

Log-linear Regression

Dr Wan Nor Arifin

Unit of Biostatistics and Research Methodology, Universiti Sains Malaysia.

wnarifin@usm.my

Last update: 13 April, 2019



Wan Nor Arifin, 2019. Sampling Distribution by Wan Nor Arifin is licensed under the Creative Commons Attribution-ShareAlike 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/4.0/>.

Outlines

Introduction.....	1
Log-linear models.....	1
Steps in Log-linear modeling.....	1
References and recommended readings.....	2

Introduction

- A type of Poisson regression.
- Modeling cell counts in contingency tables – expected cell counts.
- Focus on interaction terms – association between categorical variables.

Log-linear models

- Easiest to explain for 2-way table.
- Two categorical variables X and Y .
- Independence model (Agresti, 2002) is given as

$$\ln \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y$$

where μ_{ij} = expected frequencies, λ_i^X = row effect, λ_j^Y = column effect. Note no interaction term.

- Saturated model (Agresti, 2002) is given as

$$\ln \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

where λ_{ij}^{XY} = interaction/association term.

- Saturated model is the one that include the highest-order model term (the interaction) and all the lower-order terms (all variables that make up the interaction). This is what is called *hierarchical* model.

Steps in Log-linear modeling

- The following steps in log-linear modeling building as suggested by von Eye and Mun (2013; pg 81-84):

1. Specify models to be tested

- List down the variables.
 - Start from base model – the independence model without interaction term.
 - Lower-order models – all lower-order variables, which may include 2-way interaction terms for model with 3 variables.
 - Higher-order model – the saturated model with 3-way interaction term for model with 3 variables.
 - List down all the models – for model-to-model comparison later.
2. Estimate the models
- Fit the model specified before.
 - Look at the:
 - i. Estimated parameters.
 - ii. Expected/predicted frequencies.
 - iii. Standardized residuals
3. Hypothesis testing
- Overall GOF – by G^2 and X^2 (refer formula in Poisson Regression note).
 - Estimated parameters significance – Wald's test.
 - Standardized residuals ($\sim z$) $< |1.96|$ to $|3.89|$
 - Model-to-model comparison.
 - i. AIC
 - ii. ΔG^2 and LR test – test for conditional association and parameters significance.
4. Model interpretation
- GOF.
 - Estimated parameters. For good-fitting model.
 - Significant interaction indicates associated between variables/factors.
 - or even ORs, but less emphasized in Log-linear model.

References and recommended readings

- Agresti, A. (2002). *Categorical data analysis* (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.
- von Eye, A. & Mun, E. (2013). *Log-linear modeling: Concepts, interpretation, and application*. Hoboken, New Jersey: John Wiley & Sons.