Chapter 12
Assessing Relative Health Risks (3,000 Subjects)

General Purpose

This chapter is to assess whether interaction matrices, otherwise called contingency tables or simply crosstabs, can be used to test the effect of personal characteristics like gender, age, married status etc. on a person's health risks.

Primary Scientific Question

Can marital status affect a person's health risks.

Example

In 3,000 subjects the effect of married status on being healthy was assessed.

T.J. Cleophas, A.H. Zwinderman, *Machine Learning in Medicine - a Complete Overview*, DOI 10.1007/978-3-319-15195-3_12

ageclass	married	healthy
4,00	1	0
3,00	0	0
2,00	1	0
1,00	1	0
4,00	1	0
3,00	0	0
2,00	1	0
1,00	0	0
4,00	1	0
3,00	1	0

ageclass 1 = 30–40, 2 = 40–50, 3 = 50–60, 4 = 60–70
married 0 = no, 1 = yes
healthy 0 = no, 1 = yes

In the above table the first 10 patients are given. The entire data file is entitled "healthrisk.sav" and is in extras.springer.com. We will start the analysis by opening the data file in SPSS.

Command

Analyze....Descriptive Statistics....Crosstabs....Row(s): enter married....Column(s): enter health....Statistics: mark Observed....mark Rows....click Continue....click OK.

Married * healthy crosstabulation					
			Healthy		
			No	Yes	Total
Married	No	Count	192	1,104	1,296
		% within married	14,8 %	85,2 %	100,0 %
	Yes	Count	167	1,537	1,704
		% within married	9,8 %	90,2 %	100,0 %

The crosstab is in the output sheets. It shows that 14.8 % of the unmarried subjects were unhealthy, leaving 85,2 % being healthy. In contrast, 9.8% of the married subjects were unhealthy, 90.2% being healthy. And so, the risk of being unhealthy in this population was 14.8 % in the unmarried and 9.8% in the married subjects. The relative risk of being unhealthy in unmarried versus married subjects was, thus, 14.8/9.8 = 1.512. Similarly, the relative risk of being healthy in unmarried versus married subjects was 85.2/90.2 = 0.944.

Risk estimate			
		95% Confidence interval	
	Value	Lower	Upper
Odds ratio for married (no/yes)	1,601	1,283	1,997
For cohort healthy = no	1,512	1,245	1,836
For cohort healthy = yes	,944	,919	,971
N of valid cases	3,000		

The odds of being unhealthy in unmarried subjects was 192/1,104 = 0.1739.

The odds of being unhealthy in married subjects was 167/1,537 = 0.1087.

The ratio of the two, the odds ratio was thus 0.1739/0.1087 = 1.601, as shown in the above table. It is easy to see that this odds ratio is equal to

$$= \frac{\text{the relative risk of being unhealthy in the unmarried versus married subjects}}{\text{the relative risk of being healthy in the unmarried versus married subjects}}$$

$$= 1.512 / 0.944 = 1.601.$$

In order to assess whether this finding is robust, we will add age classes as a layer variable, and test whether different age classes have similar odds ratios.

Command

Analyze....Descriptive Statistics....Crosstabs....Row(s): enter married....Column(s): enter health....Layer 1 of 1: enter ageclass....Statistics: mark Observed....mark Rows ....mark Cochran and Mantel Haenszel Statistics....click Continue....click OK.

Married * healthy * ageclass crosstabulation

Ageclass				Healthy		
				No	Yes	Total
30–40	Married	No	Count	52	138	190
			% within married	27,4 %	72,6 %	100,0 %
		Yes	Count	53	327	380
			% within married	13,9 %	86,1 %	100,0 %
	Total		Count	105	465	570
			% within married	18,4 %	81,6 %	100,0 %
40–50	Married	No	Count	69	352	421
			% within married	16,4 %	83,6 %	100,0 %
		Yes	Count	67	593	660
			% within married	10,2 %	89,8 %	100,0 %
	Total		Count	136	945	1,081
			% within married	12,6 %	87,4 %	100,0 %
50–60	married	No	Count	28	201	229
			% within married	12,2 %	87,8 %	100,0 %
		Yes	Count	17	287	304
			% within married	5,6 %	94,4 %	100,0 %
	Total		Count	45	488	533
			% within married	8,4 %	91,6 %	100,0 %
60–70	Married	No	Count	43	413	456
			% within married	9,4 %	90,6 %	100,0 %
		Yes	Count	30	330	360
			% within married	8,3 %	91,7 %	100,0 %
	Total		Count	73	743	816
			% within married	8,9 %	91,1 %	100,0 %

* Symbol of multiplication

Risk estimate				
			95% Confidence interval	
Ageclass		Value	Lower	Upper
30–40	Odds ratio for married (no/yes)	2,325	1,511	3,578
	For cohort healthy = no	1,962	1,396	2,759
	For cohort healthy = yes	,844	,767	,929
	N of valid cases	570		
40–50	Odds ratio for married (no/yes)	1,735	1,209	2,490
	For cohort healthy = no	1,614	1,180	2,208
	For cohort healthy = yes	,931	,886	,978
	N of valid cases	1,081		
50–60	Odds ratio for married (no/yes)	2,352	1,254	4,411
	For cohort healthy = no	2,186	1,227	3,896
	For cohort healthy = yes	,930	,879	,983
	N of valid cases	533		
60–70	Odds ratio for married (no/yes)	1,145	,703	1,866
	For cohort healthy = no	1,132	,725	1,766
	For cohort healthy = yes	,988	,946	1,031
	N of valid cases	816		

In the output are the crosstabs the odds ratios of the four ageclasses. The odds ratios are pretty heterogeneous, between 1.145 and 2.352, but 95 % confidence intervals were pretty wide. Yet, it is tested whether these odds ratios are significantly different from one another.

Tests of homogeneity of the odds ratio			
	Chi-Squared	df	Asymp. Sig. (2-sided)
Breslow-Day	5,428	3	,143
Tarone's	5,422	3	,143

The above Breslow and the Tarone's tests are the heterogeneity tests. They were insignificant. The differences could, thus, be ascribed to chance findings, rather than real effects. It seems appropriate, therefore, to say that an overall odds ratio of these data adjusted for age classes is meaningful. For that purpose a Mantel Haenszel (MH) odds ratio (OR) will be calculated.

		healthy	
		no	yes
unmarried	no	a	b
	yes	c	d

Having 4 odds ratios with the above structure, it is calculated as follows (n = a + b + c + d):

$$\text{Odds Ratio}_{MH} = \frac{\Sigma\, ad / n}{\Sigma\, cd / n}$$

Tests of conditional independence			
	Chi-Squared	df	Asymp. Sig. (2-sided)
Cochran's	26,125	1	,000
Mantel-Haenszel	25,500	1	,000

Under the conditional independence assumption, Cochran's statistic is asymptotically distributed as a 1 df chi-squared distribution, only if the number of strata is fixed, while the Mantel-Haenszel statistic is always asymptotically distributed as a 1 df chi-squared distribution. Note that the continuity correction is removed from the Mantel-Haenszel statistic when the sum of the differences between the observed and the expected is 0

Mantel-Haenszel common odds ratio estimate			
Estimate			1,781
In(Estimate)			,577
Std. Error of In(Estimate)			,115
Asymp. Sig. (2-sided)			,000
Asymp. 95% confidence interval	Common odds ratio	Lower bound	1,422
		Upper bound	2,230
	In(Common odds ratio)	Lower bound	,352
		Upper bound	,802

The Mantel-Haenszel common odds ratio estimate is asymptotically normally distributed under the common odds ratio of 1,000 assumption. So is the natural log of the estimate

The Cochran's and Mantel Haenszel tests assess whether married status remains an independent predictor of health after adjustment for ageclasses. They are significantly larger than an odds ratio (OR) of 0 at $p<0.0001$. The lower graph gives the OR_{MH} is thus 1.781. This OR is adjusted, and, therefore, more adequate than the unadjusted OR of page 1 of this chapter.

Conclusion

Interaction matrices, otherwise called contingency tables or simply crosstabs, can be used to test the effect of personal characteristics like gender, age, married status etc. on a person's health risks. Results can be adjusted for concomitant effects like the effect of age classes on the relationship between married status and health status. Prior to assessment the homogeneity of the concomitant factors have to tested.

Note

More background, theoretical and mathematical information of relative risk assessments are in Statistics applied to clinical studies, Chap. 3, The analysis of safety data, pp 41–59, Edited by Springer Heidelberg Germany, 2012, from the same authors.