www.ijramr.com

ISSN : 2350-0743

International Journal of Recent Advances in Multidisciplinary Research

Vol. 04, Issue 11, pp.2986-2989, November, 2017

RESEARCH ARTICLE

STATISTICAL ESTIMATION WITH GENE FREQUENCIES IN A B O BLOOD GROUP SYSTEMS

*,[1]Madhu Sudan, G., [1]Vishnupriya, P., and [2]Prof. Abbaiah, R.

[1]Research Scholar, Department of Statistics, Sri Venkateshwara University, Tirupati-517502, India
[2]Principal and Dean SVUCS, Head Department of Statistics, Sri Venkateshwara University, Tirupati-517502

ARTICLE INFO

Article History:
Received 19th August, 2017
Received in revised form
27th September, 2017
Accepted 04th October, 2017
Published online 11th November, 2017

Keywords:
Biostatistics,
ML Method Gene frequencies,
ABO Blood group System in Human Beings.

ABSTRACT

Statistics is one of the major tools to study the quantitative aspects of Biological theories. In the literature about Biostatistics, many statisticians have applied basic biometric methods for medicinal applications where as a few statisticians have applied those for genetic applications. In the present study, an attempt has been made by estimating gene frequencies in ABO blood group system in human beings through the Maximum Likelihood method.

INTRODUCTION

Modern Statistics and Biostatistics are not mere devices for collecting numerical data but means of developing sound techniques for handling and analyzing the data and drawing inference from the collected data using various techniques and formulations. At present it is impossible to identify even a single field of human activity whether it is Biology, Agriculture, Business management, Economics, Demography, Geography, Education, and Medicine -which can do without statistics. Biostatistics or Biometrics: The application of statistics to the study and analysis a biological and medical dataQuantitative Genetics: The field of genetics that studies the mode of inheritance of Complex or Quantitative Traits. The Heredity units which are transmitted from one generation to the next generation are called GENES.Quantitative methods are research techniques and methods dealing with numbers and anything that is measurable.

Computation of Gene frequencies by Maximum Likelyhood (ML) Method

A random mating population with respect to O, A, B and AB has observed frequencies 216, 84, 223 and 57 respectively. Calculate gene frequencies by the method of Maximum Likelihood (ML) and also test the goodness of fit.

A random mating population with respect to O, A, B and AB gene frequencies r, p, and q respectively.Then, the number in the Multinomial distribution. So that probability of observing

$n_1\ O's, n_2\ A's, n_3\ B's\ and\ n_4\ AB's\ are\ given\ by$

$$L=\frac{n!}{n_1!n_2!\,n_3!\,n_4!}\left(r^2\right)^{n_1}\left(p^2+2pr\right)^{n_2}\left(q^2+2qr\right)^{n_3}\left(2pq\right)^{n_4}$$

Taking log on both sides, then we get

$$\log L=k+n_1\log r^2+n_2\log\left(p^2+2pr\right)+n_3\log\left(q^2+2qr\right)+n_4\log\left(2pq\right)$$

***Corresponding author:** **Madhu Sudan, G.,**
Research Scholar, Department of Statistics, Sri Venkateshwara University, Tirupati-517502, India

$$\log L = k + 2n_1 \log r + (\log p)(n_2 + n_4) + (\log q)(n_3 + n_4) + n_2 \log(2 - p - 2q) + n_3 \log(2 - 2p - q)$$

$$\log L = k + 2n_1 \log(1 - p - q) + n_2 \{\log p + \log(2 - p - 2q)\} + n_3 \{\log q + \log(2 - 2p - q)\} + n_4 \{\log p + \log q\}$$

$$\log L = k + 2n_1 \log(1 - p - q) + \log p(n_2 + n_4) + n_2 \log(2 - p - 2q) + n_3 \log(2 - 2p - q) \quad (1)$$

$$where\ k = \log\left(\frac{n!}{n_1!n_2!n_3!n_4!}\right);$$

$r = 1 - p - q$; $p = 1 - \sqrt{\frac{n_1 + n_3}{n}}$; $q = 1 - \sqrt{\frac{n_1 + n_2}{n}}$; $n = n_1 + n_2 + n_3 + n_4$

Maximizing equation (1), we have

$$\frac{\delta \log L}{\delta p} = 0 \Rightarrow \frac{-2n_1}{(1 - p - q)} + \frac{(n_2 + n_4)}{p} - \frac{n_2}{(2 - p - 2q)} - \frac{2n_3}{(2 - 2p - q)} = \partial_1$$

$$\frac{\delta \log L}{\delta q} = 0 \Rightarrow \frac{-2n_1}{(1 - p - q)} + \frac{(n_3 + n_4)}{q} - \frac{2n_2}{(2 - p - 2q)} - \frac{n_3}{(2 - q - 2p)} = \partial_2$$

$$\frac{\delta^2 \log L}{\delta p^2} = 0 \Rightarrow \frac{-2n_1}{(1 - p - q)^2} - \frac{(n_2 + n_4)}{p^2} - \frac{n_2}{(2 - p - q)^2} - \frac{4n_3}{(2 - 2p - q)^2} = \partial_{11}$$

$$\frac{\delta^2 \log L}{\delta q^2} = 0 \Rightarrow \frac{-2n_1}{(1 - p - q)^2} - \frac{(n_3 + n_4)}{q^2} - \frac{4n_2}{(2 - p - 2q)^2} - \frac{n_3}{(2 - 2p - q)^2} = \partial_{22}$$

$$\frac{\delta^2 \log L}{\delta p \delta q} = 0 \Rightarrow \frac{-2n_1}{(1 - p - q)^2} - \frac{2n_2}{(2 - p - 2q)^2} - \frac{2n_3}{(2 - q - 2p)^2} = \partial_{12}\ or\ \partial_{21}$$

Starting from p_0 , q_0 we may obtain new approximation to maximum likelihood solution by solving the following equations

$$\partial_{11}(p_1 - p_0) + \partial_{12}(q_1 - q_0) = \partial_1$$

$$\partial_{12}(p_1 - p_0) + \partial_{22}(q_1 - q_0) = \partial_2$$

The solutions of these equations are

$$p_1 - p_0 = \left[\frac{\partial_{22}}{\partial_{11}\partial_{22} - \partial_{12}^{\ 2}}\right]\partial_1 - \left[\frac{\partial_{12}}{\partial_{11}\partial_{12} - \partial_{12}^{\ 2}}\right]\partial_2$$

$$q_1 - q_0 = \left[\frac{-\partial_{12}}{\partial_{11}\partial_{22} - \partial_{12}^{\ 2}}\right]\partial_1 + \left[\frac{\partial_{12}}{\partial_{11}\partial_{12} - \partial_{12}^{\ 2}}\right]\partial_2$$

Now we will get the estimates of $p, q, r\ \ as\ \ \hat{p}, \hat{q}, \hat{r}\ \ respectively$

Test for Goodness of fit

To test for goodness of fit, first we state the null hypothesis as,

H_0: There is close agreement between observed and expected frequencies

To test the above null hypothesis, we use chi-square test as follows:

Blood Group	Observed Frequencies	Expected Frequencies	$(O_i - E_i)^2 / E_i$
O	n_1	nr^2	
A	n_2	$n(p^2 + 2pr)$	
B	n_3	$n(q^2 + 2qr)$	
AB	n_4	$n(2pq)$	
Total	$\sum_{i=1}^{n} O_i = n$	$\sum_{i=1}^{n} E_i = n$	$\chi^2_{cal\ val}$

Compare chi-square calculated value with chi-square table value at $\alpha \%$ level of significance.

Analysis

n_1 =O=216 n_2 =A=84 n_3 =B=223 n_4 =AB=57 n=580

$$p = 1 - \sqrt{\frac{n_1 + n_2}{n}} \Rightarrow 1 - \sqrt{\frac{439}{580}} = 1 - \sqrt{0.7569} = 0.13$$

$$q = 1 - \sqrt{\frac{n_1 + n_3}{n}} \Rightarrow 1 - \sqrt{\frac{300}{580}} = 1 - \sqrt{0.5172} = 0.2808$$

$$r = 1 - p - q \Rightarrow 1 - 0.13 - 0.2808$$

$$\partial_1 = \frac{-2n_1}{(1-p-q)} + \frac{(n_2 + n_4)}{p} - \frac{n_2}{(2-p-2q)} - \frac{2n_3}{(2-2p-q)}$$

$$\partial_1 = \frac{-2(216)}{0.5892} + \frac{141}{0.13} - \frac{84}{1.3084} - \frac{446}{1.4592}$$

$$\partial_1 = -18.4297$$

$$\partial_2 = \frac{-2n_1}{(1-p-q)} + \frac{(n_3 + n_4)}{q} - \frac{2n_2}{(2-p-2q)} - \frac{n_3}{(2-q-2p)}$$

$$\partial_2 = \frac{-2(216)}{0.5892} + \frac{280}{0.2808} - \frac{2(84)}{1.3084} - \frac{223}{1.4592}$$

$$\partial_2 = -17.2712$$

$$\partial_{11} = \frac{-2n_1}{(1-p-q)^2} - \frac{(n_2 + n_4)}{p^2} - \frac{n_2}{(2-p-q)^2} - \frac{4n_3}{(2-2p-q)^2}$$

$$\partial_{11} = \frac{-2(216)}{(0.5892)^2} - \frac{141}{(0.13)^2} - \frac{84}{(2-0.13-0.2808)^2} - \frac{4(223)}{(2-2(0.13)-0.2808)^2}$$

$$\partial_{11} = -10039.7667$$

$$\partial_{22} = \frac{-2n_1}{(1-p-q)^2} - \frac{(n_3 + n_4)}{q^2} - \frac{4n_2}{(2-p-2q)^2} - \frac{n_3}{(2-2p-q)^2}$$

$$\partial_{22} = \frac{-2(216)}{(0.5892)^2} - \frac{280}{0.0788} - \frac{4(84)}{(2-0.13-2(0.2808))^2} - \frac{223}{(2-2(0.13)-0.2808)^2}$$

$$\partial_{22} = -5098.6969$$

$$\partial_{12} \ or \ \partial_{21} = \frac{-2n_1}{(1-p-q)^2} - \frac{2n_2}{(2-p-2q)^2} - \frac{2n_3}{(2-q-2p)^2}$$

$$\partial_{12} \ or \ \partial_{21} = \frac{-2(216)}{(0.5892)^2} - \frac{2(84)}{(2-0.13-2(0.2808))^2} - \frac{2(223)}{(2-0.2808-2(0.13))^2}$$

$$\partial_{12} = -1582.2546$$

$$\partial_{11}(p_1 - p_0) + \partial_{12}(q_1 - q_0) = \partial_1$$
$$\partial_{12}(p_1 - p_0) + \partial_{22}(q_1 - q_0) = \partial_2$$

$$-10039.7667(p_1 - p_0) - 1582.2546(q_1 - q_0) = -18.4297$$

$$-1582.2546(p_1 - p_0) - 5098.6969(q_1 - q_0) = -17.2712$$

$$p_1 - p_0 = \left[\frac{\partial_{22}}{\partial_{11}\partial_{22} - \partial_{12}^2}\right]\partial_1 - \left[\frac{\partial_{12}}{\partial_{11}\partial_{12} - \partial_{12}^2}\right]\partial_2$$

$$p_1 - p_0 = 0.0001$$

$$q_1 - q_0 = \left[\frac{-\partial_{12}}{\partial_{11}\partial_{22} - \partial_{12}^2}\right]\partial_1 + \left[\frac{\partial_{12}}{\partial_{11}\partial_{12} - \partial_{12}^2}\right]\partial_2$$

$$q_1 - q_0 = -0.0136$$

Blood Group	Observed Frequencies	Expected Frequencies	$(O_i - E_i)^2 / E_i$
O	216	201	1.1191
A	84	99	2.2727
B	223	238	0.9454
AB	57	42	5.3571
Totals	580	580	9.6946

Here $\chi^2_{cal\ val} = 9.2177$ and $\chi^2_{cri\ val} = \chi^2_{n-k-1} = \chi^2_{4-2-1} = 3.84$

$\therefore\ \chi^2_{cal\ val} \le \chi^2_{cri\ val}$ *then* H_0 *is accepted at* 5% *level of significant.*

Conclusion

In the present research study the method of estimating gene frequencies in A B O Blood group systems we are determined Maximum Likelyhood method after simplification we conclude that there is close agreement between observed and expected gene frequencies.

REFERENCES

Agarwal, B.L. and Agarwal, S.P. 2007. "Statistical Analysis of Quantitative Genetics" New Age International Publishers, New Delhi.

Basu, S.B. 1996. "Quantitative Genetics Research Technique", Kalyani Publishers, New Delhi.

Fisher R.A. 1925. Theory of Statistical estimation, Proc. Camb. Phil. Soc., 22, 700-725.

Mahajan, B.R. 1989. "Methods in Biostatistics for Medical Students and Research Workers", Jayper Brothers Medical Publishers (p) Ltd, New Delhi.

****