Log-linear Models in Contingency Table

Log-linear Model

- Poisson log-linear model is used when Y = counts and all of predictors X = categories.
- Frequently, variables Y , X , and cell counts are in a form of contingency tables.
- Thus, Log-linear model is often used for modeling cell count data in contingency tables.

Objective:

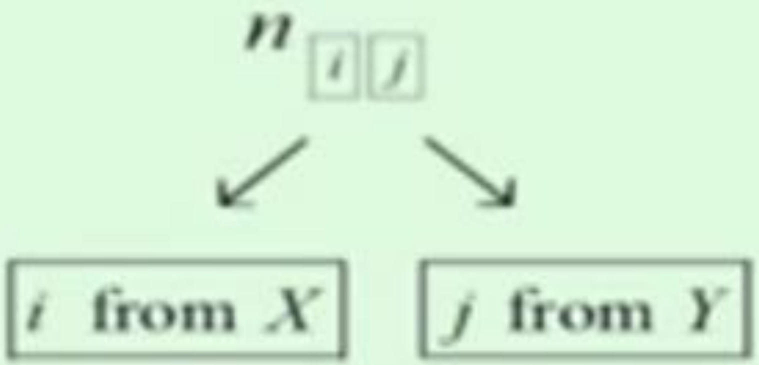
To analyze association and interaction patterns among variables in contingency table

Log-linear Models
for Two-way Table

A Log-linear Modeling Framework

Data:

X \ Y	1	0
1	n_{11}	n_{10}
0	n_{01}	n_{00}



Variables X and Y are determined for association

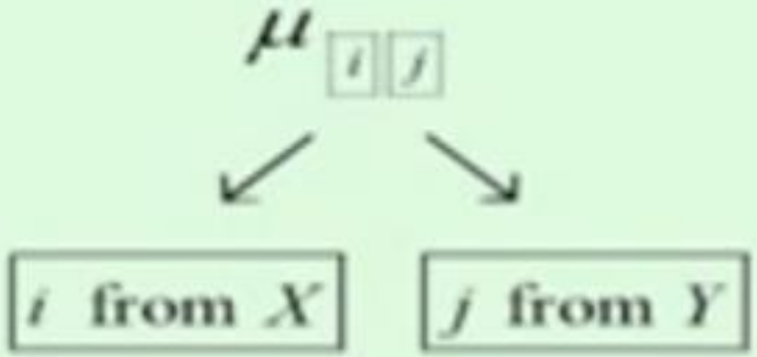
Data n_{ij} are in a form of cell counts

n_{ij} = Number of observations for cell $X = i$, $Y = j$

A Log-linear Modeling Framework

Means:

		Y: 1	Y: 0
X	1	$E(n_{11}) = \mu_{11}$	$E(n_{10}) = \mu_{10}$
	0	$E(n_{01}) = \mu_{01}$	$E(n_{00}) = \mu_{00}$



n_{ij} is assumed to be a $Poisson\,(\mu_{ij})$

Independent Model for Two Variables

Independent model: X and Y are independent

Independence: $P(X \cap Y) = P(X) \times P(Y)$

Probabilities: $p_{ij} = p_{i+} \, p_{+j}$

Let $N = n_{11} + n_{10} + n_{01} + n_{00}$

Expected counts: $\mu_{ij} = Np_{ij} = Np_{i+} \, p_{+j}$

$\log \mu_{ij} = \log N + \log p_{i+} + \log p_{+j}$

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

Independent model of two variables:

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

$$\log \mu_{00} = \lambda$$

$$\log \mu_{01} = \lambda + \lambda_1^Y$$

$$\log \mu_{10} = \lambda + \lambda_1^X$$

$$\log \mu_{11} = \lambda + \lambda_1^X + \lambda_1^Y$$

		Y = 1	Y = 0
X	1	$\log \mu_{11} = \lambda + \lambda_1^X + \lambda_1^Y$	$\log \mu_{10} = \lambda + \lambda_1^X$
	0	$\log \mu_{01} = \lambda + \lambda_1^Y$	$\log \mu_{00} = \lambda$

Association (Interaction) Model

- Usually, interaction model is used for indicating association between the row and the column factors in the two-way table

- Interaction model: $\log \mu_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$

		Y	
		1	0
X	1	$\log \mu_{11} = \lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}$	$\log \mu_{10} = \lambda + \lambda_1^X$
	0	$\log \mu_{01} = \lambda + \lambda_1^Y$	$\log \mu_{00} = \lambda$

Association (Interaction) model of two variables is obtained by adding Interaction term

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

Association Model Ignoring Indicator Variables:

$$\log \mu_{00} = \lambda$$

$$\log \mu_{01} = \lambda + \lambda_1^Y$$

$$\log \mu_{10} = \lambda + \lambda_1^X$$

$$\log \mu_{11} = \lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}$$

- Need to estimate 4 parameters: λ, $\lambda_1^X, \lambda_1^Y, \lambda_{11}^{XY}$
- Based on 4 observed cell counts: $n_{00}, n_{10}, n_{01}, n_{11}$

Association Model for Two Variables

$$\log \mu_{00} = \lambda \quad \rightarrow \mu_{00} = E(n_{00}) = \exp(\lambda)$$

$$\log \mu_{01} = \lambda + \lambda_1^Y \quad \rightarrow \mu_{01} = E(n_{01}) = \exp(\lambda + \lambda_1^Y)$$

$$\log \mu_{10} = \lambda + \lambda_1^X \quad \rightarrow \mu_{10} = E(n_{10}) = \exp(\lambda + \lambda_1^X)$$

$$\log \mu_{11} = \lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY}$$

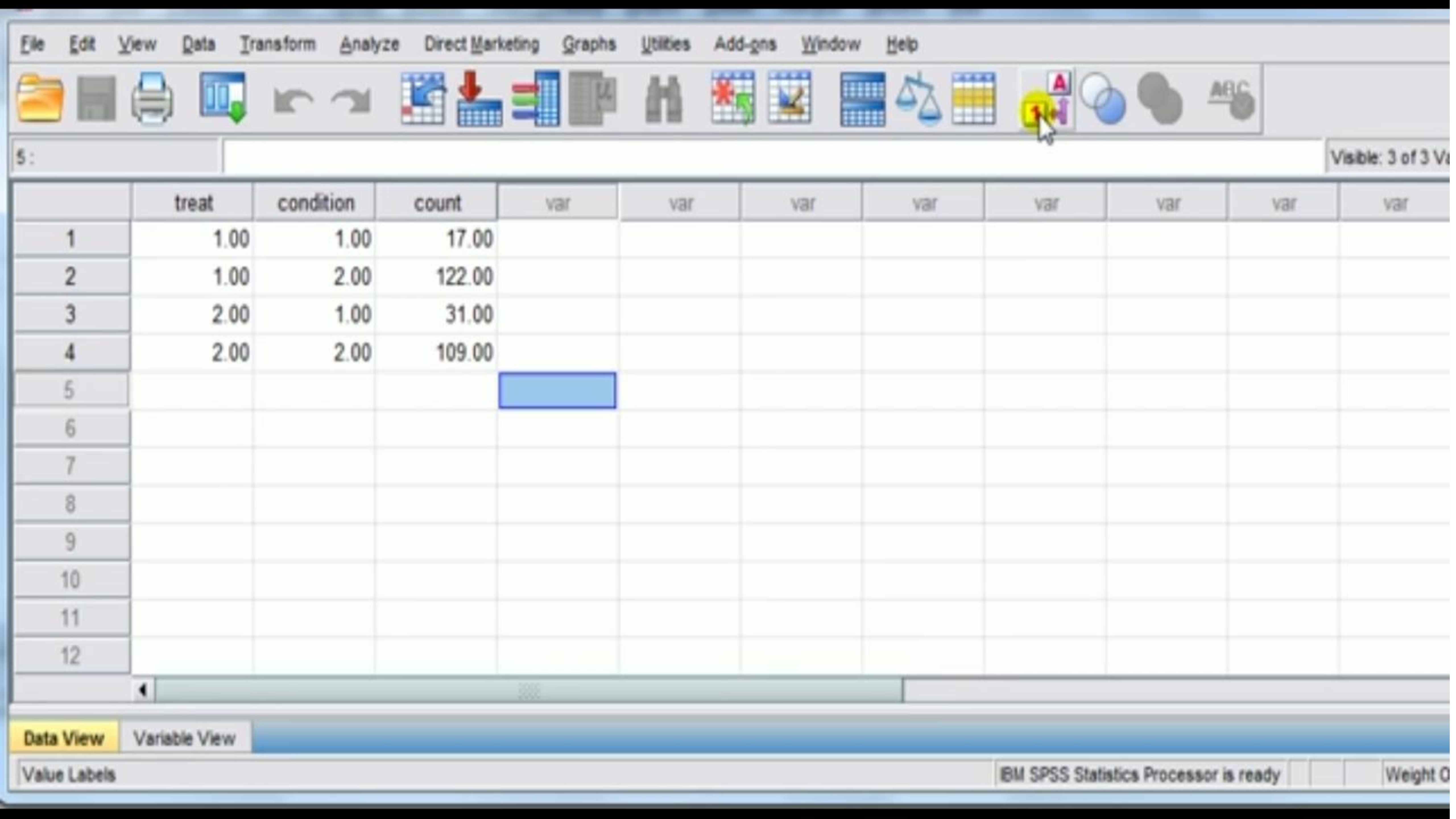
$$\rightarrow \mu_{11} = E(n_{11}) = \exp(\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})$$

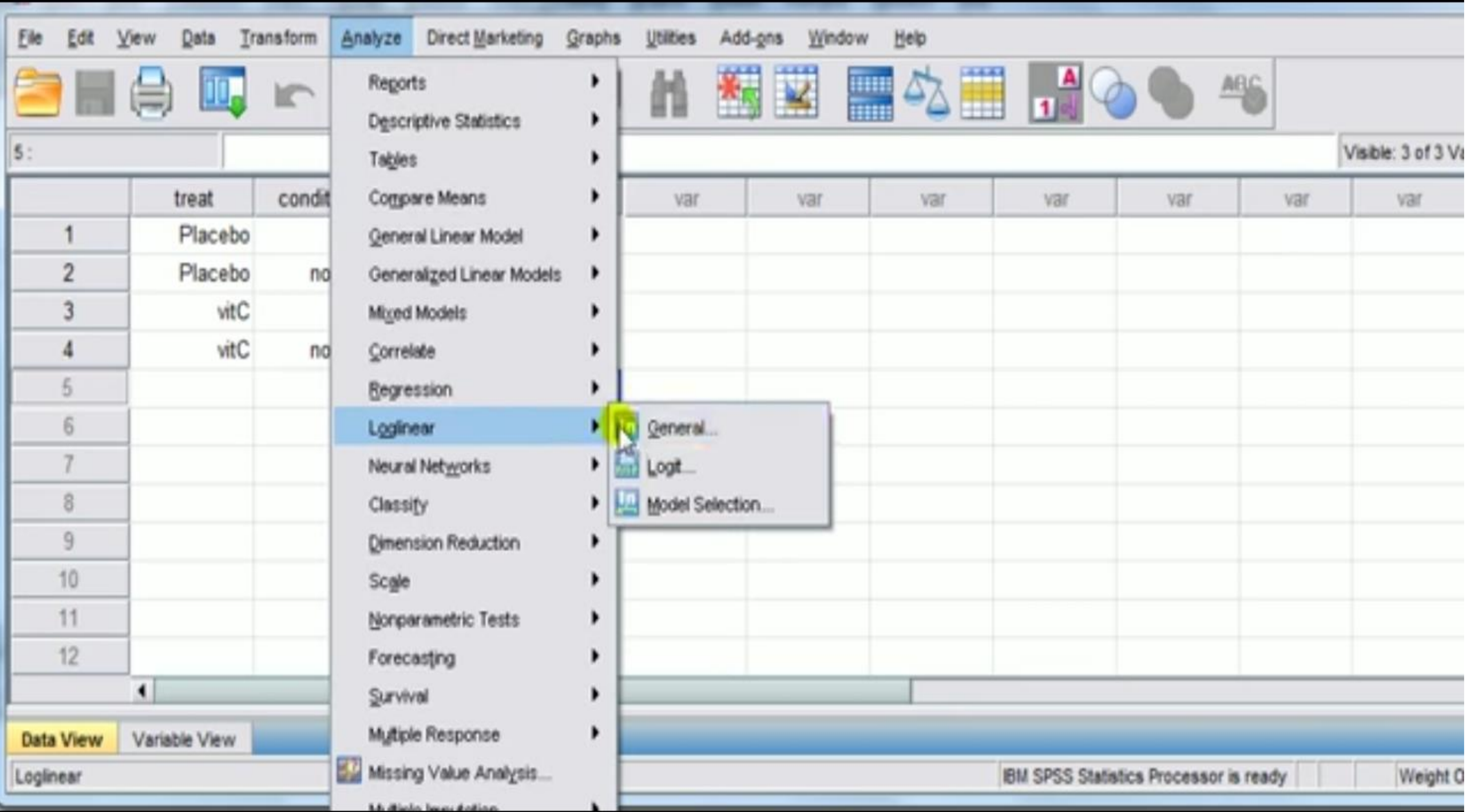
Means:

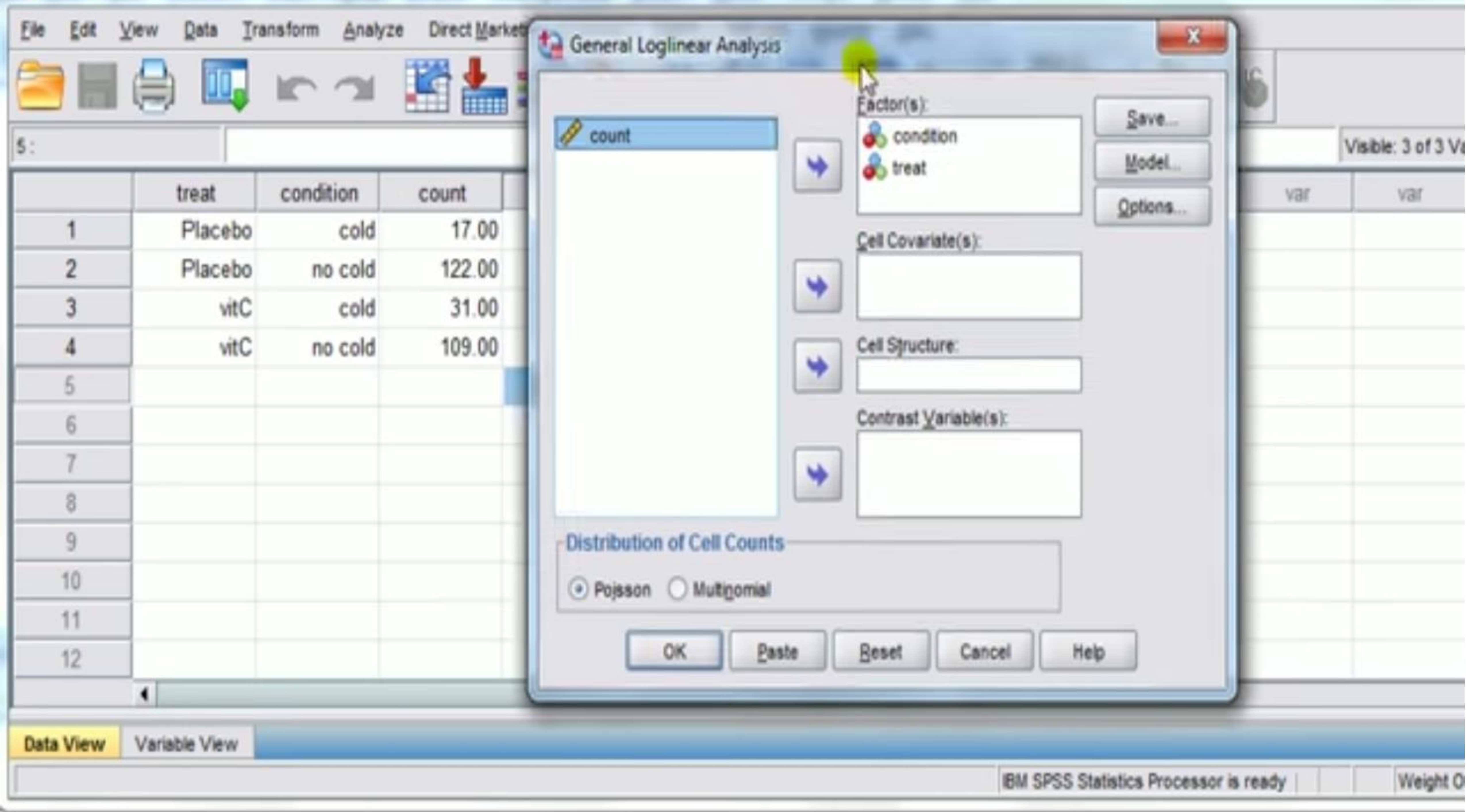
		Y = 1	Y = 0
X	1	$\mu_{11} = \exp(\lambda + \lambda_1^X + \lambda_1^Y + \lambda_{11}^{XY})$	$\mu_{10} = \exp(\lambda + \lambda_1^X)$
	0	$\mu_{01} = \exp(\lambda + \lambda_1^Y)$	$\mu_{00} = \exp(\lambda)$

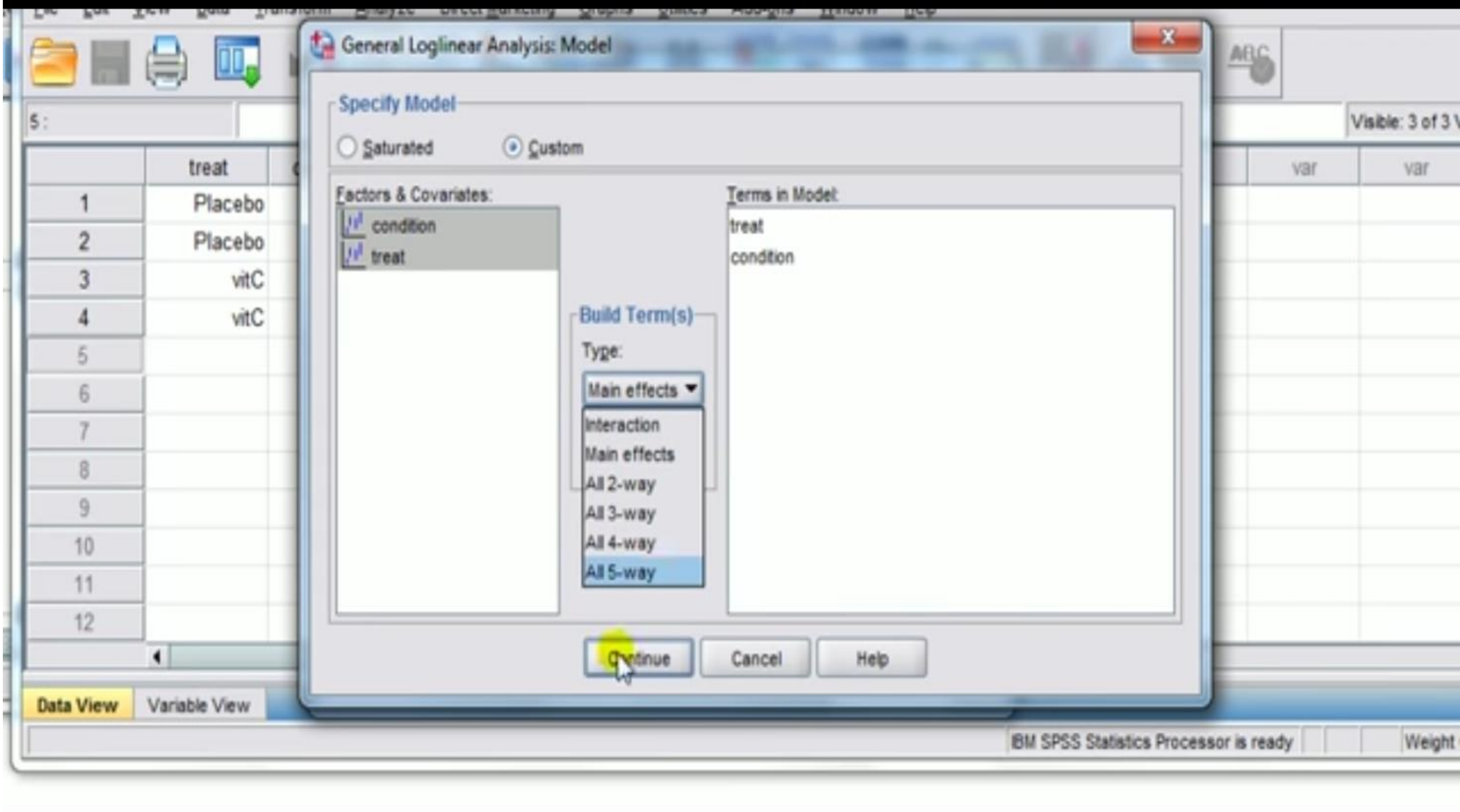
Treatment

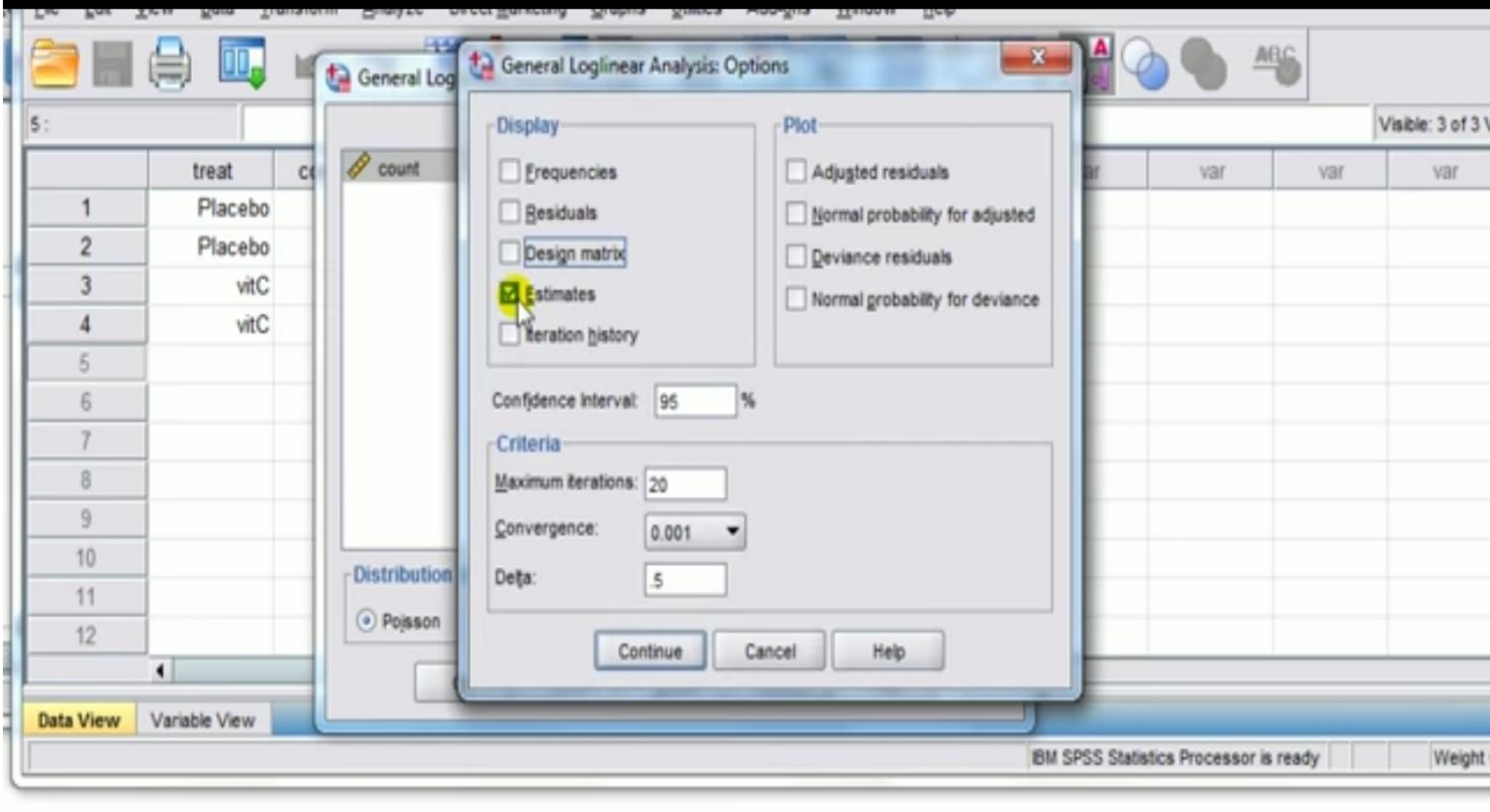
Outcome		Vitamin C	Placebo	
	Cold	17	31	48
	no cold	122	109	231
		139	109	279


File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Add-ons Window Help
5 :
Visible: 3 of 3 Va
treat condition count var var var var var var var var
1 1.00 1.00 17.00
2 1.00 2.00 122.00
3 2.00 1.00 31.00
4 2.00 2.00 109.00
5
6
7
8
9
10
11
12
Data View
Variable View
Value Labels
IBM SPSS Statistics Processor is ready
Weight O


File
Edit
View
Data
Transform
Analyze
Direct Marketing
Graphs
Utilities
Add-ons
Window
Help
Reports
Descriptive Statistics
Tables
Compare Means
General Linear Model
Generalized Linear Models
Mixed Models
Correlate
Regression
Loglinear
Neural Networks
Classify
Dimension Reduction
Scale
Nonparametric Tests
Forecasting
Survival
Multiple Response
Missing Value Analysis...
General...
Logit...
Model Selection...
5 :
Visible: 3 of 3 Va
treat
condit
var
1
2
3
4
5
6
7
8
9
10
11
12
Placebo
Placebo
vitC
vitC
no
no
Data View
Variable View
Loglinear
IBM SPSS Statistics Processor is ready
Weight O


File
Edit
View
Data
Transform
Analyze
Direct Market
General Loglinear Analysis
5 :
Visible: 3 of 3 Va
treat
condition
count
1
Placebo
cold
17.00
2
Placebo
no cold
122.00
3
vitC
cold
31.00
4
vitC
no cold
109.00
5
6
7
8
9
10
11
12
count
Factor(s):
condition
treat
Save...
Model...
Options...
Cell Covariate(s):
Cell Structure:
Contrast Variable(s):
Distribution of Cell Counts
Poisson
Multinomial
OK
Paste
Reset
Cancel
Help
var
var
Data View
Variable View
IBM SPSS Statistics Processor is ready
Weight O


General Loglinear Analysis: Model
Specify Model
Saturated
Custom
Factors & Covariates:
condition
treat
Terms in Model:
treat
condition
Build Term(s)
Type:
Main effects
Interaction
Main effects
All 2-way
All 3-way
All 4-way
All 5-way
Continue
Cancel
Help
5 :
Visible: 3 of 3
treat
Placebo
Placebo
vitC
vitC
var
var
Data View
Variable View
IBM SPSS Statistics Processor is ready
Weight


General Loglinear Analysis: Options
Display
Frequencies
Residuals
Design matrix
Estimates
Iteration history
Plot
Adjusted residuals
Normal probability for adjusted
Deviance residuals
Normal probability for deviance
Confidence Interval: 95 %
Criteria
Maximum iterations: 20
Convergence: 0.001
Delta: .5
Continue
Cancel
Help
General Log
count
Distribution
Poisson
treat
Placebo
Placebo
vitC
vitC
var
Visible: 3 of 3
Data View
Variable View
IBM SPSS Statistics Processor is ready
Weight

Likelihood Ratio	4.872	1	.027
Pearson Chi-Square	4.811	1	.028

a. Model: Poisson

b. Design: Constant + treat + condition

Parameter Estimates[b,c]

Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	4.753	.089	53.516	.000	4.579	4.927
[treat = 1.00]	-.007	.120	-.060	.952	-.242	.228
[treat = 2.00]	0[a]	.	.	.	.	.
[condition = 1.00]	-1.571	.159	-9.905	.000	-1.882	-1.260
[condition = 2.00]	0[a]	.	.	.	.	.

a. This parameter is set to zero because it is redundant.

b. Model: Poisson