

Matemática

O equilíbrio de Hardy-Weinberg no sistema sanguíneo ABO: um estudo de caso em Engenheiro Coelho – SP

The Hardy–Weinberg equilibrium in the ABO blood group system: a case study in Engenheiro Coelho – SP

Vinícius Freitas de Oliveira^I , Guilherme Augusto Pianezzer^{II} ,
Suzete Maria Silva Afonso^{III} 

^I Centro Universitário Adventista de São Paulo, Engenheiro Coelho, SP, Brasil

^{II} Centro Universitário Internacional, Curitiba, PR, Brasil. Brasil

^{III} Universidade Estadual Paulista Júlio de Mesquita Filho, Instituto de Geociências e Ciências Exatas de Rio Claro, Rio Claro, SP, Brasil

RESUMO

A genética de populações humanas é o ramo da Genética que estuda a dinâmica dos genes nas populações naturais, visando a dilucidação dos mecanismos que alteram a sua composição gênica. Dentre os fundamentos dessa ciência está o Equilíbrio de Hardy-Weinberg, que determina que as frequências gênicas permanecem inalteradas e as proporções genotípicas atingem um equilíbrio estável, obtendo a mesma relação constante entre si ao longo do tempo. Para demonstrar esse princípio, é preciso admitir que a população em estudo não esteja sujeita a fatores evolutivos nem àqueles que alteram as frequências genotípicas aumentando a homozigose. Mais especificamente, é preciso supor que a população obedeça às seguintes premissas: acasalamento aleatório, população infinita, gerações não-sobrepostas, além de ausência de mutação, seleção e migração. Mais do que recordar conceitos básicos de Genética e Estatística e promover uma interdisciplinaridade entre Biologia e Matemática, este artigo tem como objetivo descrever o Método de Bernstein para a verificação do equilíbrio gênico para tipos sanguíneos. A pesquisa é concluída com um estudo de caso na cidade de Engenheiro Coelho - SP, onde se verifica o Equilíbrio de Hardy-Weinberg para tipos sanguíneos na população.

Palavras-chave: Genética de populações; Biomatemática; Equilíbrio de Hardy-Weinberg; Modelo de Bernstein

ABSTRACT

The genetics of human populations is the branch of Genetics that studies the dynamics of genes in natural populations, aiming at the elucidation of mechanisms that alter their genetic composition. Among the fundamentals of this science is the Hardy-Weinberg Equilibrium, which determines that gene frequencies remain unchanged and genotypic proportions reach a stable balance, obtaining the same constant relation with each other over time. To demonstrate this principle, it is necessary to admit that the studied

population is not subject to evolutionary factors or to those that alter genotypic frequencies, increasing the homozygosity. More specifically, it is necessary to assume that the population obeys the following premises: random mating, infinite population, non-overlapping generations, in addition to the absence of mutation, selection and migration. More than recalling basic concepts of Genetics and Statistics, this article aims to describe the Bernstein Method for verifying the gene equilibrium for blood types. The research is concluded with a case study in the city of Engenheiro Coelho - SP, where the Hardy-Weinberg Equilibrium for blood types in the population is verified.

Keywords: Population genetics; Biomathematics; Hardy-Weinberg equilibrium; Bernstein model

1 INTRODUÇÃO

A genética de populações é a ciência que estuda o comportamento e os fenômenos de alteração das frequências alélicas e genotípicas de uma população. Com o uso de ferramentas matemáticas aliadas aos conceitos biológicos, a principal característica deste estudo é a capacidade de analisar as frequências gênicas no decorrer do tempo.

Para a compreensão dos conceitos biológicos é importante tomar por início o sistema estrutural de funcionamento dos seres vivos, a *célula*. Este sistema é composto por diversas organelas e nele está contida toda a informação genética. As células podem ser classificadas como procariontes ou eucariontes. Para o primeiro caso a célula não possui carioteca, ou seja, as informações genéticas não são organizadas em um núcleo. Já para o segundo caso, caso dos seres humanos, toda a informação hereditária é armazenada no núcleo, uma estrutura mais complexa (Oliveira, 2019).

Segundo as descobertas de Gregor Mendel (1822-1884), a herança genética é passada de pai para filho, herança esta que é chamada de *gene*. O gene é um trecho da molécula de ácido desoxirribonucleico, o *DNA*, localizado no núcleo das células eucariontes. O DNA é formado por um par de fitas paralelas em sequência compostas por apenas quatro diferentes elementos estruturais: a adenina (A), a guanina (G), a citosina (C) e a timina (T) (Griffiths et al., 2006).

Segundo Amabis e Martho (1994), no gene, através das alterações sequenciais de suas quatro diferentes estruturas, existe um código de como deverá ser

produzida uma proteína. A construção dessa proteína é o que representa a característica de um indivíduo. Sendo assim, é no DNA que estão todas as informações que representarão todas as características de um ser humano, desde a cor dos olhos ao tipo sanguíneo.

As diferentes formas nas estruturas de um gene são chamadas de *alelos*. Os alelos, são representados por letras, sendo recessivos ou dominantes. Alelos recessivos só possuem expressão quando sozinhos e são representados por letras minúsculas. Já os alelos dominantes sempre se expressarão e são representados por letras maiúsculas (Oliveira, 2019).

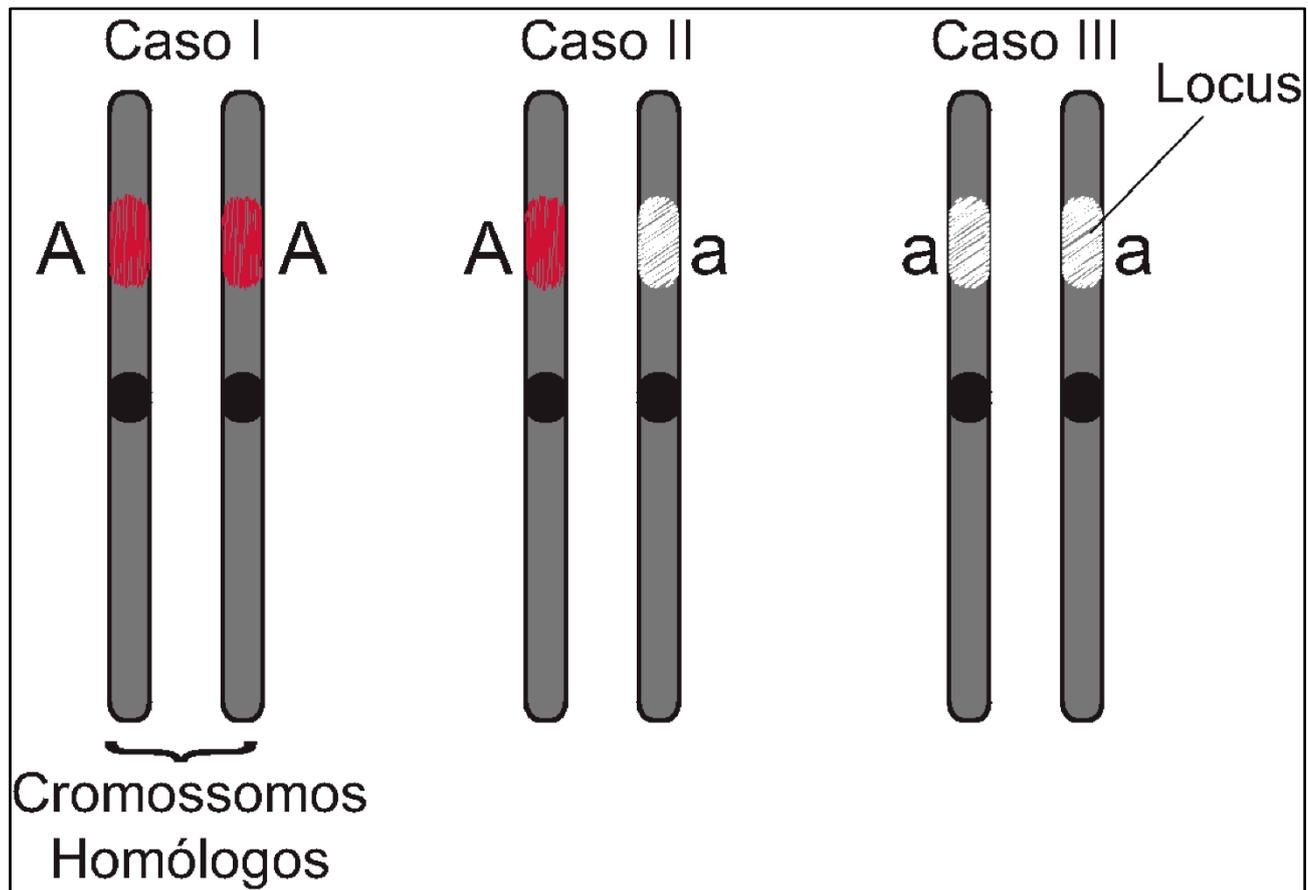
Os alelos ficam alocados no *locus* (no plural loci) em uma estrutura chamada *cromossomo*. Neste contexto, Lewis (2004) define cromossomo como um par idêntico de fitas de DNA enroladas. Quando dois cromossomos são pareados em pares homólogos, isto é, pares iguais entre si, é possível a visualização da estrutura gráfica dos alelos, os *genótipos*. Segundo Amabis e Martho (1994) são os genótipos que possuem o código para a construção da proteína e conseqüentemente a determinação de uma característica. A partir de agora essa característica será entendida como *fenótipo*.

As relações entre genótipos e fenótipos são muito mais expressivas. O genótipo em si tem apenas influência hereditária, ou seja, uma prole recebe metade dessas características do pai e a outra metade da mãe, contudo, a noção fenotípica, além de depender do genótipo também depende do meio. Por exemplo, a cor de pele é uma característica - fenótipo. Assim, existe um genótipo que irá determinar a quantidade da proteína melanina que será produzida. Porém, a exposição do indivíduo ao sol também pode alterar esta característica (Lewis, 2004).

Para uma melhor compreensão, a Figura 1 mostra três pares homólogos de cromossomos bem como os seus genótipos. No Caso I tem-se o genótipo **AA**, já no Caso II, o genótipo **Aa**, e por fim, no Caso III, o genótipo **aa**. Nos casos em que os alelos são iguais classificam-se os genótipos em *homozigotos* (Caso I e III) e para os

casos em que os alelos são diferentes, classificam-se os genótipos em *heterozigotos* (Caso II).

Figura 1 – Cromossomos Homólogos e Genótipos



Todas as noções apresentadas até aqui serão fundamentais para a contemplação dos objetivos do presente artigo, que são: descrever um método estatístico para a verificação do equilíbrio gênico (Princípio de Hardy-Weinberg) para o Sistema ABO de tipos sanguíneos e apresentar um estudo de caso para a população da cidade de Engenheiro Coelho - SP.

O artigo está organizado como segue. Princípios de genética de populações serão descritos na seção 2, na qual destaca-se o Teorema de Hardy-Weinberg. O sistema ABO de tipos sanguíneos será brevemente apresentado na seção 3. A seção 4 destinar-se-á ao Método de Bernstein (apresentado por Bernstein (1924))- o método estatístico supracitado para a verificação do equilíbrio gênico. A última

seção, Seção 5, trará uma aplicação para a pesquisa realizada, a saber, o estudo de caso que fora mencionado.

2 PRINCÍPIOS DE GENÉTICA DE POPULAÇÕES

A compreensão de uma população e de suas características genéticas dependem de diversas premissas e conceitos. Esta seção tem como objetivo descrever algumas premissas e alguns conceitos de estatística que são importantes para o estudo da genética de populações. A teoria aqui exposta também pode ser estudada em Oliveira (2019) e Hartl e Clark (2007).

2.1 Frequências Alélicas e Frequências Genotípicas

Definição 2.1. *Considere uma população qualquer e um dos possíveis genótipos, $G_1 \dots G_k$. A frequência $f_{G_1 \dots G_k}$ é definida pela seguinte equação:*

$$f_{G_1 \dots G_k} \triangleq \frac{\text{números de indivíduos com genótipo } G_1 \dots G_k}{\text{tamanho da população}} \quad (1)$$

Se o tempo for considerado, $f_{G_1 \dots G_k}(t)$ denotará a frequência genotípica no tempo t :

Exemplo 2.1. Considere a seguinte população em um estudo genotípico que assume dois alelos A e a para um único locus:

AA, AA.

Note que existem 20 (vinte) indivíduos representados nessa população, dos quais 7 (sete) são homocigóticos portadores do genótipo AA, 8 (oito) são homocigóticos portadores do genótipo aa e 5 (cinco) são heterocigóticos portadores do genótipo Aa. Dessa forma, a fórmula (1) fornece $f_{AA} = \frac{7}{20}$, $f_{AA} = \frac{8}{20} = \frac{2}{5}$ e $f_{Aa} = \frac{5}{20} = \frac{1}{4}$

Veja que a Definição 2.1 exemplificada pelo Exemplo 2.1 traz um caso particular da definição de frequência relativa da Estatística Descritiva (veja Grimmett

e Stirzaker (2001) e Magalhães e Lima (2010)). De maneira similar, para a frequência alélica temos a definição a seguir.

Definição 2.2. *Seja l um locus qualquer. Dada uma população, a coleção alélica para o locus l é a coleção de todos os alelos de genótipos da população que ocorrem no locus l . Considere o alelo particular A que pode ocorrer em l . A frequência f_A de A é definida como sua frequência relativa à coleção alélica correspondente a l da seguinte forma:*

$$f_A \triangleq \frac{\text{números de } A\text{'s na coleção alélica de } l}{\text{tamanho da coleção alélica de } l} \quad (2)$$

Exemplo 2.2. Considerando a mesma população do Exemplo 2.1 é fácil notar que na coleção alélica existem 19 (dezenove) alelos A e 21 (vinte e um) alelos a , portanto, pela relação (2) segue que $f_A = \frac{19}{40}$ e $f_a = \frac{21}{40}$.

2.2 Acasalamento Aleatório

Magalhães e Lima (2010) definem fenômeno aleatório como uma “situação ou acontecimento cujos resultados não podem ser previstos com certeza” (Magalhães e Lima, 2010, p. 49). Essa definição é exatamente o que sustenta o conceito de acasalamento aleatório. Em outras palavras, o acasalamento aleatório ocorre quando todos os indivíduos de uma população têm chances iguais de se acasarem e ocorrer um sucesso reprodutivo. A Definição 2.3 descreve formalmente esse conceito.

Definição 2.3. *O acasalamento aleatório é a criação de um novo indivíduo pela união de dois gametas, um óvulo e um espermatozoide, escolhidos ao acaso na população. Em suma, cada gameta, seja masculino ou feminino, é cedido de maneira aleatória por seus pais.*

Segundo Magalhães e Lima (2010), em algumas situações as frequências relativas são iguais às suas respectivas probabilidades. Os autores ainda observam

que na biologia este fato é muito comum. O próximo resultado estabelece que esse fenômeno se verifica para as frequências alélicas

Lema 2.1. *Sejam S uma população da qual um pai é escolhido e A um alelo. Considere a experiência de escolher primeiro um pai por amostragem aleatória de S e, em seguida, escolher um dos seus gametas aleatoriamente. Se p_A^S é a probabilidade desse gameta possuir o alelo A e f_A^S é a frequência desse alelo na população S , então*

$$p_A^S = f_A^S \quad (3)$$

O Lema 2.1 será de suma importância para o estudo desenvolvido neste trabalho. Sua prova pode ser analisada em Oliveira (2019).

2.3 Premissa da População Infinita

Outro princípio importante é a premissa da população infinita. Este princípio consiste em assumir que o tamanho da população tende ao infinito e, embora tão simples, traz um resultado significativo, similar ao Lema 2.1.

Lema 2.2. *Considere uma população gerada por acasalamentos aleatórios. Seja $f_{G_1 \dots G_k}$ a frequência de um genótipo na população de descendentes e seja $p_{G_1 \dots G_k}$ a probabilidade de que o genótipo é produzido em um único acasalamento aleatório. A premissa de população infinita fundamenta-se em impor que, para todo genótipo, é válida a seguinte igualdade*

$$f_{G_1 \dots G_k} = p_{G_1 \dots G_k} \quad (4)$$

O Lema 2.2 complementa o Lema 2.1 ao estabelecer que para as frequências genotípicas também vale a igualdade com as probabilidades; sua demonstração também pode ser encontrada em Oliveira (2019). A identidade (4) representa o que deveria acontecer no limite quando o tamanho da população tende ao infinito e sua compreensão é facilitada por meio de um resultado não elementar da estatística: a Lei

Forte dos Grandes Números, que pode ser consultada em Grimmett e Stirzaker (2001). Na prática, o infinito pressuposto da população diz que a população em estudo é tão grande que a equação (4) é uma excelente aproximação para a relação entre as frequências genotípicas e as probabilidades.

2.4 Mutaç o, Seleç o e Migraç o

Segundo Hartl e Clark (2007), os tr s principais fatores que mais alteram as frequ ncias g nicas no decorrer do tempo   a mutaç o, a seleç o natural e a migraç o. Segundo Lewis (2004), a mutaç o ocorre quando h  uma mudanç a na sequ ncia da estrutura de um  nico gene. Essa mudanç a provoca diferenç as al licas e genot picas e pode ocorrer por meio de um erro no processo de reproduç o ou por consequ ncias de radiaç o, por exemplo.

J  a seleç o natural, para Griffiths et al. (2006), ocorre quando diferentes indiv duos t m diferentes chances de sobreviv ncia ou sucesso reprodutivo. Na populaç o humana esse fen meno geralmente ocorre em pa ses subdesenvolvidos devido   fome, m s condiç es de viv ncia, saneamento b sico prec rio, entre outros fatores.

Por fim, a migraç o, segundo Hartl e Clark (2007), ocorre quando indiv duos entram e saem sem ordenaç o do grupo de poss veis pais de acasalamento aleat rio. Esse fen meno geralmente ocorre nas cidades universit rias, onde os alunos chegam para cursar o Ensino Superior, entram no grupo de poss veis acasaladores, se formam e deixam as cidades depois de um certo tempo. Um efeito da migraç o   a anulaç o do Lema 2.1 devido  s diverg ncias genot picas no decorrer do tempo.

2.5 Geraç es N o-sobrepostas

O conceito de geraç es n o-sobrepostas, segundo Hartl e Clark (2007), consiste em assumir que indiv duos de uma geraç o t qualquer acasalam para produzir a geraç o $t + 1$ e depois n o se acasalam mais. Ap s isso, os membros da geraç o $t + 1$ acasalam entre si para gerar a geraç o $t + 2$ e assim, sucessivamente.

2.6 Equilíbrio de Hardy-Weinberg

O matemático inglês Godfrey H. Hardy e o médico alemão Wilhelm Weinberg, em 1908, concluíram que, se não houvesse fator evolutivo atuando sobre uma população que satisfizesse determinadas condições, as frequências de seus alelos permaneceriam inalteradas ao longo das gerações. Esse princípio ficou conhecido como *Princípio (ou Equilíbrio) de Hardy-Weinberg* e será descrito nas próximas linhas.

Definição 2.4. *As frequências genotípicas f_{AA} , f_{Aa} e f_{aa} , com $f_{AA} + f_{Aa} + f_{aa} = 1$, estão no equilíbrio de Hardy-Weinberg se existir $p \in \mathbb{R}$, com $0 \leq p \leq 1$, de tal modo que*

$$f_{AA} = p^2, \quad f_{Aa} = 2p(1 - p), \quad e \quad f_{aa} = (1 - p)^2 \quad (5)$$

A Definição 2.4 possibilita o entendimento do Teorema de Hardy-Weinberg, que será enunciado a seguir.

Teorema 2.1. *Se uma população obedecer às seguintes premissas:*

(P1) *Acasalamento aleatório;*

(P2) *Gerações não-sobrepostas;*

(P3) *População infinita;*

(P4) *Sem mutação, seleção e migração,*

então pode-se afirmar que as frequências alélicas são constantes e, para qualquer geração $t \geq 1$, as frequências genotípicas f_{AA} , f_{Aa} e f_{aa} estão no Equilíbrio de Hardy-Weinberg, com $p = f_A(0)$.

O obedecimento das premissas do teorema acima fundamentam a validade do Equilíbrio de Hardy-Weinberg, uma vez que a evidência de gerações não-sobrepostas garante que as gerações estão bem definidas; a ausência de mutação garante que as frequências gênicas permaneçam inalteradas; e a inexistência de mutação, seleção e migração asseguram o acasalamento aleatório e a população infinita justificando a utilização dos Lemas 2.1 e 2.2.

A partir de deduções lógicas das premissas (P1)-(P4) também é possível a construção do chamado modelo básico, composto por três equações de diferenças. Hardy e Weinberg foram os primeiros pesquisadores a encontrar soluções para este modelo, a partir do Teorema 2.1. Para não fugirmos do tema do artigo, não nos ateremos aos detalhes dessa modelagem. No entanto, ao leitor interessado, informamos que estes podem ser analisados em Oliveira (2019).

Mesmo que aparentemente o modelo acima seja quase inaplicável na natureza, os resultados das pesquisas de Hardy e Weinberg trazem resultados aproximados de extrema importância. Exemplos desses resultados são: a possibilidade da verificação da solidez ou não de hipóteses sobre mecanismos de herança; conhecimento das frequências dos vários genótipos, mesmo quando não reconhecidos pelos fenótipos; possibilidade de verificar os efeitos da seleção, migração e mutação; dentre outros (Freire-Maia, 1974). O exemplo abaixo mostra um caso utópico em que a amostra encontra-se perfeitamente no equilíbrio gênico.

Exemplo 2.3. *Suponha que deseja-se estudar uma determinada espécie de borboletas que possuem asas de cor azul ou de cor amarela, sabendo que a determinação deste fenótipo ocorre a partir de um gene que possui dois alelos, B e Y. Uma amostra de 150 borboletas foi colhida e os resultados são vistos pela tabela abaixo.*

Genótipo	Frequência Absoluta Observada
BB	24
BY	72
YY	54

Primeiramente, calcular-se-á as frequências genotípicas. Por 1, $f_{BB} = 0,16$, $f_{BY} = 0,48$, e $f_{YY} = 0,36$. Assim, é fácil notar que a população está no Equilíbrio de Hardy-Weinberg, pois existe $p = 0,4$ tal que

$$\begin{cases} f_{AA} = 0,4^2 = 0,16 \\ f_{BY} = 2 \cdot 0,4 \cdot (1 - 0,4) = 0,48 \\ f_{YY} = (1 - 0,4)^2 = 0,36 \end{cases} \quad (6)$$

É importante dizer que se as frequências genotípicas observadas não aderirem ao Equilíbrio de Hardy-Weinberg, não é imediato que todas as suposições do modelo são inválidas (Hartl e Clark, 2007). Daí, torna-se importante testes estatísticos para averiguar, dado um nível de significância, o quão próximas as frequências estão do modelo, ressaltando ainda mais a importância da interdisciplinaridade.

Assim, quando não é possível demonstrar que uma amostra representa uma população em equilíbrio de Hardy e Weinberg, deve-se investigar a(s) causa(s) desse desvio entre os fatores evolutivos e entre aqueles que alteram as frequências genotípicas, tais como os casamentos consanguíneos (Beiguelman, 2008, p. 17).

Além dos casos de dialelismo – um par de alelos em um locus autossômico, Beiguelman (2008) afirma ainda que o equilíbrio encontrado pelos pesquisadores ocorre também em casos de polialelismo autossômico, ou seja, também se enquadra no sistema sanguíneo ABO clássico. Este caso será melhor discutido na Seção 4 com o uso do Método de Bernstein.

3 SISTEMA ABO

No final do século XIX, o médico austríaco Karl Landsteiner (1868 - 1943) percebeu que ao misturar amostras de sangue de algumas pessoas, as hemácias se juntavam e aglomerados semelhantes a coágulos eram formados. Karl Landsteiner concluiu que algumas pessoas possuem sangues incompatíveis, o que revelou a divergência de diferentes tipos sanguíneos na população (Batissoco e Novaretti, 2003).

Através das descobertas de Landsteiner notou-se a presença de quatro tipos de sangue: *A*, *B*, *AB* e *O*. A presença ou não de certas substâncias diferem esses tipos sanguíneos. Segundo Beiguelman (2008), 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*.

Devido aos antígenos, alguns tipos sanguíneos não podem se misturar. Assim, as doações de sangue só podem ocorrer da seguinte maneira: o tipo *A* só pode

receber de *O* ou *A*, o tipo *B* só pode receber de *O* ou *B*, o tipo *AB* recebe de todos e o tipo *O* só recebe dele mesmo.

Geneticamente, os alelos *A* e *B* possuem uma relação de codominância entre si, ou seja, ambos são dominantes de igual forma, de modo que os heterozigotos *AB* são responsáveis pela formação do grupo sanguíneo *AB*. Desta forma, ambos alelos dominam o alelo *O*. A Tabela 1 mostra a relação Fenótipo x Genótipo do Sistema ABO.

Tabela 1 – Relação Fenótipo x Genótipo do Sistema ABO

Fenótipo (Tipo Sanguíneo)	Genótipo
A	AA
	AO
B	BB
	BO
AB	AB
O	OO

4 MÉTODO DE BERNSTEIN

A verificação do Equilíbrio de Hardy-Weinberg para tipos sanguíneos no sistema ABO pode ser observada através do Método de Bernstein que será descrito nesta seção, tomando como base Bernstein (1924). Para inserir os dados no método, as frequências dos alelos *A*, *B* e *O* serão designadas por *p*, *q* e *r*, respectivamente. Esses resultados permitem uma expansão da Tabela 1 para a Tabela 2.

Note que as frequências fenotípicas em equilíbrio geral estável são calculadas por intermédio da expansão do trinômio $(p + q + r)^2$, descrito na última coluna da Tabela 2. Veja também que se a amostra representar uma população em equilíbrio de Hardy-Weinberg segue que $(p + q + r)^2 = 1$, conforme ampliação da Definição 2.4.

Os indivíduos do grupo O possuem o genótipo OO , logo, sua frequência é obtida por $r \cdot r = r^2$. Sendo assim, para se obter a frequência do alelo O basta extrair a raiz quadrada da frequência observada de indivíduos do grupo O . Portanto, $r = \sqrt{r^2} = \sqrt{O}$. Já para o alelo A , segue que:

$$\begin{aligned} (p + q + r)^2 = 1 &\Rightarrow p = 1 - (q + r) \\ &= 1 - \sqrt{(q + r)^2} \\ &= 1 - \sqrt{q^2 + 2qr + r^2} \\ &= 1 - \sqrt{B - O} \end{aligned} \quad (7)$$

onde a última igualdade vem dos resultados da Tabela 2. De maneira análoga, tem-se para o alelo B que $q = 1 - \sqrt{A + O}$.

Levando em conta que a soma dessas estimativas dificilmente fornece um valor igual a 1, essas frequências alélicas são denominadas *estimativas preliminares*, representadas por \hat{p} , \hat{q} e \hat{r} . As equações (8), (9) e (10) descrevem essas estimativas.

$$\hat{p} = 1 - \sqrt{B + O} \quad (8)$$

$$\hat{q} = 1 - \sqrt{A + O} \quad (9)$$

$$\hat{r} = \sqrt{O} \quad (10)$$

O próximo passo trata-se de uma correção por meio do cálculo de um desvio (D). Esse desvio é encontrado pela diferença entre 1 e a soma das estimativas preliminares, ou seja, o desvio mostra o quão distante, inicialmente, está a amostra do equilíbrio

Tabela 2: Relação Fenótipo x Genótipo do Sistema ABO

Fenótipo (Tipo Sanguíneo)	Genótipo	Freq. Genotípica	Freq. Fenotípica
A	AA	p^2	$p^2 + 2pr$
	AO	$2pr$	
B	BB	q^2	$q^2 + 2qr$
	BO	$2qr$	
AB	AB	$2pq$	$2pq$
O	OO	r^2	r^2

gênico. Sendo assim,

$$D = 1 - (\hat{p} + \hat{q} + \hat{r}) \quad (11)$$

Posto isto, de forma que a soma das estimativas seja de fato igual a 1, as estimativas preliminares \hat{p} , \hat{q} e \hat{r} devem ser corrigidas. Essas correções p , q e r são determinadas em seguida, pelas Equações (12), (13) e (14).

$$p = \hat{p} \left(1 + \frac{D}{2}\right) \quad (12)$$

$$q = \hat{q} \left(1 + \frac{D}{2}\right) \quad (13)$$

$$r = \left(\hat{r} + \frac{D}{2}\right) \left(1 + \frac{D}{2}\right) \text{ ou } r = 1 - (p + q) \quad (14)$$

A fim de verificar se a distribuição dos fenótipos na população estudada está de acordo com a hipótese genética, isto é, se a população encontra-se de fato no Equilíbrio de Hardy-Weinberg, Bernstein sugeriu o teste do qui-quadrado. Conforme Magalhães e Lima (2010), o cálculo do χ^2 ocorre por meio da Equação 15 a seguir

$$\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i} \quad (15)$$

onde k representa o número de classes, o_i a frequência observada em cada classe e e_i a frequência esperada para aquela classe. Adaptando (15) para o Sistema ABO tem-se a Equação (16):

$$\chi^2 = \frac{[A - (p^2 + 2pr)n]^2}{(p^2 + 2pr)n} + \frac{[B - (q^2 + 2qr)n]^2}{(q^2 + 2qr)n} + \frac{(AB - 2pqn)^2}{2pqn} + \frac{(O - r^2n)^2}{r^2n} \quad (16)$$

onde A , B , AB e O são os números observados de indivíduos em cada um desses grupos sanguíneos e n é o tamanho da amostra.

Por fim, a verificação do equilíbrio gênico vem por meio da Região Crítica (RC). Dado um nível de significância α , verifica-se o valor crítico q_c por meio da Tabela do

qui-quadrado (veja (Magalhães e Lima, 2010, p. 373)) e define-se RC da seguinte maneira:

$$RC = \{\chi^2 : \chi^2 > q_c\} \quad (17)$$

Cabe aqui recordar que o *nível de significância* (α) significa o risco de se rejeitar uma hipótese verdadeira e deve ser estabelecido antes da análise de dados; o valor de χ^2 ao nível de significância α é denominado *qui-quadrado crítico ou tabelado* e denotado por q_c . Ademais, denomina-se *grau de liberdade* à diferença entre o número de classes de resultados e o número de informações da amostra que são necessários ao cálculo dos valores esperados nessas classes.

Se $\chi^2 \notin RC$, então a população estará no Equilíbrio de Hardy-Weinberg. Cabe ainda salientar que o qui-quadrado para o sistema ABO terá apenas um grau de liberdade visto que existem quatro classes fenotípicas e são necessárias apenas três informações da amostra (tamanho da amostra e as frequências de dois genes) para calcular os valores esperados.

5 ESTUDO DE CASO

Esta seção se propõe a cumprir o último objetivo deste artigo: apresentar um estudo de caso para verificar o Equilíbrio de Hardy-Weinberg para os tipos sanguíneos na população da cidade de Engenheiro Coelho - SP. A cidade em questão, localizada na região metropolitana de Campinas, possui uma população estimada de 20.773 habitantes, dados do IBGE (2020).

5.1 Metodologia

Para a coleta dos dados, o Laboratório Municipal de Análises Clínicas da cidade forneceu as informações dos 267 exames de tipagem sanguínea realizados no ano de

2019, representando uma amostra aleatória da população. Com a amostra obtida, o Método de Bernstein será aplicado de forma a obter e discutir os resultados.

5.2 Resultados e Análise dos Resultados

Através dos resultados fornecidos pelo Laboratório Municipal de Análises Clínicas de Engenheiro Coelho - SP e da Equação (1) construiu-se a Tabela 3, tabela de frequência para os tipos sanguíneos. Cabe ressaltar que o fator RH é irrelevante para o presente estudo.

Tabela 3 – Tabela de Frequência para os Tipos Sanguíneos

Fenótipo (Tipos Sanguíneos)	n_i	f_i
A	101	0,37828
B	37	0,13858
AB	9	0,03371
O	120	0,44944
Total	267	1,0

Com efeito, por intermédio das frequências fenotípicas e das Equações (8), (9) e (10), calcula-se as estimativas preliminares das frequências alélicas:

$$\hat{p} = 1 - \sqrt{0,13858 + 0,44944} = 0,23318$$

$$\hat{q} = 1 - \sqrt{0,37828 + 0,44944} = 0,09021$$

$$\hat{r} = \sqrt{0,44944} = 0,67040$$

Por conseguinte, tem-se, por (11), $D = 1 - (0,23318 + 0,09021 + 0,67040) = 0,00621$. Em consequência, corrige-se as estimativas preliminares, por (12), (13) e (14), obtendo

$$p = 0,23318 + \frac{0,00621}{2} = 0,23390 \quad (12)$$

$$q = 0,09021 + \frac{0,00621}{2} = 0,09049 \quad (13)$$

$$r = 1 - (0,23390 + 0,09049) = 0,67561 \quad (14)$$

Com as estimativas corrigidas é possível calcular o χ^2 por meio da Equação (16). Assim,

$$\chi^2 = \frac{(101 - 98,99353)^2}{98,99353} + \frac{(37 - 34,8333)^2}{34,8333} + \frac{(9 - 11,30273)^2}{11,30273} + \frac{(120 - 121,87)^2}{121,87} = 0,67329 \quad (16)$$

Utilizando $\alpha = 0,05$ obtém-se $q_c = 3,841$ (Magalhães e Lima, 2010, p. 373). Portanto, $RC = \{\chi^2 : \chi^2 > 3,841\}$. Como $\chi^2 \notin RC$, segue que a população de Engenheiro Coelho - SP está em Equilíbrio de Hardy-Weinberg para tipagem sanguínea.

O resultado acima comprova a veracidade das premissas relatadas no Teorema 2.1. Como as três primeiras premissas são notoriamente verdadeiras nessa população e também não há presença de seleção natural, visto que em via de regra nenhum indivíduo escolhe seu parceiro sexual ao observar o tipo sanguíneo, a comprovação do Equilíbrio de Hardy-Weinberg tem como principal relevância a percepção de que a população não possui mutação genética para os genes de tipos sanguíneos.

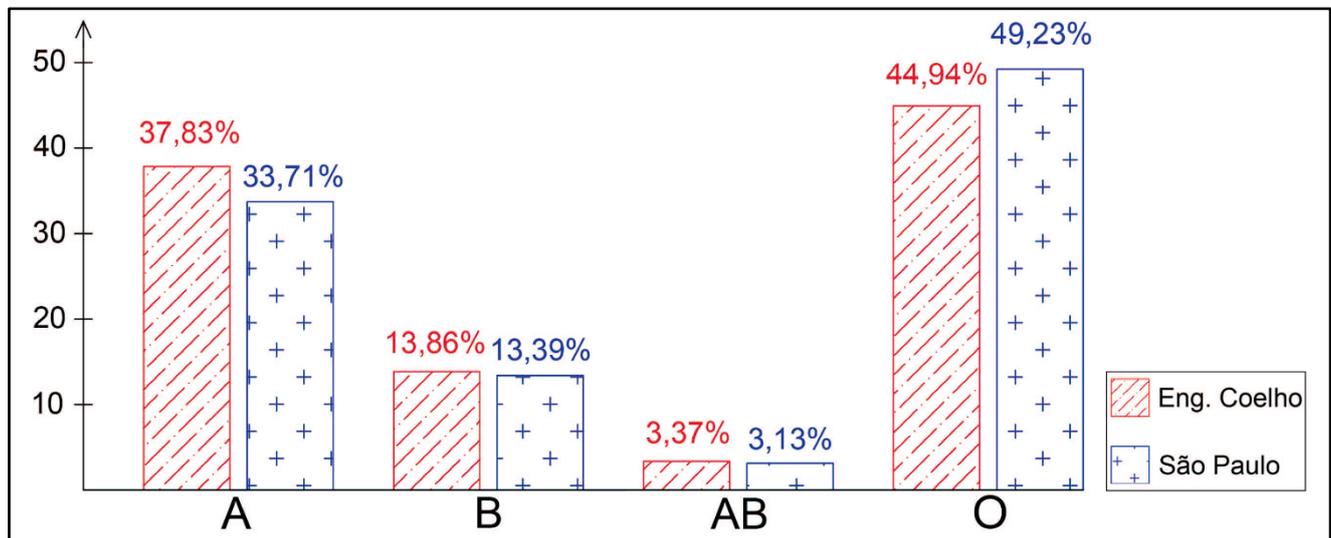
Outra análise que pode ser feita é a relação entre as frequências dos tipos sanguíneos dos 2.462 doadores de sangue da Fundação Pró-Sangue/Hemocentro de São Paulo. Segundo Batissoco e Novaretti (2003), no ano de 2003, 33,71% da amostra era tipo A, 13,39% era tipo B, 3,13% era tipo AB e 49,23% era tipo O. Com esses resultados a Figura 2 pode ser obtida, sendo fácil notar que as frequências entre a capital paulista e a cidade de Engenheiro Coelho se mantêm próximas.

6 CONCLUSÃO

Por meio das inúmeras aplicações da genética de populações, é possível concluir que há importantes relações interdisciplinares entre as áreas da Matemática e da Biologia. Com simples noções de Genética e de Estatística, este

artigo trouxe uma aplicação prática do Equilíbrio de Hardy-Weinberg, equilíbrio tão importante para a verificação de premissas como a mutação genética.

Figura 2 – Comparação das Frequências Fenotípicas



O estudo de caso na população de Engenheiro Coelho - SP mostrou que essa população encontra-se no Equilíbrio de Hardy-Weinberg para tipagem sanguínea, prevalecendo as premissas de acasalamento aleatório, população infinita, gerações não-sobrepostas, e população sem mutação, seleção e migração. O nível de significância $\alpha = 0,05$ mostrou a baixíssima probabilidade de erro no teste estatístico, validando ainda mais a veracidade da pesquisa.

Cabe também destacar que ao observar o desvio $D = 0,00621$ pôde-se notar que, mesmo sem corrigir as estimativas preliminares, os dados já mostravam o quão próximo $(p' + q' + r')^2$ estava de 1, ou seja, esse resultado preliminarmente expôs a proximidade que essa população encontrava-se do equilíbrio de Hardy-Weinberg para tipagem sanguínea.

Ademais, ao observar os resultados da aplicação da Equação (17), notou-se também que embora a diferença entre as frequências observadas e esperadas não tenha sido significativa, a maior discrepância ocorreu no fenótipo AB com 0,46914 e a menor decorreu do fenótipo O com 0,02871.

Por fim, constata-se que o artigo trouxe resultados importantes, não apenas para o meio acadêmico, com possibilidades de abrangência para Matemática e Biologia, mas também para a cidade de Engenheiro Coelho - SP com a observância do equilíbrio gênico para tipos sanguíneos.

REFERÊNCIAS

Amabis, J. M., Martho, G. R. (1994). *Biologia das populações: genética, evolução e ecologia*, 1. ed. **Editora Moderna**, São Paulo.

Batissoco, A. C., Novaretti, M. C. (2003). Aspectos moleculares do sistema sanguíneo ABO. **Revista Brasileira de Hematologia e Hemoterapia**, v. 25(n. 1), p.47-58.

Beiguelman, B. (2008). **Genética de populações humanas**. SBG, Ribeirão Preto.

Bernstein, F. (1924). Ergebnisse einer biostatistischen zusammenfassenden betrachtung über die erblichen blutstrukturen des menschen. **Klin Wschr**, v. 3, p.1495-1497.

Freire-Maia, N. (1974). **Genética de populações humanas**. HUCITEC, Editora da Universidade de São Paulo, São Paulo.

Griffiths, A. J. F., Wessler, S. R., Lewontin, R. C., Gelbart, W. M., Suzuki, D. T., Miller, J. H. (2006). **Introdução à Genética**. 8. ed. Guanabara Koogan S.A., Rio de Janeiro.

Grimmett, G., Stirzaker, D. (2001). **Probability and Random Processes**. Texts from Oxford University Press, OUP Oxford, New York.

Hartl, D. L., Clark, A. G. (2007). **Principles of population genetics**, 4. ed. Sinauer Associates, Sunderland.

IBGE (2020). **Diretoria de pesquisas, coordenação de população e indicadores sociais: estimativas da população residente com data de referência 1 de julho de 2019**.

Lewis, R. (2004). **Genética humana: conceitos e aplicações**, 5. ed. Guanabara Koogan S.A., Rio de Janeiro.

Magalhães, M. N., Lima, A. C. P. (2010). **Noções de probabilidade e estatística**, 7. ed. Editora da Universidade de São Paulo, São Paulo.

Oliveira, V. F. (2019). **Modelagem matemática em genética de populações**. Dissertação de Mestrado, Instituto de Geociências e Ciências Exatas, UNESP, Rio Claro, p.89.

Contribuições de autoria

1 - Vinícius Freitas de Oliveira

Possui mestrado em matemática e, atualmente, é professor do Centro Universitário Adventista de São Paulo.

<https://orcid.org/0000-0002-7256-9349> - viniciusfreitasoliveira@hotmail.com

Contribuição: conceitualização, análise formal, investigação, metodologia, escrita, revisão e edição.

2 - Guilherme Augusto Pianezzer

Possui doutorado em Métodos Numéricos em Engenharia e, atualmente, é professor do Centro Universitário Internacional e da Faculdade da Lapa.

<https://orcid.org/0000-0002-8366-2063> - guilherme.pi@uninter.com

Contribuição: conceitualização, análise formal, investigação e revisão.

3 - Suzete Maria Silva Afonso

Possui doutorado e pós-doutorado em Matemática e, atualmente, é Professora Associada da Universidade Estadual Paulista "Júlio de Mesquita Filho".

<https://orcid.org/0000-0003-3070-5856> - s.afonso@unesp.br

Contribuição: conceitualização, análise formal, investigação e revisão.

Como citar este artigo

OLIVEIRA, V. F.; PIANEZZER, G. A.; AFONSO, S. M. O equilíbrio de Hardy-Weinberg no sistema sanguíneo ABO: um estudo de caso em Engenheiro Coelho – SP. *Ciência e Natura*, Santa Maria, v. 43, e16, p. 1-20, 2021. Disponível em:

<https://doi.org/10.5902/2179460X43422>. Acesso em: dia, mês (abreviado), ano.