

# Log Linear Models for Contingency Tables

By Kanlaya Vanichbuncha



# Loglinear Model

- Loglinear model are useful for the potential complex relationships among the variables in a multiway crosstabulation.
- In loglinear models , all variables that are used for for classification are independent variables , and the dependent variable is the number of cases in a cell of the classification.

การผลิตปุ๋ยอินทรีย์ * การไปทัศนศึกษา Crosstabulation				
Count				
		การไปทัศนศึกษา		Total
		ไม่เคยไป	เคยไป	
การผลิตปุ๋ยอินทรีย์	ไม่ได้ผลิตใช้เอง	101	31	132
	ผลิตใช้เอง	93	88	181
Total		194	119	313

# Loglinear Models for 2-WAY Tables

## $I \times J$ contingency table

- Multinomial sample of  $n$  subjects on 2 categorical response

Variable X	Variable Y			
	1	2	...	J
1				
2				
⋮	⋮	⋮	⋮	⋮
I				

$I \times J$  cells

# Loglinear Models for GLM

## ➤ Loglinear models for Two-Dimensional Tables

Loglinear Model	Symbol
1. $\log(\mu_{ij}) = \mu + \lambda_i^x + \lambda_j^y$	(X,Y)
2. $\log(\mu_{ij}) = \mu + \lambda_i^x + \lambda_j^y + \lambda_{ij}^{xy}$	(XY)

The loglinear model for cell (I, J) is

$$\ln(n_{ij}) = \mu + \lambda_i^{\text{การผลิต}} + \lambda_j^{\text{ทัศนศึกษา}} + \lambda_{ij}^{\text{การผลิต และทัศนศึกษา}}$$

where

$\mu$  = grand mean in ANOVA

= average of log of frequencies in all table cells

$\lambda$  = increment or decrements from the base value ( $\mu$ )  
for particular combinations of values of the row and  
column variables

# Loglinear Models for GLM

In general, the model for log of observed frequencies in the  $i^{\text{th}}$  row and  $j^{\text{th}}$  column is

$$\ln(n_{ij}) = \mu + \lambda_i^{\text{การผลิต}} + \lambda_j^{\text{ทัศนศึกษา}} + \lambda_{ij}^{\text{การผลิต และทัศนศึกษา}}$$

Where

$n_{ij}$  = observed frequency in cell (i, j)

$\lambda_i^{\text{การผลิต}}$  = อิทธิพลหลักของการผลิต ที่  $i$

$\lambda_j^{\text{ทัศนศึกษา}}$  = อิทธิพลหลักของการไปทัศนศึกษา ที่  $j$

$\lambda_{ij}^{\text{การผลิต และทัศนศึกษา}}$  = อิทธิพลร่วมของการผลิต ที่  $i$  และ การไปทัศนศึกษา ที่  $j$



การผลิตปุ๋ยอินทรีย์ * การไปทัศนศึกษา Crosstabulation				
Count				
		การไปทัศนศึกษา		Total
		ไม่เคยไป	เคยไป	
การผลิตปุ๋ยอินทรีย์	ไม่ได้ผลิตใช้เอง	101	31	132
	ผลิตใช้เอง	93	88	181
Total		194	119	313

## Saturated Model

$$\ln(101) = \mu + \lambda_{\text{การผลิตไม่ได้ผลิต}} + \lambda_{\text{การไปไม่เคยไป}} + \lambda_{\text{การผลิตไม่ได้ผลิต การไปไม่ได้ไป}}$$

So, the number of cases in a cell is a function of the values of the row and column variables and their interactions

# Assumptions

---

## Two distributions

### I. Poisson Distribution

- The total sample size is not fixed before the study , or the analysis is not conditional on the total sample size.
- The event of an observation being in a cell is statistically Independent of the cell counts of the other cells.



# Assumptions

---

Two distributions

## II. Multinomial Distribution

- The total sample size is fixed before the study , or the analysis is conditional on the total sample size.
- The cell counts are not statistically independent.

# Loglinear Models for 2-WAY Tables

- For Maximum Likelihood

$$\hat{\mu}_{ij} = \frac{n_{i+} n_{+j}}{n}$$

- The estimated expected Frequencies for **Chi-Squared Tests** of independence

Use Chi-square Test

I. Pearson Chi-Square

$$\chi_{(I-1)(J-1)}^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}}$$

Constraints:

$$\sum_{i=1}^I \pi_{i+} = \sum_{j=1}^J \pi_{+j} = 1$$



# Loglinear Models for 2-WAY Tables

## 2. Likelihood-Ratio Chi-Squared Test

For multinomial Sampling

$H_0$  : Independence

$$\left( \hat{\pi}_{ij} = \hat{\pi}_{i+} \hat{\pi}_{+j} = \frac{n_{i+} n_{+j}}{n^2} \right)$$

In general case,

$$\hat{\pi}_{ij} = \frac{n_{ij}}{n}$$



# Loglinear Models for 2-WAY Tables

- The ratio of the likelihood is

$$\lambda = \frac{\prod_{i=1}^I \prod_{j=1}^J (n_{i+} n_{+j})^{n_{ij}}}{n^n \prod_{i=1}^I \prod_{j=1}^J n_{ij}^{n_{ij}}}$$

- The likelihood-ratio Chi-Squared Statistics is

$$G^2 = -2 \log \lambda = 2 \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \left( \frac{n_{ij}}{\hat{\mu}_{ij}} \right)$$

where  $\hat{\mu}_{ij} = \frac{n_{i+} n_{+j}}{n}$

# Loglinear Models for 2-WAY Tables

- ▶  $G^2$  is not good when  $\frac{n}{IJ} < 5$
- ▶ When  $I$  or  $J$  is large, use  $\chi^2$  when some expected frequencies are as small as 1 but most exceed 5.

# Loglinear Models for GLM

➤ In general,

➤ main effect of  $i^{\text{th}}$  category, is estimated as

$$\hat{\lambda}_i = \hat{\mu}_i - \hat{\mu} \quad ; i = 1, 2, \dots, I$$

$\mu_i$  = mean of the logs in the  $i^{\text{th}}$  category

$\mu$  = grand mean

➤ main effect of  $j^{\text{th}}$  category, is estimate

$$\hat{\lambda}_j = \hat{\mu}_j - \hat{\mu} \quad ; j = 1, 2, \dots, J$$

# Loglinear Models for Three-WAY Tables

## ➤ Three Categorical Variables (X, Y, Z)

$$\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$

Loglinear Model	Symbol
1. $\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z$	(X, Y, Z)
2. $\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY}$	(XY, Z)
3. $\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ}$	(XY, YZ)
4. $\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ}$	(XY, YZ, XZ)
5. $\log(\mu_{ijk}) = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$	(XYZ)

# SPSS Loglinear Analysis Procedure

---

## I. Model Selection

- Use for hierarcmodels, it produce parameters estimate for saturated model only
  - Most efficient

## 2. General Loglinear

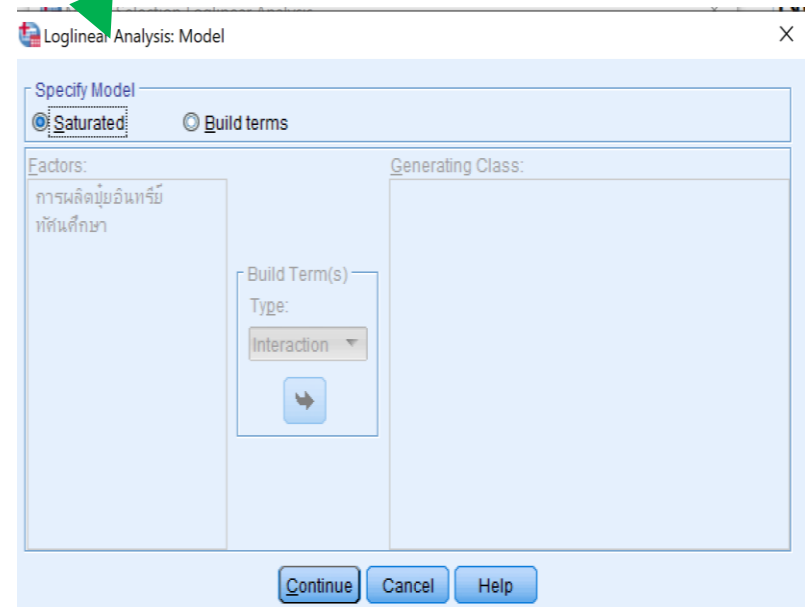
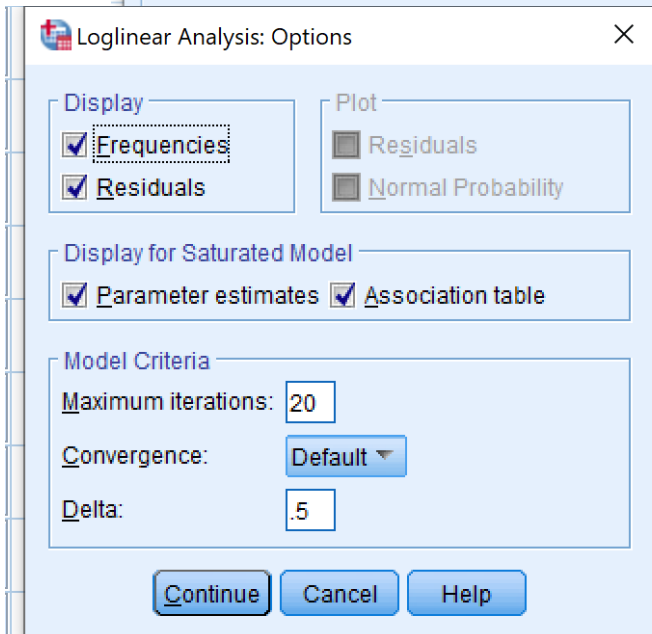
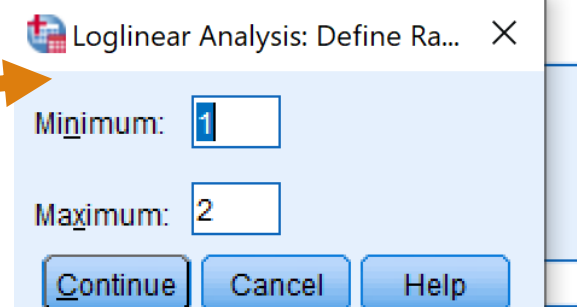
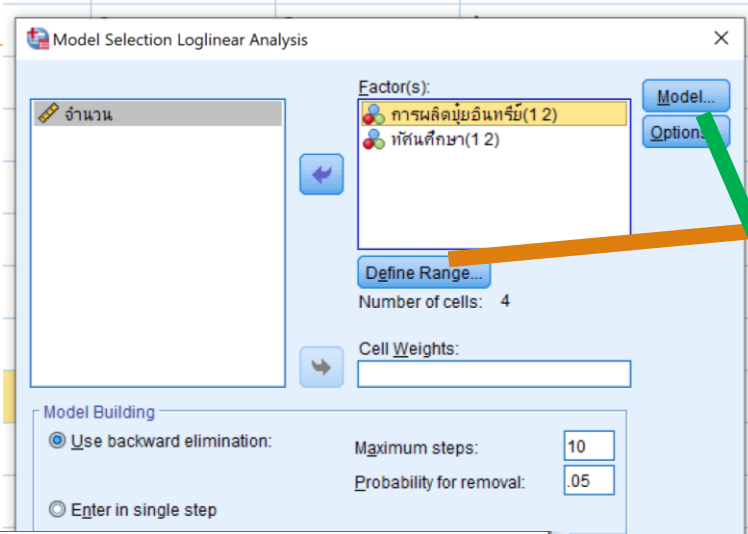
- Estimate parameters for a variety of models , including nonhierarchical model

## 3. Logit Loglinear

- Use for logit model



# SPSS : Analyze → Loglinear → Model Selection



## Warnings

For Design 1, .500 has been added to all observed cells for this saturated model, This value may be changed by using the CRITERIA = DELTA subcommand.

Data Information		
		N
Cases	Valid	4
	Out of Range <sup>a</sup>	0
	Missing	0
	Weighted Valid	313
Categories	การผลิตปุ๋ยอินทรีย์	2
	การไปทัศนศึกษา	2

a. Cases rejected because of out of range factor values.

Cell Counts and Residuals							
การผลิตปุ๋ยอินทรีย์	การไปทัศนศึกษา	Observed		Expected		Residuals	Std. Residuals
		Count <sup>a</sup>	%	Count	%		
ไม่ได้ผลิตใช้เอง	ไม่เคยไป	101.500	32.4%	101.500	32.4%	.000	.000
	เคยไป	31.500	10.1%	31.500	10.1%	.000	.000
ผลิตใช้เอง	ไม่เคยไป	93.500	29.9%	93.500	29.9%	.000	.000
	เคยไป	88.500	28.3%	88.500	28.3%	.000	.000

a. For saturated models, .500 has been added to all observed cells.

## K-Way and Higher-Order Effects

	K	df	Likelihood Ratio		Pearson		Number of Iterations
			Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects <sup>a</sup>	1	3	46.933	.000	39.141	.000	0
	2	1	21.084	.000	20.463	.000	2
K-way Effects <sup>b</sup>	1	2	25.850	.000	18.678	.000	0
	2	1	21.084	.000	20.463	.000	0

b. Tests that k-way effects are zero.

## Partial Associations

Effect	df	Partial Chi-Square	Sig.	Number of Iterations
การผลิตปุ๋ยอินทรีย์	1	7.703	.006	2
ทัศนศึกษา	1	18.147	.000	2

Parameter Estimates

Effect	Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
การผลิตปุ๋ยอินทรีย์*ทัศนศึกษา	I	.279	.063	4.422	.000	.155	.402
การผลิตปุ๋ยอินทรีย์	I	-.238	.063	-3.771	.000	-.361	-.114
ทัศนศึกษา	I	.306	.063	4.858	.000	.183	.430