Loglinear Regression

In loglinear regression analysis is used to describe the pattern of data in a contingency table. A model is constructed to predict the natural log of the frequency of each cell in the contingency table. For a 2x2 table, that means the model is $\ln f' = b_r * row + b_c * col + b_i * int + a$

We will focus on two aspects of that model: 1) how well it fits the data and 2) how to use the parameters of that model to describe the pattern of data in the contingency table.

Main Effects

Main effects look at the marginal means of each variable and test whether the conditions of that variable are equiprobable -- that is if the conditions occur with the same frequency. This is similar to computing a goodness-of-fit X^2 for the marginal frequencies of that variable. While they might not seem very interesting, main effects are a simple place to start and can provide important information for a complete description of the .

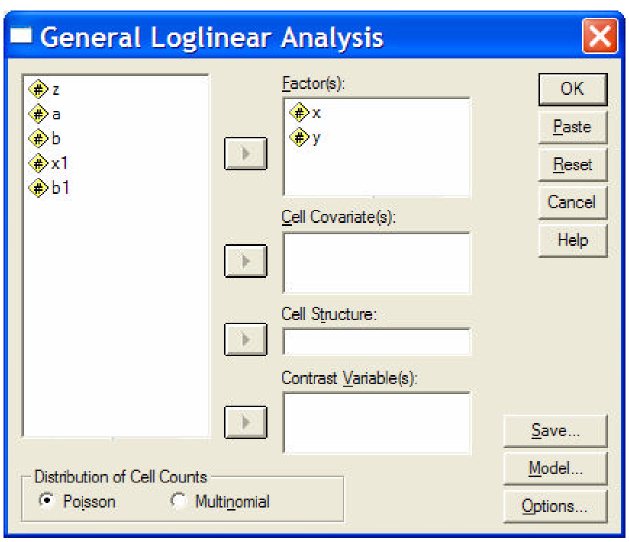
Let's start by looking at how qualitative variables are coded and how those codes are interpreted. As in other models using qualitative variables (e.g., GLM) SPSS builds a set of dummy codes for each qualitative variable. The highest coded group is set as the "comparison" or "control" group and each of the dummy codes compares the frequency of each successive group to the frequency of the comparison group.

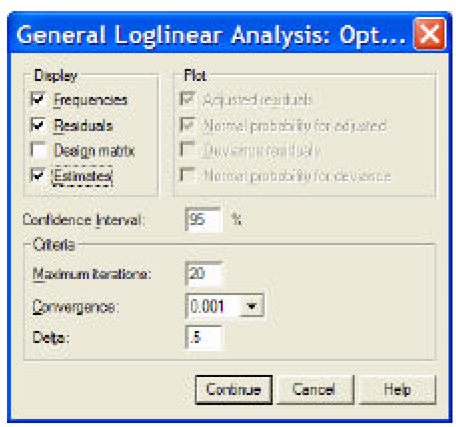
For example, if the variable has the pet categories turtles, rats & dogs (coded 1, 2 & 3, respectively), SPSS will compute three dummy-coded parameters, as shown below. You might find parameter3 confusing?!? There are only 2 degrees of freedom among three groups, and so you might reasonably expect 2 dummy codes/parameters. However, it is easier to do the related computations with all of the codes. So, SPSS computes the last, redundant or "alias" code and you have to remember that it is always going to be 0.

			dummy codes	
Category	original code	parameter1	parameter2	parameter3
Turtles	1	1	0	0
Rats	2	0	1	0
Dogs	3	0	0	1

Some examples…

Analyze → Loglinear → General





Be sure to check "Estimates" in the Options window

Move the variables you want to define the contingency table into "Factor(s)" window and chose the distribution model (Poisson or Multinomial) model.

I don't like the way the contingency table is given in the output, so I usually also run a cross-tabs to get a "regular looking" table, using… Analyze → Descriptive Statistics → Crosstabs

Y * X Crosstabulation

Count

		X 1.00	X 2.00	Total
Y	1.00	40	40	80
	2.00	10	10	20
Total		50	50	100

There is no "main effect" of X -- the marginal frequencies are equal.

There is a "main effect" of Y -- more in the "1" group than in the "2" group

Also, there is no relationship between X and Y or no "interaction".

Parameter	Aliased	Term
1		Constant
2		[X = 1.00]
3	x	[X = 2.00]
4		[Y = 1.00]
5	x	[Y = 2.00]
6		[X = 1.00]*[Y = 1.00]
7	x	[X = 1.00]*[Y = 2.00]
8	x	[X = 2.00]*[Y = 1.00]
9	x	[X = 2.00]*[Y = 2.00]

Note: 'x' indicates an aliased (or a redundant) parameter. These parameters are set to zero.

Here are the parameters SPSS built do use in this analysis.

- p1 is always the constant
- p2 & p3 represent "X" -- p2 compares X1 to X2 and p3 is redundant
- p4 & p5 represent "Y" -- p4 compares Y1 to Y2 and p5 is redundant
- p6-p9 represent the interaction or relationship between "X" and "Y" -- p6 represents the single degree or freedom and p7-p9 are redundant

So, we will be looking at three of these parameters to describe the data pattern in the table:
p2 -- main effect of "X"
p4 -- main effect of "Y"
p6 -- interaction or relationship between "X" and "Y"

Goodness-of-fit Statistics

	Chi-Square	DF	Sig.
Likelihood Ratio	.0000	0	.
Pearson	.0000	0	.

SPSS gives an overall model fit -- this will always be $X^2=0$ because the full model (called a "saturated model") will always fit the data perfectly.

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper
1	2.3514	.3086	7.62	1.75	2.96
2	-3.015E-15	.4364	-6.908E-15	-.86	.86
3	.0000	.	.	.	.
4	1.3499	.3463	3.90	.67	2.03
5	.0000	.	.	.	.
6	7.980E-16	.4898	1.629E-15	-.96	.96
7	.0000	.	.	.	.
8	.0000	.	.	.	.
9	.0000	.	.	.	.

p2 -- we expected no main effect of "X" and this parameter is very close to 0.00
p4 -- we expected a "Y" main effect and this parameter confirms this
-- the positive parameter value tells us that the target group (Y=1) has a higher frequency than the comparison (Y=2)
p6 -- we expected no relationship between "X" and "Y" -- this parameter is also very close to 0.00

How does significance testing work? There are two (equivalent) tests provided for each parameter

- The Z-value can be used -- a two-tailed test with p=.05 uses 1.96 as the critical value
- If the confidence interval includes 0.00, there one would retain the H0: that the parameter = 0.00

Here's a slightly more interesting one…

X * Z Crosstabulation

Count

		Z				Total
		1.00	2.00	3.00	4.00	
X	1.00	30	10	20	20	80
	2.00	30	10	20	20	80
Total		60	20	40	40	160

No "X" main effect & no "interaction"

"Y" main effect
Y1 > Y4 (comparison group)
Y2 < Y4
Y3 = Y4

Para-Meter	Alias	Term		
1		Constant		
2		[X = 1.00]	←	x
3	x	[X = 2.00]		
4		[Z = 1.00]	←	y1
5		[Z = 2.00]	←	y2
6		[Z = 3.00]	←	y3
7	x	[Z = 4.00]		
8		[X = 1.00]*[Z = 1.00]	←	int1
9		[X = 1.00]*[Z = 2.00]	←	int2
10		[X = 1.00]*[Z = 3.00]	←	int3
11	x	[X = 1.00]*[Z = 4.00]		
12	x	[X = 2.00]*[Z = 1.00]		
13	x	[X = 2.00]*[Z = 2.00]		
14	x	[X = 2.00]*[Z = 3.00]		
15	x	[X = 2.00]*[Z = 4.00]		

Parameter	Estimate	SE	Z-value	Lower	Upper		
1	3.0204	.2209	13.68	2.59	3.45		
2	-4.257E-15	.3123	-1.363E-14	-.61	.61	←	nsig x = 0
3	.0000	.	.	.	.		
4	.3973	.2856	1.39	-.16	.96	←	sig y1 > y4
5	-.6690	.3795	-1.76	-1.41	.07	←	sig y2 < y4
6	-4.607E-15	.3123	-1.475E-14	-.61	.61	←	sig y3 = y4
7	.0000	.	.	.	.		
8	1.942E-15	.4039	4.809E-15	-.79	.79	←	no interaction
9	1.661E-16	.5367	3.095E-16	-1.05	1.05	←	
10	2.109E-15	.4417	4.773E-15	-.87	.87	←	
11	.0000	.	.	.	.		
12	.0000	.	.	.	.		
13	.0000	.	.	.	.		
14	.0000	.	.	.	.		
15	.0000	.	.	.	.		

2-way Interactions

Parameters for 2-way interactions are formed as the product of the corresponding main effect parameters (same as interactions terms in regression). The number of non-redundant interaction parameters corresponds to the number or unique combinations of the main effect parameters. For example…

- A 2x2 table would have 1 non-redundant parameter for each main effect and 1 non-redundant parameter for the interaction.
- A 2x3 table would have 1 non-redundant parameter for the row effect, 2 non-redundant parameters for the column effect and 2 non-redundant parameters for the interaction.
- A 4x3 table would have 3 non-redundant parameters for the row effect, 2 non-redundant parameters for the column effect and 6 non-redundant parameters for the interaction.

An example… Pay attention -- this is where collinearity raises its head !!!

X * Y Crosstabulation

Count

		Y		Total
		1.00	2.00	
X	1.00	10	40	50
	2.00	40	10	50
Total		50	50	100

Looks like no “main effect” of X -- the marginal frequencies are equal.

Looks like no “main effect” of Y -- the marginal frequencies are equal.

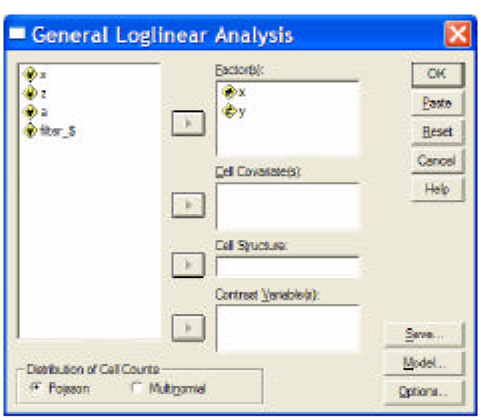
Looks like there is an “interaction” or a relationship between X and Y -- if x=1 y1 < y2 however if x=2 y1 > y2

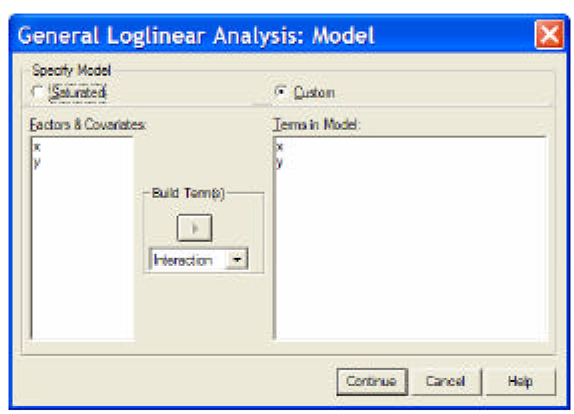
Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper		
1	2.3514	.3086	7.62	1.75	2.96		
2	1.3499	.3463	3.90	.67	2.03	← “X” main effect	Huh?
3	.0000	.	.	.	.		
4	1.3499	.3463	3.90	.67	2.03	← “y” main effect	Huh?
5	.0000	.	.	.	.		
6	-2.6999	.4898	-5.51	-3.66	-1.74	← xy interaction	ok!
7	.0000	.	.	.	.		
8	.0000	.	.	.	.		
9	.0000	.	.	.	.		

What happened with the main effects?

Remember that this is a multivariate model -- the parameters represent the unique contribution of each term in the model after controlling for all the others. The collinearity among the variables will influence how that contribution gets “shared” among the terms.

Watch what happens when we fit a model that includes only the main effects …





Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper	
1	3.2189	.1732	18.59	2.88	3.56	
2	-3.721E-16	.2000	-1.861E-15	-.39	.39	← no “X” effect
3	.0000	.	.	.	.	
4	-3.721E-16	.2000	-1.861E-15	-.39	.39	← no “y” effect
5	.0000	.	.	.	.	

So, the pattern of the interaction us such that when the interaction is “controlled for” there are main effect differences -- this is the important difference between interpreting the parameters of a model fit to the data and interpreting the pattern of the data! You must consider the “unique effect” described by the parameter estimate corresponds to the “visible” pattern in the raw data, and inform your reader appropriately!

Another…

X * Z Crosstabulation

Count

		Z 1.00	Z 2.00	Total
X	1.00	10	40	50
	2.00	40	40	80
Total		50	80	130

Looks like a “main effect” of X -- x1 < x2

Looks like a “main effect” of Y -- y1 < y2

Also, looks like there is an “interaction” or a relationship between X and Y -- if x=1 y1 < y2 however if x=2 y1 = y2

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper		
1	3.7013	.1571	23.55	3.39	4.01		
2	1.467E-15	.2222	6.602E-15	-.44	.44	← no “X” effect	huh?
3	.0000	.	.	.	.		
4	1.962E-15	.2222	8.829E-15	-.44	.44	← no “Y” effect	huh?
5	.0000	.	.	.	.		
6	-1.3499	.4115	-3.28	-2.16	-.54	← interaction	
7	.0000	.	.	.	.		
8	.0000	.	.	.	.		
9	.0000	.	.	.	.		

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper	
1	3.8965	.1316	29.62	3.64	4.15	
2	-.4700	.1803	-2.61	-.82	-.12	← expected “X” effect
3	.0000	.	.	.	.	
4	-.4700	.1803	-2.61	-.82	-.12	← expected “Y” effect
5	.0000	.	.	.	.	

So, as before, the “pattern of the marginal means” is not the same as “the pattern of the marginal means after correcting for the interaction?

When this happens → is usually the “apparent interaction pattern” from the table that matches the parameters, while the lower-order main effect terms get “jumbled”.

3-way Interactions

Parameters for 3-way interactions are computed as the product of the three corresponding main effect terms. A significant 3-way interaction tells us that the pattern of a 2-way is different at different values of the 3rd variable.

An Example…

X * Y * Z Crosstabulation

Count

Z			Y 1.00	Y 2.00	Total
1.00	X	1.00	10	40	50
		2.00	40	10	50
	Total		50	50	100
2.00	X	1.00	10	40	50
		2.00	40	40	80
	Total		50	80	130

Apparent effects..

X main effect 100 < 130
Y main effect 100 < 130
Z main effect 100 < 130

XY	y1	y2
x1	20	80
x2	80	50

XZ	z1	z2
x1	50	50
x2	50	80

ZY	y1	y2
z1	50	50
z2	50	80

XYZ different XY pattern for z1 & z2

Parameter	Aliased	Term
1		Constant
2		[X = 1.00]
3	x	[X = 2.00]
4		[Z = 1.00]
5	x	[Z = 2.00]
6		[Y = 1.00]
7	x	[Y = 2.00]
8		[X = 1.00]*[Z = 1.00]
9	x	[X = 1.00]*[Z = 2.00]
10	x	[X = 2.00]*[Z = 1.00]
11	x	[X = 2.00]*[Z = 2.00]
12		[X = 1.00]*[Y = 1.00]
13	x	[X = 1.00]*[Y = 2.00]
14	x	[X = 2.00]*[Y = 1.00]
15	x	[X = 2.00]*[Y = 2.00]
16		[Z = 1.00]*[Y = 1.00]
17	x	[Z = 1.00]*[Y = 2.00]
18	x	[Z = 2.00]*[Y = 1.00]
19	x	[Z = 2.00]*[Y = 2.00]
20		[X = 1.00]*[Z = 1.00]*[Y = 1.00]
21	x	[X = 1.00]*[Z = 1.00]*[Y = 2.00]
22	x	[X = 1.00]*[Z = 2.00]*[Y = 1.00]
23	x	[X = 1.00]*[Z = 2.00]*[Y = 2.00]
24	x	[X = 2.00]*[Z = 1.00]*[Y = 1.00]
25	x	[X = 2.00]*[Z = 1.00]*[Y = 2.00]
26	x	[X = 2.00]*[Z = 2.00]*[Y = 1.00]
27	x	[X = 2.00]*[Z = 2.00]*[Y = 2.00]

Parameter	Estimate	SE	Z-value	Asymptotic 95% CI Lower	Upper		
1	3.7013	.1571	23.55	3.39	4.01		
2	-6.229E-16	.2222	-2.803E-15	-.44	.44	← x	didn't get expected ME
4	-1.3499	.3463	-3.90	-2.03	-.67	← z	didn't get expected ME
6	-2.375E-15	.2222	-1.069E-14	-.44	.44	← y	didn't get expected ME
8	1.3499	.4115	3.28	.54	2.16	← xz	expected 2-way
12	-1.3499	.4115	-3.28	-2.16	-.54	← xy	expected 2-way
16	1.3499	.4115	3.28	.54	2.16	← zy	expected 2-way
20	-1.3499	.6397	-2.11	-2.60	-.10	← xzy	expected 3-way

Again we see the differences between the "main effect pattern" and the "contribution of the main effect to the model".

Also, notice that it is the lower order terms that tend to be "distorted" by considering their unique contribution to the model, rather than their pattern. This is one reason that some give the advice the "ignore" lower order terms when higher order terms contribute to the model.