Statistical Method To Determine The Effect Of Sex And Age Variables On Some Eye Diseases

Mushtak A.K. Shiker
Department of mathematics – college of education - Ibn Hayan

Abstract
This research aims to analyze the impact of variables of sex and age on the incidence of some eye diseases, to identify some of the causes for these diseases and the possibility to avoid them early. We studied the impact of factors sex and age together on the incidence of two types of eye disease (cornea inflammation and iris inflammation) which is most common among people. I used for this purpose log-linear model to analyze the classification data in three-dimensional contingency tables.

Introduction
In 1971 Muller and Mayhall adjusted log-linear hierarchical models of cells frequencies in multi-Dimensional Contingency tables. In 1980 Sikkle concluded that the log-linear analysis is a convenient way to track the interactions of the higher levels. In 1981 Goodman L.A. analyzed the two-dimensional contingency table that contain two variables by using the log-linear models. In Iraq, and in the year 2001 there was an important study by researcher Qutaiba Nabil Nayef AL-Kazzaz about the analysis of classification data in the contingency tables by using bize method, he used the results of log-linear analyzed by using the maximum likelihood method and compares it with the results of log-linear analyzed by using bize method. The eye diseases still the most common diseases among the people and the most impact on their daily lives because it affects the most important sense blessed by Allaah to humans, namely, the grace of look. With scientific and civilization progress of humanity and the increasing of suffering of the people with these diseases, we find the science stands unable to find full treatment for such diseases, but it was noted that there are some factors related to the emergence of some eye diseases, whether these factors have directly relationship or when they are interacting with each other. So the research aims to explore the relationship between the incidence of disease (cornea inflammation and iris inflammation) and the factors of age and sex. It is no doubt that the use of statistical method and advanced contribute significantly and effectively to identify the most influential factor on eye disease, and thus try to confront the disease early, before it becomes too late so can not be cured. Since it may be not only one factor causes the eye diseases, but several factors may cause them, each one of the factors has alone effect, or there effect may be as a result of interaction among themselves in different degrees. The researcher found that the statistical method is appropriate and the most accurate in identifying those factors is the log-linear model. In this paper we used Q-basic program.
Sample

I took a real data from the records of Ibn Al-Haytham Hospital for eye disease – Baghdad – Iraq, of patients in 2001 and the sample size: \( N = 2724 \) patients were classified according to:

1. Type of eye disease: I take it here as dependent variable as follows:
   - A - cornea inflammation.
   - B - iris inflammation.
2. Sex: which was divided into two categories:
   - A - male
   - B - female
3. Age in years: which was divided into four categories:
   - A - less than 15
   - B - 15-44
   - C - 45-64
   - D - more than 64

The number of patients within the age groups above as follows:
- Age group (less than 15): 995 infected, a rate of 36.53%
- Age group (15-44): 755 patients, a rate of 27.72%
- Age group (45-64): 691 patients, a rate of 25.36%
- Age group (over 64): 283 infected, a rate of 10.39%

These data have developed in a table with a three-dimensional contingency table.

**Table(1)**

Cornea inflammation and Iris inflammation classified according to age and sex

<table>
<thead>
<tr>
<th>Age</th>
<th>Less than 15</th>
<th>15 – 44</th>
<th>45 - 64</th>
<th>More than 64</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>420</td>
<td>393</td>
<td>349</td>
<td>330</td>
</tr>
<tr>
<td>female</td>
<td>393</td>
<td>393</td>
<td>338</td>
<td>313</td>
</tr>
<tr>
<td>cornea inflam</td>
<td>349</td>
<td>349</td>
<td>330</td>
<td>313</td>
</tr>
<tr>
<td>iris inflam</td>
<td>338</td>
<td>338</td>
<td>330</td>
<td>313</td>
</tr>
<tr>
<td>Total</td>
<td>520</td>
<td>475</td>
<td>385</td>
<td>356</td>
</tr>
</tbody>
</table>

**Log-linear model:** [Kazaz, 2001] [Yoshimasa, Jun and Sophia, 2009]

In two-way contingency table and in both of independence test (between the variables of rows and columns) and the homogeneity test between the categories of rows variable (or columns variable), then the estimate of the expected value of the cell located in row \( i \) and column \( j \) where \( : i = 1, \ldots , r ; j = 1, \ldots , c \) Expressed as the following:

\[
m_{ij} = \frac{(x_i)(x_j)}{N} \quad \ldots \ldots \quad (1)
\]

Where:
- \( N \) is the sum of observed frequencies \( N = X.. \)
- \( X_i \) is the sum of observed frequencies in row \( i \).
- \( X_j \) is the sum of observed frequencies in column \( j \).

By taking the Logarithm we get:

\[
\log m_{ij} = \log x_i + \log x_j - \log x..
\]
From this we can write the Logarithm of expected value in the cell \((i, j)\) in contingency table \(I \times J\) in the form: 
\[
\log m_{ij} = u + u_{1i} + u_{2j} + u_{12ij} \ldots \ldots (2)
\]
Where \(u\) represents the arithmetic average of the Logarithm.

The model (2) is called the saturated model in two-dimensional contingency table, where all the variables and its effects of all grades appear in this model, starting with the effect of the first variable \(u_{1i}\), and the effect of the second variable \(u_{2i}\), then the effect between them \(u_{12ij}\). With respect to this research we have three variables, so we will have three-dimensional contingency table, therefore the logarithm saturated model will be in the form: 
\[
\log m_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk} + u_{123ijk} \ldots \ldots (3)
\]
where \(m_{ijk}\) represent the expected frequency of the number of patients with type \(I\) from eye disease \((I = 1, 2)\) of sex \(j\) \((j = 1, 2)\) of age group \(k\) \((k = 1, 2, 3, 4)\).

**Maximum Likelihood Estimate:** [Stephan, 2000]

This method of estimate used to find the cell expected in different contingency tables, it’s easy to calculate and possible to apply, and one of its important method is the direct method, which can used even if the data contain cells with zero values.

**Testing Goodness of Fit:** [Agresti, 1990]

After configuring the log-linear models under different assumptions in terms of independence or dependence between variables, comes the stage of testing these models. Then we test the goodness of fit of the model for data by using Likelihood-Ratio statistic \(G^2\) which its general form is:

\[
G^2 = \sum (\text{observed}) \log \left[ \frac{\text{observed}}{\text{expected}} \right] \ldots \ldots (4)
\]

Where \(\Sigma\) here is for all cells in the contingency table.

**Choosing the best model:** [Figueroa and Lopez, 2009], [Sufi, 1985]

In this work we used the Goodman way by using the method of partition the \(G^2\) to associative components to choose the best model. To illustrate this method, we must first modeling nested hierarchy models as in hierarchy order as the following:

- \(\log m_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij}\)
- \(\log m_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{23jk}\)
- \(\log m_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{13ik} + u_{23jk}\)
- \(\log m_{ijk} = u + u_{1i} + u_{2j} + u_{3k} + u_{12ij} + u_{13ik} + u_{23jk}\)

Then we calculate the value of \(G^2\) for each model, the following condition must be satisfied:

\[
G^2(D1) \geq G^2(D2) \geq G^2(D3) \geq G^2(D4) \ldots \ldots (5)
\]

We identify the level of morale given \((0.05)\), then there will be a series of goodness of fit tests corresponding to the partitioned compounds, we start with the compound corresponding with the most complicated model, the test will stop when the difference between two respectively models morally under the null hypothesis, i.e hypothesis that the proposed model is correct.

**Application on the research sample:**

The expected values and \(G^2\) and the degrees of freedom as shown in the following tables:
From the tables above we note that the requirement for test models verified: As: $G^2(D1) \geq G^2(D2) \geq G^2(D3) \geq G^2(D4)$

Where: $61.0191 > 60.5515 > 1.6511 > 0.6822$

From the results above we conclude that the fourth model is the best model to represent observed data and show the relationship between the type of eye disease (dependent variable) and the other two variables, where the value of the Likelihood-Ratio test $G^2 = 0.6822$ under level of morale given (0.05), That is, we do not reject the hypothesis that the above model fits the relationship between the variables (and we reject the hypothesis that says the opposite), and this requires the presence of bilateral interactions in the model.

Conclusions
According to the fourth model above:
1. We find that the hypothesis: $H_0: u_{12ij} = 0$ for all $i, j$ Rejected under level of morale given (0.05), that is the result of $G^2$ is very morale, that's mean the eye diseases is not independent of the patient sex and the patient age, spirits too and this means that a disease eye is not independent of the sex of the patient, so we note that the injuries in males grater than its in females in all age groups, and we note that the injuries in first age group grater than the next one, and so on, without exception, and this confirms the close relationship between eye diseases with gender from side and with age from another side. 2. For the same reason as set out in point (1) we find that the hypothesis: $H_0: u_{13ik} = 0$ for all $i, k$
Rejected at the level of morale 0.05, and this means that there is a close relationship between the type of eye disease and the patient's age, that is the eye diseases (Cornea inflammation and Iris inflammation) concentrated in first age group and then it decreases in next age groups ,which clearly shows us the possibility of diagnosing the disease early and prevent aggravation with age.

**Recommendations**

I recommend the need to facilitate the task of researchers and to provide maximum services to them in order to facilitate the completion of their research, because I suffered many in order to obtain correct and accurate data from the real sources. I also recommend the competent health authorities need to record all the detailed and accurate data for people with eye disease in order to provide for students.

**References**


Sufi, Abdul Majid Rashid , 1985 "Kai-test 2 (² X) and its uses in statistical analysis", published by Dar al struggle for printing and publishing and distribution.