

## CAPÍTULO 2. EXTENSÃO DA LEI DE HARDY E WEINBERG AO POLIALELISMO AUTOSSÔMICO

A lei de Hardy e Weinberg não está restrita apenas à ocorrência de um par de alelos em um loco autossômico, isto é, aos casos de *dialelismo*. Ela pode ser generalizada para os casos de *polialelismo autossômico*, ou seja, para aqueles em que consideramos vários alelos pertencentes a um loco autossômico. Em outras palavras, nos casos de polialelismo autossômico também se pode demonstrar que, se uma população obedecer as oito condições estabelecidas para a população teórica do capítulo anterior, *as freqüências gênicas dos diferentes alelos se manterão inalteradas e as proporções genótípicas atingirão equilíbrio estável após uma única geração de panmixia*. Assim, se um loco autossômico possuir  $n$  alelos ( $A, a, a_1, a_2, \dots, a_n$ ), que ocorrem em uma população teórica como a descrita no capítulo anterior com freqüências  $p, q, r, s, \dots, x$ , essas freqüências se manterão inalteradas e o equilíbrio estável dos genótipos será atingido após uma geração de panmixia, com os genótipos se distribuindo segundo:

$$(p + q + r + s + \dots + x)^2$$

Para demonstrar que esse equilíbrio se estabelece após uma única geração de panmixia consideremos um exemplo de trialelismo em uma população teórica que obedeça às oito premissas estabelecidas para a demonstração do equilíbrio de Hardy e Weinberg. Assim, se na geração inicial encontrarmos apenas indivíduos com os genótipos  $AA, aa$  e  $a_1a_1$ , determinados pelos alelos autossômicos  $A, a$  e  $a_1$  e se as freqüências desses indivíduos forem  $AA = 0,30, aa = 0,10$  e  $a_1a_1 = 0,60$ , tanto entre os homens, quanto entre as mulheres, é claro que as freqüências dos três alelos em questão serão  $A = p = 0,30, a = q = 0,10$  e  $a_1 = r = 0,60$ .

Evidentemente, a geração inicial não está em equilíbrio segundo a lei de Hardy e Weinberg, pois, para tanto, seria necessário que os genótipos decorrentes dos alelos  $A, a$  e  $a_1$  se distribuíssem segundo  $(p + q + r)^2 = 1$ . Em outras palavras, seria necessário observar as freqüências genótípicas abaixo, as quais, de acordo com a Tabela 1.2 são alcançadas após uma única geração em panmixia.

$$\begin{aligned} AA &= p^2 = 0,30 \times 0,30 = 0,09 \\ aa &= q^2 = 0,10 \times 0,10 = 0,01 \\ a_1a_1 &= r^2 = 0,60 \times 0,60 = 0,36 \\ Aa &= 2pq = 2 \times 0,30 \times 0,10 = 0,06 \\ Aa_1 &= 2pr = 2 \times 0,30 \times 0,60 = 0,36 \end{aligned}$$

$$aa_1 = 2qr = 2 \times 0,10 \times 0,60 = 0,12$$

Tabela 1.2. Distribuição dos genótipos decorrentes de trialelismo após uma única geração de uma população teórica em panmixia. A geração inicial é constituída por indivíduos homocigotos  $AA$  (30%),  $aa$  (10%) e  $a_1a_1$  (60%).

Casais (geração inicial)		Primeira Geração Filial					
Tipo	Frequência	$AA$	$aa$	$a_1a_1$	$Aa$	$Aa_1$	$aa_1$
$AA \times AA$	$0,30 \times 0,30 = 0,09$	0,09	-	-	-	-	-
$AA \times aa$	$2 \times 0,30 \times 0,10 = 0,06$	-	-	-	0,06	-	-
$AA \times a_1a_1$	$2 \times 0,30 \times 0,60 = 0,36$	-	-	-	-	0,36	-
$aa \times aa$	$0,10 \times 0,10 = 0,01$	-	0,01	-	-	-	-
$aa \times a_1a_1$	$2 \times 0,10 \times 0,60 = 0,12$	-	-	-	-	-	0,12
$a_1a_1 \times a_1a_1$	$0,60 \times 0,60 = 0,36$	-	-	0,36	-	-	-
Total	1,00	0,09 ( $p^2$ )	0,01 ( $q^2$ )	0,36 ( $r^2$ )	0,06 ( $2pq$ )	0,36 ( $2pr$ )	0,12 ( $2qr$ )

A Tabela 2.2 mostra a distribuição de famílias em uma população teórica que está em equilíbrio de Hardy e Weinberg em relação aos genótipos decorrentes de três alelos autossômicos  $A$ ,  $a$  e  $a_1$  com frequências  $p$ ,  $q$  e  $r$ .

Tabela 2.2. Distribuição das famílias em uma população teórica que está em equilíbrio genético em relação aos genótipos determinados por três alelos autossômicos  $A$ ,  $a$  e  $a_1$  cujas frequências são  $p$ ,  $q$  e  $r$ .

Casais		Filhos					
Tipo	Frequência	$AA$	$aa$	$a_1a_1$	$Aa$	$Aa_1$	$aa_1$
$AA \times AA$	$p^2 \cdot p^2 = p^4$	$p^4$	-	-	-	-	-
$AA \times aa$	$2p^2 \cdot q^2 = 2p^2q^2$	-	-	-	$2p^2q^2$	-	-
$AA \times a_1a_1$	$2p^2 \cdot r^2 = 2p^2r^2$	-	-	-	-	$2p^2r^2$	-
$AA \times Aa$	$2p^2 \cdot 2pq = 4p^3q$	$2p^3q$	-	-	$2p^3q$	-	-
$AA \times Aa_1$	$2p^2 \cdot 2pr = 4p^3r$	$2p^3r$	-	-	-	$2p^3r$	-
$AA \times aa_1$	$2p^2 \cdot 2qr = 4p^2qr$	-	-	-	$2p^2qr$	$2p^2qr$	-
$aa \times aa$	$q^2 \cdot q^2 = q^4$	-	$q^4$	-	-	-	-
$aa \times a_1a_1$	$2q^2 \cdot r^2 = 2q^2r^2$	-	-	-	-	-	$2q^2r^2$
$aa \times Aa$	$2q^2 \cdot 2pq = 4pq^3$	-	$2pq^3$	-	$2pq^3$	-	-
$aa \times Aa_1$	$2q^2 \cdot 2pr = 4pq^2r$	-	-	-	$2pq^2r$	-	$2pq^2r$
$aa \times aa_1$	$2q^2 \cdot 2qr = 4q^3r$	-	$2q^3r$	-	-	-	$2q^3r$
$a_1a_1 \times a_1a_1$	$r^2 \cdot r^2 = r^4$	-	-	$r^4$	-	-	-
$a_1a_1 \times Aa$	$2r^2 \cdot 2pq = 4pqr^2$	-	-	-	-	$2pqr^2$	$2pqr^2$
$a_1a_1 \times Aa_1$	$2r^2 \cdot 2pr = 4pr^3$	-	-	$2pr^3$	-	$2pr^3$	-
$a_1a_1 \times aa_1$	$2r^2 \cdot 2qr = 4qr^3$	-	-	$2qr^3$	-	-	$2qr^3$
$Aa \times Aa$	$2pq \cdot 2pq = 4p^2q^2$	$p^2q^2$	$p^2q^2$	-	$2p^2q^2$	-	-
$Aa \times Aa_1$	$2 \cdot 2pq \cdot 2pr = 8p^2qr$	$2p^2qr$	-	-	$2p^2qr$	$2p^2qr$	$2p^2qr$
$Aa \times aa_1$	$2 \cdot 2pq \cdot 2qr = 8pq^2r$	-	$2pq^2r$	-	$2pq^2r$	$2pq^2r$	$2pq^2r$
$Aa_1 \times Aa_1$	$2pr \cdot 2pr = 4p^2r^2$	$p^2r^2$	-	$p^2r^2$	-	$2p^2r^2$	-
$Aa_1 \times aa_1$	$2 \cdot 2pr \cdot 2qr = 8pqr^2$	-	-	$2pqr^2$	$2pqr^2$	$2pqr^2$	$2pqr^2$
$aa_1 \times aa_1$	$2qr \cdot 2qr = 4q^2r^2$	-	$q^2r^2$	$q^2r^2$	-	-	$2q^2r^2$
Total	$(p+q+r)^4 = 1$	$p^2$	$q^2$	$r^2$	$2pq$	$2pr$	$2qr$

## CÁLCULO DA FREQUÊNCIA GÊNICA EM CASOS DE POLIALLELISMO AUTOSSÔMICO QUANDO EXISTE CODOMINÂNCIA

Nos casos de poliallelismo em que não existe relação de dominância e recessividade entre os fenótipos, as frequências gênicas podem ser facilmente calculadas por contagem. Assim, por exemplo, num caso de trialelismo, levando em conta os alelos  $A$ ,  $a$  e  $a_1$ , teríamos que considerar a existência de seis genótipos  $AA$ ,  $aa$ ,  $a_1a_1$ ,  $Aa$ ,  $Aa_1$ ,  $aa_1$ . Em uma amostra de  $n$  indivíduos podemos calcular as frequências de um gene após somar o número de vezes em que ele aparece nos diferentes genótipos e dividir a soma pelo número total de genes, isto é, por  $2n$ . Tem-se, desse modo, que as frequências  $p$ ,  $q$  e  $r$  dos alelos  $A$ ,  $a$  e  $a_1$  podem ser obtidas por intermédio de:

$$p = \frac{2AA + Aa + Aa_1}{2n}$$

$$q = \frac{2aa + Aa + aa_1}{2n}$$

$$r = \frac{2a_1a_1 + Aa_1 + aa_1}{2n}$$

Se em vez de trabalhar com o número de genótipos quisermos trabalhar com a sua frequência relativa, assinaladas em **negrito**, o denominador dessas fórmulas se reduzirá ao número 2, porque, nesse caso, o total da amostra passa a ser a unidade. Nesse caso, as frequências  $p$ ,  $q$  e  $r$  dos alelos  $A$ ,  $a$  e  $a_1$  são obtidas por intermédio de:

$$p = \frac{2AA + Aa + Aa_1}{2}$$

$$q = \frac{2aa + Aa + aa_1}{2}$$

$$r = \frac{2a_1a_1 + Aa_1 + aa_1}{2}$$

Tomemos um exemplo numérico a respeito da fosfatase ácida eritrocitária, cujos alelos mais frequentes são os genes  $P^a$ ,  $P^b$  e  $P^c$  responsáveis pelos fenótipos A, B, C, AB, AC e BC os quais correspondem, respectivamente, aos genótipos  $P^aP^a$ ,  $P^bP^b$ ,  $P^cP^c$ ,  $P^aP^b$ ,  $P^aP^c$ ,  $P^bP^c$ . Assim, consideremos que numa amostra de 480 indivíduos caucásios observou-se a seguinte distribuição fenotípica, com as frequências em porcentagem assinaladas entre parênteses:

$$A = 35 (7,3); B = 256 (53,3); C = 0 (0); AB = 161 (33,5); AC = 8 (1,7); BC = 20 (4,2)$$

Se levássemos em conta o número de indivíduos, as frequências  $p$ ,  $q$  e  $r$  dos alelos  $P^a$ ,  $P^b$  e  $P^c$  seriam calculadas como abaixo.

$$p = P^a = \frac{(2 \times 35) + 161 + 8}{2 \times 480} = 0,249 \text{ ou } 24,9\%$$

$$q = P^b = \frac{(2 \times 256) + 161 + 20}{2 \times 480} = 0,722 \text{ ou } 72,2\%$$

$$r = P^c = \frac{8 + 20}{2 \times 480} = 0,029 \text{ ou } 2,9\%$$

Levando em conta as frequências relativas dos fenótipos, que correspondem aos genótipos, chegaríamos aos mesmos resultados a partir das operações abaixo:

$$p = P^a = \frac{(2 \times 0,073) + 0,335 + 0,017}{2} = 0,249 \text{ ou } 24,9\%$$

$$q = P^b = \frac{(2 \times 0,533) + 0,335 + 0,042}{2} = 0,722 \text{ ou } 72,2\%$$

$$r = P^c = \frac{0,017 + 0,042}{2} = 0,029 \text{ ou } 2,9\%$$

Para investigar se a amostra estudada representa uma população em equilíbrio de Hardy e Weinberg basta calcular o número de indivíduos nas seis classes fenotípicas esperadas segundo  $A = np^2$ ;  $B = nq^2$ ;  $C = nr^2$ ;  $AB = n2pq$ ;  $AC = n2pr$ ;  $BC = n2qr$ . O qui-quadrado resultante tem três graus de liberdade, porque para calcular as seis classes esperadas são necessárias três informações, ou seja, o total da amostra e a frequência de dois dos três alelos, visto que, conhecendo as frequências de dois deles, a do terceiro alelo será automaticamente conhecida, pois  $p+q+r = 1$ . Em decorrência do qui-quadrado obtido (2,396) com três graus de liberdade, pode-se concluir que a amostra analisada representa uma população em equilíbrio de Hardy e Weinberg quanto aos genótipos da fosfatase ácida eritrocitária.

Valor	A	B	C	AB	AC	BC	Total
Observado ( $o$ )	35	256	0	161	8	20	480
Esperado ( $e$ )	29,8	250,2	0,4	172,6	6,9	20,1	480
$\frac{(o - e)^2}{e}$	0,907	0,134	0,4	0,780	0,175	0,0005	$\chi^2_{(3)} = 2,396$ $0,30 < P < 0,50$

### CÁLCULO DA FREQUÊNCIA GÊNICA EM CASOS DE POLIALELISMO AUTOSSÔMICO QUANDO EXISTE DOMINÂNCIA E RECESSIVIDADE

Dentre os casos de polialelismo autossômico, a situação que parece mais interessante para ser aqui discutida é a de trialelismo, com um dos alelos se manifestando apenas em homozigose, como ocorre no sistema sanguíneo ABO clássico. Nesse sistema levamos em conta os alelos  $A$ ,  $B$  e  $O$ , sendo que o gene  $A$  determina a

produção do antígeno A, o gene *B* a produção do antígeno B, enquanto o gene *O*, quando em homozigose, é responsável pela falta dos antígenos A e B. Esses alelos, sabe-se atualmente, pertencem a um loco situado no braço inferior do cromossomo 9, mais precisamente em *9q31.3-qter* (Cook *et al.*, 1978; Narahara *et al.*, 1986). Os alelos *A* e *B* mantêm uma relação de codominância entre si, de modo que os heterozigotos *AB* são responsáveis pelo grupo sanguíneo AB. Esses dois alelos apresentam relação de dominância sobre o alelo *O*. Por isso, os indivíduos homozigotos *AA* ou os heterozigotos *AO* apresentam grupo sanguíneo A, aqueles com genótipos homozigoto *BB* e heterozigoto *BO* apresentam grupo sanguíneo B, enquanto os indivíduos do grupo sanguíneo O são os homozigotos *OO*.

Designando as freqüências dos alelos *A*, *B* e *O* por, respectivamente, *p*, *q* e *r*, as proporções genótípicas e fenotípicas do sistema ABO esperadas nas populações em equilíbrio genético estável são aquelas calculadas por intermédio da expansão do trinômio  $(p + q + r)^2$ , o que permite apresentar um quadro como o da Tabela 3.2.

Tabela 3.2 Freqüências genótípicas e fenotípicas do sistema ABO clássico que são esperadas em populações em equilíbrio genético estável, sendo *p*, *q* e *r* as freqüências dos genes *A*, *B* e *O*.

Grupo Sanguíneo (Fenótipo)	Genótipos	Freqüência Genotípica	Freqüência Fenotípica
A	<i>AA</i>	$p^2$	$p^2 + 2pr$
	<i>AO</i>	$2pr$	
B	<i>BB</i>	$q^2$	$q^2 + 2qr$
	<i>BO</i>	$2qr$	
AB	<i>AB</i>	$2pq$	$2pq$
O	<i>OO</i>	$r^2$	$r^2$

Com base nos elementos fornecidos na Tabela 3.2 é possível estimar as freqüências populacionais dos alelos do sistema ABO, a partir do conhecimento da distribuição fenotípica, isto é, de seus grupos sanguíneos. As melhores estimativas dessas freqüências são obtidas pelo método dos escores de verossimilhança máxima, com o auxílio de programas computacionais (Cabello e Krieger, 1997). Essas estimativas, entretanto, também podem ser obtidas de modo fácil com o emprego de métodos algébricos, dentre os quais, o que fornece valores muito semelhantes aos calculados pelo método dos escores de verossimilhança máxima é o proposto por Bernstein (1924,1925,1930). O próprio Stevens (1950) que, pela primeira vez, aplicou o método da verossimilhança máxima para estimar as freqüências dos genes do sistema ABO (Stevens, 1938), reconheceu que o método de Bernstein é muito eficiente, apesar

de sua simplicidade.

No método de Bernstein parte-se do princípio de que, se a amostra representar uma população em equilíbrio de Hardy e Weinberg, ter-se-á  $(p+q+r)^2 = 1$  e, para obter a frequência do alelo  $O$ , basta extrair a raiz quadrada da frequência observada de indivíduos do grupo  $O$ , pois,  $r = \sqrt{r^2} = \sqrt{O}$ . Depois disso, deve-se calcular:

$$p = 1 - \sqrt{B + O} = 1 - \sqrt{q^2 + 2qr + r^2} = 1 - (q + r)$$

$$q = 1 - \sqrt{A + O} = 1 - \sqrt{p^2 + 2pr + r^2} = 1 - (p + r)$$

Visto que a soma dessas estimativas raramente fornece um valor exatamente igual a 1, elas são chamadas de *estimativas preliminares* e representadas por  $p'$ ,  $q'$  e  $r'$ . Para corrigi-las deve-se calcular o desvio ( $D$ ) entre a unidade e a soma das estimativas, isto é, é necessário obter:

$$D = 1 - (p' + q' + r')$$

e calcular as *estimativas corrigidas*  $p$ ,  $q$  e  $r$  por intermédio de:

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

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

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

Aplicamos essas fórmulas aos dados a respeito de uma amostra de 2.571 brasileiros caucasóides que apresentaram a seguinte distribuição de grupos sanguíneos do sistema ABO clássico, com os valores em porcentagem entre parênteses:

$$A = 1001 (38,93); B = 250 (9,72); AB = 92 (3,58); O = 1228 (47,76)$$

Com base nesses dados, calcularíamos:

$$p' = 1 - \sqrt{B + O} = 1 - \sqrt{0,0972 + 0,4776} = 0,2418$$

$$q' = 1 - \sqrt{A + O} = 1 - \sqrt{0,3893 + 0,4776} = 0,0689$$

$$r' = \sqrt{O} = \sqrt{0,4776} = 0,6911$$

$$\text{Total} = 1,0018$$

$$D = 1 - 1,0018 = -0,0018$$

$$\frac{D}{2} = -0,0009$$

$$1 + \frac{D}{2} = 0,9991$$

$$p = 0,2418 \times 0,9991 = 0,242$$

$$q = 0,0689 \times 0,9991 = 0,069$$

$$r = 1 - (0,242 + 0,069) = 0,689$$

Para verificar se a distribuição dos fenótipos na amostra estudada está de acordo com a hipótese genética, isto é, se a amostra representa, realmente, uma população em equilíbrio de Hardy e Weinberg, existe a alternativa de calcular um qui-quadrado a partir da razão entre o quadrado do desvio  $D$  e a sua variância. Visto que a variância do desvio  $D$  é  $\sigma^2_{(D)} = \frac{1}{2n\left(1 + \frac{r}{pq}\right)}$  onde  $n$  é o tamanho amostral, o qui-quadrado pode ser

obtido como abaixo:

$$\chi^2 = 2n\left(1 + \frac{r}{pq}\right)D^2$$

Esse qui-quadrado terá apenas um grau de liberdade porque existem quatro classes fenotípicas (grupos A, B, AB e O) e são necessárias três informações da amostra para calcular os valores esperados nessas classes, isto é, o tamanho da amostra e as frequências de dois genes. É importante salientar que nessa fórmula de qui-quadrado não é necessário empregar as estimativas corrigidas das frequências gênicas, podendo-se partir das estimativas preliminares. Usando as estimativas preliminares das frequências dos alelos  $A$ ,  $B$  e  $O$  da amostra caucasóide de nosso exemplo obtemos a indicação de que ela representa uma população em equilíbrio de Hardy e Weinberg, pois:

$$\chi^2 = 2 \times 2.571 \left(1 + \frac{0,6911}{0,2418 \times 0,0689}\right) 0,00000324 = 0,707; 1 \text{ G.L.}; 0,30 < P < 0,50.$$

Uma outra alternativa é a de calcular um qui-quadrado depois de comparar os números observados com os esperados em cada classe, isto é, a partir de:

$$\chi^2 = \frac{[A - (p^2 + 2pr)n]^2}{(p^2 + 2pr)n} + \frac{[B - (q^2 + 2qr)n]^2}{(q^2 + 2qr)n} + \frac{(AB - 2pqn)^2}{2pqn} + \frac{(O - r^2n)^2}{r^2n}$$

onde A, B, AB e O são os números observados de indivíduos em cada um desses grupos sanguíneos. Assim, teríamos:

$$\chi^2 = \frac{(1001 - 1007,93)^2}{1007,93} + \frac{(250 - 256,70)^2}{256,70} + \frac{(92 - 85,86)^2}{85,86} + \frac{(1228 - 1220,51)^2}{1220,51} = 0,707$$

Como se pode constatar, os dois valores de qui-quadrado obtidos são idênticos, o que indica que qualquer das duas alternativas para calcular o qui-quadrado pode ser aceita. A primeira alternativa é menos trabalhosa, mas, nos casos em que a amostra não representa uma população em equilíbrio genético, a segunda alternativa é melhor do que a primeira porque permite detectar qual(ais) o(s) grupo(s) sanguíneo(s) que mostram

uma diferença significativa entre o observado e o esperado. Assim, por exemplo, se a amostra estudada não estivesse em equilíbrio por causa de um excesso de indivíduos do grupo sanguíneo AB, a aplicação da fórmula da primeira alternativa indicaria apenas que a amostra não representa uma população em equilíbrio genético, mas nada informaria a respeito do número excessivo observado de indivíduos do grupo sanguíneo AB.

Um bom expediente para facilitar o cálculo do qui-quadrado usando a segunda alternativa é o de trabalhar apenas com as freqüências relativas, deixando de multiplicá-las pelo tamanho da amostra ( $n$ ) *mas não se deve esquecer, ao final das operações, de multiplicar o valor do qui-quadrado obtido pelo tamanho amostral*, porque o valor obtido é igual ao do qui-quadrado dividido por  $n$ . Assim, usando os dados de nosso exemplo, teríamos:

$$\frac{(0,3893 - 0,3920)^2}{0,3920} + \frac{(0,0972 - 0,0998)^2}{0,0998} + \frac{(0,0358 - 0,0334)^2}{0,0334} + \frac{(0,4776 - 0,4747)^2}{0,4747} = 0,000276$$

e, multiplicando esse resultado por 2.571, obtém-se um valor de qui-quadrado muito próximo aos calculados anteriormente, pois  $0,000276 \times 2.571 = 0,7096$ ;  $0,30 < P < 0,50$ .

Quando os subgrupos sanguíneos  $A_1$ ,  $A_2$ ,  $A_1B$  e  $A_2B$  são levados em conta, o sistema ABO passa a incluir seis fenótipos explicados à custa de dez genótipos que, numa população em equilíbrio genético se distribuem como na Tabela 4.2.

Tabela 4.2. Freqüências genótípicas e fenótípicas do sistema ABO em uma população em equilíbrio genético, quando se consideram os subgrupos  $A_1$ ,  $A_2$ ,  $A_1B$  e  $A_2B$ .

Fenótipo	Genótipo	Freqüência Genotípica	Freqüência Fenotípica
$A_1$	$A_1A_1$	$p_1^2$	$p_1^2 + 2p_1(p_2+r)$
	$A_1A_2$	$2p_1p_2$	
	$A_1O$	$2p_1r$	
$A_2$	$A_2A_2$	$p_2^2$	$p_2(p_2+2r)$
	$A_2O$	$2p_2r$	
B	$BB$	$q^2$	$q(q+2r)$
	$BO$	$2qr$	
$A_1B$	$A_1B$	$2p_1q$	$2p_1q$
$A_2B$	$A_2B$	$2p_2q$	$2p_2q$
O	$OO$	$r^2$	$r^2$

As freqüências preliminares  $p_1'$ ,  $p_2'$ ,  $q'$  e  $r'$  dos alelos  $A_1$ ,  $A_2$ ,  $B$  e  $O$  são estimadas a partir das freqüências fenótípicas  $A_1$ ,  $A_2$ ,  $A_2B$  e  $O$  como abaixo:

$$p_1' = 1 - \sqrt{A_2 + B + A_2B + O}$$

$$p_2' = \sqrt{A_2 + B + A_2B + O} - \sqrt{B + O}$$

$$q' = 1 - \sqrt{A_1 + A_2 + O}$$

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

Depois de determinar o desvio (D) entre a unidade e a soma das estimativas preliminares, isto é, depois de calcular:

$$D = 1 - (p_1' + p_2' + q' + r')$$

as estimativas corrigidas  $p_1$ ,  $p_2$  e  $q$  das frequências dos genes  $A_1$ ,  $A_2$  e  $B$  serão obtidas pela multiplicação das estimativas preliminares  $p_1'$ ,  $p_2'$  e  $q'$  por  $\left(1 + \frac{D}{2}\right)$ . A estimativa corrigida  $r$  do alelo  $O$  pode ser obtida simplesmente pela diferença  $1 - (p_1 + p_2 + q)$  ou calculando  $\left(r' + \frac{D}{2}\right)\left(1 + \frac{D}{2}\right)$ .

Para estimar as frequências dos genes do sistema ABO em populações com o alelo  $A_{int}$  pode-se fazer o cálculo das estimativas preliminares como abaixo:

$$\begin{aligned} p_1' &= 1 - \sqrt{A_{int} + A_2 + B + A_{int}B + A_2B + O} \\ p_{int}' &= \sqrt{A_{int} + A_2 + B + A_{int}B + A_2B + O} - \sqrt{A_2 + B + A_2B + O} \\ p_2' &= \sqrt{A_2 + B + A_2B + O} - \sqrt{B + O} \\ q' &= 1 - \sqrt{A_1 + A_2 + A_{int} + O} \\ r' &= \sqrt{O} \end{aligned}$$

Depois de calcular o desvio  $D = 1 - (p_1' + p_{int}' + p_2' + q' + r')$ , todas as estimativas corrigidas, com exceção de  $r$ , serão obtidas pela multiplicação das estimativas preliminares por  $\left(1 + \frac{D}{2}\right)$ . A estimativa corrigida  $r$  será obtida por intermédio de  $1 - (p_1 + p_{int} + p_2 + q)$  ou por intermédio de  $\left(r' + \frac{D}{2}\right)\left(1 + \frac{D}{2}\right)$ .

## ANOMALIAS RECESSIVAS E CÁLCULO DAS FREQUÊNCIAS GÊNICAS

Não é raro termos que lidar com anomalias recessivas resultantes da homozigose ou heterozigose de diferentes mutações. Em outras palavras, não é raro nos depararmos com várias mutações, que poderemos designar por  $a$ ,  $a_1$ ,  $a_2$ ,  $a_3$  etc. de um gene “normal”  $A$  e que tanto os homozigotos  $aa$ ,  $a_1a_1$ ,  $a_2a_2$  ou  $a_3a_3$  quanto os heterozigotos  $aa_1$ ,  $aa_2$ ,  $aa_3$ ,  $a_1a_2$ ,  $a_1a_3$  ou  $a_2a_3$  determinem um quadro patológico semelhante. Em situações como essa, em que os diferentes genótipos podem ser detectados por técnicas de genética molecular, mas que, do ponto de vista clínico, resultam em um quadro semelhante, podemos, para fins de epidemiológicos, reunir os homozigotos  $aa$ ,  $a_1a_1$ ,

$a_2a_2$ ,  $a_3a_3$  e os heterozigotos  $aa_1$ ,  $aa_2$ ,  $aa_3$ ,  $a_1a_2$ ,  $a_1a_3$  e  $a_2a_3$  em uma única categoria, designada por  $aa$ .

Para exemplificar, consideremos o caso da *fibrose cística*, que é o defeito genético mais freqüente em populações caucasóides, decorrente de muitas mutações caracterizáveis por técnicas de estudo de DNA. Considerando, porém, que essas mutações em homozigose ou em heterozigose dos diferentes mutantes resulta num quadro patológico de fibrose cística, podemos tratá-las em conjunto para fins epidemiológicos. Assim, numa população caucasóide na qual a fibrose cística ocorre em 4 de cada 10.000 nascimentos, poderíamos estimar que, dentre cerca de 25 indivíduos dessa população, um seria portador (heterozigoto) de um alelo mutante responsável pela fibrose cística. Realmente, a partir do conhecimento da incidência de fibrose cística, calcularíamos  $q = \sqrt{0,0004} = 0,02$  e  $p = 1 - 0,02 = 0,98$ . Isso nos permitiria estimar a freqüência de portadores como sendo  $2pq = 0,0392$  ou  $1: 25,5 \cong 1: 25$ .

## COMPARAÇÃO DE VÁRIAS AMOSTRAS

Quando lidamos com polialelismo com codominância e queremos comparar várias amostras, a fim de investigar se elas podem ser reunidas em uma única ou se isso não deve ser feito em decorrência de uma diferença significativa entre elas, podemos levar em conta a distribuição gênica, como nos casos de dialelismo. Contudo, se as amostras representarem populações em equilíbrio de Hardy e Weinberg poder-se-á usar a alternativa de comparar as distribuições genótípicas dessas amostras. Para exemplificar, consideremos que em amostras de três populações estudadas quanto à fosfatase ácida eritrocitária tivéssemos encontrado os resultados apresentados na Tabela 5.2. Com esses dados pode-se elaborar a Tabela 6.2 e, pelo valor de qui-quadrado encontrado, concluir a favor da hipótese de que as três amostras podem ser reunidas como se fossem procedentes de uma mesma população, já que as diferenças entre as proporções dos três alelos estudados não diferem significativamente.

Tabela 5.2 Distribuição de três amostras populacionais quanto aos genótipos de fosfatase ácida eritrocitária.

Amostra	A	B	C	AB	AC	BC	Total
População 1	19	54	0	53	1	3	130
População 2	10	22	0	16	0	2	50
População 3	11	34	0	41	2	6	94

Tabela 6.2. Comparação das três amostras da Tabela 5.2 quanto às proporções dos alelos  $P^a$ ,  $P^b$  e  $P^c$ .

Amostra	$P^a$	$P^b$	$P^c$	Total
População 1	92	164	4	260
População 2	36	62	2	100
População 3	65	115	8	188

$$\chi^2_{(4)} = 3,402; 0,30 < P < 0,50$$

Nos casos de polialelismo em que existem relações de dominância e recessividade, a comparação de várias amostras pode ser feita usando o recurso de comparar as suas distribuições fenotípicas. No caso de trialelismo, como o dos grupos sanguíneos do sistema ABO clássico, Stevens (1950) idealizou um método mais preciso, que denominou de *método dos círculos equivalentes*, o qual permite a comparação da frequência dos alelos. Esse método foi publicado quando o eminente professor inglês, precocemente falecido, ministrava aulas de Estatística na Universidade de São Paulo. Ele leva em conta o sistema ABO, mas pode ser estendido para outros sistemas análogos

O método em questão é gráfico e permite a comparação rápida de um grande número de amostras, cada uma das quais é representada por um círculo, cujo raio é função da frequência  $r$  do gene  $O$ , ou da soma das frequências  $p$  e  $q$  dos alelos  $A$  e  $B$ . As coordenadas retangulares que servem para determinar a posição de cada círculo são dadas pelas raízes quadradas das frequências dos alelos  $A$  e  $B$  ( $\sqrt{p}$  e  $\sqrt{q}$ ).

No método de Stevens (1950) a hipótese de que duas ou mais amostras não diferem quanto às frequências gênicas do sistema ABO é aceita quando os círculos que as representam se interpenetram. Quando isso não ocorre, rejeita-se tal hipótese e aceita-se a alternativa de que as amostras têm frequências gênicas que diferem significativamente.

Para a obtenção dos raios dos círculos em função de  $r$ , Stevens (1950) preparou uma tabela que fornece os elementos para cálculo de raios de círculos que incluem diferentes probabilidades, parte da qual está apresentada na Tabela 7.2. Com base nos dados dessa tabela pode-se achar o raio do círculo equivalente se dividirmos pela raiz quadrada do tamanho amostral ( $\sqrt{n}$ ) o valor tabulado em função da frequência  $r$ . Assim, por exemplo, se o valor estimado de  $r$  em uma amostra de 100 indivíduos for 0,7, o raio do círculo de 80% de probabilidade será  $\frac{0,607}{10} = 0,0607$  e o de 95% de

probabilidade será  $\frac{0,829}{10} = 0,0829$ . Tais valores serão representados graficamente na mesma escala utilizada para indicar os valores de  $\sqrt{p}$  e de  $\sqrt{q}$ .

Tabela 7.2. Elementos para o cálculo de raios de círculos equivalentes que incluem 80% e 95% de probabilidade em função da frequência  $r$  do gene  $O$  do sistema ABO.

$r$	80%	95%
0,5	0,581	0,793
0,6	0,596	0,813
0,7	0,607	0,829
0,8	0,617	0,842
0,9	0,626	0,854
1,0	0,634	0,865

Se o valor de  $r$  não estiver representado na Tabela 7.2, poder-se-á fazer uma interpolação para obter o raio do círculo equivalente. Assim, por exemplo, se em uma amostra de 100 indivíduos o valor de  $r$  for estimado em 0,65 considerar-se-á que, quando  $r$  é igual a 0,6 o valor tabulado para obter o raio do círculo de 80% de probabilidade é 0,596, sendo igual a 0,607 quando  $r$  é igual a 0,7. Assim, a um aumento de 0,1 no valor de  $r$  corresponde um aumento de 0,011 no valor tabulado. Portanto, o valor 0,0055 deve corresponder a um aumento de 0,05, isto é, à diferença entre 0,65 e 0,6. Desse modo, o raio do círculo equivalente de 80% de probabilidade será  $\frac{0,596 + 0,0055}{10} = 0,0602$ . No caso de se querer obter o raio do círculo equivalente de 95% de probabilidade sendo o valor de  $r$  igual a 0,65 ter-se-ia, pelos mesmos motivos  $\frac{0,813 + 0,008}{10} = 0,0821$ . Do ponto de vista prático, contudo, as interpolações para a obtenção dos valores dos raios dos círculos equivalentes constituem apenas refinamentos. Portanto, do ponto de vista prático, quando  $r$  não estiver representado na Tabela 7.2 bastará tomar o seu valor mais próximo.

Para entender melhor a aplicação do método de Stevens (1950), consideremos os dados apresentados na Tabela 8.2 a respeito de cinco amostras, distribuídas segundo os grupos sanguíneos do sistema ABO clássico (Beiguelman e Marchi, 1962; Beiguelman, 1963). Os dados dessa tabela permitem obter os elementos necessários à construção dos círculos equivalentes, de modo que, a partir deles pode-se preparar uma outra tabela como a Tabela 9.2 e desenhar os gráficos das Figuras 1.2 e 2.2.

Tabela 8.2. Distribuição de cinco amostras segundo os grupos sanguíneos do sistema ABO clássico e estimativas corrigidas das frequências dos alelos  $A$ ,  $B$  e  $O$ . Entre parênteses estão assinalados os valores em porcentagem. 1. Brasileiros caucásoides com ancestrais brasileiros até bisavós; 2. brasileiros negróides; 3. italianos e brasileiros descendentes não-miscigenados de italianos; 4. espanhóis e brasileiros descendentes não miscigenados de espanhóis; 5. japoneses e descendentes não miscigenados de japoneses.

Amostra	A	B	AB	O	Total	$p$	$q$	$r$
1	228 (37,44)	52 (8,54)	15 (2,46)	314 (51,56)	609 (100)	0,225	0,057	0,718
2	56 (31,11)	23 (12,78)	8 (4,44)	93 (51,67)	180 (100)	0,196	0,090	0,714
3	276 (38,76)	69 (9,69)	33 (4,64)	334 (46,91)	712 (100)	0,247	0,074	0,679
4	47 (35,60)	13 (9,85)	5 (3,79)	67 (50,76)	132 (100)	0,221	0,071	0,708
5	111 (37,00)	68 (22,67)	40 (13,33)	81 (27,00)	300 (100)	0,293	0,198	0,509

Tabela 9.2. Elementos para a construção dos círculos equivalentes de 80% e 95% de probabilidade das amostras descritas na Tabela 7.

Amostra	$\sqrt{p}$	$\sqrt{q}$	$\sqrt{n}$	Raio (80%)	Raio (95%)
1	0,474	0,239	24,678	0,025	0,034
2	0,443	0,300	13,416	0,046	0,062
3	0,497	0,274	26,683	0,023	0,031
4	0,470	0,266	11,489	0,054	0,072
5	0,713	0,455	17,321	0,034	0,046

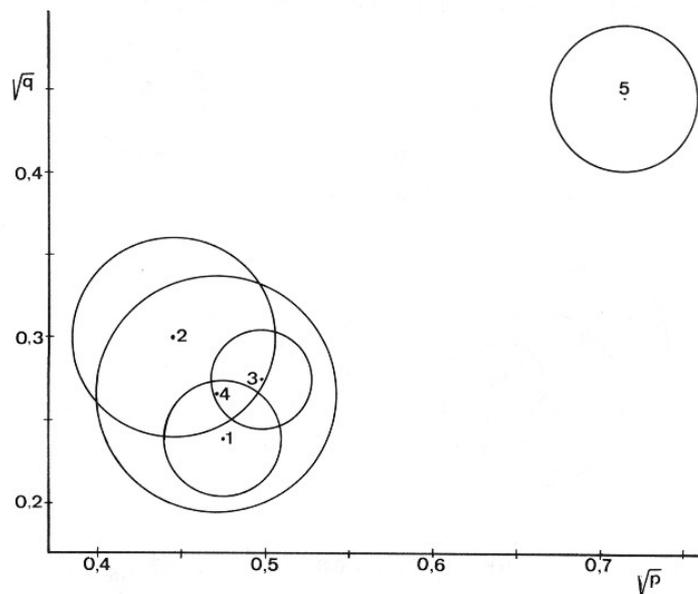


Fig.1.2. Círculos equivalentes de 95% de probabilidade das cinco amostras relacionadas na Tabela 8.2.

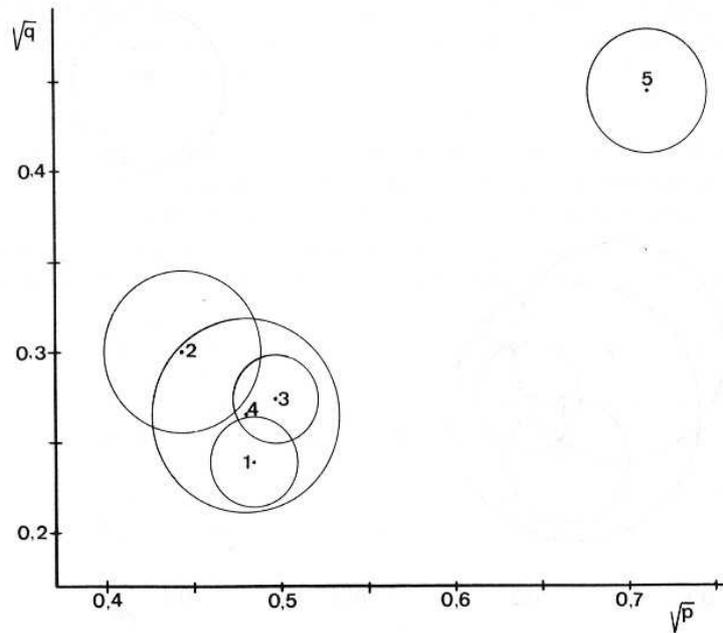


Fig.2.2. Círculos equivalentes de 80% de probabilidade das cinco amostras relacionadas na Tabela 8.2.

A análise da Fig.1.2 permite concluir que, no concernente às frequências gênicas do sistema ABO, as amostras 1, 2, 3 e 4 não diferem significativamente entre si, mas todas elas são bem diferentes da amostra 5, constituída por japoneses e seus descendentes não miscigenados. Um analista mais rigoroso construiria o gráfico da Fig.2.2 e, com base nos círculos equivalentes de 80% de probabilidade aceitaria que a amostra 2 difere da amostra 1 e da amostra 5, e que as amostras 1, 3 e 4 não diferem entre si, mas diferem da amostra 5.

## QUESTÕES E RESPOSTAS

**Q 1.** Um sistema genético é constituído por quatro alelos autossômicos designados pelas letras *A*, *B*, *C* e *D*. Se em uma população as frequências desses alelos forem  $A = 10\%$ ,  $B = 20\%$ ,  $C = 30\%$  e  $D = 40\%$  quais as frequências genótípicas esperadas em equilíbrio genético?

**R 1.**

$$\begin{array}{ll}
 AA = 0,01 & AC = 0,06 \\
 BB = 0,04 & AD = 0,08 \\
 CC = 0,09 & BC = 0,12 \\
 DD = 0,16 & BD = 0,16 \\
 AB = 0,04 & CD = 0,24
 \end{array}$$

**Q 2.** Os grupos sanguíneos do sistema ABO foram estudados em duas amostras de indivíduos caucasóides do Estado de São Paulo. Os resultados percentuais foram os apresentados abaixo. Com base neles:

- calcular as estimativas preliminares das frequências dos alelos *A*, *B* e *O*;
- verificar se as amostras representam populações em equilíbrio genético;
- calcular as estimativas corrigidas das frequências dos alelos *A*, *B* e *O*.

Amostra	No.	A	B	AB	O
1	712	38,76	9,69	4,64	46,91
2	609	37,44	8,54	2,46	51,56

**R 2. a)** Amostra 1

$$p' = 1 - \sqrt{0,0969 + 0,4691} = 0,2477$$

$$q' = 1 - \sqrt{0,3876 + 0,4691} = 0,0744$$

$$r' = \sqrt{0,4691} = 0,6849$$

$$D = -0,0070$$

Amostra 2

$$p' = 1 - \sqrt{0,0854 + 0,5156} = 0,2248$$

$$q' = 1 - \sqrt{0,3744 + 0,5156} = 0,0566$$

$$r' = \sqrt{0,5156} = 0,7181$$

$$D = 0,0005$$

b) As duas amostras representam populações em equilíbrio genético porque:

$$\text{Na amostra 1: } \chi^2_{(1)} = 2 \times 712 \left( 1 + \frac{0,5849}{0,2477 \times 0,0744} \right) 0,000049 = 2,667 ; 0,10 < P < 0,20.$$

$$\text{Na amostra 2: } \chi^2_{(1)} = 2 \times 609 \left( 1 + \frac{0,7181}{0,2248 \times 0,0566} \right) 0,00000025 = 0,017 ; 0,80 < P < 0,90$$

c) Amostra 1:  $p = 0,247$ ;  $q = 0,074$ ;  $r = 0,079$

Amostra 2:  $p = 0,225$ ;  $q = 0,057$ ;  $r = 0,718$

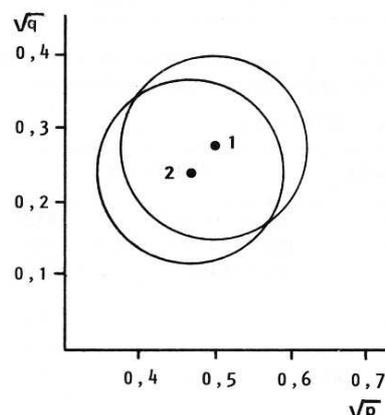
**Q 3.** Verificar graficamente se as duas amostras da questão anterior não diferem significativamente quanto às frequências dos genes do sistema ABO. Se elas não diferirem significativamente, calcular as estimativas combinadas das frequências dos alelos *A*, *B* e *O* nas amostras reunidas em uma única.

**R 3.** Tendo em vista a inexistência de diferença significativa entre as amostras estudadas pode-se calcular:

$$p = \frac{(712 \times 0,247) + (609 \times 0,225)}{1.321} = 0,237$$

$$q = \frac{(712 \times 0,074) + (609 \times 0,057)}{1.321} = 0,066$$

$$r = \frac{(712 \times 0,679) + (609 \times 0,718)}{1.321} = 0,697$$



**Q 4.** As amostras de sangue de praticamente todos os indivíduos (180) de uma comunidade ribeirinha de Rondônia (Portuchuelo, RO) foram investigadas eletroforeticamente para estudar a distribuição genotípica da fosfatase ácida eritrocitária, tendo-se observado o seguinte resultado:

$$AA = 2; BB = 118; CC = 0; AB = 50; AC = 1; BC = 9$$

Quer-se saber: as estimativas das freqüências dos alelos A, B e C, e se a distribuição genotípica observada pode ser considerada como representativa de uma população em equilíbrio de Hardy e Weinberg.

**R 4.** As freqüências  $p$ ,  $q$  e  $r$  dos alelos A, B e C podem ser obtidas como abaixo:

$$p = \frac{4 + 50 + 1}{360} = 0,153 \quad q = \frac{236 + 50 + 9}{360} = 0,819 \quad r = \frac{0 + 1 + 9}{360} = 0,028$$

Essas freqüências permitem calcular os valores esperados em cada classe genotípica e, com base no valor do qui-quadrado obtido (2,070) com três graus de liberdade, concluir que as proporções observadas não diferem significativamente daquelas esperadas em equilíbrio de Hardy e Weinberg

Valores	AA	BB	CC	AB	AC	BC
Obs. ( $o$ )	2	118	0	50	1	9
Esp. ( $e$ )	4,2	120,7	0,1	45,1	1,5	8,3
$\frac{(o - e)^2}{e}$	1,152	0,060	0,100	0,532	0,167	0,059

$$\chi^2_{(3)} = 2,070; 0,50 < P < 0,70.$$

**Q 5.** As hemácias de 862 indivíduos da população de Monte Negro, RO, foram tratadas com os anti-soros anti-Fy<sup>a</sup> e anti-Fy<sup>b</sup> do sistema sanguíneo Duffy, tendo sido observado o seguinte resultado:

Fenótipo	Fy(a+b-)	Fy(a-b+)	Fy(a+b+)	Fy(a-b-)
Genótipo	$Fy^a Fy^a$ ou $Fy^a Fy$	$Fy^b Fy^b$ ou $Fy^b Fy$	$Fy^a Fy^b$	$Fy Fy$
No.	236	328	243	55
%	27,38	38,05	28,19	6,38

Quer-se saber as estimativas das freqüências dos alelos  $Fy^a$ ,  $Fy^b$  e  $Fy$  e se a amostra estudada representa uma população em equilíbrio genético.

**R 5.** Aplicando o método de Bernstein para a análise dos dados do sistema Duffy, obtemos, inicialmente, as estimativas preliminares  $p'$ ,  $q'$  e  $r'$  dos alelos  $Fy^a$ ,  $Fy^b$  e  $Fy$ , bem como o desvio D:

$$\begin{aligned}
 p' &= 0,3334 \\
 q' &= 0,4190 \\
 r' &= 0,2526 \\
 D &= 1 - 1,005 = - 0,005
 \end{aligned}$$

Com esses elementos obtemos as estimativas corrigidas  $p$ ,  $q$  e  $r$ :

$$\begin{aligned}
 p &= 0,3326 \\
 q &= 0,4180 \\
 r &= 0,2494
 \end{aligned}$$

e, pelo baixo valor do qui-quadrado ( $\chi^2_{(1)} = 0,120$ ;  $0,70 < P < 0,80$ ) concluímos que a amostra de Monte Negro representa uma população em equilíbrio genético em relação aos grupos sanguíneos do sistema Duffy.

## REFERÊNCIAS

- Beiguelman, B. Grupos sanguíneos e lepra. *Rev. Bras. Lepr.* 31: 34-44, 1963.
- Beiguelman, B. & Marchi, A. Blood groups among Japanese in Brazil. *Am. J. Phys. Anthrop.* 20: 29-31, 1962
- Bernstein, F. Ergebnisse einer biostatistischen zusammenfassenden Betrachtung über die erblichen Blutstrukturen des Menschen. *Klin. Wschr.* 3: 1495-1497, 1924.
- Bernstein, F. Zusammenfassende Betrachtungen über die erbliche Blutstrukturen des Menschen. *Z. indukt. Abstamm. u. VererbLehre.* 37: 237-270, 1925.
- Bernstein, F. Fortgesetzte Untersuchungen aus der Theorie der Blutgruppen. *Z. indukt. Abstamm. u. VererbLehre* 56: 233-237, 1930.
- Cabello, P.H. & Krieger, H. *Genioc. Sistema para análises de dados de Genética*. Publicação Técnica, Departamento de Genética, Instituto Oswaldo Cruz-FIOCRUZ, Rio de Janeiro, 1997.

- Cook, P.J., Robson, E.B., Buckton, K.E., Slaughter, C.A., Gray, J.E., Blank, C.E., James, F.E., Ridler, M.A., Insley, J. & Hulten, M. Segregation of ABO, AK(1) and ACONs in families with abnormalities of chromosome 9. *Ann. Hum. Genet.* 41: 365-378, 1978.
- Narahara, K., Takahashi, Y., Kikkawa, K., Wakita, Y., Kimura, S. & Kimoto, H. Assignment of ABO locus to 9q31.3qter by study of a family in which an intrachromosomal shift involving chromosome 9 is segregating. *Japan. J. Hum. Genet.* 31: 289-296, 1986.
- Stevens, W.L. Estimation of blood-group gene frequencies. *Ann. Eugenics* 8: 362-375, 1938.
- Stevens, W.L. Statistical analysis of the ABO blood groups. *Human Biol.* 22:191-217, 1950.