

## **Prática 1 - Simulação de Deriva Genética**

**LISTA 1 - para ser entregue dia 22/09/2014 (Diurno) até 14:05 e 23/09/2014 (Noturno) até 19:05**

### **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”. Como veremos posteriormente, essas simulações são muito semelhantes, do ponto de vista lógico, às feitas por computadores. A deriva genética é um processo aleatório (ver Quadro 6.1 do Ridley). Para simular a deriva genética com diferentes frequências alélicas, faremos simulações envolvendo sorteio de feijões.

Primeiramente, precisaremos definir alguns parâmetros populacionais. Nos exercícios de hoje esses parâmetros são 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).

Também assumiremos alguns pressupostos sobre a nossa população:

- Não há migração
- Não há mutação
- Não há sobreposição de gerações
- Indivíduos haplóides
- Número infinitamente grande de gametas

### **As simulações com feijões**

Considere que cada indivíduo produz muitos gametas e que a população tem tamanho constante. O genótipo desses indivíduos será determinado ao acaso a partir das frequências alélicas na geração anterior. Simularemos três cenários diferentes, variando um dos parâmetros iniciais em cada caso. Iremos estabelecer uma população inicial de acordo com o tamanho populacional e as frequências alélicas iniciais propostas. Sortearemos um feijão desse conjunto inicial. Esse será o genótipo do primeiro indivíduo da próxima geração. Antes de realizar outro sorteio, devemos devolver o feijão ao copo. Repetiremos então 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 da próxima geração vamos colocar no copo o número de feijões de cada cor que corresponda à frequência de pretos da geração atual. Por exemplo, se na etapa anterior o preto foi sorteado 3 vezes e o marrom apenas 1, a frequência atual é 3/4. Então, colocaremos no copo 3 feijões pretos e 1 marrom. 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. $N = 4$ , $p_0 = 1/2$**

**Para realizar a simulação, siga o algoritmo abaixo:**

1. Coloque 2 feijões pretos e 2 feijões marrons no copo.
2. Mexa bem e sorteie um feijão. Anote sua cor na Tabela 1: A para pretos, ou V para marrom.
3. Devolva o feijão 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 feijões pretos.
5. Se o número de feijões pretos for 0 ou 4, ou se você tiver atingido a geração 12, vá para a Cenário 2. Caso contrário, prossiga para o próximo passo.
6. Esvazie o copinho.
7. Coloque no copinho o número de feijões pretos igual ao número total de feijões pretos que foi sorteado. Complete com feijões marrons até totalizar 4 feijões.
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 pretos =  $\frac{1}{4}$ . Para tanto, repita o algoritmo do cenário 1. Anote os resultados na Tabela 2.

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

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

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

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

**Compilação dos resultados**

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

1. Abra a planilha do link:

**DIURNO:** <https://docs.google.com/spreadsheets/d/1jZbxJzstOdgaNADw6n1ov4pKo0l0dAD1uEcefAlFpII/edit#gid=0>

**NOTURNO:** [https://docs.google.com/spreadsheets/d/1Ohp7eJ9RQnDUsmamCgX\\_hvuPLSIQLAGrktFVR-p8Vyl/edit#gid=0](https://docs.google.com/spreadsheets/d/1Ohp7eJ9RQnDUsmamCgX_hvuPLSIQLAGrktFVR-p8Vyl/edit#gid=0)

e informe 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.

2. Observe alguns histogramas dos resultados. Você pode rodar o resultados no R a qualquer momento ou acompanhar pela projeção feita em sala. Para rodar a rotina em R, abra o arquivo pratica2.R no R (você pode baixar este programa da página da disciplina: <http://biologia.ib.usp.br/bio208/bio208/static/pdfs/pratica1/2014-pratica1-hist.R>). Para rodar no R, você deve iluminar as linhas da rotina e apertar "Control+r".

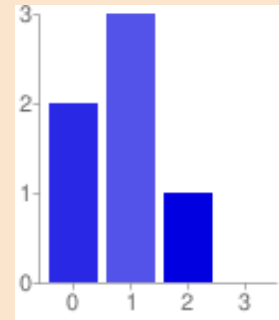
As linhas que começam com "PlotHistograms" no programa são as linhas que desenham os histogramas. Rode uma por vez para poder ver os resultados de cada cenário. Se você quiser saber mais sobre a rotina de programação, pergunte aos monitores.

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:**

0, 0, 1, 1, 1, 2

Nesse caso temos um histograma de valores inteiros e cada barra representa quantas vezes o valor ocorreu.



## Análise dos resultados

### Parte 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. Por isso, reuniremos os resultados de todos os grupos para analisar o padrão geral do efeito da deriva genética sobre as frequências alélicas.

Analise os gráficos 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 marrons?

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 marrons?

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 marrons? 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 marrons? 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?

**Parte II. Média e variância das frequências alélicas e taxa de heterozigose**

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

Veja nos quadros abaixo como são calculados a variância e a taxa de heterozigose.

**Variância**

Variância ( $\sigma^2$ ) 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.

### 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 heterizigose por:

$$H = 2p_1p_2$$

ou, mais comumente:

$$H = 2pq$$

Note que essas são apenas duas notações diferentes para a mesma equação.

Veja mais sobre esse tema no Quadro 6.3 do Ridley.

### Procedimento para análise dos dados:

1. Acesse planilha de resultados através do link:

**DIURNO:** <https://docs.google.com/spreadsheets/d/1jZbxJzstOdgaNADw6n1ov4pKo0l0dAD1uEcefAIFpII/edit#gid=0>

**NOTURNO:** [https://docs.google.com/spreadsheets/d/1Ohp7eJ9RQnDUsmamCgX\\_hvuPLSIQLAGrktFVR-p8Vyl/edit#gid=0](https://docs.google.com/spreadsheets/d/1Ohp7eJ9RQnDUsmamCgX_hvuPLSIQLAGrktFVR-p8Vyl/edit#gid=0)

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.

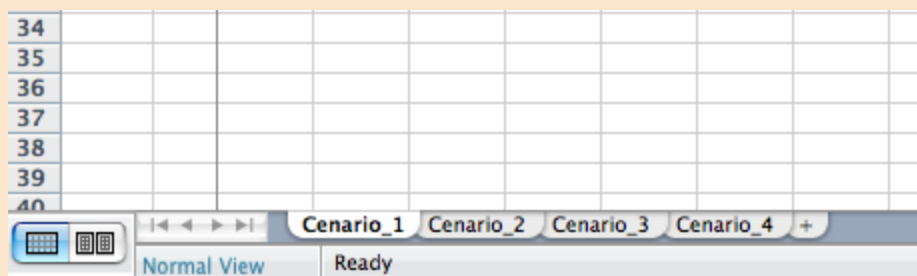
## Entendendo a planilha de dados

Em cada linha estão apresentados os dados de cada grupo (identificados na coluna B). O conjunto de dados da sala para cada geração se agrupa em cada coluna.

A	B	C	D	E	F
Cenário 1	Grupo	geração 0	geracao 1	geracao 2	geracao 3
N=4	1	2	4	4	4
p=1/2	2	2	3	3	3
	3	2	3	4	4
	4	2	2	2	3
	5	2	2	4	4

Por exemplo, na coluna D da figura ao lado, estão os dados enviados por todos os grupos referentes a Geração 1 do primeiro cenário (N = 4, p = 1/2).

Note que todos os cenários 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 figura abaixo). Fique atento para não misturar os diferentes cenários na hora de interpretar a 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 pretos. Logo, a frequência de pretos poderá ser calculada da seguinte forma:

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

Para auxiliar nos cálculos, sugerimos que você crie uma tabela ao lado para transformar os dados absolutos para valores de frequência. Anote na célula Q1 que os dados a seguir serão os valores de p. Copie as células referentes aos números dos grupos na coluna B para a coluna R. Para seu auxílio, copie também as referências das gerações que estão na Linha 1, de C a M, e cole na mesma Linha 1, agora de R a AE. 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	W	X
p	Grupo	geração 0	geracao 1	geracao 2	geracao 3	geracao 4	geracao 5
	1						
	2						
	3						
	4						
	5						
	6						
	7						
	8						

Para fazer isso, use o sinal de igual (=) e faça a divisão da observação do número de pretos feita pelo grupo 1 na geração 0 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 0, a referência do número de pretos observados está em C2.

Q	R	S	T	U	V
p	Grupo	geração 0	geracao 1	geracao 2	geracao 3
	1	=C2/4			
	2				
	3				
	4				
	5				

Q	R	S	T	U
p	Grupo	geração 0	geracao 1	geracao 2
	1	0,5		
	2	0,5		
	3	0,5		
	4	0,5		
	5	0,5		
	6			

Após inserir a fórmula da frequência referente ao grupo 1 na geração 0, é possível copiar esta fórmula para as linhas abaixo arrastando para baixo o pequeno quadrado/cruz que aparece no canto inferior direito da célula.

Não esqueça de fazer os ajustes necessários para os caso dos cenários 3 e 4!

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)

Todos os valores da coluna de frequência alélicas devem ser selecionados (o exemplo ao lado não contém todos os grupos)

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

Q	R	S	T
p	Grupo	geração 0	geracao
	1	0,5	
	2	0,5	
	3	0,5	
	4	0,5	
	5	0,5	
	6	0,5	
	7	0,5	
	8	0,5	
		=MÉDIA(S2:S9)	

## 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(...)**

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

Sugerimos que você faça este cálculo na linha baixo à linha da média. MAS CUIDADO, todos os valores da coluna de frequência alélicas devem ser selecionados, exceto a média.

## 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 AG1 que os dados a seguir serão os valores de taxa de heterozigose (sugerimos que você use a letra H). Copie as células referentes aos números dos grupos na coluna B para a coluna AH. 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 AI a AU. Desta forma

podemos começar a calcular os valores de taxa de heterozigoze 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 R2, teremos:

$$= 2 * R2 * (1 - R2)$$

Da mesma forma, você poderá arrastar essa 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.

### Responda às seguintes questões:

1. Faça gráficos de cada estatística pelo tempo para cada cenário. Serão doze gráficos no total, sendo 3 para cada cenário:

- 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 (coluna D), na geração 2 (coluna G), na geração 3 e assim por diante. É possível selecionar células não adjacentes no Excel mantendo a tecla Ctrl pressionada. (SE FICAR COMO ESTÁ, ESSA DICA NÃO SERÁ NECESSÁRIA).

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, Mark, ed. *Evolução*. 3ª ed. Artmed, 2006.

Freeman, Scott e Herron, Jon. Análise Evolutiva 4<sup>a</sup> ed. Artmed, 2009.