

PRÁTICA 1

Simulação de Deriva Genética

Objetivo

O objetivo desta aula prática é analisar o efeito da deriva genética sobre a mudança de frequências alélicas. Investigaremos também o processo que leva um alelo a se fixar, isto é, atingir frequência de 100% em uma população. Nossas simulações irão explorar de que modo as diferenças nos tamanhos populacionais e nas frequências alélicas iniciais afetam o processo de deriva genética.

O modelo da simulação

Para compreender a deriva genética, serão realizadas simulações “manuais” que, do ponto de vista lógico, são muito semelhantes às simulações feitas por computadores. Como a deriva genética é um processo aleatório (ver **Quadro 6.1 do Ridley**), faremos simulações envolvendo sorteio de bolinhas para simular a deriva genética com diferentes frequências alélicas. Primeiramente, precisaremos definir alguns parâmetros e pressupostos populacionais. Nos exercícios de hoje, teremos como parâmetros o **tamanho da população (N)** e a **frequência inicial dos alelos (p_0)**; onde o zero ao lado do p indica que trata-se da frequência alélica na geração inicial) e assumiremos os seguintes pressupostos sobre a nossa população:

- Não há migração
- Não há mutação
- Não há sobreposição de gerações
- Os indivíduos são haplóides
- Existe um número infinitamente grande de gametas

As simulações com bolinhas

Para a simulação com bolinhas, devemos considerar que cada indivíduo produz muitos gametas e que a população tem tamanho constante. O genótipo dos indivíduos na população será determinado ao acaso a partir das frequências alélicas na geração anterior. Simularemos **três cenários diferentes**, variando um dos parâmetros iniciais em cada caso e iremos estabelecer uma população inicial de acordo com o tamanho populacional (quantidade de bolinhas dentro do copo) e as frequências alélicas (frequência de cores das bolinhas dentro do copo) iniciais propostas.

Para iniciar a simulação, sortearmos uma bolinha desse conjunto inicial. Esse será o genótipo do primeiro indivíduo da próxima geração. Antes de realizar outro sorteio, **devolveremos a bolinha ao copo** e, em seguida, repetiremos o sorteio para determinar o genótipo do segundo indivíduo. Esse processo será repetido até que a nova geração atinja o mesmo número de indivíduos da geração original.

Sorteio com reposição

Um dos nossos pressupostos é que a população produz um número infinitamente grande de gametas. Em nosso sorteio, estamos determinando o genótipo dos gametas que irão contribuir para a geração seguinte. Para simular um cenário no qual a quantidade de

gametas é infinita, é preciso que a retirada de um gameta não altere as frequências originais. Por isso, devemos fazer um **sorteio com reposição**.

Para os sorteios das próximas gerações, vamos colocar no copo o número de bolinhas de cada cor que corresponde à frequência de pretas e brancas da geração atual. Por exemplo, se na etapa anterior a bolinha preta foi sorteada 3 vezes e a branca apenas 1, a frequência de bolinhas pretas atual é 3/4. Então, colocaremos no copo 3 bolinhas pretas e 1 branca. O processo será repetido por 12 gerações ou até que todos os indivíduos tenham a mesma cor. Como estamos simulando populações completamente isoladas e sem mutação, uma vez que um alelo é perdido ele não é mais recuperado. De maneira arbitrária, definimos que iremos simular no máximo 12 gerações. Assim, se não ocorrer a fixação de nenhum dos alelos até a geração 12, paramos a simulação.

Cenário 1: $N = 4$, $p_0 = \frac{1}{2}$

Para realizar a simulação, siga os passos abaixo:

1. Coloque 2 bolinhas pretas e 2 bolinhas brancas no copo.
2. Mexa bem e sorteie uma bolinha. Anote sua cor na Tabela 1 (impressa): "P" para preta, ou "B" para branco.
3. Devolva a bolinha ao copo.
4. Repita os passos 2 e 3 mais três vezes, até completar os 4 indivíduos da próxima geração. Anote na mesma linha o número total de bolinhas pretas.

5. Se o número de bolinhas pretas for 0 ou 4, ou se você tiver atingido a geração 12, vá para o Cenário 2. Caso contrário, prossiga para o próximo passo.

6. Esvazie o copinho.

7. Coloque no copinho o número de bolinhas pretas igual ao número total de bolinhas pretas que foi sorteado. Complete com bolinhas brancas até totalizar 4 bolinhas.

8. Volte para o passo 2, e simule mais uma geração.

Cenário 2. $N = 4$, $p_0 = \frac{1}{4}$

Simule uma população com 4 indivíduos ($N = 4$) e frequência alélica inicial de pretas = $\frac{1}{4}$. Para tanto, repita os passos do cenário 1. Anote os resultados na Tabela 2 (impressa).

Cenário 3. $N = 8$, $p_0 = \frac{1}{2}$

Simule uma população com 8 indivíduos ($N = 8$) e frequência alélica inicial de pretas = $\frac{1}{2}$. Anote os resultados na Tabela 3 (impressa).

Cenário 4. $N = 16$, $p_0 = \frac{1}{4}$

Simule uma população com 16 indivíduos ($N = 16$) e frequência alélica inicial de pretas = $\frac{1}{4}$. Anote os resultados na Tabela 4 (impressa).

Compilação dos resultados

Para analisar as simulações, iremos juntar os dados de toda turma.

1. Acesse a planilha respectiva à sua turma.

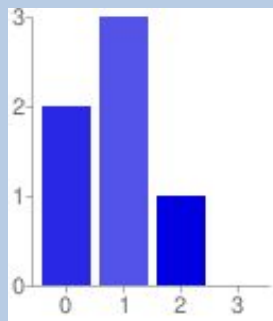
- [Planilha DIURNO](#)
- [Planilha NOTURNO](#)

2. Informe, na planilha, os resultados dos sorteios. Para os casos em que o alelo se perdeu ou fixou antes da geração 12, repita o resultado final em todas as gerações posteriores.

3. Observe alguns histogramas dos resultados sendo formados na parte inferior da planilha. Estes histogramas se atualizam à medida que os grupos inserem os seus dados na planilha, de modo que o resultado final será observado quando todos os grupos terminarem essa parte do trabalho.

Histograma

Histograma é um gráfico que representa a distribuição de frequências de um conjunto de valores. Normalmente, é um gráfico de barras no qual cada barra representa o número ou proporção de ocorrências dentro de um intervalo de valores. Exemplo: no histograma abaixo, as barras representam a frequência em que cada um dos valores inteiros (0, 1, 2 ou 3) ocorreu.



Análise dos resultados

ATENÇÃO!

Ao final desta aula prática, vocês deverão responder alguns exercícios que serão entregues na próxima semana. Anote no caderno as respostas das perguntas das etapas a seguir e use-as como guia para os exercícios finais. **As respostas destas perguntas não precisam ser entregues.**

Etapa I. Histogramas das frequências alélicas a cada geração

Como a deriva genética é um processo aleatório, não é possível prever a direção da mudança das frequências alélicas provocadas por deriva em **uma** população. No entanto, o comportamento médio de um grande número de populações apresenta padrões bem definidos. Assim, para estudar o padrão geral do efeito da deriva genética sobre as frequências alélicas, é preciso considerar os dados de todos os grupos (populações) da sala. Analise os histogramas junto com o resultado final que se encontra na parte inferior da planilha e responda às seguintes questões.

Cenário 1 (observe os gráficos da geração 1 a 12):

- 1.1. Qual a proporção de populações com todos os indivíduos pretos na geração 12? E com todos os indivíduos brancos?
- 1.2. Na geração 0 todas as populações eram idênticas. O que aconteceu com a variação **entre** populações com o passar das gerações?

1.3. Como é a variação **dentro** de cada população isolada na geração 12?

Cenário 2 (observe os gráficos da geração 1 a 12):

2.1. Qual a proporção de populações com todos os indivíduos pretos no gráfico da geração 12?

E com todos os indivíduos brancos?

2.2. Relacione as respostas das questões 1.1 e 2.1 com as condições iniciais desses dois cenários.

Cenário 3 (observe os gráficos da geração 1 a 12):

3.1. Qual a proporção de populações com todos os indivíduos pretos na geração 12? E com todos os indivíduos brancos? Como essas proporções podem ser comparadas aos resultados dos cenários 1 e 2?

3.2. O que aconteceu com a variação **dentro** e **entre** as populações com o passar das gerações? Como esses valores se comparam aos do cenário 1?

Cenário 4 (observe os gráficos da geração 1 a 12):

4.1. Qual a proporção de populações com todos os indivíduos pretos na geração 12? E com todos os indivíduos brancos? Como essas proporções podem ser comparadas aos resultados dos cenários 1, 2 e 3?

4.2. O que aconteceu com a variação **dentro** e **entre** as populações com o passar das gerações? Como esses valores se comparam aos do cenário 2?

Etapa II. Previsões

Para avaliar quantitativamente os resultados das simulações, iremos estimar três estatísticas para cada geração.

- média das frequências alélicas
- variância das frequências alélicas
- taxa de heterozigose média

Dada a sua análise anterior dos histogramas, responda:

1. O que você espera que aconteça com a média da frequência de alelos pretas na **população** ao longo do tempo?
2. O que você espera que aconteça com a variância da frequência de alelos pretas ao longo do tempo?

Variância

Variância (σ^2) é uma medida de dispersão de um conjunto de valores que indica o quão distante eles estão da média.

$$\sigma^2 = \frac{(x_1 - \text{media})^2 + (x_2 - \text{media})^2 + \dots + (x_N - \text{media})^2}{N}$$

N = número de observações. Nesse caso, cada população simulada é uma observação.

3. O que você espera que aconteça com a taxa de heterozigose (veja o quadro abaixo) ao longo do tempo? Como isso difere entre os três cenários simulados?

Taxa de heterozigose em populações haplóides?

Como a própria palavra sugere, a heterozigose está ligada à chance de ser formado um indivíduo heterozigoto na população. Como vimos em aula, ela é calculada por:

$$H = 1 - \sum_{i=1}^k p_i^2$$

Note, no entanto, que o seu cálculo depende apenas das frequências alélicas (p_i), o que permite aplicá-lo a populações não-diplóides. Nesse caso, a taxa de heterozigose nos informa a chance de que dois alelos tomados ao acaso sejam diferentes. É uma medida da variação genética por loco em uma população.

Para um gene bialélico, podemos também calcular facilmente a taxa de heterozigose por:

$$H = 2 p_1 p_2$$

ou, mais comumente:

$$H = 2 p q$$

Note que essas são apenas duas notações diferentes para a mesma equação. Veja mais sobre esse tema no **Quadro 6.3 do Ridley**.

Etapa III. Média e variância das frequências alélicas e taxa de heterozigose

Procedimento para análise dos dados:

1. Acesse planilha de resultados respectiva à sua turma.
2. Clique em “File” → “Download as” → “Microsoft Excel (.xlsx)”, ou “Arquivo” → “Fazer Download como” → “Microsoft Excel (.xlsx)”

3. Abra o arquivo que foi salvo no computador. Pode ser que ele abra em “Modo de Exibição”. Se for o caso, clique em “Habilitar Edição” antes de prosseguir.

Entendendo a planilha de dados

Em cada linha estão apresentados os dados das populações simuladas por cada grupo (identificados na coluna B). O conjunto de dados da sala para cada geração se agrupa nas colunas consecutivas (denominados F0 - F12).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1		Grupo/População	F0	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
2	CENARIO 1	1	2	2	1	1	1	0	0	0	0	0	0	0	0
3	N=4	2	2	4	4	4	4	4	4	4	4	4	4	4	4
4	p=1/2	3	2	0	0	0	0	0	0	0	0	0	0	0	0
5		4	2	2	2	2	2	1	1	2	4	4	4	4	4
6		5	2	3	3	3	3	3	3	2	2	2	1	2	1
7		6	2	2	2	2	1	1	1	2	3	2	3	2	2
8		7	2	2	3	4	4	4	4	4	4	4	4	4	4
9		8	2	1	1	1	0	0	0	0	0	0	0	0	0
10		9	2	2	2	1	1	2	2	2	4	4	4	4	4
11		10	2	1	1	0	0	0	0	0	0	0	0	0	0
12		11	2	4	4	4	4	4	4	4	4	4	4	4	4
13		12	2	1	1	0	0	0	0	0	0	0	0	0	0

Por exemplo, na coluna D da figura acima, estão os dados enviados por todos os grupos referentes à Geração 1 (F1) do primeiro cenário ($N = 4$, $p = 1/2$). Note que todos os resultados de cada cenário estão em planilhas diferentes do mesmo arquivo. Os resultados dos outros cenários estarão nas outras planilhas e você pode mudar de planilha clicando na aba correspondente no canto inferior esquerdo (veja a parte inferior da figura apresentada acima). Fique atento para não misturar os diferentes cenários na hora de interpretar as análises.

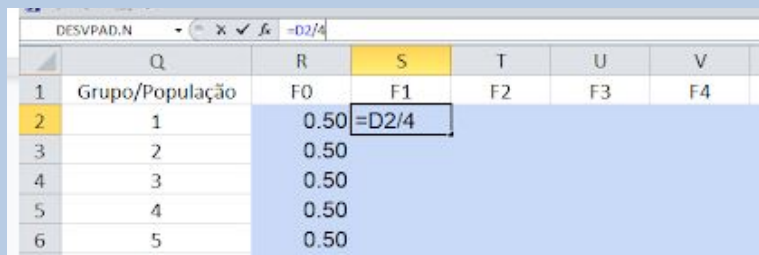
4. Calcule a média das frequências alélicas para cada geração.

Calculando as frequências alélicas

Primeiro devemos calcular as frequências alélicas em cada geração. Os dados na tabela correspondem a valores absolutos de números de indivíduos/bolinhas pretas. Logo, a frequência de bolinhas pretas poderá ser calculada da seguinte forma:

$$p = \frac{\text{contagem de pretas}}{N}$$

Para auxiliar nos cálculos, sugerimos que você crie uma tabela ao lado para transformar os dados absolutos para valores de frequência. Anote na célula P1 que os dados a seguir serão os valores de p. Copie as células referentes aos números dos grupos na coluna B para a coluna Q. Para seu auxílio, copie também as referências das gerações que estão na Linha 1, de C a N, e cole essas informações na mesma Linha 1, agora de R a AC. Veja a figura abaixo para maiores detalhes. Desta forma, podemos começar a transformar os dados para as frequências em cada geração, medidos por cada grupo.



	Q	R	S	T	U	V
1	Grupo/População	F0	F1	F2	F3	F4
2	1	0.50	=D2/4			
3	2	0.50				
4	3	0.50				
5	4	0.50				
6	5	0.50				

Para fazer isso, use o sinal de igual (=) e faça a divisão da observação do número de pretas feita pelo grupo 1 na geração 1 pelo total de indivíduos da população deles (N=4). Use a referência da célula na matriz de observações ao invés de escrever o número observado.

Para o grupo 1 na geração 1, a referência do número de pretas observados está na coluna D2.

Após inserir a fórmula da frequência referente ao grupo 1 na geração 1, é possível copiar esta fórmula para as linhas abaixo arrastando para baixo o pequeno quadrado/cruz que aparece no canto inferior direito da célula (veja figura abaixo).

	Q	R	S	T	U
1	Grupo/População	F0	F1	F2	F3
2	1	0.50	0.50		
3	2	0.50	1.00		
4	3	0.50	0.00		
5	4	0.50	0.50		
6	5	0.50	0.75		
7	6	0.50	0.50		
8	7	0.50	0.50		
9	8	0.50	0.25		
10	9	0.50	0.50		
11	10	0.50	0.25		
12	11	0.50	1.00		
13	12	0.50			
14	13	0.50			
15	14	0.50			
16	15	0.50			
17	16	0.50			

Não esqueça de fazer os ajustes necessários para no caso dos cenários 3 e 4! A mesma lógica se aplica à geração inicial (F0).

5. Calcule a média das frequências alélicas para cada geração.

Calculando a média das frequências alélicas

Para calcular uma média no Excel podemos usar a função que já faz parte do programa:

=MÉDIA(...)

onde "... " é o conjunto de células com as valores a serem utilizadas (cada coluna)

	Q	F0	F1	T
1	Grupo/População			
16	15	0.50	0.25	
17	16	0.50	0.50	
18	17	0.50	0.50	
19	18	0.50	0.75	
20	19	0.50	0.25	
21	20	0.50	0.25	
22	21	0.50	0.75	
23	22	0.50	0.50	
24	23	0.50	0.25	
25	24	0.50	1.00	
26	25	0.50	0.25	
27	26	0.50	1.00	
28	27	0.50	1.00	
29	Médias	0.50	=MÉDIA[S2:S28]	
30				

Todos os valores da coluna de frequência alélicas devem ser selecionados

. Sugerimos que a célula abaixo dos valores de frequência de cada geração seja usada para o cálculo das médias.

6. Calcule a variância das frequências alélicas para cada geração.

Calculando a variância

Calcularemos a variância de modo análogo. A função correspondente no Excel é a seguinte:

=VAR(...) ou =VAR.A(...)

onde "... " é o conjunto de células com os valores a serem utilizados (cada coluna). Sugerimos que você faça este cálculo na linha abaixo à linha da média. MAS CUIDADO, todos os valores da coluna de frequência alélicas devem ser selecionados, exceto a média.

	Q	R	S	T
1	Grupo/População	F0	F1	F2
16	15	0.50	0.25	
17	16	0.50	0.50	
18	17	0.50	0.50	
19	18	0.50	0.75	
20	19	0.50	0.25	
21	20	0.50	0.25	
22	21	0.50	0.75	
23	22	0.50	0.50	
24	23	0.50	0.25	
25	24	0.50	1.00	
26	25	0.50	0.25	
27	26	0.50	1.00	
28	27	0.50	1.00	
29	Médias	0.50	0.54	
30	Variância	0.00	=VAR.A(S2:S28)	
31				

7. Calcule a média da taxa de heterozigose para cada geração.

Calculando a média da taxa de heterozigose

Para calcular a média da taxa de heterozigose iremos fazer uma tabela auxiliar assim como fizemos para as frequências. Anote na célula AE1 que os dados a seguir serão os valores de H. Copie as células referentes aos números dos grupos na coluna B para a coluna AF. Para seu auxílio, copie também as referências das gerações que estão na Linha 1, de C a N, e cole na mesma Linha 1, agora de AG a AR. Desta forma, podemos começar a calcular os valores de taxa de heterozigose medidos por cada grupo em cada geração.

Primeiro devemos calcular a heterozigose em cada simulação (linha). Como o gene é

bialélico nesse caso, lembre que: $q = 1 - p$.

Logo,

$$H_e = 2pq = 2p(1-p)$$

Por exemplo, para o primeiro resultado de frequência da geração 1 do grupo 1 que está na célula S2, teremos:

$$= 2 * S2 * (1 - S2)$$

Você poderá arrastar esta fórmula para o cálculo da taxa de heterozigose dos outros grupos e das outras gerações. Com a tabela de taxas de heterozigose pronta, podemos calcular a média da taxa de heterozigose em cada geração de modo semelhante ao que fizemos para as frequências.

Relatório das análises de deriva genética

ATENÇÃO!

As perguntas a seguir **devem ser entregues no início da próxima aula**. Utilize suas anotações das etapas I, II e III como apoio.

1. Faça gráficos de cada parâmetro pelo tempo para cada um dos cenários simulados. Ao final, você deverá obter 12 combinações de parâmetros com cenários (1, 2, 3 e 4), que podem ser apresentados em 12 gráficos diferentes ou em três gráficos, um para cada parâmetro. Os parâmetros são:

- média das frequências alélicas pelo tempo (gerações)
- variância das frequências alélicas pelo tempo (gerações)
- média da taxa de heterozigose pelo tempo (gerações)

Dica: Os dados que serão usados em cada gráfico não estão em células adjacentes na planilha. Por exemplo, o gráfico da média da frequência alélica irá incluir a média das frequências na geração 1, na geração 2, na geração 3 e assim por diante. É possível selecionar células não adjacentes no Excel mantendo a tecla Ctrl pressionada.

2. O que aconteceu com a variação genética em cada população com o passar do tempo? E no conjunto total de populações?

3. Por que não usamos a frequência alélica média em cada geração para calcular a taxa de heterozigose?

Bibliografia

- Ridley, M. *Evolução*. 3ª ed. Artmed, 2006.
- Freeman, S. & J. Herron. *Análise Evolutiva*. 4ª ed. Artmed, 2009.