© Kamla-Raj 2008 Anthropologist, 10(2): 139-141 (2008)

Statistical Analysis of Gene Frequencies for Mothers and Babies

K. Senthamaraikannan[#], D. Nagarajan[*] and P. Arumugam[#]

 # Manonmanium Sundaranar University, Department of Statistics, Tirunelveli, India
** St. Jopseph's College of Engineering, Chennai 119, Tamil Nadu, India*
E-mail: dnrmsu2002@yahoo.com

KEYWORDS Bernstein's Methods, Maximum Likelihood Estimation, Score Methods.

ABSTRACT This paper analysis the estimation of the gene frequencies and stationary distributions. This paper also compares the findings with maximum likelihood estimates and score methods. The homogeneity of mothers and babies blood groups are examined whether the frequencies in each groups are in accordance with Bernstein's hypothesis and test or the gene frequencies are the same in all population are also analysed.

INTRODUCTION

A blood type describes a set of 26 proteins on the surface of red blood cells and other tissues which varies among individuals. These proteins are important because they contain specific stretches of amino acid and carbohydrate antigens which can be recognized by the immune system. Upon recognition by the immune system, antibodies against these antigens are produced. These antibodies attach to the antigens, often leading to destruction of the cells by other components of the immune system. Knowledge of a person's blood type allows identification of appropriate blood and tissue for transfusion and transplantation.

DATABASE

To accomplish the above said objective, initially data's were collected from 1004 hospital records from Governmental hospital .Tirunelveli, Tamil Nadu . delivering singleton live birth babies, makes the experiment to study. The frequencies A=158,B=5058,AB=121 and O=220 were collected during Feb,2006 to April 2006.

Description of Model

If r, p and q are gene frequencies of O, A and B, then the expected probabilities of the six genotypes (four phenotypes) in random mating. Following the Bernstein and Maximum likelihood estimate. We have obtained the allele frequencies.

Address for correspondence: D. Nagarajan, Department of Mathematics, St.Joseph's College of Engineering Chennai 119, Tamil Nadu, India
E-mail: dnrmsu2002@yahoo.com

Table 1: Model description

Phenotype	*Genotype*	*Probability*
O	OO	r^2
A	AAAO	$P^2 + 2pr$
B	BBBO	$q^2 + 2qr$
AB	AB	2pq

given $\overline{O}, \overline{A}, \overline{B}$ and $\overline{AB}$ are observed frequen-cies and estimate the gene frequencies p,q and r (Rao,1973).

Table 2: Model description

	Probability, (π)	*Derivatives* $\frac{\partial \pi}{\partial p}$	$\frac{\partial \pi}{\partial q}$
O	r^2	-2r	-2r
A	$P^2 + 2pr$	2r	-2p
B	$q^2 + 2qr$	-2q	2r
AB	2pq	2q	2p

The frequencies A=158,B=505,AB=121, O=220

Approximate solutions obtained by Bernstein's method are

p' =0.1502, q'= 0.3864, r'= 0.4681 and D= -0.0047

Better estimation due to Bernstein are obtained as follows:

$\hat{p} = 0.1498$, $\hat{q} = 0.3855$, $\hat{r} = 0.4623$

The probability and coefficients for the calculation of efficient scores at the approximate value obtained in table 3.

Table 3: Stationary distribution of the study frequencies

	Probability (π)	$\frac{1}{\pi}\frac{\partial \pi}{\partial p}$	$\frac{1}{\pi}\frac{\partial \pi}{\partial q}$	*Observed frequencies*
O	0.2137	-4.3266	-4.3266	220
A	0.1609	5.7464	-1.8620	158
B	0.5050	-1.5267	1.8309	505
AB	0.1155	6.6753	2.5940	121
				1004

The value of π show that the blood group of mothers B, O, A and AB have probability of 0.5050,

0.2137, 0.1609 and 0.1155 with the blood group of mothers and babies, respectively. It suggest that the Blood group of B is the highest probability of among four blood group.

The scores are

$$Q_p = -7.193 \quad Q_q = -7.5695$$

The information matrix for a single observation is

$$I = \begin{bmatrix} 15.6372 & 2.8671 \\ 2.8671 & 7.0282 \end{bmatrix}$$

The inverse of the information matrix per single observation is

$$I' = \begin{bmatrix} 0.0691 & -0.0282 \\ -0.0282 & 0.1538 \end{bmatrix}$$

$$i^{pp} = 0.0691, i^{pq} = -0.0282, i^{qq} = 0.1538, i^{qp} = -0.0282$$

The solutions for correction are

$$\delta_p = -0.00028, \quad \delta_q = -0.00095$$

The corrections are very small. If the corrections are not small, the whole process has to be repeated with the second approximation. It is the information matrix need not be recalculated for each approximation. Only the new scores have to be calculated at each stage and used in conjunction with the same inverse information matrix to obtain closer approximations. When convergence is achieved, the information matrix and scores may be calculated for obtaining the last approximate values.

The maximum likelihood estimates and the variance are

$$\hat{p} = 0.1498, \quad \hat{q} = 0.3855, \hat{r} = 0.4623$$

$$v(\hat{p}) = 0.00068, v(\hat{q}) = 0.000153, v(\hat{r}) = 0.000165$$

References may be made to Finney(1952), Biswas (1993) ,Fisher(1950) , Senthamaraikannan et al (2005) and Rao(1973).

In case of O,A, and B blood group data there is another algorithm which yields the maximum likelihood estimates. Let p_0, q_0 and r_0 be provision estimates and compute $P=p_0/r_0$ and $Q=q_0/r_0$. Let us represent by P_k and Q_k the k^{th} approximation of p/r and q/r. the $(k+1)^{th}$ approximation are found from the formula,

the iteration may be repeated until stable values of P and Q are obtained from which the estimates p,q and r are computed as

$$\hat{r} = \frac{1}{P+Q+1} \quad \hat{P} = \hat{r}P, \ \hat{q} = \hat{r}Q$$

by adopting the provisional estimates

$$\hat{r} = 0.4634, \hat{P} = 0.1501, \hat{q} = 0.3864$$

with P= 0.3240, Q= 0.8339

The maximum likelihood estimates p, q and r are, $\hat{r} = 0.4634$, $\hat{P} = 0.1501$, $\hat{q} = 0.3864$

Which agree closely with the estimates obtained by the iterative procedure using the information matrix.

Table 4: Classification of individuals of blood Group

Community	*O*	*A*	*B*	*AB*	*Total*
I	111	78	255	58	502
II	109	80	250	63	502
Total	220	158	505	121	1004

Community mother is I and baby is II

The distribution in four blood group classes O,A,B and AB of 502 individuals belonging to community –I, 502 individuals of community II (Table 4).

$p = \frac{community - I}{Total}$ is 0.5045, 0.4936, 0.5049, 0.4793 and 0.5000

$$\sum n_i p_i - n\,p = 0.0492$$

$$p(1-p)=0.2500$$

$$\chi^2 = \frac{0.0492}{0.2500} = 0.1968$$

The object is to establish difference between communities on three degree of freedom not significance at 5% level indicating difference between communities. The intrinsic difference, if any between the communities, is in the relative frequencies of the *O, A* and *B* genes. A test of the hypothesis of equality of gene frequencies is more fundamental than that of the equality of frequencies of the phenotype without recognizing that the latter are know function of the former. Only when such functional dependence on more intrinsic characters like the gene frequencies is unknown, a test of type

$$\frac{\sum n_i p_i - n_i p}{p(1-P)}$$

Involving the phonotypical frequencies valid (Cochran,1954).

From the table 5 the estimation of p, q and r for I , II and combined samples are obtained. Using these estimates, the expected value E_1, E_2 on the basis of individual estimates and combined estimates are obtained for each community show in the following tables.

Table 5: Maximum likelihood estimate Community

Community	$\hat{p}$	$\hat{q}$	$\hat{r}=(1-\hat{p}-\hat{q})$
I	0.1459	0.3858	0.4683
II	0.1538	0.3851	0.4611
Combined sample	0.1498	0.38550	0.4647

Expected Values Under Hypothesis

H_0: All the gene frequencies are equal

Table 6: Expected values under the hypothesis for community-I

Classes	*Probability*	*Community-I*		
		Observed	E_1 *(Individual expectation)*	E_2 *(Combined expectation)*
O	r^2	111	110.09	108.38
A	$P^2 + 2pr$	78	79.27	81.17
B	$q^2 + 2qr$	255	256.12	254.46
AB	2pq	58	56.52	57.99
		502	502	502

χ^2 for goodness of fit

$$\chi^2 = \frac{\sum(O-E_1)^2}{E_1} = 0.07152$$

From the values for this community-I is very small, which indicates that the specification of cell probability interms of gene frequencies are valid (Table 6).

Table 7: Expected values under the hypothesis for community-II

Classes	*Probability*	*Community-II*		
		Observed	E_1 *(Individual expectation)*	E_2 *(Combined expectation)*
O	r^2	109	106.73	108.38
A	$P^2 + 2pr$	80	83.08	81.17
B	$q^2 + 2qr$	250	252.71	254.46
AB	2pq	63	59.48	57.99
		502	502	502

χ^2 for community –II , $\chi^2 = \frac{\sum(O-E_1)^2}{E_1} = 0.40112$

From the values for this community-II is very small, which indicates that the specification of cell probability interms of gene frequencies are valid.

χ^2 for Homogeneity of Communities

From the value for homogeneity of community, which indicate that the specification of cell probability is valid of expected value. The calculated values for the individual community are smaller. Compare to these χ^2 table values which indicate the specification of cell probability interms of gene frequency are valid. As the result of the analysis conclude that the last χ^2 table establish the difference between the community better way than the χ^2 of 0.25346 on 3 degree of freedom obtain for homogeneity method (Table 8). Individuals with group B blood have the opposite arrangement: antigen B is on their cells, and antibodies against antigen A are produced in their serum. Therefore, a group B person can only receive blood from people in groups B or O, preferably B.

Table 8: Expected values under the hypothesis

Community-I		*Community-II*	
E_1	E_2	E_1	E_2
110.09	108.38	106.73	108.38
79.27	81.17	83.08	81.17
256.12	254.46	252.71	254.46
56.52	57.99	59.48	57.99

$$\chi^2 = \sum\sum \frac{(E_1 - E_2)^2}{E_2} = 0.25346$$

CONCLUSION

From the above results show that the blood group of mothers B, O, A and AB have probability of 0.5050, 0.2137, 0.1609 and 0.1155 with the blood group of mothers and babies, respectively. It suggests that the Blood group of B is the highest probability of among four blood groups in this area. These findings will be very useful for health policy makers, health administrated in planning and mobilization of health care management.

REFERENCES

Biswas, S.: *Applied Stochastic Processes*. New Age International Publications, Wiley Eastern Ltd., New York (1993).

Cochran, W.G.: Some methods for strenon tests. *Biometrics*, **10:** 417-451 (1954).

Finney, D.J.: *Probit Analysis.* Second Edition, Cambridge University Press, Cambridge , Mass (1952).

Fisher, R.A.: *Statistical Methods for Research Workers.* Eleventh Edition. Oliver and Boyd , Edinburgh (1950).

Radhakrisha Rao, C.: Linear statistical inference and its applications. John Wiley & Sons Inc., New York (1973).

Senthamarai Kannan, K, Nagarajan, D. and Tamizh, Chelvam T.: The Stochastic model of human gestation period for mature and premature live births-Part II, *Tamkang Journal of Mathematics*, **36(2)**: 143-158 (2005).