International Journal of Engineering & Technology, 7 (4.10) (2018) 508-510

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Estimation of Gene Frequencies in Clinical Research

N. Vijayalakshmi [1], Dr .P.Sekhar[2], Dr.G.Mokesh Rayalu[3*]

[1]*Research Scholar ,Department of Mathematics, Madras University,Tamilnadu*
[2]*Dean,Science and Humanities, SRM Institute of science and technology, Ramapuram, Tamilnadu*
[3]*Assistant Professor, Department of Mathematics, VIT, Vellore, Tamilnadu.*
**Corresponding author E-mail: mokesh.g@vit.ac.in*

Abstract

Biometrics is a branch of statistics in which various mathematical and statistical techniques can be applied to biological research problems. These are two main areas of specialization of Biometry namely, Bioassays and Quantitative Genetics. Genetics concerns with Heredity and variation. Quantitative Genetics is concerned with the inheritances of quantitative differences between individuals.
The essence of Quantitative Genetics is to estimate the genetic parameters such as Gene frequencies, segregation Ratios, Recombination of Genes and so on. Among them, the estimation of Gene Frequencies in the population is an important one. The proportion or percentage of genes in the population is called gene Frequency. In the present research articles, the ABO blood group system of man has been described by discussing the multiple alleles; genotypes, Frequencies and phenotypes of blood groups. The various estimation methods for estimating gene frequencies have gene presents in the present study.

Keywords: *Biostatistics, Bioassays, Quantitavive genetics, variance covariance matrix, linear pseudo model, least squares estimator, degrees of freedom.*

1. Introduction

Biostatistics or Biometrics is a branch of either Biology or Applied statistics which deals with the application of statistical techniques to study various research problems in the Biology. Biometrical Genetics is a branch of Genetics that uses various statistical concepts and methods to study the genetical research problems sometimes it is known as statistical genetics or mathematical genetics. This may be divided into two major parts namely (i) Quantitative Genetics and ii) Population Genetics. Population Genetics deals with the Gene frequencies, and Genotypes and phenotypes in the mendelian population. These populations have two main features namely a) Random mating and b) Equal survival of all genotypes.
Under the inferential aspects of Genetic parameters, one may consider the problems of estimation of genetic parameters, for instance, segregation ratios, Gene frequencies, Recombination values, Heritability coefficients, Repeatability coefficients, Genetic correlation coefficients and so on. Then one may turn to the examination of goodness of fit with the estimated genetic parameters and to the comparison of sets of data.
The proposed study deals with the problems of estimation of some genetic parameters connected with the O,A,B blood group system of Human being.

2. A B O Blood Group System

Multiple alleles can be found in A B O blood group system of a man, where the allele I^A for the A antigen is codominant with the allele I^B for the B antigen. Both I^A and I^B are completely dominant to the allele i which fails to specify any detectable antigenic structure. The genotypes and phenotypes are given by:

Genotypes	Genotypic frequencies	Phenotypes Blood groups
I^AI^A	p^2	A
I^{Ai}	2 p r	
I^BI^B	q^2	B
I^{Bi}	2 q r	
I^AI^B	2 p q	AB
i i	r^2	O

Here p = frequency of I^A allele, q = frequency of I^B allele,r = frequency of i allele.
The A B O blood group system is governed by a multiple allelic system in which some codominant relationships exist. Three alleles I^A, I^B and i form the dominance hierarchy $(I^A=I^B)>i$.
The genotypic and phenotypic expectations for this blood group locus from a population in genetic equilibrium and a formula for use in finding the allelic frequencies at the A B O blood group locus can be derived as follows.
Let p = frequency of I^A allele, q= frequency of I^B allele, r =frequency of i allele. The expansion of $(p+q+r)^2$ gives the zygotic ratio expected under random mating.
Let $\bar{A}$, $\bar{B}$ and $\bar{O}$ represent the phenotypic frequencies of A, B and O respectively.
One may have $r^2 = \bar{o} \Rightarrow \boxed{r = \sqrt{\bar{o}}}$ = frequency of recession allele i
Consider,

$$p^2 + 2pr + r^2 = \left(\bar{A} + \bar{O}\right)$$

$$\Rightarrow (p+r)^2 = \bar{A} + \bar{O} \Rightarrow p + r = \sqrt{\bar{A} + \bar{O}}$$

$$\Rightarrow p = \sqrt{\bar{A} + \bar{O}} - r = \sqrt{\bar{A} + \bar{O}} - \sqrt{\bar{O}}$$

$$\Rightarrow \boxed{p = \sqrt{\bar{A} + \bar{O}} - \sqrt{\bar{O}}}$$ = frequency of I^A allele again, consider

$$q^2 + 2qr + r^2 = \left(\bar{B} + \bar{O}\right)$$

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$\Rightarrow (q+r)^2 = \bar{B} + \bar{O}$

$\Rightarrow q+r = \sqrt{\bar{B}+\bar{O}}$

$\Rightarrow \boxed{q = \sqrt{\bar{B}+\bar{O}} - \sqrt{\bar{O}}}$ = frequency of I^B allele.

One also set q as q=1-p-r. One may obtain,

$\sqrt{\bar{A}+\bar{O}} - \sqrt{\bar{O}} + \sqrt{\bar{B}+\bar{O}} - \sqrt{\bar{O}} + \sqrt{\bar{O}} = 1.0$

(or) p+q+r = 1.0

$\Rightarrow p = 1-(q+r) = 1-\sqrt{\bar{B}+\bar{O}}$

$q = 1-(p+r) = 1-\sqrt{\bar{A}+\bar{O}}$

$r = \sqrt{\bar{O}}$

3. Estimation of Gene Frequencies in A B O Blood Group System

Suppose p, q and r denote the gene frequencies for A, B, O, in A B O blood group system. Assume that the mating is random mating. We have

Type	Probability	Observed number	Observed proportion
O	r^2	n_1	n_1/n
A	$p^2+2\,p\,r$	n_2	n_2/n
B	$q^2+2\,q\,r$	n_3	n_3/n
AB	$2\,p\,q$	n_4	n_4/n
Total	1	n	l

Method 1:

A procedure to estimate r is $\hat{r} = \left(\frac{n_1}{n}\right)^{1/2}$ an estimate for p as $\hat{p}$, which is a solution of quadratic equation in p namely

$$p^2 + 2\,p\,\hat{r} - \frac{n_2}{n} = 0. \quad (1)$$

Similarly an estimate for q as $\hat{q}$, which is a solution of quadratic equation in q as

$$q^2 + 2\,p\hat{r} - \frac{n_3}{n} = 0 \quad (2)$$

Method 2:

According to Bernstein: 1930, the estimates are given by $\hat{r} = \left(\frac{n_1}{n}\right)^{1/2}$,

One may have from the table,

$$r^2 + p^2 + 2pr = \left(\frac{n_1+n_2}{n}\right)^{1/2}$$

$$\Rightarrow (p+r)^2 = E\left(\frac{n_1+n_2}{n}\right)$$

Since q = 1-(p+r), $\hat{q}$ is given by

$$\hat{q} = 1-\left(\frac{n_1+n_2}{n}\right)^{1/2} \quad (3)$$

Similarly, from the table,

$$q^2 + 2qr + r^2 = \left(\frac{n_1+n_3}{n}\right)$$

$$\Rightarrow (q+r)^2 = E\left(\frac{n_1+n_3}{n}\right) \quad (4)$$

Since, $p = 1-(q+r)$, $\hat{p}$ is given by

$$\hat{p} = 1-\left(\frac{n_1+n_3}{n}\right)^{1/2}$$

Here, it should be noted that p+q+r = 1,

But, in general, $\hat{p}+\hat{q}+\hat{r} \neq 1$.

Method 3:

Bernstein: 1930, suggested the following estimation method, which is asymptotically efficient:

Consider the initial estimates for p, q,r as

$$p' = 1-\left(\frac{n_1+n_3}{n}\right)^{1/2}$$

$$q' = 1-\left(\frac{n_1+n_2}{n}\right)^{1/2}$$

$$r' = \left(\frac{n_1}{n}\right)^{1/2}$$

Define D = 1- $(p'+q'+r')$ then the estimates are given by

$$\hat{p} = p'\left(1+\frac{D}{2}\right)$$

$$\hat{q} = q'\left(1+\frac{D}{2}\right)$$

$$\hat{r} = \left(r'+\frac{D}{2}\right)\left(1+\frac{D}{2}\right) \quad (5)$$

It should be noted that

$$\hat{p}+\hat{q}+\hat{r} = (p'+q')\left(1+\frac{D}{2}\right) + \left(r'+\frac{D}{2}\right)\left(1+\frac{D}{2}\right)$$

$$= \left(p'+q'+r'+\frac{D}{2}\right)\left(1+\frac{D}{2}\right)$$

$$= \left(1-D+\frac{D}{2}\right)\left(1+\frac{D}{2}\right)$$

$$\because D = 1-p'-q'-r'$$

$$= \left(1-\frac{D}{2}\right)\left(1+\frac{D}{2}\right)$$

$$\boxed{\hat{p}+\hat{q}+\hat{r} = 1-\frac{D^2}{4}} \quad (6)$$

By using the goodness of fit, χ^2 is given by

$$\chi^2 = 2\,n\left(1+\frac{\hat{r}}{\hat{p}\,\hat{q}}\right)D^2 \quad (7)$$

Method 4: Method of Minimum χ^2

It is known that χ^2 is a function of the parameters and the method of minimum χ^2 estimates are obtained by minimizing χ^2 w.r.t the parameters.

χ^2 goodness of fit criterian is given by

$$\chi^2 = \frac{\left(n_1 - np^2\right)^2}{np^2} + \frac{\left[n_2 - 2np(1-p)\right]^2}{2np(1-p)} + \frac{\left[n_3 - n(1-p)^2\right]^2}{n(1-p)^2} \qquad (8)$$

I.

By minimizing χ^2 w.r.t p gives the minimum χ^2 estimate for p.

Method 5: Scoring Method

Consider the case of unselected data on a sample of individuals, some of whom are related and assume that there is one parameter to be estimated.

First write the, joint likelihood function (L) of the whole sample and then find the solution to the equation

$$\frac{d(\log L)}{d\theta} = 0$$

Where θ is the parameter to be estimated?

Under the method of scoring, the scores may be assigned to individuals based on the way they occur in the sample. It is equivalent to solving the likelihood equation and differs from it only in that the scores of the unrelated individuals are forced to be zero or unity, as they would be in a simple enumeration of unrelated individuals.

If the individuals in the populations are denoted by i=1,2,…, n with true probabilities p_i, the maximum likelihood equation is

$$\frac{d\log L}{d\theta} = 0 \Rightarrow \sum_{i=1}^{n} \frac{d}{d\theta}(\log p_i) = 0$$

Assuming the initial value of θ as θ_0, then, by the Taylor series expansion, one may have

$$\sum \frac{d}{d\theta}(\log p_i) = \sum \frac{d}{d\theta}(\log p_i)\Big|_{\theta=\theta_0} + (\theta - \theta_0)\sum \frac{d^2}{d\theta^2}(\log p_i)\Big|_{\theta=\theta_0} + R$$

Where R is the remainder.

If the L.H.S to be zero and ignore R, one may obtain,

$$\hat{\theta} = \theta_0 + \frac{\sum \frac{d}{d\theta}(\log p_i)\Big|_{\theta=\theta_0}}{-\sum \frac{d^2}{d\theta^2}(\log p_i)\Big|_{\theta=\theta_0}}$$

$$\mathrm{var}(\hat{\theta}) = \frac{1}{-\sum \frac{d^2}{d\theta^2}(\log p_i)\Big|_{\theta=\theta_0}}$$

By writing w_i as $-\frac{d^2}{d\theta^2}(\log p_i)\Big|_{\theta=\theta_0}$, the estimate of θ under the method of scoring is given by

$$\hat{\theta} = \frac{\sum \left[w_i \theta_0 \frac{d}{d\theta}(\log p_i)\Big|_{\theta=\theta_0} \right]}{\sum w_i} \qquad (9)$$

Here, W_i is the weight contributed by the individual;

The individual term in the numerator is known as the score of the i^{th} individual.

The aforementioned scores may be modified by using scores of 0 and 1 and weight $\left[\theta(1-\theta)\right]^{-1}$ for individuals whose probabilities are (1-θ) and θ respectively. One may eliminate the possibility of negative weights by using expected weights w_i in place of w_i^* where

$$w_i^* = \frac{1}{p_i^2}\left(\frac{dp_i}{d\theta}\right)^2$$

4. Conclusions

Statistical Genetics is concerned with the inheritances of quantitative differences between individuals. The theory of Quantitative or Biometrical Genetics came into existence on the basis of experiments conducted on pea plants by Johann Gregor Mendel, an Australian Biologist in year 1866. People called Mendel as 'Father of Genetics' and he died in year 1884 and never lived to see that he had open the water gates to a new branch of Biology called 'Genetics'. The theoretical basis of Statistical Genetics was established around about 1920 by the works of Fisher: 1918, Haldane: 1932 and Wright: 1921. Bio-statistical Techniques have a wide number of practical applications in Genetics and plant breeding. The essence of quantitative genetics is to estimate the genetic parameters such as gene frequencies, segregation ratios, recombination of genes and soon. Among them, the estimation of gene frequencies in the population is an important one in the literature.

In the present study, the ABO blood group system of man has been described by discussing the multiple alleles; genotypes, Gene Frequencies and phenotypes of blood groups. The various estimation methods for estimating gene frequencies have been presented in the study.

References

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

[2] Basu, S.B. (1996), "Quantitative Genetics Research Technique", Kalyani Publishers, New Delhi.

[3] Becker, W.A. (1975), "Manual of Quantitative Genetics", Student Book corporation, Pullman, Washington, U.S.A.

[4] Boyd, William C. 1956. Variances of gene frequency estimates. Amer. J. Hum. Genet. 8, 24- 38.

[5] Cotterman, C.W. 1954. Estimation of gene frequencies in non-experimental populations. Chapter 35 of Statistics and mathematics in biology, edited by O. Kempthore et al. Iowa State College Press. Ames.

[6] Fisher, R.A. (1952). Statistical methods in genetics, Heredity, 6, 1-12.

[7] Kempthorne, O. (1957), "An Introduction to Genetic Statistics", John Wiley and Sons, New York.

[8] Mather, K., and J.L. Jinks (1982): Biometrical Genetics, 3rd ed., Chapman and Hall, London.

[9] Narain, P. (1990): Statistical Genetics. Wiley Eastern, New Delhi.

[10] Rao, C.R. (1952), "Advanced Statistical Methods in Biometric Research", Wiley and sons.

[11] Stevens, W.L. (1938), "Estimation of blood-group gene frequencies", Ann. Eugen. Lond., 8, 362-375.

[12] Wright, Sewall, (1931), "Statistical methods in biology", J. Amer. Statist. Ass. 26, Suppl., 155-163.