Open Access Medical Statistics

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Analysis of tooth decay data in Japan using asymmetric statistical models

This article was published in the following Dove Press journal:
Open Access Medical Statistics
17 November 2012

Kouji Yamamoto[1]
Sadao Tomizawa[2]

[1]Department of Medical Innovation, Osaka University Hospital, Osaka, [2]Department of Information Sciences, Faculty of Science and Technology, Tokyo University of Science, Noda City, Chiba, Japan

Background: The aim of the present paper was to develop two new asymmetry probability models to analyze data for tooth decay from 363 women and 349 men aged 18–39 years who visited a dental clinic in Sapporo City, Japan, from 2001 to 2005.

Methods: We analyzed the probability relationship between grade of upper and lower tooth decay for men and women using the two new models, and tested goodness-of-fit for the models.

Results: The probability that a woman's (man's) grade of lower tooth decay is i ($i = 1,2$) and her (his) grade of upper tooth decay is $j(>i)$, ($j = 2,3$) is estimated to be at most 13.52 (10.23) times higher than the probability that the woman's (man's) grade of upper tooth decay is i and grade of lower tooth decay is j.

Conclusion: From the data on tooth decay, decay of the upper teeth is worse than of the lower teeth in women and men, and the tendency becomes stronger as the numbers of decayed upper and lower teeth increase.

Keywords: distance-proportional symmetry, asymmetry, square contingency table, teeth

Introduction

Consider the data in Table 1A and B.[1] Table 1B shows data for teeth decay from 363 women and 349 men aged 18–39 years who visited a dental clinic in Sapporo City, Japan, from 2001 to 2005. Table 1A and B show the numbers of decayed lower and upper teeth.[2] The categories are 0–4 decayed teeth; 5–8 decayed teeth; and >9 decayed teeth. From the data presented in Table 1A and B, it is likely that decay in the upper teeth is worse than in the lower teeth in both men and women. We are interested in using these data to determine the probability that a woman's (man's) grade of lower tooth decay is i and her (his) grade of upper tooth decay grade is $j(>i)$ is higher than the probability that her (his) grade of upper tooth decay is i and her (his) grade of lower tooth decay is j. For example, what is the probability that a woman's (man's) grade of lower tooth decay is 1 and her (his) grade of upper tooth decay grade is 3 is higher than the probability that her (his) grade of upper tooth decay grade is 1 and her (his) grade of lower tooth decay is 3?

Consider an $r \times r$ square contingency table with the same row and column classifications, as shown in Table 1. Let p_{ij} denote the probability that an observation will fall in the ith row and jth column of the table, and let n_{ij} denote the observed frequency in the ith row and jth column of the table ($i = 1, \ldots, r$; $j = 1, \ldots, r$). Also, let $\hat{p}_{ij} = n_{ij}/n$ and $n = \Sigma n_{ij}$.

Table 2 gives (1) the values of $(i - 1) + (j - 1)$ (say, d_{ij}), $i < j$, which is the sum of the distance of row value i from the base category 1 and the distance of column value j from the base category 1 for Table 1A, and the values of $(j - i)(i - 1 + j - 1)$

Correspondence: Kouji Yamamoto
Department of Medical Innovation, Osaka University Hospital, 2-15 Yamadaoka, Suita, Osaka, 565-0871, Japan
Tel +816 6879 5681
Fax +816 6879 5686
Email yamamoto-k@hp-crc.med.osaka-u.ac.jp

http://dx.doi.org/10.2147/OAMS.S35009

Table 1 Data on tooth decay for women and men in Japan

Decayed lower teeth (n)	Decayed upper teeth (n) 0–4 (1)	5–8 (2)	9+ (3)	Total
(A) Women with lower and upper tooth decay				
0–4				
(1)	97	62	15	174
	(97.0)[a]	(63.4)	(14.8)	
	(97.0)[b]	(62.7)	(15.8)	
5–8				
(2)	20	63	75	158
	(18.6)	(63.0)	(73.8)	
	(19.3)	(63.0)	(73.5)	
9+				
(3)	2	6	23	31
	(2.2)	(7.2)	(23.0)	
	(1.2)	(7.5)	(23.0)	
Total	119	131	113	363
(B) Men with lower and upper tooth decay				
0–4				
(1)	115	55	25	195
	(115.0)[a]	(55.7)	(22.9)	
	(115.0)[b]	(54.8)	(24.2)	
5–8				
(2)	16	49	60	125
	(15.3)	(49.0)	(61.4)	
	(16.2)	(49.0)	(61.0)	
9+				
(3)	1	7	21	29
	(3.1)	(5.6)	(21.0)	
	(1.8)	(6.0)	(21.0)	
Total	132	111	106	349

Note: [a,b]Maximum likelihood estimates of expected frequencies for the DIPS-I and DIPS-II models, respectively.

(say, s_{ij}), $i < j$, which is the product of the difference between the column value j and the row value i and the sum $(i - 1) + (j - 1)$, for Table 1B, (2) the values of $\{\hat{p}_{ij}/\hat{p}_{ji}\}$, $i < j$, for Table 1A and B, obtained under the saturated model, and (3) the values of $\{\hat{p}_{ij}/(\hat{p}_{ji}d_{ij})\}$, $i < j$, for Table 1A and the values of $\{\hat{p}_{ij}/(\hat{p}_{ji}s_{ij})\}$, $i < j$, for Table 1B, obtained under the saturated model.

For Table 1A, it is likely that the odds $\{p_{ij}/p_{ji}\}$, $i < j$, may be proportional to the sum of the distance, ie, $(i - 1) + (j - 1)$. Also for Table 1B, it is likely that the odds $\{p_{ij}/p_{ji}\}$, $i < j$, may be proportional to the product of the difference $(j - i)$ and the sum $(i - 1) + (j - 1)$. So we are interested in considering models which have such structures. Therefore, the present paper proposes new asymmetry models and analyzes the data in Table 1 using these new models.

Materials and methods

Reviews of statistical models

Being models which indicate the structure of asymmetry probability, the conditional symmetry model[3] and the diagonals-parameter symmetry model[4] are considered. The linear column-parameter symmetry model[1] is considered and used for the data in Table 1.

The diagonals-parameter symmetry model is given as:

$$\frac{p_{ij}}{p_{ji}} = \delta_{j-i} \quad (i < j). \tag{1}$$

A special case of the diagonals-parameter symmetry model is derived by making $\delta_1 = \ldots = \delta_{r-1}$ $(=\delta)$, ie, the conditional symmetry model, and another is derived by making $\delta_1 = \ldots = \delta_{r-1} = 1$, ie, the symmetry model.[5,6]

Also, the linear column-parameter symmetry model is given as:

$$\frac{p_{ij}}{p_{ji}} = \xi^{j-1} \quad (i < j). \tag{2}$$

For the data in Table 1, the linear column-parameter symmetry model indicates that the probability that a woman's (man's) grade of lower tooth decay is i and her (his) grade of upper tooth decay is $j(>i)$ is ξ^{j-1} times higher than the probability that her (his) grade of upper tooth decay is i and her (his) grade of lower tooth decay is j.

New models

We shall propose two new models. First, consider a model as follows:

$$\frac{p_{ij}}{p_{ji}} = d_{ij}\theta \quad (i < j), \tag{3}$$

where $d_{ij} = (i - 1) + (j - 1)$. This indicates that the odds, $\{p_{ij}/p_{ji}\}$ for $i < j$ are proportional to the sum of the distance of row value i from the base category 1 and the distance of column value j from the base category 1. Therefore, we will refer to this model as the distance-proportional symmetry I (DIPS-I) model.

For the data in Table 1, the DIPS-I model indicates that the probability that a woman's (man's) grade of lower tooth decay is i and her (his) grade of upper tooth decay is $j(>i)$ is $d_{ij}\,\theta$ times higher than the probability that the woman's (man's) grade of upper tooth decay is i and her (his) grade of lower tooth decay is j. If $\theta > 1$, we can see that a woman's (man's) upper teeth are worse than her (his) lower teeth, and this tendency becomes stronger as the numbers of decayed teeth increase.

Secondly, consider a model as follows:

$$\frac{p_{ij}}{p_{ji}} = s_{ij}\gamma \quad (i < j), \tag{4}$$

Table 2 Values of (1) $\{d_{ij}\}$ ($i < j$) for Table 1A and $\{s_{ij}\}$ ($i < j$) for Table 1B, (2) $\{\hat{p}_{ij}/\hat{p}_{ji}\}(i < j)$ for Table 1A and B, and (3) $\{\hat{p}_{ij}/(\hat{p}_{ji}d_{ij})\}$ ($i < j$) for Table 1A and $\{\hat{p}_{ij}/(\hat{p}_{ji}s_{ij})\}$ ($i < j$) for Table 1B

(1) $\{d_{ij}\}$ and $\{s_{ij}\}$					
$\{d_{ij}\}$ for Table 1A			$\{s_{ij}\}$ for Table 1B		
	$j = 2$	3		$j = 2$	3
$i = 1$	1	2	$i = 1$	1	4
2	–	3	2	–	3
(2) $\{\hat{p}_{ij}/\hat{p}_{ji}\}$					
For Table 1A			For Table 1B		
	$j = 2$	3		$j = 2$	3
$i = 1$	3.1	7.5	$i = 1$	3.44	25
2	–	12.5	2	–	8.57
(3) $\{\hat{p}_{ij}/(\hat{p}_{ji}d_{ij})\}$ and $\{\hat{p}_{ij}/(\hat{p}_{ji}s_{ij})\}$					
$\{\hat{p}_{ij}/(\hat{p}_{ji}d_{ij})\}$ for Table 1A			$\{\hat{p}_{ij}/(\hat{p}_{ji}s_{ij})\}$ for Table 1B		
	$j = 2$	3		$j = 2$	3
$i = 1$	3.1	3.75	$i = 1$	3.44	6.25
2	–	4.17	2	–	2.86

where $s_{ij} = (j - i)\,\{(i - 1) + (j - 1)\}$. This indicates that the odds $\{p_{ij}/p_{ji}\}$ for $i < j$, are proportional to the product of the difference between the column value j and the row value i and the sum, $(i - 1) + (j - 1)$. We shall denote this model as the distance-proportional symmetry II (DIPS-II) model.

For the data in Table 1, if $\gamma > 1$ under the DIPS-II model, we can see that a woman's (man's) upper teeth are worse than her (his) lower teeth, and the tendency becomes stronger as the difference between the number of decayed upper teeth and number of decayed lower teeth increases, and as the total numbers of decayed teeth increase.

Statistical analysis

Let m_{ij} denote the expected frequency of $\{n_{ij}\}$. Assume that $\{n_{ij}\}$ has a multinomial distribution. The maximum likelihood estimates of expected frequencies $\{m_{ij}\}$ under each model could be obtained, for example, using the Newton-Raphson method for log-likelihood equations. Each model can be tested for goodness-of-fit by, eg, the likelihood ratio Chi-squared statistic G^2 with the corresponding degrees of freedom defined by:

$$G^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{r} n_{ij}\log\left(\frac{n_{ij}}{\hat{m}_{ij}}\right), \tag{5}$$

where $\hat{m}_{ij}$ is the maximum likelihood estimate of m_{ij} under the model.[7] The numbers of degrees of freedom for the DIPS-I and DIPS-II models are both $(r + 1)(r - 2)/2$, which equal those for the conditional symmetry and linear column-parameter symmetry models.

Results

We analyzed the tooth decay data in Table 1 using the new models. Table 3 gives the values of the likelihood ratio test statistic G^2 for each model. From Table 3 we can see that the symmetry, conditional symmetry, and diagonals-parameter symmetry models fit these data poorly; however, the linear column-parameter symmetry, DIPS-I, and DIPS-II models fit the data very well. For the data in Table 1A (ie, for women), the DIPS-I model fits somewhat better than the DIPS-II model. For the data in Table 1B (ie, for men), the DIPS-II model fits better than the DIPS-I model.

For the DIPS-I model, the maximum likelihood estimate θ is $\hat{\theta} = 3.41$ using the data in Table 1A. Therefore, the probability that a woman's grade of lower tooth decay is i and her grade of upper tooth decay is $j(>i)$ is estimated to be $d_{ij} \times 3.41$ times higher than the probability that her grade of upper tooth decay is i and her grade of lower tooth decay is j. From the data in Table 1A, we see from $\hat{\theta} > 1$ that a woman's upper teeth are worse than her lower teeth, and the tendency becomes stronger as the total number of upper and lower teeth with decay increases.

Next, applying the DIPS-II model to the data in Table 1B, the maximum likelihood estimate of γ is $\hat{\gamma} = 3.38$. Therefore, the probability that a man's grade of lower tooth decay is i and his grade of upper tooth decay is $j(>i)$ is estimated to be $s_{ij} \times 3.38$ times higher than the probability that his grade of upper tooth decay is i and his grade of lower tooth decay is j. From the data in Table 1B, we see from $\hat{\gamma} > 1$ that a man's upper tooth decay is worse than his lower tooth decay, and that the tendency becomes stronger as the difference between the number of decayed upper teeth and that of decayed lower teeth increases, and as the total numbers of decayed teeth increase.

Discussion

Readers may be interested in guidelines for choosing between the DIPS-I and DIPS-II models. We might want to choose

Table 3 Values of likelihood ratio statistic G^2 applied to data in Table 1

Models used	Degrees of freedom	G^2	
		Table 1A	Table 1B
S	3	103.33*	98.24*
CS	2	9.40*	7.44*
DPS	1	9.18*	3.72*
LCPS	2	0.39	1.22
DIPS-I	2	0.39	2.57
DIPS-II	2	0.87	0.61

Note: *Significant at 0.05 level.
Abbreviations: S, symmetry; CS, conditional symmetry; DPS, diagonals-parameter symmetry; LCPS, linear column-parameter symmetry; DIPS-I, distance-proportional symmetry I; DIPS-II, distance-proportional symmetry II.

the DIPS-I model if we can assume that the odds $\{p_{ij}/p_{ji}\}$, $i < j$, are proportional to the sum of the distance of a row value from the base category 1 and the distance of a column value from the base category 1. On the other hand, we might want to choose the DIPS-II model if we can assume that the odds $\{p_{ij}/p_{ji}\}$, $i < j$, are proportional to the product of the difference between a column value and a row value, and the sum.

Some readers may also be interested in knowing the age distribution and other clinical characteristics of patients in Table 1 in addition to data on tooth decay. However, we took the data in Table 1 directly from another paper,[1] so cannot investigate those items in detail. Therefore, it is important to perform clinical research to collect the data on decayed teeth and related characteristics in the future and to analyze and compare them with the data presented in this paper in more detail.

Conclusion

We have proposed two kinds of asymmetry models, namely, the DIPS-I and DIPS-II. These models are useful for seeing the structure of asymmetry of cell probabilities $\{p_{ij}\}$, especially for the data on tooth decay as shown in Table 1. Upper tooth decay appears worse than lower tooth decay in both men and women, and the tendency becomes stronger as the sum of numbers of upper and lower decayed teeth increases.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Tomizawa S, Miyamoto N, Iwamoto M. Linear column-parameter symmetry model for square contingency tables: application to decayed teeth data. *Biometrical Letters*. 2006;43:91–98. Available at: http://www.up.poznan.pl/biometrical.letters/index.php?p=abstract&a=2006.43.2.3.
2. Kurakami H, Yamamoto K, Tomizawa S. Generalized exponential symmetry model and orthogonal decomposition of symmetry for square tables. *The Open Statistics and Probability Journal*. 2011;3:1–6. Available at: http://www.benthamscience.com/open/tospj/articles/V003/1TOSPJ.pdf.
3. McCullagh P. A class of parametric models for the analysis of square contingency tables with ordered categories. *Biometrika*. 1978;65:413–418. Available at: http://www.stat.uchicago.edu/~pmcc/pubs/paper3.pdf.
4. Goodman LA. Multiplicative models for square contingency tables with ordered categories. *Biometrika*. 1979;66:413–418. Available at: http://biomet.oxfordjournals.org/content/66/3/413.short.
5. Bowker AH. A test for symmetry in contingency tables. *J Am Stat Assoc*. 1948;43:572–574.
6. Bishop YMM, Fienberg SE, Holland PW. *Discrete Multivariate Analysis: Theory and Practice*. Cambridge, MA: MIT Press; 1975.
7. Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*. 1938;9: 60–62. Available at: http://www.jstor.org/stable/2957648?seq=1.