The Average Odds Ratios as a Technique for Categorical Simplification

Ildefonso Marqués-Perales
University of Seville
UPO - Universidad Pablo de Olavide

15 Conferencia Española de Usuarios de Stata
22 May 2025, 07:44 PM CEST

Table of Contents

Introduction

- Mobility tables express the association of fathers and children in terms of social class.
- This association is considered an index of equality of opportunities (post facto).
- Open societies vs Closed Societies: More Fluid vs More Rigid.
- Log-linear models were adopted with marginal assumptions and log-multiplicative interaction structure.

Uniform Association Model: Unidiff Equation

The unidiff model, formulated by Xie and Goldthorpe & Erikson (1992):

$$\log F_{ijk} = \mu + \lambda_i^R + \lambda_j^C + \lambda_k^C + \lambda_{ik}^{RL} + \lambda_{jk}^{RL} + \beta_{jk}^{RL}$$

Where:

- F_{ijk}: Expected frequency in cell (i, j, k).
- μ: Overall mean.
- λ_i^R: Row marginal effect.
- λ_j^C: Column marginal effect.

Uniform Association Model: Unidiff Equation (cont.)

- λ_k^C: Layer marginal effect.
- λ_{ik}^{RL}: Row-layer interaction.
- λ_{jk}^{RL}: Column-layer interaction.
- β_k: Strength by layer.
- ϕ_{ij}: Association parameter.

Stata users: Two ados: Unidiff Pisati & Udiff Ben Jann. 20 m.

RC-II Model

The RC-II model, formulated by Goodman:

$$\log F_{ij} = \lambda + \lambda_i + \lambda_j + \sum_{m=1}^{M} \phi_m \mu_m \nu_{im}$$

Where:

- F_{ij}: Expected frequency for cell (i, j) in the contingency table.
- λ: Overall constant.
- λ_i: Main effect for row i.
- λ_j: Main effect for column j.

RC-II Model (cont.)

- $\phi_m \mu_m \nu_{im}$: Interaction effects, with M being the number of latent dimensions, μ_m the latent scores for row i, and ν_{jm} the latent scores for column j.

STATA: No ados. Extremely difficult estimation. For one table: John Hendrickx.

Mutual Information Index

Based on information theory principles:

$$I(X_i;Y) = \sum_i \sum_j p_{ij} \log \left(\frac{p_{ij}}{p_i p_j} \right)$$

Where:

- $I(X_i;Y)$: Mutual Information Index between variables X and Y.
- p_{ij}: Joint probability of being in cell (i,j).
- p_i: Marginal probability of row i.
- p_j: Marginal probability of column j.

Stata Users: Mindex Ado by Ben Jann. 6/10.

Average Odds Ratios: AGOR

Uses log-linear models (e.g., mlogit) to compute Average Global Odds Ratios (AGOR):

$$\text{AGOR} = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \frac{\left(\frac{F_{ij} F_{i+1,j+1}}{F_{i,j+1} F_{i+1,j}} \right)}{\left(\frac{F_{ij} F_{i+1,j+1}}{F_{i,j+1} F_{i+1,j}} \right)}$$

Where:

- AGOR = Average Global Odds Ratio.
- I: Number of rows in the mobility table.
- J: Number of columns in the mobility table.

Average Odds Ratios: AGOR (cont.)

- F_{ij}: Observed frequency in cell (i, j).
- $F_{i+1,j+1}$: Odds ratio for the 2×2 subtable formed by cells $(i, j), (i, j+1), (i+1, j)$, and $(i+1, j+1)$.

Average Global ORs calculated for different mobility scenarios. Mean AGOR for Clase Trabajadora vs. Clase de Servicio: 1.416 (SE: [incomplete]).

Alternative Average Odds Ratios

An alternative formulation using a simple mean of odds ratios:

$$\theta_{ij} = \left(\frac{F_{ij} F_{i+1,j+1}}{F_{i,j+1} F_{i+1,j}} \right), \quad \bar{\theta} = \frac{1}{N} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \theta_{ij}$$

When we have two means of odds ratios, we can subject them to a t-test if we have the errors.

Odds Ratios Comparison

Table: Odds Ratios Comparison Between Countries

Statistic	Value
Mean OR for country 1	2.4119868
Mean OR for country 2	2.0455212
Difference	0.36646562
Standard Error of difference	0.17482364
t-statistic	2.0962018
p-value	0.03606429

Stata Program: Setup

```
* Stata Meeting, Sevilla (2025)
* Average Odds Ratios

* Load the dataset from the specified path
use "C:\Users\imarq\Downloads\Cyprus_2019.dta"
```

Stata Program: Country 1 Analysis (1/3)

```
* COUNTRY 1 ANALYSIS

* 1. Run multinomial logit model
* This fits a multinomial logistic regression model with egp3 as the dependent variable
* and egp3f as a categorical predictor, restricted to country==1, using outcome 1 as base
mlogit egp3 i.egp3f if country==1, baseoutcome(1)
matrix b = e(b)  * Stores the model coefficients in matrix b

* 2. Define outcomes and categories
* Defines local macros for equation names and category numbers
local eqs `" "IIIab" "Clases_trabajadoras__IVabc_V_VI_" "'
```

Stata Program: Country 1 Analysis (2/3)

```
* 3. Calculate Odds Ratios
* Loops through equations and categories to compute odds
   ratios using exponentiation
local i = 1
foreach eq of local eqs {
    foreach cat of local cats {
        scalar or`i' = exp(_b[`eq':`cat'.egp3f])
        local i = `i' + 1
    }
}

* 4. Calculate mean OR
* Computes the average of the four odds ratios
scalar sum_or = or1 + or2 + or3 + or4
scalar mean_or = sum_or / 4
```

Stata Program: Country 1 Analysis (3/3)

```
* 6. Calculate SE of mean OR
* Computes variance and standard error for the mean odds
    ratio
scalar se_or1 = or1 * se_1
scalar se_or2 = or2 * se_2
scalar se_or3 = or3 * se_3
scalar se_or4 = or4 * se_4

scalar var_or1 = se_or1^2
scalar var_or2 = se_or2^2
scalar var_or3 = se_or3^2
scalar var_or4 = se_or4^2

scalar var_mean_or = (var_or1 + var_or2 + var_or3 +
    var_or4) / (4^2)
scalar se_mean_or = sqrt(var_mean_or)
```

Stata Program: Country 2 Analysis (1/3)

```
* COUNTRY 2 ANALYSIS

* 1. Run multinomial logit model
* Fits the model for country 2
mlogit egp3 i.egp3f if country==2, baseoutcome(1)
matrix b = e(b)

* 2. Define outcomes and categories
* Same categories as country 1
local eqs `" "IIIab" "Clases_trabajadoras__IVabc_V_VI_"
    "'
local cats "2 3"

* 3. Calculate Odds Ratios
* Computes odds ratios for country 2
local i = 1
```

Stata Program: Country 2 Analysis (2/3)

```
* 4. Calculate mean OR
* Computes the average odds ratio for country 2
scalar sum_or2 = or2_1 + or2_2 + or2_3 + or2_4
scalar mean_or2 = sum_or2 / 4

* 5. Calculate standard errors
* Extracts standard errors for country 2 coefficients
local i = 1
foreach eq of local eqs {
    foreach cat of local cats {
        scalar se2_`i' = _se[`eq':`cat'.egp3f]
        local i = `i' + 1
    }
}

* 6. Calculate SE of mean OR
```

Stata Program: Comparison Between Countries

```
* COMPARISON BETWEEN COUNTRIES
* Calculates the difference in means and performs a t-test
scalar diff_means = mean_or - mean_or2
scalar se_diff = sqrt(se_mean_or^2 + se_mean_or2^2)
scalar t_stat = diff_means/se_diff
scalar p_value = 2*(1-normal(abs(t_stat)))

* DISPLAY KEY RESULTS
* Displays the comparison results
display "Mean OR for country 1: " mean_or
display "Mean OR for country 2: " mean_or2
display "Difference: " diff_means
display "Standard Error of difference: " se_diff
display "t-statistic: " t_stat
display "p-value: " p_value
```