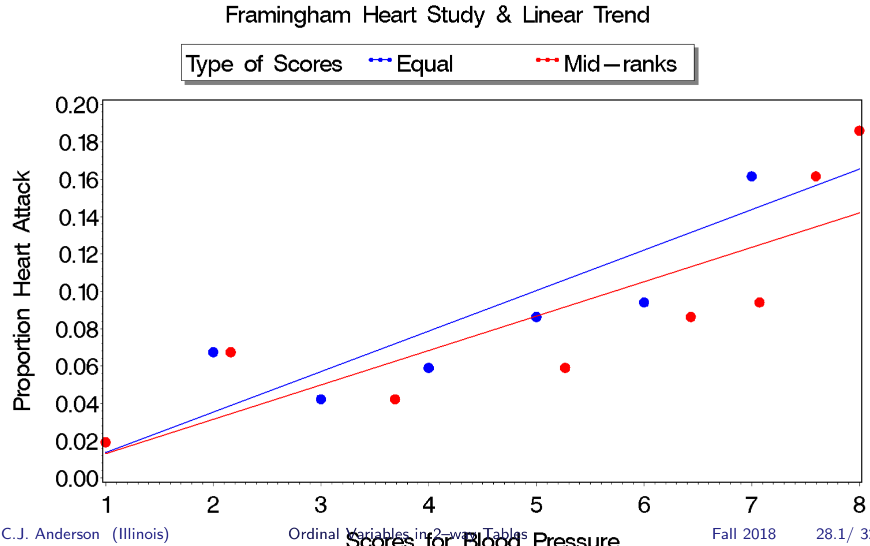
Situation: Polytomous ordinal X with scores $u_1, \ldots, u_I$ and binary Y ($v_1 = 0$ and $v_2 = 1$).

- This test detects whether the proportion classified as (for example) Y_1 increases (or decreases) linearly with X.
- Cochran–Armitage trend test is the $I \times 2$ version of M^2. You can specify choice of scores (SAS default: scores=table).
- Example: The Framingham heart study from Cornfield (1962). 40–59 year old males from Framingham, MA were classified on several factors. At a 6 year follow-up,

Blood pressure	Heart disease Present	(%)	Absent	Total
< 117	3	(.02)	153	156
117–126	17	(.07)	235	252
127–136	12	(.04)	272	284
137–146	16	(.06)	255	271
147–156	12	(.09)	127	139
157–166	8	(.09)	77	85
167–186	16	(.16)	83	99
> 186	8	(.19)	35	43

- Is there is significant linear trend?

Look at the Data



Final Comments: Cochran–Armitage Trend Test

- Cochran–Armitage trend test is analogous to testing the slope in a linear (probability) regression model:

$$\pi_i = \alpha + \beta(\text{category score})_i + \epsilon_i$$

- Cochran–Armitage trend test is the "score test" for β.
- Let $z \sim \mathcal{N}(0, 1)$,

$$\chi^2(\text{independence}) = z^2 + \chi^2(\text{lack of linear trend}).$$

The Cochran–Armitage trend test statistic equals z.

SAS

The data

```
data frame;
input bp $ heart $ count bpguess @@ ;
label bp='Blood Pressure'
heart='Heart Disease Present?';
cards;
```

< 117	yes	3	1	< 117	no	153	1
117 – 126	yes	17	2	117 – 126	no	235	2
127 – 136	yes	12	3	127 – 136	no	272	3
137 – 146	yes	16	4	137 – 146	no	255	4
147 – 156	yes	12	5	147 – 156	no	127	5
157 – 166	yes	8	5.5	157 – 166	no	77	5.5
167 – 186	yes	16	8	167 – 186	no	83	8
> 186	yes	8	10	> 186	no	35	10

SAS continued

```
title 'I X 2 linear trend test -- Cochran-Armitage
(equally spaced scores)';
proc freq order=data; weight count;
tables heart*bp /chisq nopercent norow trend ;

title 'I X 2 linear trend test -- Cochran-Armitage
(scores=midranks)';
proc freq order=data; weight count;
tables heart*bp /chisq nopercent norow trend
score=ridit;
run;

title 'I X 2 linear trend test -- Cochran-Armitage
(crude guess of scores)';
proc freq order=data; weight count;
tables heart*bpguess /chisq nopercent norow trend ;
```

R

```
# Needed for Cochran-Armitage trend test
library(DescTools)

# Read in data as data frame
hs ←
read.table("framingham_heart_data.txt",header=TRUE)

# Need table data for the test
hs.tab ← xtabs(count ~ bp + heart,data=hs)

CochranArmitageTest(hs.tab, alternative =
c("two.sided", "increasing", "decreasing"))
```