

# Chapter 10

## Predictions from Nominal Clinical Data (450 Patients)

### General Purpose

In Chap. 9 the typology of medical data was reviewed. Nominal data are discrete data without a stepping function like genders, age classes, family names. They can be assessed with pie charts, frequency tables and bar charts. Statistical testing is not of much interest. Statistical testing becomes, however, interesting, if we want to know whether two nominal variables like treatment modality and treatment outcome are differently distributed between one another. An interaction matrix of these two nominal variables could, then, be used to test, whether one treatment performs better than the other.

### Primary Scientific Question

This chapter assesses the relationship between four treatment modalities, and, as outcome, five levels of quality of life (qol). Can an interaction matrix, otherwise called contingency table or crosstab, be used to assess whether some treatment modalities are associated with a better qol score than others, and to assess the directions of the differences in distribution of the variables.

### Example

In 450 patients with coronary artery disease four complementary treatment modalities, including cardiac fitness, physiotherapy, wellness, and hydrotherapy, were assessed for quality of life scores. The first 10 patients are in the table underneath.

The entire data file is entitled “Qol.sav”, and is in extras.springer.com. The example is also used in the Chap. 11. SPSS is applied for analysis.

treatment	counseling	qol	sat doctor
3	1	4	4
4	0	2	1
2	1	5	4
3	0	4	4
2	1	2	1
2	0	1	4
4	0	4	1
3	0	4	1
4	1	4	4
2	1	3	4

treatment=treatment modality (1=cardiac fitness, 2=physiotherapy, 3=wellness, 4=hydrotherapy, 5=nothing)  
 counseling=counseling given (0=no, 1=yes)  
 qol=quality of life score (1=very low, 5=vey high)  
 sat doctor=satisfaction with doctor (1=very low, 5=very high)

Start by opening the data file in SPSS statistical software.

**Command**

Analyze....Descriptive Statistics....Crosstabs....Rows: enter “treatment”.... Columns: enter “qol score”....click Statistics....mark Chi-square....click Continue....click OK.

In the output sheets the underneath tables are given.

Treatment * qol score crosstabulation		Qol score					Total
Count		Very low	Low	Medium	High	Very high	
Treatment	Cardiac fitness	21	21	16	24	36	118
	Physiotherapy	22	20	18	20	20	100
	Wellness	23	14	12	30	25	104
	Hydrotherapy	20	18	25	35	30	128
Total		86	73	71	109	111	450

Both hydrotherapy and cardiac fitness produce highest qol scores.

Chi-Square tests			
	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	12,288 <sup>a</sup>	12	,423
Likelihood ratio	12,291	12	,423
Linear-by-Linear Association	,170	1	,680
N of valid cases	450		

<sup>a</sup>0 cells (,0 %) have expected count less than 5. The minimum expected count is 15,78

However, the cells are not significantly different from one another, and so the result is due to chance. We have clinical arguments that counseling may support the beneficial effects of treatments, and, therefore, perform an analysis with two layers, one in the patients with and one in those without counseling.

**Command**

Analyze....Descriptive Statistics....Crosstabs....Rows: enter “treatment”.... Columns: enter “qol score”....Layer 1 of 1: enter “counseling”....click Statistics ....mark Chi-square....mark Contingency coefficient....mark Phi and Cramer’s V....mark Lambda....mark Uncertainty coefficient....click Continue....click OK.

The underneath tables are in the output sheets.

Treatment * qol score * counseling crosstabulation								
Count								
Counseling			Qol score					Total
			Very low	Low	Medium	High	Very high	
No	Treatment	Cardiac fitness	19	16	8	8	14	65
		Physiotherapy	8	8	7	7	15	45
		Wellness	23	8	6	15	9	61
		Hydrotherapy	15	14	9	10	11	59
	Total		65	46	30	40	49	230
Yes	Treatment	Cardiac fitness	2	5	8	16	22	53
		Physiotherapy	14	12	11	13	5	55
		Wellness	0	6	6	15	16	43
		Hydrotherapy	5	4	16	25	19	69
	Total		21	27	41	69	62	220

Chi-Square tests				
Counseling	Value	df	Asymp. Sig. (2-sided)	
No	Pearson Chi-Square	14,831	12	,251
	Likelihood ratio	14,688	12	,259
	Linear-by-Linear Association	,093	1	,760
	N of valid cases	230		

(continued)

Chi-Square tests				
Counseling		Value	df	Asymp. Sig. (2-sided)
Yes	Pearson Chi-Square	42,961	12	,000
	Likelihood ratio	44,981	12	,000
	Linear-by-Linear Association	,517	1	,472
	N of valid cases	220		

Obviously, if we assess the subjects who received counseling, then the high scores appear to appear very significantly more often in the hydrotherapy and cardiac fitness patients than in the physiotherapy and wellness groups.

Symmetric measures				
Counseling			Value	Approx. Sig.
No	Nominal by nominal	Phi	,254	,251
		Cramer's V	,147	,251
		Contingency coefficient	,246	,251
	N of Valid Cases		230	
Yes	Nominal by nominal	Phi	,442	,000
		Cramer's V	,255	,000
		Contingency coefficient	,404	,000
	N of valid cases		220	

Also the phi value, which is the ratio of the computed Pearson chi-square value and the number of observations, are statistically significant. They support that the differences observed in the yes-counseling group are real findings, not chance findings. Cramer's V and contingency coefficient are rescaled phi values, and furthermore support this conclusion.

Directional measures							
Counseling				Value	Asymp. Std. Error	Approx T	Approx Sig.
No	Nominal by nominal	Lambda	Symmetric	,061	,038	1,570	,116
			Treatment dependent	,079	,061	1,238	,216
			Qol score dependent	,042	,028	1,466	,143
	Goodman and Kruskal tau		Treatment dependent	,021	,011		,277
			Qol score dependent	,018	,009		,182
	Uncertainty coefficient		Symmetric	,022	,011	1,933	,259
			Treatment dependent	,023	,012	1,933	,259
			Qol score dependent	,020	,010	1,933	,259

(continued)

Directional measures				Value	Asymp. Std. Error	Approx T	Approx Sig.
Counseling							
Yes	Nominal by nominal	Lambda	Symmetric	,093	,050	1,806	,071
			Treatment dependent	,132	,054	2,322	,020
			Qol score dependent	,053	,063	,818	,414
		Goodman and Kruskal tau	Treatment dependent	,065	,019		,000
			Qol score dependent	,042	,013		,000
		Uncertainty coefficient	Symmetric	,071	,018	3,839	,000
			Treatment dependent	,074	,019	3,839	,000
			Qol score dependent	,067	,017	3,839	,000

The lambda value is also given. It shows the percentages of misclassifications in the row if you would know the column values, is also statistically significant in the yes-counseling subgroup at  $p=0.020$ . The value of 0.132 would mean 1.32 % reduction of misclassification, which is, however, not very much. Goodman and uncertainty coefficients serve similar purpose and are also statistically significant.

## Conclusion

In conclusion, many high qol levels are in the hydrotherapy and physiotherapy groups, and, correspondingly, very few low qol levels are a major factor for the overall result of this study assessing the effects of treatment modalities on qol scores. The interaction matrix can be used to assess whether some treatment modalities are associated with a better qol score than others, and to assess the directions of the differences in distribution of the variables.

## Note

More background, theoretical and mathematical information of crosstabs is given in Statistics applied to clinical studies 5th edition, Chap. 3, The analysis of safety data, pp 41–59, Edited by Springer Heidelberg Germany, 2012, from the same authors.