Statistical Inferences for Ordinal Variables in 2–way Tables
Edpsy/Psych/Soc 589

Carolyn J. Anderson

Department of Educational Psychology

© Board of Trustees, University of Illinois

Spring 2017
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.

<table>
<thead>
<tr>
<th>Item 2</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Neither</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strongly Agree</td>
<td>97</td>
<td>96</td>
<td>22</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Agree</td>
<td>102</td>
<td>199</td>
<td>48</td>
<td>38</td>
<td>5</td>
</tr>
<tr>
<td>Disagree</td>
<td>42</td>
<td>102</td>
<td>25</td>
<td>36</td>
<td>7</td>
</tr>
<tr>
<td>Strongly Disagree</td>
<td>9</td>
<td>18</td>
<td>7</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>415</td>
<td>102</td>
<td>101</td>
<td>16</td>
</tr>
</tbody>
</table>
GSS Example

<table>
<thead>
<tr>
<th>Statistic</th>
<th>df</th>
<th>Value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-square</td>
<td>12</td>
<td>47.576</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-square</td>
<td>12</td>
<td>44.961</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

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.
- $r = -1$ 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 $H_0 : \rho = 0$)

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 $H_0$.
- If $H_0$ 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 $H_0 : \rho = 0$

Try integer (Likert) scores for our categories:

<table>
<thead>
<tr>
<th>Rows</th>
<th>Response</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_1 = 1$</td>
<td>Strongly Agree</td>
<td>$v_1 = 1$</td>
</tr>
<tr>
<td>$u_2 = 2$</td>
<td>Agree</td>
<td>$v_2 = 2$</td>
</tr>
<tr>
<td></td>
<td>Neither</td>
<td>$v_3 = 3$</td>
</tr>
<tr>
<td>$u_3 = 3$</td>
<td>Disagree</td>
<td>$v_4 = 4$</td>
</tr>
<tr>
<td>$u_4 = 4$</td>
<td>Strongly Disagree</td>
<td>$v_5 = 5$</td>
</tr>
</tbody>
</table>

$r = .203$ 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.

```sas
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.

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

On the output:
Extra Power with Ordinal Test

<table>
<thead>
<tr>
<th>Statistic</th>
<th>$df$</th>
<th>Value</th>
<th>$p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-square</td>
<td>12</td>
<td>47.576</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-square</td>
<td>12</td>
<td>44.961</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-square</td>
<td>1</td>
<td>36.261</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

- $X^2$ 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 $df$.  

C.J. Anderson (Illinois)  
Ordinal Variables in 2-way Tables  
Spring 2017  
11.1/28
Power for Chi-square Tests: $G^2$

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

Null and Alternative Chi-Square Distributions

$df = 12$, $omega = Gsq = 44.961$

Null (central Chi-Square)

Light Grey = alpha
Dark Grey = power

Alternative (non-central Chi-Square)
Power for $M^2$

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

Null and Alternative Chi-Square Distributions 
$df=1$, $\omega = (M^2) = 36.261$

![Graph showing null and alternative chi-square distributions with power indicated.](image-url)
Computing Power

- $\pi_{ij}$ = probabilities under the alternative model (which we’ll take as the “saturated” model).
- $\pi^*_{ij}$ = 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: $\uparrow N \implies \uparrow \omega \implies \uparrow$ Power
Power and Sample Size

Power Curves for G2 and M2 Based on GSS Example

![Graph showing power curves for G2 and M2 based on GSS Example.](image)
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).

<table>
<thead>
<tr>
<th>Gender</th>
<th>Prestige Level of Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40–49</td>
</tr>
<tr>
<td>Women</td>
<td>22</td>
</tr>
<tr>
<td>Men</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>5</td>
<td>24.640</td>
<td>0.001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>5</td>
<td>27.372</td>
<td>0.001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>19.840</td>
<td>0.001</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>.421</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

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

- e.g., Farmer et al data:

- 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

<table>
<thead>
<tr>
<th>Scoring</th>
<th>$M^2$</th>
<th>$p$</th>
<th>Pearson $r$</th>
<th>ASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Midranks (Ridits)</td>
<td>19.142</td>
<td>&lt; .01</td>
<td>.413</td>
<td>0.081</td>
</tr>
<tr>
<td>Equally spaced</td>
<td>19.840</td>
<td>&lt; .01</td>
<td>.421</td>
<td>0.077</td>
</tr>
<tr>
<td>Unequal spacing*</td>
<td>18.281</td>
<td>&lt; .01</td>
<td>.404</td>
<td>0.078</td>
</tr>
<tr>
<td>Unequal spacing†</td>
<td>21.664</td>
<td>&lt; .01</td>
<td>.440</td>
<td>0.076</td>
</tr>
</tbody>
</table>

* Column scores were $-4, -2, -1, 1, 2,$ and $4$
† Column scores were $-4, -3, -0.5, 0.5, 3,$ and $4$

Didn’t really make much of a difference...now for one where scores do matter.
School of Psychiatric Thought

Wrong ordering of scores:

<table>
<thead>
<tr>
<th>Scores</th>
<th>Frequency</th>
<th>SCHOOL</th>
<th>ORIGIN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>eclectic</td>
<td>1 bio</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>90</td>
<td>12</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>19</td>
<td>13</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>4</td>
<td>22.378</td>
<td>0.001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>4</td>
<td>23.036</td>
<td>0.001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>10.736</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Pearson Correlation 0.195 (ASE=0.056)
A Better Ordering of Categories

Uniform Scores for row and column with good ordering:

<table>
<thead>
<tr>
<th>Frequency</th>
<th>bio</th>
<th>env</th>
<th>comb</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>eclectic</td>
<td>2</td>
<td>90</td>
<td>12</td>
<td>78</td>
</tr>
<tr>
<td>medical</td>
<td>1</td>
<td>13</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>psychan</td>
<td>3</td>
<td>19</td>
<td>13</td>
<td>50</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>4</td>
<td>22.378</td>
<td>0.001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>4</td>
<td>23.036</td>
<td>0.001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>20.260</td>
<td>0.001</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td></td>
<td>0.269 (ASE=0.056)</td>
<td></td>
</tr>
</tbody>
</table>
A Better Ordering and Scores: *RC* Model

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

<table>
<thead>
<tr>
<th>Statistic</th>
<th>DF</th>
<th>Value</th>
<th>Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>4</td>
<td>22.378</td>
<td>0.001</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>4</td>
<td>23.036</td>
<td>0.001</td>
</tr>
<tr>
<td>Mantel-Haenszel Chi-Square</td>
<td>1</td>
<td>22.042</td>
<td>0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>ASE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>0.280</td>
<td>0.055</td>
</tr>
</tbody>
</table>
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
  \[
  \bar{v}(i = 2) = \sum_j v_j n_{2j}/n_{2+}
  \]
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,

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

- Is there is significant linear trend?
Look at the Data

Framingham Heart Study & Linear Trend

Type of Scores

- Equal
- Mid-ranks

Proportion Heart Attack

C.J. Anderson (Illinois)

Ordinal Variables in 2-way Tables

Spring 2017
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 \( \beta \).

- Let \( z \sim N(0, 1) \),

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

The Cochran–Armitage trend test statistic equals \( z \).