Ordinal Association in Contingency Tables

Grinnell College

November 9, 2026

Ordered Categories

So far, our loglinear models have treated categories as nominal.

- Independence allows no association.
- A full two-way interaction allows any pattern of association.

For ordered categories, we can ask a more specific question:

As one response increases, does the other tend to increase?

We will use category scores to describe this trend with one parameter.

Opinions in the General Social Survey

Two questions ask about attitudes toward:

- Premarital sex: always wrong through not wrong at all.
- Making birth control available to teenagers: strongly disagree through strongly agree.

Both variables have four ordered categories.

We study their association without designating either as a response variable.

Observed Counts

Premarital sex	Teenage birth control			
	Strongly disagree	Disagree	Agree	Strongly agree
Always wrong	81	68	60	38
Almost always wrong	24	26	29	14
Wrong only sometimes	18	41	74	42
Not wrong at all	36	57	161	157

Higher categories mean greater tolerance of premarital sex and greater support for access to birth control.

What pattern would you expect if these attitudes were positively associated?

Independence Misses the Corners

Under independence, $\hat{\mu}_{ij} = n_{i+}n_{+j}/n$.

Corner cell	Observed	Independence fit
Always wrong / strongly disagree	81	42.4
Always wrong / strongly agree	38	67.0
Not wrong / strongly disagree	36	70.6
Not wrong / strongly agree	157	111.4

The two corners pairing low with low and high with high have excess counts. The opposite corners have deficits.

This suggests a positive ordinal association.

A Model Between Independence and Saturation

In this 4×4 table:

Model	Parameters	Residual df
Independence	$1 + 3 + 3 = 7$	9
One ordinal association term	$7 + 1 = 8$	8
Saturated	16	0

A nominal interaction adds nine association parameters.

An ordinal model uses the ordering to describe association with just one.

Category Scores

Assign increasing scores to the rows and columns:

$$(u_1, u_2, u_3, u_4) = (1, 2, 3, 4), \qquad (v_1, v_2, v_3, v_4) = (1, 2, 3, 4).$$

Here u_i scores the premarital-sex opinion and v_j scores the birth-control opinion.

- Scores record the order of categories.
- Equal spacing treats each adjacent step as the same distance.
- Ordering alone does not tell us that the distances are equal.

The Linear-by-Linear Association Model

$$\boxed{\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \beta u_i v_j}$$

- The row and column main effects remain categorical.
- The score product $u_i v_j$ supplies one association term.
- $\beta = 0$ gives the independence model.

At a fixed column score v_j, the added term is linear in u_i. At a fixed row score u_i, it is linear in v_j.

This explains the name **linear-by-linear association**.

Direction of Association

With both sets of scores increasing in the stated category order:

- $\beta > 0$: higher row categories tend to occur with higher column categories.
- $\beta < 0$: higher row categories tend to occur with lower column categories.
- $\beta = 0$: independence.

The main effects also adjust when we fit the model. To interpret the strength of association, we will form odds ratios.

A Local Odds Ratio

Choose two adjacent rows and two adjacent columns.

	Column j	Column $j+1$
Row i	μ_{ij}	$\mu_{i,j+1}$
Row $i+1$	$\mu_{i+1,j}$	$\mu_{i+1,j+1}$

$$\theta_{ij} = \frac{\mu_{ij}\mu_{i+1,j+1}}{\mu_{i,j+1}\mu_{i+1,j}}.$$

It compares the odds of column $j+1$ versus j in the higher row with those odds in the lower row.

These four cells define the comparison. We do not collapse other categories into them.

Why the Main Effects Cancel

Take the log of the cross-product ratio:

$$\log \theta_{ij} = \log \mu_{ij} + \log \mu_{i+1,j+1} - \log \mu_{i,j+1} - \log \mu_{i+1,j}.$$

Each row effect and each column effect appears once with each sign. Only the association terms remain:

$$\begin{aligned} \log \theta_{ij} &= \beta[u_i v_j + u_{i+1} v_{j+1} - u_i v_{j+1} - u_{i+1} v_j] \\ &= \beta(u_{i+1} - u_i)(v_{j+1} - v_j). \end{aligned}$$

With scores $1, 2, 3, 4$, both differences equal 1:

$$\boxed{\theta_{ij} = e^{\beta}.}$$

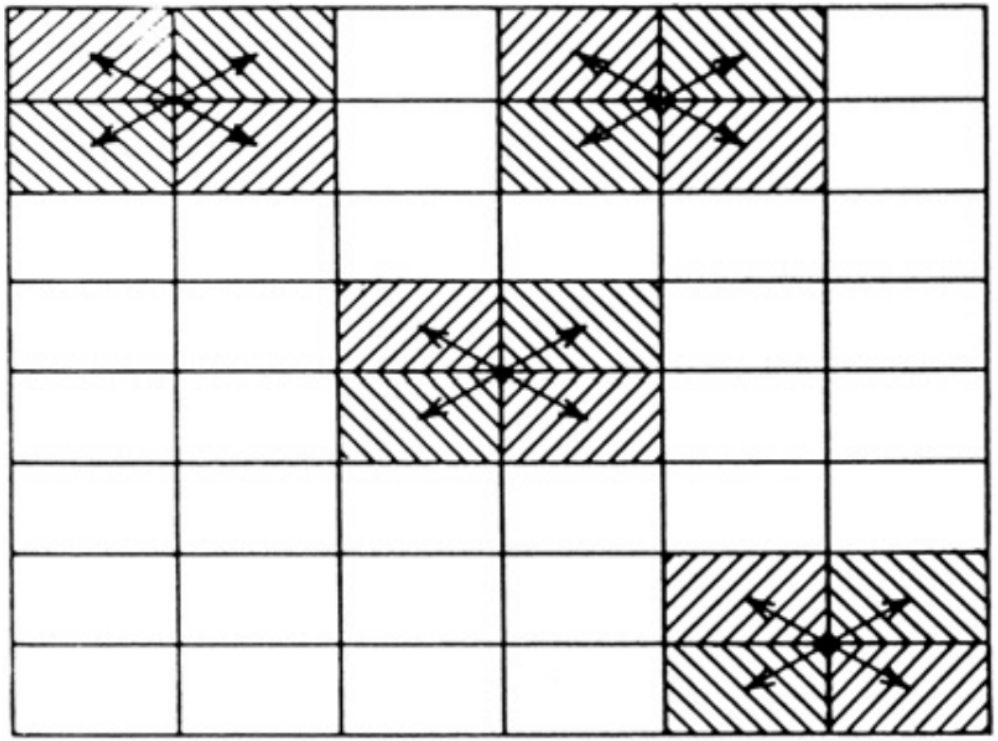
Uniform Association

With unit-spaced scores, every adjacent 2×2 comparison has the same **model-based** odds ratio e^{β}.

A 4×4 table has nine such comparisons. The model summarizes them with one common value.

- This special case is called **uniform association**.
- The observed local odds ratios need not be exactly equal.
- Model fit tells us whether a common local OR is an adequate summary.

Local Odds Ratios Across a Table



Each shaded block uses two adjacent rows and two adjacent columns.

Under uniform association, every such block has the same fitted odds ratio. With unit-spaced scores, that value is e^{β}.

Fitting the Model in R

One row of `d` represents one cell. The columns `sex` and `birth` contain scores $1, 2, 3, 4$.

```
d$sex_group <- factor(d$sex)
d$birth_group <- factor(d$birth)
d$score_product <- d$sex * d$birth

ind <- glm(count ~ sex_group + birth_group,
           family = poisson, data = d)

ordinal <- update(ind, . ~ . + score_product)
```

The factor main effects allow separate category effects. The numeric product adds exactly one coefficient.

Estimated Ordinal Association

Term	Estimate	SE	Wald z
Score product	0.28584	0.02824	10.12

$$\hat{\theta}_{\text{local}} = e^{0.28584} = 1.33.$$

For example, compare rows "always wrong" and "almost always wrong," using columns "disagree" and "agree."

The model estimates that the odds of agree versus disagree are **1.33 times as large** in the higher row.

The same fitted OR applies to every adjacent row/column comparison.

Comparisons Farther Apart

For rows $a < c$ and columns $b < d$,

$$\theta = \frac{\mu_{ab}\mu_{cd}}{\mu_{ad}\mu_{cb}} = \exp\{\beta(u_c - u_a)(v_d - v_b)\}.$$

The four corner cells have score differences of 3 in both directions:

$$\hat{\theta}_{\text{corners}} = e^{0.28584(3)(3)} = 13.10.$$

The odds of strongly agree versus strongly disagree are about 13 times as large for "not wrong at all" as for "always wrong."

A modest local OR can produce a much larger comparison across the full range.

Observed and Fitted Odds Ratios

For the upper-left 2×2 block, the observed OR is

$$\frac{81(26)}{68(24)} = 1.29.$$

For the four corner cells, the observed OR is

$$\frac{81(157)}{38(36)} = 9.30.$$

The fitted ORs are 1.33 and 13.10, respectively.

Fitted Counts Under Uniform Association

Premarital sex	Birth-control opinion score 1	2	3	4
Always wrong	80.9	67.7	69.4	29.1
Almost always wrong	20.8	23.1	31.5	17.6
Wrong only sometimes	24.4	36.2	65.7	48.8
Not wrong at all	33.0	65.1	157.4	155.5

The model largely captures the corner pattern that independence missed.

How can we assess whether the remaining discrepancies are larger than sampling variability would explain?

Deviance and Degrees of Freedom

Model	Deviance	Residual df
Independence	127.65	9
Ordinal	11.53	8
Saturated	0.00	0

- Independence versus ordinal: does the score-product term improve fit?
- Ordinal versus saturated: does the ordinal restriction fit adequately?

Both use likelihood-ratio tests, but they test different hypotheses.

Testing for an Ordinal Trend

Within the linear-by-linear model,

$$H_0 : \beta = 0 \qquad \text{versus} \qquad H_A : \beta \neq 0.$$

```
> anova(ind, ordinal, test = 'LRT')
Analysis of Deviance Table

Model 1: count ~ sex_group + birth_group
Model 2: count ~ sex_group + birth_group + score_product
  Resid. Df Resid. Dev Df Deviance            Pr(>Chi)
1         9      127.7
2         8       11.5  1      116 <0.0000000000000002 ***
```

With $p < 0.0001$, there is strong evidence of association. The positive coefficient indicates its direction.

Checking the Ordinal Model's Fit

The residual deviance compares the ordinal model with saturation:

```
> summary(ordinal)

Call:
glm(formula = count ~ sex_group + birth_group + score_product,
    family = poisson, data = d)

Coefficients:
              Estimate Std. Error z value             Pr(>|z|)
(Intercept)     4.1068     0.0895   45.88 < 0.0000000000000002
.
score_product   0.2858     0.0282   10.12 < 0.0000000000000002
---

    Null deviance: 431.078  on 15  degrees of freedom
Residual deviance:  11.534  on  8  degrees of freedom
```

We do not find evidence of lack of fit. Uniform association is a reasonable summary of this table.

The Benefit and the Restriction

The full nominal test allows any association pattern and uses 9 df. The ordinal trend test adds just one parameter.

- When the association follows the specified trend, focusing on it can improve power.
- A nonmonotone pattern can escape the one-parameter summary.
- A nonsignificant trend test alone does not establish independence.

The fit check assesses what the chosen ordinal model leaves unexplained.

Choosing the Score Spacing

Scores $(1, 2, 3, 4)$ assume equal adjacent distances.

Scores $(1, 2, 4, 5)$ allow a larger gap between categories 2 and 3. Using these scores for both variables gives

$$G^2 = 8.85 \quad \text{on 8 df},$$

compared with 11.53 for equal spacing.

- Use the meaning of the categories to motivate scores.
- Check whether the main conclusion depends on a plausible alternative.
- These two score choices give models that are not nested. Their deviance difference is not a standard nested-model LRT.

Changing Units Versus Changing Spacing

If we double every row score, then

$$\beta u_i v_j = \left(\frac{\beta}{2}\right)(2u_i)v_j.$$

The fitted counts and odds ratios stay the same. The coefficient halves because its units change.

With arbitrary increasing scores, a local OR is

$$\exp\{\beta(u_{i+1} - u_i)(v_{j+1} - v_j)\}.$$

With unit spacing in both variables, it simplifies to e^{β}.

Changing relative gaps, as in $(1, 2, 4, 5)$, can change the fit.

Discussion

1. Local OR: $e^{0.20} = 1.22$.
2. Two steps in each direction: $e^{0.20(2)(2)} = 2.23$.
3. Doubling row scores halves β to 0.10. The fitted ORs and counts stay unchanged.
4. Independence versus ordinal tests for a trend. Ordinal versus saturated checks fit.

The scores specify the association pattern. The data estimate its strength and show how well it fits.

Extensions

For a three-way table, we can model a common ordinal association between two variables at each level of the third.

If one variable is specifically an ordinal response, an ordinal logistic model can address that response directly.

The loglinear approach here treats both classifications symmetrically and models the cell counts.