Lecture 9: Ordinal Associations of $I \times J$ Contingency Tables

Dipankar Bandyopadhyay, Ph.D.

BMTRY 711: Analysis of Categorical Data Spring 2011

Division of Biostatistics and Epidemiology

Medical University of South Carolina

Motivation for the Lecture

Recall, in this course we planned to discuss

1. Nominal variables
2. Ordinal variables
3. Discrete interval variables having relatively few values
4. Continuous variables grouped into a small number of categories

Up to this point in time, we have largely concentrated on Nominal variables

Loss of Power

- In some cases, we have ignored the natural ordering of the data.
- This, as previously stated, “throws away” data and reduces the power.
- We will want to develop methods that can detect Monotone Trends in the data
  1. Monotonically Increasing: As levels of X increase, the levels of the response, Y, increases
  2. Monotonically Decreasing: As levels of X increase, the levels of the response, Y, decreases
- We want to develop a single measure, similar to a correlation, that summarizes these trends.

Introduction to Examples

- We are going to look at two examples today.
- The first example is the Cross-Classification of Job Satisfaction by Income that we used when we first introduced gamma
- The second example is a study in which we will compare drug toxicities by drug dose

Job Satisfaction Example

Cross-Classification of Job Satisfaction by Income

	Job Satisfaction			
	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
< $15,000	1	3	10	6
$15,000 - $25,000	2	3	10	7
$25,000 - $40,000	1	6	14	12
> $40,000	0	1	9	11

We want to summarize how job satisfaction and income relate.

Our general hypothesis may be that higher pay relates to higher job satisfaction.

Drug Toxicity Example

Drug Toxicity by Dose Level				
	Toxicity			
Dose	None	Mild	Severe	Total
1	92	6	2	100
10	90	7	3	100
100	89	8	3	100
1000	88	3	9	100
Total	359	24	17	400

Our general hypothesis is that higher doses of the drug are related to increased level of toxicity.

Recall the Gamma Statistic

The estimator of gamma is based only on the number of concordant and discordant pairs of observations. It ignores tied pairs. Recall the Definitions:

1. A pair of subjects is *Concordant* if the subject ranked higher on X and also ranks higher on Y (or ranks lower on X and ranks lower on Y).
2. Denote C as the number of concordant pairs
3. A pair of subjects is *Discordant* if the subject ranked higher on X but ranks lower on Y (or ranked lower on X but ranks higher on Y)
4. Denote D as the number of discordant pairs
5. The pair is *tied* if both rank the same on X and/or Y

Then, Gamma (Goodman and Kruskal 1954) is defined as

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

Recall the analysis of the Job Satisfaction Data

```
proc freq;
 tables i*j/measures;
 weight count;
 run;
```

Selected Results

```
Statistic                       Value       ASE
-------------------------------------------------
Gamma                          0.2211    0.1172 <--- Our result
```

$\widehat{\gamma} = 0.2211$ with SE =0.1172, so an approximately 95% confidence interval can be calculated as

$$CI_{95\%} = 0.2211 \pm 1.96(0.1172) = (-0.0086, 0.4508)$$

Therefore at the $\alpha = 0.05$ level, there is insufficient evidence to support the hypothesis that a linear trend exists in the data.

This analysis ignored the relative differences in the categories and only exploited the ordinal nature by considering the concordant and discordant nature of the data.

We will want to extend our methods to better describe the data.

Linear Association and Correlations

- Under the assumption of a monotonic trend, one could consider a correlation measure.
- Since Ordinal Scales by definition do not have an underlying metric, we will need to assign, arbitrary values (called SCORES) to the ordinal categories
- We will examine several different methods of assigning scores
- In general, you want the score values to be similar to the hypothesized underlying metric
- For example,
  Soft drinks at your favorite Fast Food restaurant may come in a medium, large and super size. The volume in fluid ounces would be a reasonable metric.

Example: Clinical Trial for Toxicity

Drug Toxicity by Dose Level				
	Toxicity			
Dose	None	Mild	Severe	Total
1	92	6	2	100
10	90	7	3	100
100	89	8	3	100
1000	88	3	9	100
Total	359	24	17	400

- Give 100 animals (this is an unethical design in humans) different doses of a new drug, and see the toxicities at the different doses:
- The rows are numerically ordered, and the columns are not, but ‘none’ is better than ‘Mild’, which is better than ‘Severe’.
- Even though the rows are Interval, we can consider both the rows and columns are ordinal.

Correlation Coefficient after Assigning Scores

- Suppose both the rows and columns are ordinal.
- We want to assign 'scores' to the rows and columns, and then take the CORRELATION COEFFICIENT between the scores.
- We denote the assigned row scores by $(s_1, s_2, ..., s_I)$:
  the scores are such that

$$s_1 < s_2 < \cdots < s_I \text{ (or } s_1 > s_2 > \cdots > s_I)$$

- The column scores are denoted by $(u_1, u_2, ..., u_J)$: where

$$u_1 < u_2 < \cdots < u_J \text{ (or } u_1 > u_2 > \cdots > u_J)$$

The SCORES could be:

1. The actual numerical value (e.g., dose)
2. The midpoint of the interval if X is a crude grouping of an underlying continuous variable, such as AGE with three levels [20,30), [30,40), [40,50), in which scores 25,35,45 are assigned.
3. If the variable is ordinal, but not numerical, such as

   (NONE, MILD, SEVERE),

   investigators often just use (1,2,3) or scores that capture the relative weight such as (1, 2, 5).
4. Or, the ranks of the observations (which is similar to the Wilcoxon rank sum test)

Example

Drug Toxicity by Dose Level				
	Toxicity			
Dose	None	Mild	Severe	Total
1	92	6	2	100
10	90	7	3	100
100	89	8	3	100
1000	88	3	9	100
Total	359	24	17	400

- Looking at the table, we expect the toxicity to increase as the dose increases.
- Then, we would expect higher doses to correspond to higher toxicity.
- Then, we assign the row scores

$$(s_1, s_2, s_3, s_4) = (1, 10, 100, 1000)$$

and the column scores

$$(u_1, u_2, u_3) = (1, 2, 3)$$

Correlation Coefficient

- To describe the correlation coefficient, suppose we let let S_i be the row score for the cell that the i^{th} individual is in, and U_i be the column score for the cell that the i^{th} individual is in, $(i = 1, ..., n)$, where $n = y_{\cdot\cdot}$
- For each subject, we have the pair of random variables

$$(S_i, U_i), \quad i = 1, ..., n$$

- For the above toxicity example, the scores:

Drug Toxicity by Dose Level Grouped by Scores			
		Toxicity	
Dose	None ($U = 1$)	Mild ($U = 2$)	Severe ($U = 3$)
1 ($S = 1$)	(1,1)	(1,2)	(1,3)
10 ($S = 10$)	(10,1)	(10,2)	(10,3)
100 ($S = 100$)	(100,1)	(100,2)	(100,3)
1000 ($S = 1000$)	(1000,1)	(1000,2)	(1000,3)

- In general, the true correlation between S_i and U_i is

$$\begin{aligned}\rho &= Corr(S_i, U_i) \\ &= \frac{E\{[S_i - E(S_i)][U_i - E(U_i)]\}}{\sqrt{Var(S_i)Var(U_i)}} \\ &= \frac{E(S_iU_i) - E(S_i)E(U_i)}{\sqrt{Var(S_i)Var(U_i)}}\end{aligned}$$

- For any set of scores (even any set that doesn't make sense), under independence of row and column,

$$E(S_iU_i) = E(S_i)E(U_i)$$

so that

$$\rho = 0$$

- Thus, if we want to test for independence of row and column, we can use the sample correlation coefficient:

$$\widehat{\rho} = \frac{\sum_{i=1}^{n}(S_i - \bar{S})(U_i - \bar{U})}{\sqrt{\sum_{i=1}^{n}(S_i - \bar{S})^2 \sum_{i=1}^{n}(U_i - \bar{U})^2}}$$

where

$$\bar{S} = n^{-1}\sum_{i=1}^{n} S_i \qquad \bar{U} = n^{-1}\sum_{i=1}^{n} U_i$$

are the sample means.

- Note, in terms of the cell counts Y_{ij} and the row and column scores $(s_1, s_2, ..., s_I)$ and $(u_1, u_2, ..., u_J)$, we can also write the sample correlation coefficient as

$$\widehat{\rho} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} [(s_i - \bar{S})(u_j - \bar{U}) Y_{ij}]}{\sqrt{\sum_{i=1}^{I} Y_{i\cdot}(s_i - \bar{S})^2 \sum_{j=1}^{J} Y_{\cdot j}(u_j - \bar{U})^2}}$$

where the the sample means are the same as before, but can be calculated as

$$\bar{S} = n^{-1} \sum_{i=1}^{I} Y_{i\cdot} s_i \qquad \bar{U} = n^{-1} \sum_{j=1}^{J} Y_{\cdot j} u_j$$

Test Statistic Based on Correlation Coefficient

- Suppose we want to test the null of independence, i.e.,

$$H_0 : \text{independence}$$

versus one sided alternatives

$$H_A : \rho > 0$$

or

$$H_A : \rho < 0,$$

- A valid test statistic is the usual test statistic for testing that the correlation coefficient equals 0, as if the data are bivariate normal:

$$T = \frac{\widehat{\rho}\sqrt{n-2}}{\sqrt{1-\widehat{\rho}^2}} \sim t_{n-2}$$

which is approximately distributed as a t with $(n-2)$ degrees-of-freedom. For large n, people sometimes assume $T \sim N(0,1)$

- For a two sided alternative of non-independence, one can use

$$T^2 = \frac{(n-2)\widehat{\rho}^2}{1-\widehat{\rho}^2} \sim \chi_1$$

  a chi-square with 1 df.

- A variant of this test statistic sets ρ equal to its value under the null of independence $(\rho = 0)$ in the denominator:

$$X^2 = (n-2)\widehat{\rho}^2 \sim \chi_1$$

- In fact, the 'Mantel-Haenzel' (MH) statistic for testing for no association between ordinal rows and columns is

$$M^2 = X^2_{MH} = \frac{n-1}{n-2} X^2 = (n-1)\widehat{\rho}^2 \sim \chi_1$$

- For large n, there is practically no difference between X^2 and X^2_{MH}.
- Also, if there are only two rows and two columns, and we use the scores 1,2 for the rows, and 1,2 for the columns, one can show that Pearson's chi-square equals:

$$X^2_P = n\widehat{\rho}^2$$

Actually, as we will see below, any scores for the rows and columns of a (2×2) table will give the same test statistic.

Thoughts about choosing scores

- Suppose instead of using the scores (s_i, u_j), we choose a new set of scores in which we multiply the old scores by a constant, and add a constant, i.e.,

$$s_i^* = a + bs_i,$$

and

$$u_j^* = c + du_j,$$

- Note, this is a linear transformation of the score.
- Since the correlation coefficient is the same for any transformation like this, we get the same test statistic that we got with the original scores s_i and u_j.
- A valid question is:
- Will any set of scores be OK ?
- The answer is, since, under the null of independence, $\rho = 0$ for any set of scores, any set of scores will give you a valid test (Type I error OK under the null).

- However, some scores are more powerful to detect departures from independence than others.
- In particular, suppose there is a set of true ‘scores’ out there that describe the real relationship between row and column.
- The closer we choose our ‘scores’ to the true scores, the higher the power.
- If we choose a set of scores that are really off base, say, we choose linear scores, when the relationship is quadratic, our test statistic will still have the correct Type I error (and thus, be valid), but could have 0 power.
- If you choose a set of scores, and get a significant result, then this suggests that row and column are not independent. However, you shouldn’t keep choosing a lot of different scores until you get a significant result.
- Agresti’s quotes Cochran (1954) as saying “any set of scores gives a *valid* test, provided that they are constructed without consulting the results of the experiment”.

Example

Drug Toxicity by Dose Level				
	Toxicity			
Dose	None	Mild	Severe	Total
1	92	6	2	100
10	90	7	3	100
100	89	8	3	100
1000	88	3	9	100
Total	359	24	17	400

Questions of Interest:

1. Does toxicity increase as a result of dose ?
2. Are toxicity and dose independent?

SAS Proc Freq

```
data one;
  input dose tox count;
  dose2=1+log10(dose) ;
  tox2 = tox;
  if tox=3 then tox2=10;
cards;
   1 1  92
   1 2   6
   1 3   2
  10 1  90
  10 2   7
  10 3   3
 100 1  89
 100 2   8
 100 3   3
1000 1  88
1000 2   3
1000 3   9
;
```

Listing of Data

dose	dose2	tox	tox2	count
1	1	1	1	92
1	1	2	2	6
1	1	3	10	2
10	2	1	1	90
10	2	2	2	7
10	2	3	10	3
100	3	1	1	89
100	3	2	2	8
100	3	3	10	3
1000	4	1	1	88
1000	4	2	2	3
1000	4	3	10	9

```
PROC FREQ;
 TABLE (dose dose2)*(tox tox2)/ chisq measures;
 WEIGHT count;
run;
```

Note that this code will produce 4 tables.

1. dose * tox
2. dose2 * tox
3. dose * tox2
4. dose2 * tox2

with

Table	Row Scores (s_1, s_2, s_3, s_4)	Column Scores (u_1, u_2, u_3)
dose * tox	(1, 10, 100, 1000)	(1,2,3)
dose2 * tox	(1, 2, 3, 4)	(1,2,3)
dose * tox2	(1, 10, 100, 1000)	(1,2,10)
dose2 * tox2	(1, 2, 3, 4)	(1,2,10)

Results: DOSE BY TOX (1, 10, 100, 1000) $\times$ (1,2,3)

```
/* SELECTED OUTPUT */

Statistics for Table of DOSE BY TOX

Statistic                      DF       Value      Prob
------------------------------------------------------
Chi-Square                      6      9.6661    0.1394
Likelihood Ratio Chi-Square     6      9.0101    0.1730
Mantel-Haenszel Chi-Square      1      2.8136    0.0935***



Statistic                                Value       ASE
-------------------------------------------------------
Pearson Correlation                     0.0840    0.0572
```

Note: $2.8136 = (400 - 1)(0.0840)^2$

Results: DOSE2 BY TOX (1, 2, 3, 4) $\times$ (1,2,3)

Statistics for Table of dose2 by tox

Statistic	DF	Value	Prob
Chi-Square	6	9.6661	0.1394
Likelihood Ratio Chi-Square	6	9.0101	0.1730
Mantel-Haenszel Chi-Square	1	2.7590	0.0967***

Statistics for Table of dose2 by tox

Statistic	Value	ASE
Pearson Correlation	0.0832	0.0503

Results: DOSE BY TOX2 (1, 10, 100, 1000) $\times$ (1,2,10)

```
Statistics for Table of dose by tox2

Statistic                       DF       Value      Prob
------------------------------------------------------------
Chi-Square                       6      9.6661    0.1394
Likelihood Ratio Chi-Square      6      9.0101    0.1730
Mantel-Haenszel Chi-Square       1      6.4410    0.0112***

Statistics for Table of dose by tox2

Statistic                                Value       ASE
------------------------------------------------------------
Pearson Correlation                     0.1271    0.0589
```

Results: DOSE2 BY TOX2 (1, 2, 3, 4) $\times$ (1,2,10)

Statistics for Table of dose2 by tox2

Statistic	DF	Value	Prob
Chi-Square	6	9.6661	0.1394
Likelihood Ratio Chi-Square	6	9.0101	0.1730
Mantel-Haenszel Chi-Square	1	4.9414	0.0262***

Statistics for Table of dose2 by tox2

Statistic	Value	ASE
Pearson Correlation	0.1113	0.0501

SAS Proc Corr

```
proc corr data=one;
 var tox tox2 ;
 with dose dose2;
 freq count;
run;
```

```
Pearson Correlation Coefficients, N = 400
        Prob > |r| under H0: Rho=0

                 tox          tox2

dose         0.08397       0.12705
              0.0935        0.0110

dose2        0.08315       0.11128
              0.0968        0.0260
```

Tests of Independence for different scores

	SCORES		$H_0: \rho = 0$	
			$p-$value	
	(s_1, s_2, s_3, s_4)	(u_1, u_2, u_3)	MH	$T-$stat
(1)	(1,2,3,4)	(1,2,3)	0.0967	.0968
(2)	(1,2,3,4)	(1,2,10)	0.0262	.0260
(3)	(1,10,100,1000)	(1,2,3)	0.0935	.0935
(4)	(1,10,100,1000)	(1,2,10)	0.0112	.0110

Note, you can 'lie' with statistics, using different scores.

Here, if we artificially consider a severe toxicity as 10 times worse than no toxicity and 5 times worse than a mild toxic event, then we can show significance.

However, we would need to justify this assumption (*a priori*).

Scores Based on other Measures

Lets revisit the Job Satisfaction Data and consider using the average for the salary classifications.

Cross-Classification of Job Satisfaction by Income

	Job Satisfaction			
	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
< $15,000	1	3	10	6
$15,000 - $25,000	2	3	10	7
$25,000 - $40,000	1	6	14	12
> $40,000	0	1	9	11

- The selection of the scores of $< \$15,000$ and $> \$40,000$ is arbitrary unless you have the raw (un-aggregated data)
- Suppose the median salary for the $< \$15,000$ group is $\$10,000$. Then, we could set the scores for $< \$15,000$ at $\$10,000$.
- Similarly, we could set the scores for $> \$40,000$ at $\$60,000$.

```
data two;
 input salary satisfaction count;
 salary2 = salary /1000 - 10;
cards;
10000 1 1
10000 2 3
10000 3 10
10000 4 6
20000 1 2
20000 2 3
20000 3 10
20000 4 7
32500 1 1
32500 2 6
32500 3 14
32500 4 12
60000 1 0
60000 2 1
60000 3 9
60000 4 11
;
run;
```

SAS PROC Statements

```
proc freq data=two;
 tables salary*satisfaction /chisq measures;
 weight count;
run;


proc corr data=two;
 FREQ count;
 var satisfaction;
 with salary salary2;
run;
```

Selected Results

Statistics for Table of salary by satisfaction

Statistic	DF	Value	Prob	
Chi-Square	9	5.9655	0.7434	<-|Ignores Order
Likelihood Ratio Chi-Square	9	6.7641	0.6617	<-|
Mantel-Haenszel Chi-Square	1	3.9401	0.0471	<-Our result

Statistic	Value	ASE	
Gamma	0.2211	0.1172	<- Same as before, where scores are not used in calculation (Page 8)

Empirical Evidence of Linear Transformation

```
                                  Simple Statistics

Variable                   N           Mean       Std Dev

salary                    96          30964         17537
salary2                   96       20.96354      17.53707
satisfaction              96        3.15625       0.81212


Pearson Correlation Coefficients, N = 96 <---Note: Check N to make
                                            sure you have FREQ
       Prob > |r| under H0: Rho=0           specified correctly

              satisfaction

salary             0.20365  <---|
                    0.0466      |Note: these are equal
                                |      as expected
salary2            0.20365  <---|
                    0.0466
```

Additional Concepts of Scoring

- Previously, our selection of the scores was relatively arbitrary.
- For example, Assigning a score of 3 or 10 to ‘severe’ toxicity may be hard to justify.
- Using a Pearson’s correlation on these scores may not be valid since the scores may not be multivariate normal
- We should be interested in developing some nonparametric equivalents
- The easiest approach is to implement a non-parametric approach to score selection

Mid-ranks and Ridits

- A non-parametric approach to score selection is to consider the Mid-ranks of each row (or column)
- The mid-ranks are calculated as
  For ROWS
  $$R1_i = \sum_{k<i} n_{k\cdot} + (n_{i\cdot} + 1)/2 \quad i = 1, \ldots, I$$
  and
  FOR COLUMNS
  $$C1_i = \sum_{l<j} n_{\cdot l} + (n_{\cdot j} + 1)/2 \quad j = 1, \ldots, J$$
- Note, these represent the midpoint of each row when you rank the observations (sort them) and assign 1 to the first record, 2 to the second, etc.
- Thus, for the first column, the ranks would range from 1 to $n_{i\cdot}$ with the mean (or mid) being $(1 + n_{i\cdot})/2$.

Ridits and Modified Ridits

- One of the more common types of Scores is the “Ridits” scores (Bross 1958)
- Ridits are standardized midranks
- To calculate the ridits,
  FOR ROWS

$$R2_i = R1_i/n$$

  and
  FOR COLUMNS

$$C2_i = C1_i/n$$

- Modified Ridits divide by $(n+1)$

Cross-Classification of Job Satisfaction by Income				
	Job Satisfaction			
	Very Dissatisfied	Little Dissatisfied	Moderately Satisfied	Very Satisfied
< $15,000	1	3	10	6
$15,000 - $25,000	2	3	10	7
$25,000 - $40,000	1	6	14	12
> $40,000	0	1	9	11

For this data,

$$
\begin{aligned}
R1_1 &= 0 + (20 + 1)/2 = 21/2 = 10.5 \\
R1_2 &= 20 + (22 + 1)/2 = 33 \\
&\dots \\
C1_1 &= 0 + (4 + 1)/2 = 2.5 \\
&\dots
\end{aligned}
$$

In SAS

In SAS you specify the SCORES = {rank, ridit, modridit } to fit the mid rank, ridits or modified ridits to the data.

```
data two;
 input row salary satisfaction count;
 salary2 = salary /1000 - 10;
cards;
1 10000 1 1
1 10000 2 3
1 10000 3 10
1 10000 4 6
2 20000 1 2
2 20000 2 3
2 20000 3 10
2 20000 4 7
3 32500 1 1
3 32500 2 6
3 32500 3 14
3 32500 4 12
4 60000 1 0
4 60000 2 1
4 60000 3 9
4 60000 4 11
;
run;
```

```
 proc freq data=two;
 weight count;
tables (row salary salary2)*satisfaction/chisq
                                         measures
                                         scores=rank;
run;
```

Selected Results

Statistics for Table of row by satisfaction

Statistic	DF	Value	Prob
Chi-Square	9	5.9655	0.7434
Likelihood Ratio Chi-Square	9	6.7641	0.6617
MH Chi-Square (Rank Scores)	1	2.9726	0.0847

Statistics for Table of row by satisfaction

Statistic	Value	ASE
Gamma	0.2211	0.1172
Pearson Correlation (Rank Scores)	0.1769	0.0950
Spearman Correlation	0.1769	0.0955

Selected Results

Statistics for Table of salary by satisfaction

Statistic	DF	Value	Prob
Chi-Square	9	5.9655	0.7434
Likelihood Ratio Chi-Square	9	6.7641	0.6617
MH Chi-Square (Rank Scores)	1	2.9726	0.0847

Statistics for Table of salary by satisfaction

Statistic	Value	ASE
Gamma	0.2211	0.1172
Pearson Correlation (Rank Scores)	0.1769	0.0950
Spearman Correlation	0.1769	0.0955

Selected Results

Statistics for Table of salary2 by satisfaction

Statistic	DF	Value	Prob
Chi-Square	9	5.9655	0.7434
Likelihood Ratio Chi-Square	9	6.7641	0.6617
MH Chi-Square (Rank Scores)	1	2.9726	0.0847

Statistics for Table of salary2 by satisfaction

Statistic	Value	ASE
Gamma	0.2211	0.1172
Pearson Correlation (Rank Scores)	0.1769	0.0950
Spearman Correlation	0.1769	0.0955

Summary of Non-parametric Methods

- Note that Pearson's correlation has been replaced by its nonparametric equivalent
- Note that MH Chi-Square is equal (because the scores are all linear transformation of the ranks)
- The primary advantage is that we will be less criticized for our selection of our scores.
- However, the non-parametric methods are better when the variables are truly ordinal (only an arbitrary metric can be assigned)

Special Case; Cochran-Armitage Trend test

- For an $(I \times 2)$ table, the correlation test is called the 'Cocharan-Armitage' trend test:

Dose	Toxicity		
	None	Some	Total
1	92	8	100
10	90	10	100
100	89	11	100
1000	88	12	100
Total	359	41	400

- Here, we are looking for does the probability of "some" toxicity increase as a function of dose.

Hypotheses

The null hypothesis for the Cochran-Armitage Trend test is that

$$H_0 = p_{i1} = n_{i1}/n_{i\cdot} = p \ \ \forall i$$

To test this in SAS, you need to specify the TREND option in the table statement.

```
data one;
  input dose tox count;
  dose2=1+log10(dose) ;
  tox2 = tox;
  if tox=3 then tox2=10;
  tox3 = tox;
   if tox > 1 then tox3 = 2;
cards;
...
;
run;
proc freq data=one;
 weight count;
 tables dose*tox3 / trend;
 run;
```

```
Cochran-Armitage Trend Test
---------------------------
Statistic (Z)       -0.7165
One-sided Pr <  Z    0.2368
Two-sided Pr > |Z|   0.4737
```

For this example, we would fail to reject H_0 of equal proportion of no toxicity for each level of the drug.

In the future, we will look at additional methods of testing this hypothesis.