Chapter 24
General Loglinear Models for Identifying Subgroups with Large Health Risks (12 Populations)

1 General Purpose

Data files that assess the effect of discrete predictors on frequency counts of morbidities/mortalities can be assessed with multiple linear regression. However, the results do not mean too much, if the predictors interact with one another. In that case they can be cross-classified in tables of multiple cells using general loglinear modeling.

2 Schematic Overview of Type of Data File

predictor discrete	predictor discrete	predictor discrete	frequency count	cell structure variable
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.

Linear regresssion with frequency count as continuous outcome can test whether the predictors are independent determinants of the outcome. However, they do not tell you whether one predictor is significantly different from the other and whether interaction between the predictors is in the data.

T.J. Cleophas, A.H. Zwinderman, *SPSS for Starters and 2nd Levelers*,
DOI 10.1007/978-3-319-20600-4_24

3 Primary Scientific Question

Can general loglinear modeling identify subgroups with significantly larger incident risks than other subgroups.

4 Data Example

In patients at risk of infarction with little soft drink consumption, and consumption of wine and other alcoholic beverages the incident risk of infarction equals 240/930 = 24.2 %, in those with lots of soft drinks, no wine, and no alcohol otherwise it is 285/1043 = 27.3 %.

Soft drink (1 = little)	Wine (0 = no)	Alc beverages (0 = no)	Infarcts number	Population number
1,00	1,00	1,00	240	993
1,00	1,00	,00	237	998
2,00	1,00	1,00	236	1016
2,00	1,00	,00	236	1011
3,00	1,00	1,00	221	1004
3,00	1,00	,00	221	1003
1,00	,00	1,00	270	939
1,00	,00	,00	269	940
2,00	,00	1,00	274	979
2,00	,00	,00	273	966
3,00	,00	1,00	284	1041
3,00	,00	,00	285	1043

We wish to identify the subgroups with particular high risks. The data file is entitled “chapter24generalloglinear”, and is in extras.springer.com.

5 Traditional Linear Regression

Start by opening the data file in SPSS. For analysis the statistical model Linear in the module Regression is required.

Command:

Analyze....Linear Regression....Dependent: infarcts....Independent(s): soft drink, wine, other alc (alcoholic) beverages....WLS Weight: population....click OK.

ANOVA[a,b]

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	5,937E9	3	1.979E9	39174,044	,000[c]
	Residual	6,025E8	11927	50514,056		
	Total	6,539E9	11930			

[a]Dependent Variable: infarcts
[b]Weighted Least Squares Regression-Weighted by population
[c]Predictors: (Constant), other alc beverages, soft drink, wine

Coefficients[a,b]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	277,397	,201		1381,647	,000
	Soft drink	−,657	,080	−,023	−8,213	,000
	Wine	−44,749	,131	−,953	−342,739	,000
	Other alc beverages	,569	,130	,012	4,364	,000

[a]Dependent Variable: infarcts
[b]Weighted Least Squares Regression – Weighted by population

The above tables show that the three discrete predictors soft drink, wine, and other alc beverages are very strong independent predictors of infarcts adjusted for population size. We will now add interaction variables to the data.

wine * other alc beverages
soft drink * wine
soft drink * other alc beverages

Command:
The same commands as above with interaction variables as additional predictors.

ANOVA[a,b]

Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	5,941E9	6	9.902E8	19757,821	,000[c]
	Residual	5.976E8	11924	50118,453		
	Total	6.539E9	11930			

[a]Dependent Variable: infarcts
[b]Weighted Least Squares Regression-Weighted by population
[c]Predictors: (Constant), soft *alc, wine, soft drink, soft*wine, wine*alc, other alc beverages

Coefficients[a,b]

Model		Unstandardized coefficients		Standardized coefficients	t	Sig.
		B	Std. error	Beta		
1	(Constant)	275,930	,272		1013,835	,000
	Soft drink	,118	,115	,004	1,026	,305
	Wine	−45,619	,224	−,972	−203,982	,000
	Other alc beverages	3,762	,407	,080	9,240	,000
	Wine*alc	,103	,269	,002	,385	,700
	Soft*wine	,487	,096	,024	5,088	,000
	Soft*alc	−1,674	,175	−,084	−9,561	,000

[a]Dependent Variable: infarcts
[b]Weighted Least Squares Regression – Weighted by population

The output sheets now show that soft drink is no longer a significant predictor of infarcts, while several interactions were very significant. This leaves us with an inconclusive analysis. Due to the interactions the meaning of the former discrete predictors have no further meaning.

6 General Loglinear Modeling

The general loglinear model computes cell counts in cross-classification tables, and can be simultaneously analyzed after logarithmic transformation in the form of analysis of variance data (see also the Chaps. 51 and 52). In this way an overall analysis of subgroup differences can be produced, and the significant differences can be identified. For analysis the statistical model General Loglinear Analysis in the module Loglinear is required.

Command:

Analyze....Loglinear ....General Loglinear Analysis....Factor(s): enter softdrink, wine, other alc beverages....click “Data” in the upper textrow of your screen.... click Weigh Cases....mark Weight cases by....Frequency Variable: enter “infarcts”....click OK....return to General Loglinear Analysis....Cell structure: enter “population”.... Options ....mark Estimates....click Continue....Distribution of Cell Counts: mark Poisson....click OK.

Parameter estimates[a,b]

Parameter	Estimate	Std. error	Z	Sig.	95 % Confidence interval	
					Lower bound	Upper bound
Constant	−1,513	,067	−22,496	,000	−1,645	−1,381
[softdrink = 1,00]	,095	,093	1,021	,307	−,088	,278
[softdrink = 2,00]	,053	,094	,569	,569	−,130	,237
[softdrink = 3,00]	0[c]					
[wine = ,00]	,215	,090	2,403	,016	,040	,391
[wine = 1,00]	0[c]					
[alcbeverages = ,00]	,003	,095	,029	,977	−,184	,189
[alcbeverages = 1,00]	0[c]					
[softdrink = 1,00] * [wine = ,00]	−,043	,126	−,345	,730	−,291	,204
[softdrink = 1,00] * [wine = 1,00]	0[c]					
[softdrink = 2,00] * [wine = ,00]	−,026	,126	−,209	,834	−,274	,221
[softdrink = 2,00] * [wine = 1,00]	0[c]					
[softdrink = 3,00] * [wine = ,00]	0[c]					
[softdrink = 3,00] * [wine = 1,00]	0[c]					
[softdrink = 1,00]* [alcbeverages = ,00]	−,021	,132	−,161	,872	−,280	,237
[softdrink = 1,00]* [alcbeverages = 1,00]	0[c]					
[softdrink = 2,00]* [alcbeverages = ,00]	,003	,132	,024	,981	−,256	,262
[softdrink = 2,00]* [alcbeverages = 1,00]	0[c]					
[softdrink = 3,00]* [alcbeverages = ,00]	0[c]					
[softdrink = 3,00] * [alcbeverages = 1,00]	0[c]					
[wine = ,00] * [alcbeverages = ,00]	−,002	,127	−,018	,986	−,251	,246
[wine = ,00] * [alcbeverages = 1,00]	0[c]					
[wine = 1,00] * [alcbeverages = ,00]	0[c]					
[wine = 1,00] * [alcbeverages = 1,00]	0[c]					
[softdrink = 1,00] * [wine = ,00] * [alcbeverages = ,00]	,016	,178	,089	,929	−,334	,366
[softdrink = 1,00] * [wine = ,00] * [alcbeverages = 1,00]	0[c]					
[softdrink = 1,00] * [wine = 1,00] * [alcbeverages = ,00]	0[c]					
[softdrink = 1,00] * [wine = 1,00] * [alcbeverages = 1,00]	0[c]					

(continued)

Parameter	Estimate	Std. error	Z	Sig.	95 % Confidence interval Lower bound	Upper bound
[softdrink = 2,00] * [wine = ,00] * [alcbeverages = ,00]	,006	,178	,036	,971	−,343	,356
[softdrink = 2,00] * [wine = ,00] * [alcbeverages = 1,00]	0[c]					
[softdrink = 2,00] * [wine = 1,00] * [alcbeverages = ,00]	0[c]					
[softdrink = 2,00] * [wine = * [alcbeverages = 1.00]	0[c]					
[softdrink = 3,00] * [wine = 00] * [alcbeverages = ,00]	0[c]					
[softdrink = 3,00] * [wine = ,00] * [alcbeverages = 1,00]	0[c]					
[softdrink = 3,00] * [wine = 1,00] * [alcbeverages = ,00]	0[c]					
[softdrink = 3,00] * [wine = 1,00] * [alcbeverages = 1.00]	0[c]					

[a]Model: Poisson
[b]Design: Constant + softdrink + wine + alcbeverages + softdrink * wine + softdrink * alcbeverages + wine * alcbeverages + softdrink * wine * alcbeverages
[c]This parameter is set to zero because it is redundant

The above pretty dull table gives some wonderful information. The soft drink classes 1 and 2 are not significantly different from zero. These classes have, thus, no greater risk of infarction than class 3. However, the regression coefficient of no wine is greater than zero at p = 0,016. No wine drinkers have a significantly greater risk of infarction than the wine drinkers have. No “other alcoholic beverages” did not protect from infarction better than the consumption of it. The three predictors did not display any interaction effects.

7 Conclusion

Data files that assess the effects of discrete predictors on frequency counts of morbidities / mortalities can be classified into multiple cells with varying incident risks (like ,e.g., the incident risk of infarction) using general loglinear modeling.

They can identify subgroups with significantly larger or smaller incident risks than other subgroups. Linear regression can also be used for the purpose. However, possible interactions between the predictors require that interaction variables are computed and included in the linear model. Significant interaction variables render the linear regression model pretty meaningless (see also Chap. 23).

8 Note

More background, theoretical and mathematical information of loglinear models are given in the Chaps 51 and 52. Interaction effects are reviewed in the Chap. 23.