

M. A. A. Moussa,
A. M. Karnik, Y. T. Omar,
F. F. Fenech

On the Analysis of Three-Dimensional Contingency Tables – Application to the Influence of Allergy and Age on Cancer

(From the Faculty of Medicine and the Cancer Control Centre, Kuwait)

1. Introduction

In clinical studies, as the number of variables measured on each patient increases, the resulting multidimensional contingency tables become more unwieldy. The investigator is inclined to give up on multiway tables and look instead at large numbers of two-way tables derived by adding over the categories of all except two variables.

In this paper, where emphasis is placed on three-way cross-classification, a procedure is described for the analysis of $I \times J \times K$ tables. This comprises selecting the best model and searching for interactions through iterative proportional fitting of the data to the appropriate configuration of log-linear models. Situations are considered where collapsing three-way tables to two dimensions is valid without violating the structure of the table. The relation of the Mantel-Haenszel test to log-linear models is also discussed.

These methods are applied to a data set based on a study recently undertaken in the Cancer Control Centre of Kuwait by Karnik et al. [3] with the objective of investigating the effect of allergy on the occurrence of cancer. All newly diagnosed cancer patients attending the outpatient department of the Cancer Control Centre during January to December 1982 were included in the study. During that period 468 cancer patients were examined and investigated by the oncologist, and a final diagnosis

Summary

A procedure is described for the analysis of $I \times J \times K$ cross-classifications. This includes 1. selecting the appropriate model by iterative proportional fitting of the data to relevant configurations of log-linear models; 2. assumptions underlying reduction of three-way tables to two dimensions; 3. the Mantel-Haenszel statistic with reference to its relation to log-linear models.

The methods are applied to a data set based on a study recently conducted at the Cancer Control Centre, Kuwait to investigate the relation between allergy and cancer. It was concluded that the allergy rate is significantly lower in the cancer group which suggests that allergy could be a protective factor against cancer.

Key-Words: Three-Way Contingency Tables, Log-Linear Models, Mantel-Haenszel Statistic

Über die Analyse dreidimensionaler Kontingenztafeln – Anwendung auf den Einfluß von Allergie und Alter auf Krebs

Ein Verfahren zur Analyse von $I \times J \times K$ Kreuzklassifikationen wird beschrieben. Dies beinhaltet 1. die Auswahl des geeigneten Modells durch iterative proportionale Anpassung der Daten an relevante Konfigurationen log-linearer Modelle; 2. Annahmen, die der Reduktion von Dreiwegtabellen auf zwei Dimensionen zugrunde liegen; 3. den Mantel-Haenszel-Test in bezug auf seine Beziehung zu log-linearen Modellen.

Die Methoden werden auf einen Datensatz angewendet, der auf einer kürzlich im Cancer Control Centre von Kuwait durchgeführten Studie zur Untersuchung der Beziehung zwischen Allergie und Krebs basiert. Man kam zu der Schlußfolgerung, daß die Allergierate in der Gruppe der Krebspatienten erheblich niedriger ist, was zu der Annahme führt, daß Allergie ein Schutzfaktor gegen Krebs sein könnte.

Schlüssel-Wörter: Dreiwegkontingenztafeln, log-lineare Modelle, Mantel-Haenszel-Test

regarding the site and type of the malignancy was reached. Meanwhile, they were seen in the same outpatient department by a research assistant who completed a special form about the history of bronchial asthma, hay fever, eczema, drug or food allergy and urticaria. The drug history was reliably ascertained by showing a chart with various medicines stuck on it. For comparative purposes, 519 non-cancer patients belonging to the same age groups from the general medical outpatient department were interviewed and had their allergy ascertained in the same way as the cancer group.

2. The Procedure

2.1. Models for Three-Dimensional Arrays

The appropriate saturated log-linear model (Goodman [2]) for any rectangular array in three dimensions is

$$v_{ijk} = \mu + \zeta_i^A + \zeta_j^B + \zeta_k^C + \zeta_{ij}^{AB} + \zeta_{ik}^{AC} + \zeta_{jk}^{BC} + \zeta_{ijk}^{ABC} \quad (1)$$

where

$v_{ijk} = \ln(m_{ijk})$, m_{ijk} being the expected count, $i = 1, \dots, I$; $j = 1, \dots, J$; $k = 1, \dots, K$.

The indexed terms on the right hand side of (1) sum to zero. These terms correspond to the overall average, the effect of variable A, variable B, variable C, the interaction between the variables A and B, A and C, B and C, A and B and C respectively. The main effects and the interactions refer to the natural logarithms of cell probabilities. The parameters ζ have superscripts to show to which variables they refer and subscripts to show to which category of that variable they apply.

A variety of unsaturated models are considered, searching for the most appropriate model that adequately fits the data:

1. For testing the null hypothesis of no three-factor interaction

$$H_0 = \zeta^{ABC} = 0.$$

$$\mu_{ijk} = \mu + \zeta_i^A + \zeta_j^B + \zeta_k^C + \zeta_{ij}^{AB} + \zeta_{ik}^{AC} + \zeta_{jk}^{BC}. \quad (2)$$

2. For testing the hypothesis of no two-factor interaction between A and B

$$H_0 = \zeta^{ABC} = \zeta^{AB} = 0.$$

$$\mu_{ijk} = \mu + \zeta_i^A + \zeta_j^B + \zeta_k^C + \zeta_{ik}^{AC} + \zeta_{jk}^{BC}. \quad (3)$$

3. For testing the hypothesis of no two-factor interaction between A and C

$$H_0 = \zeta^{ABC} = \zeta^{AC} = 0$$

$$\mu_{ijk} = \mu + \zeta_i^A + \zeta_j^B + \zeta_k^C + \zeta_{ij}^{AB} + \zeta_{jk}^{BC}. \quad (4)$$

2.2. Reduction of Two Dimensions

In a rectangular three-dimensional table a variable is collapsible (i.e. its categories can be combined without violating the structure of the table) with respect to the interaction between the other two variables if it is at least conditionally independent of one of the other two variables given the third (Theorem 2.4 – 1, Bishop et al. [1] p. 39). For example, variable C is only collapsible with respect to ζ^{AB} if the $I \times J \times K$ table is described by an unsaturated model with $\zeta^{ABC} = 0$ and either $\zeta^{AC} = 0$ or $\zeta^{BC} = 0$ or both.

2.3. The Mantel-Haenszel Statistic

The Mantel-Haenszel [4] statistic was used to assess the degree of

Table 1 Three-dimensional array relating cancer to allergy in six age intervals

Age Groups (years)	Allergy	Cancer		Exact Test P
		+	-	
10-19	+	2	14	0.1114
	-	21	36	
20-29	+	4	20	0.0053
	-	50	51	
30-39	+	7	31	0.0252
	-	77	124	
40-49	+	7	32	0.0019
	-	75	89	
50-59	+	10	15	0.0282
	-	101	53	
≥60	+	9	9	0.1526
	-	105	45	

association based on a weighted average of the K differences between proportions arising from the subtables of dimension $I \times J$. The value of this statistic is effectively equivalent to the difference between the goodness-of-fit likelihood ratio of the no 3-factor interaction model (2) and the no 2-factor interaction model (3).

3. Application

Table 1 is a cross-classification of cancer (variable A) according to allergy (variable B) and age (variable C). The Fisher-Irwin "exact" test indicates significant association of cancer and allergy in four age categories. Consequently, we may not safely collapse variable C without violating the structure of the 3-dimensional array. Another evidence of the inability to reduce the table dimension to two is the argument in section (2.1) that variable C is only collapsible with respect to ζ^{AB} if the $I \times J \times K$ table is described by an unsaturated model with $\zeta^{ABC} = 0$, (2), and either $\zeta^{AC} = 0$, (4) or $\zeta^{BC} = 0$ or both. Fitting the models in section (2.1) to Table 1, we found that the simultaneous test for the second-order interactions is highly significant (the likelihood ratio, $G^2 = 117.18$ at 11 degrees of freedom (d.f.), $p < 0.005$); but the third-order interaction, (model 2), was not significant and fits the data well ($G^2 = 1.42$ at 5 d.f., $p = 0.92$). Model (3) was not acceptable ($G^2 = 38.68$ at 6 d.f.,

$p < 0.005$), neither was model (4) ($G^2 = 67.44$ at 10 d.f., $p < 0.005$). Accordingly, we conclude that the hypothesis of no three-factor effect between cancer, allergy and age is acceptable. Neither of the other two-factor models fits the data, i.e. cancer occurrence is affected by allergy and by age but these two factors operate independently in their influence on cancer.

Model (2) with $\zeta^{ABC} = 0$ gives G^2 (1) = 1.42 with 5 d.f.; model (3) with $\zeta^{AB} = \zeta^{ABC} = 0$ gives G^2 (2) = 38.68 at 6 d.f. The difference G^2 (2) - G^2 (1) = G^2 (2/1) = 37.26 is distributed as χ^2 with 1 d.f. The Mantel-Haenszel chi-square statistic with continuity correction is 34.59 at 1 d.f. ($p < 0.005$) which is close to G^2 (2/1) apart from the continuity correction difference. Both values are less than the χ^2 value of 40.66 obtained from the condensed table by combining variable C (age). We conclude that we may not consider $\zeta^{AB} = 0$ because of the significant Mantel-Haenszel statistic which is based on a weighted average of the differences between proportions of allergy in cancer and non-cancer patients in the six age categories. This supports the evidence that we cannot collapse the table over variable C and that the role of age in the development of cancer cannot be ignored.

REFERENCES

- [1] Bishop, Y. M. M., Fienberg, S. E., Holland, P. W.: Discrete Multivariate Analysis: Theory and Practice. (Cambridge/Mass.: M. I. T. Press 1975).
- [2] Goodman, L. A.: The multivariate analysis of qualitative data: interactions among multiple classifications. J. Amer. Statist. Ass. 65 (1970) 226-256.
- [3] Karnik, A. M., Moussa, M. A. A., Omar, Y. T., Fenech, F. F.: Allergy and Cancer. A negative association (In press).
- [4] Mantel, N., Haenszel, W.: Statistical aspects of the analysis of data from retrospective studies of disease. J. nat. Cancer Inst. 22 (1959) 719-748.

Addresses of the authors:

Dr. M. A. A. Moussa,
A. M. Karnik,
F. F. Fenech,
Faculty of Medicine, Kuwait University,
P.O. Box 24923, Safat, Kuwait;
Dr. Y. T. Omar,
Cancer Control Centre, Kuwait.