

# Exercício de parentesco GHEP 2026: Nível avançado

Desenvolvido por: Magnus Dehli Vigeland

## Introdução

Este exercício é de escolha múltipla com 20 perguntas. Para cada pergunta, existe apenas uma alternativa correta. Pode utilizar qualquer software que desejar, mas lembre-se de que alguns programas possuem convenções integradas (por exemplo, arredondamento) que podem afetar o resultado obtido. Se a sua resposta não coincide totalmente com nenhuma das alternativas, deverá escolher a opção que mais se aproxime.

O tema do desafio deste ano é endogamia (*inbreeding*): estimar coeficientes de endogamia a partir de dados genotípicos e o efeito da endogamia nas LR's em casos de parentesco com incesto.

## Pressupostos ao longo de todo o exercício

- Ausência de desvios ao equilíbrio de Hardy-Weinberg
- Todos os marcadores são independentes (*unlinked*) e não se encontram em desequilíbrio de ligamento (*linkage disequilibrium*)
- Ausência de *drop-outs*, *drop-ins*, alelos silenciosos ou mutações

## Arquivos necessários

- *db35.txt*: Base de dados de frequências para 35 STRs
- *mother\_child.txt*: Genótipos da mãe e da criança
- *father.txt*: Genótipos do pai

## Algumas definições

Alelos homólogos são *idênticos por descendência* (IBD) se tiverem a mesma origem dentro de um determinado pedigree. O *coeficiente de endogamia*  $f$  de um membro do pedigree é a probabilidade de, para um *locus* aleatório, os dois alelos serem IBD. O coeficiente de endogamia de uma criança é igual ao coeficiente de parentesco  $\varphi$  entre os pais:  $f_{\text{child}} = \varphi_{\text{parents}}$ .

Os coeficientes  $\kappa$  ( $\kappa_0, \kappa_1, \kappa_2$ ) entre dois indivíduos não consanguíneos correspondem às probabilidades de estes partilharem 0, 1 e 2 alelos IBD, respectivamente, num *locus* autossómico aleatório. Por exemplo, para um par de irmãos completos  $(\kappa_0, \kappa_1, \kappa_2) = (0.25, 0.5, 0.25)$ . Os  $\kappa$ 's estão relacionados com os coeficientes de parentesco através da fórmula  $\varphi = \kappa_1/4 + \kappa_2/2$ .

## Leitura adicional

- Vigeland, 2024. *Measures of relatedness*. (notas de curso)
- Ritland, 1996. *Estimators for pairwise relatedness and individual inbreeding coefficients*. *Genetical Research*, 67(2), 175-185.

**BOA SORTE!**

## Prólogo

O detetive Vargas já não estava a gostar daquele caso. Começou mal e provavelmente acabaria pior. Teve pena da jovem sentada à sua frente, com uma criança a dormir num carrinho ao seu lado.

“Não há pai”, disse ela, puxando a manga por cima de um hematoma que estava a desaparecer. “O meu filho não precisa de pai. Ele tem-me a mim.”

Era evidente que ela estava a fugir de alguém. Mas de quem? Quando ele perguntou pela família, ela fechou-se em silêncio, com os olhos fixos no carrinho. Protetora, mas também assustada.

O detetive Vargas examinou novamente o relatório genético. Era a mãe da criança, sem dúvida. Mas algo nos dados o confundia. Vargas não era especialista, mas o perfil da criança parecia demasiado simples. Muitos números repetidos. Um pensamento começou a formar-se na sua mente. Se a mulher não identificasse o pai, tudo bem. Ele já estava ali, inscrito no DNA da criança. Metade do perfil, se não mais...?

Ele precisava de alguém que pudesse extrair segredos do DNA. E aquele especialista que deu uma ajuda no ano passado? Com um sorriso irónico, pegou no telefone...

## Exercícios

Considere um STR com  $m$  alelos e frequências  $p_1, \dots, p_m$ , tal que  $p_1 + \dots + p_m = 1$ .

1. Seja  $H$  a homozigotia esperada para esse marcador, i.e, a probabilidade de um indivíduo ao acaso, não consanguíneo, ser homozigótico. Assim  $H$  é igual a

- a)  $p_1^2 + \dots + p_m^2$
- b)  $1 - p_1^2 - \dots - p_m^2$
- c)  $(1 - p_1)^2 + \dots + (1 - p_m)^2$
- d)  $1 - (1 - p_1)^2 - \dots - (1 - p_m)^2$
- e)  $1/m$

2. Suponha que uma criança com um coeficiente de endogamia  $f$  é tipada para esse marcador. A probabilidade de a criança ser homozigótica é

- a)  $fH$
- b)  $f + (1 - f)H$
- c)  $f(1 - H) + (1 - f)H$
- d)  $(1 - f)H$
- e)  $1 - fH$

3. Suponha que a criança é genotipada para  $L$  marcadores independentes, com homozigotia esperada  $H_1, \dots, H_L$ , respectivamente. O número esperado de genótipos homozigóticos é

- a)  $f(H_1 + \dots + H_L)$
- b)  $(1 - f)(H_1 + \dots + H_L)$
- c)  $fL + (1 - f)(H_1 + \dots + H_L)$
- d)  $fL + (1 - 2f)(H_1 + \dots + H_L)$
- e)  $\frac{L}{m}(H_1 + \dots + H_L)$

Seja  $H^*$  a soma  $H^* = H_1 + \dots + H_L$ . Das perguntas anteriores, pode demonstrar-se que um estimador do coeficiente de endogamia  $f$  pelo método dos momentos (*method-of-moments*) é dado por

$$\hat{f}_1 = \frac{N_{\text{hom}} - H^*}{L - H^*},$$

onde  $N_{\text{hom}}$  é o número observado de marcadores homozigóticos na criança.

Para os exercícios seguintes, utilize os dados fornecidos em *mother\_child.txt* e *db35.txt*.

4. O estimador  $\hat{f}_1$  fornece um coeficiente de endogamia estimado para a criança de aproximadamente

- a) 0,10
- b) 0,14
- c) 0,17
- d) 0,21
- e) 0,28

Outro estimador do coeficiente de endogamia muito usado é o estimador de Ritland (Ritland, 1996), definido por

$$\hat{f}_2 = \frac{(T_1 + \dots + T_L) - L}{(m_1 + \dots + m_L) - L}.$$

Aqui,  $m_j$  é o número de alelos do marcador  $j$ , e  $T_j$  é igual a  $1/p_a$  se o genótipo para o marcador  $j$  é  $a/a$ , e 0 para os outros casos.

5. O estimador de Ritland fornece um coeficiente de endogamia estimado para a criança de aproximadamente

- a) -0,08
- b) 0,14
- c) 0,16
- d) 0,22
- e) 0,28

6. O estimador de Ritland é muito sensível a alelos raros. Suponha (só para este exercício!) que o genótipo da criança para o marcador *PentaD* é alterado de 7/9 para 7/7. Neste caso, o estimador de Ritland para o coeficiente de endogamia seria aproximadamente

- a) -0,08
- b) 0,18
- c) 0,38
- d) 0,58
- e) 0,78

Quando estão disponíveis dados genotípicos para a criança e para um dos progenitores, existem outros estimadores  $f_{\text{child}}$  disponíveis. É possível demonstrar que, para qualquer marcador, temos probabilidades iguais

$$P(\text{mãe e criança terem o mesmo genótipo}) = P(\text{criança ser homozigótica}).$$

(Esta curiosa igualdade é interessante por si só, e encorajamo-lo a tentar prová-la!). Isto leva a uma terceira estimativa de  $f_{\text{child}}$ , substituindo  $N_{\text{hom}}$  na equação de  $\hat{f}_1$ , pelo número  $N_{\text{eq}}$  de marcadores para os quais a mãe e a criança têm genótipos idênticos:

$$\hat{f}_3 = \frac{N_{\text{eq}} - H^*}{L - H^*}.$$

7. O estimador  $\hat{f}_3$  fornece um coeficiente de endogamia estimado para a criança de aproximadamente

- a) 0,14
- b) 0,17
- c) 0,21
- d) 0,24
- e) 0,28

Como  $\hat{f}_3$  se baseia nas mesmas probabilidades que  $\hat{f}_1$ , geralmente estes têm propriedades semelhantes. No entanto, obtemos um estimador mais preciso ao combiná-los, definido por

$$\hat{f}_4 = \frac{\hat{f}_1 + \hat{f}_3}{2}.$$

8. Usando  $\hat{f}_4$ , o coeficiente de endogamia estimado para a criança é de aproximadamente

- a) 0,15
- b) 0,17
- c) 0,19
- d) 0,21
- e) Nenhuma das anteriores

Os gráficos da Figura 1 mostram as distribuições dos quatro estimadores  $\hat{f}_1, \hat{f}_2, \hat{f}_3, \hat{f}_4$ , obtidas pela simulação de 10.000 pares mãe-filho em dois cenários: quando os pais não são aparentados e quando os pais são irmãos completos.

9. Com base na Figura 1, qual das seguintes afirmações **não** é correta?

- a) No cenário de ausência de consaguinidade (*outbred*), o estimador de Ritland  $\hat{f}_2$  apresenta a maior precisão.
- b) No cenário de irmãos completos (*full-sibs*), o estimador de Ritland  $\hat{f}_2$  é o menos preciso.
- c) Os estimadores de momento  $\hat{f}_1$  e  $\hat{f}_3$  apresentam distribuições quase idênticas em ambos os cenários.
- d) No cenário de irmãos completos (*full-sibs*), o estimador  $\hat{f}_4$  apresenta a maior precisão.
- e) A altura dos picos das distribuições determina a precisão dos estimadores.

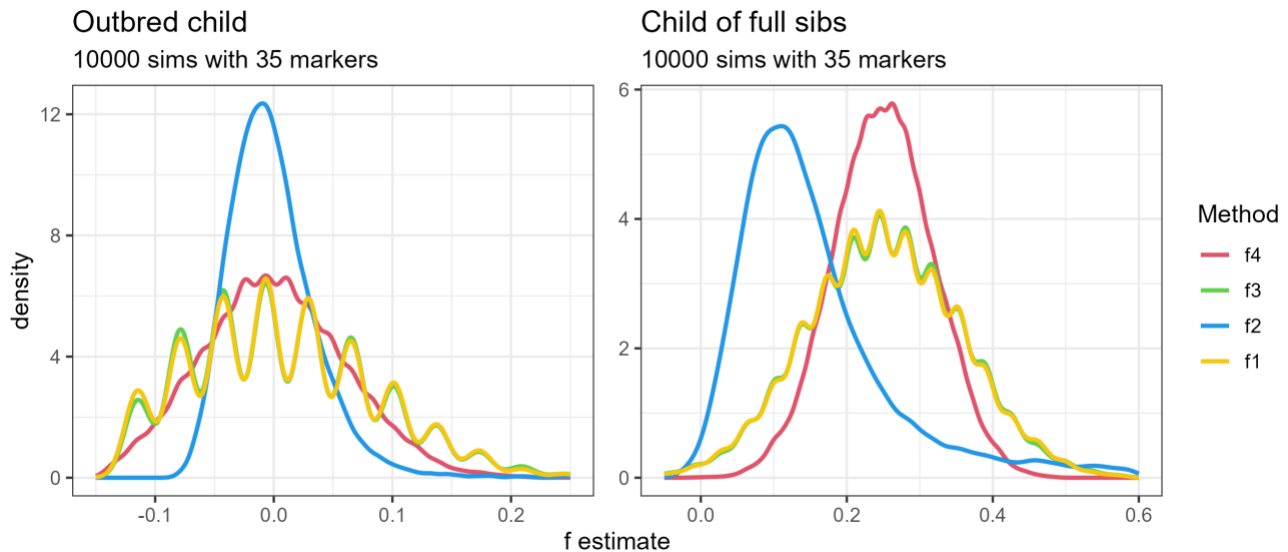


Figura 1: Distribuições das estimativas do coeficiente de endogamia com base em 10.000 simulações em cada cenário.

Considere as seguintes hipóteses de coeficiente de endogamia da criança:

$$H_0: f = 0 \quad (\text{ausência de endogamia})$$

$$H_A: f > 0 \quad (\text{endogamia}).$$

10. Seja  $c$  a estimativa observada  $\hat{f}_4$ , i.e., o valor obtido no Exercício 8. Em 10 000 simulações com pais não consanguíneos, 9 resultaram em  $\hat{f}_4 \geq c$ . Da mesma forma, das 10 000 simulações com pais irmãos completos, 7706 resultaram em  $\hat{f}_4 \geq c$ . Com base nisto, um valor de  $p$  ( $p$ -value) aproximado para testar a hipótese nula  $H_0$ , contra a alternativa  $H_A$ , é

- a) 0,00045
- b) 0,0009
- c) 0,3853
- d) 0,7706
- e) Nenhuma das anteriores

Agora, vamos analisar os dados com base em valores de LR. Considere as seguintes hipóteses:

- $U$ : Os pais são não relacionados
- $H$ : Os pais são meios irmãos
- $S$ : Os pais são irmãos completos
- $P$ : Os pais são pai-filha

11. Usando apenas os dados da criança, a LR comparando  $S$  com  $U$ , é aproximadamente

- a) 1,02
- b) 38
- c) 43
- d) 44
- e) 3076

**12.** Usando apenas os dados da criança, as LR's das hipóteses satisfazem

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $1 < LR_{H:U} < LR_{S:U} = LR_{P:U}$
- d)  $1 < LR_{S:U} = LR_{P:U} < LR_{H:U}$
- e)  $1 < LR_{P:U} < LR_{S:U} < LR_{H:U}$

**13.** Usando os dados da mãe e da criança, a LR comparando  $P$  com  $U$  é aproximadamente

- a) 1,02
- b) 38
- c) 156
- d) 3076
- e) 48 335

**14.** Usando os dados da mãe e da criança, as LR's satisfazem

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $1 < LR_{H:U} < LR_{S:U} = LR_{P:U}$
- d)  $1 < LR_{S:U} = LR_{P:U} < LR_{H:U}$
- e)  $1 < LR_{P:U} < LR_{S:U} < LR_{H:U}$

Algum tempo depois, um homem foi detido após um arrombamento no prédio da mãe. Ela continuou a recusar-se a falar, mas o DNA do homem foi recolhido e genotipado. O resultado foi reportado com uma conclusão clara: era o pai biológico da criança. Mas qual era o parentesco dele com a mãe?

Para os próximos exercícios, utilize os dados anteriores juntamente com os genótipos do progenitor em *father.txt*.

**15.** Usando todos os dados disponíveis, o número de marcadores que mostram incompatibilidade com a hipótese  $P$  é

- a) 0
- b) 1
- c) 2
- d) 3
- e) Nenhuma das anteriores

**16.** Usando todos os dados disponíveis, as LR's satisfazem

- a)  $1 < LR_{H:U} < LR_{S:U} < LR_{P:U}$
- b)  $1 < LR_{H:U} < LR_{P:U} < LR_{S:U}$
- c)  $LR_{H:U} < LR_{P:U} < 1 < LR_{S:U}$
- d)  $LR_{P:U} < LR_{H:U} < 1 < LR_{S:U}$
- e)  $LR_{P:U} < 1 < LR_{H:U} < LR_{S:U}$

17. O LR comparando a hipótese mais provável com a hipótese  $U$ , é aproximadamente

- a)  $8,5 \cdot 10^9$
- b)  $3,1 \cdot 10^6$
- c)  $1,9 \cdot 10^6$
- d)  $7,0 \cdot 10^5$
- e)  $2,3 \cdot 10^3$

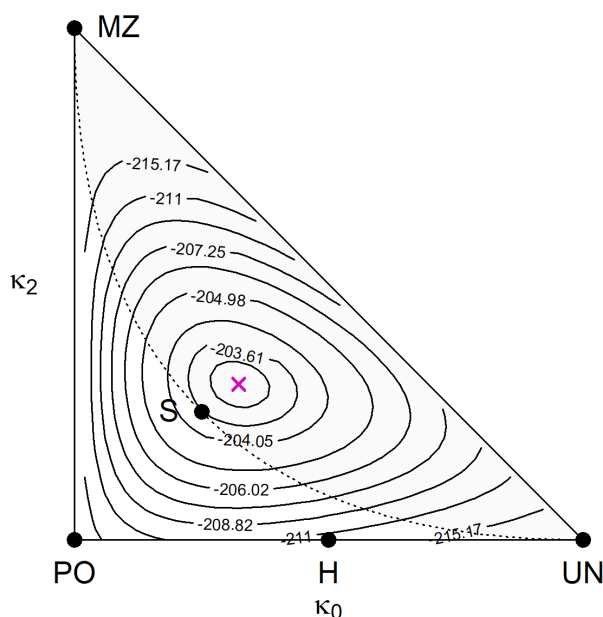


Figura 2: Gráfico de contorno da função de log-probabilidade do parentesco entre os pais da criança. Parentescos indicados: não aparentados (UN), pai-filho (PO), gémeos monozigóticos (MZ), irmãos completos (S) e meios-irmãos (H). Cada eixo varia de 0 a 1.

18. A Figura 2 mostra os contornos da função de log-probabilidade da relação entre o pai e a mãe, em função dos seus coeficientes  $\kappa$  ( $\kappa_0, \kappa_2$ ). Com base na figura, a estimativa de máxima verossimilhança ( $\hat{\kappa}_0, \hat{\kappa}_1, \hat{\kappa}_2$ ) é aproximadamente

- a) (0,32; 0,38; 0,30)
- b) (0,25; 0,39; 0,36)
- c) (0,52; 0,45; 0,03)
- d) (0,29; 0,53; 0,08)
- e) Nenhuma das anteriores

Recorde-se que o coeficiente de parentesco  $\varphi$  entre dois indivíduos é determinado pelos seus coeficientes  $\kappa$  pela fórmula  $\varphi = \frac{1}{4}\kappa_1 + \frac{1}{2}\kappa_2$ . Como o coeficiente de consanguinidade de uma pessoa é igual ao coeficiente de parentesco entre os progenitores, isto dá origem a outro estimador do coeficiente de consanguinidade do filho:

$$\hat{f}_5 = \frac{1}{4}\hat{\kappa}_1 + \frac{1}{2}\hat{\kappa}_2$$

19. Usando  $\hat{f}_5$ , o coeficiente de endogamia estimado para a criança é de aproximadamente

- a) 0,13
- b) 0,17
- c) 0,25
- d) 0,27
- e) 0,28

20. Considere um trio composto por uma criança e ambos os progenitores, genotipados para um conjunto de marcadores autossômicos. Seja  $\mathcal{R}$  a relação entre os progenitores. Qual das seguintes afirmações **não** é verdadeira?

- a) Os dados da criança, por si só, contêm informação sobre o  $\mathcal{R}$ .
- b) Os dados da criança e de um dos pais contêm mais informação sobre o  $\mathcal{R}$  do que os dados da criança sozinha.
- c) Os dados da criança e de um dos pais não permitem distinguir se  $\mathcal{R}$  é irmão(ã) ou filho(a) de um dos pais.
- d) Considerando os dados da criança e de um dos pais, o outro progenitor não fornece informações adicionais sobre o  $\mathcal{R}$ .
- e) Considerando os dados de ambos os pais, a criança não fornece informação adicional sobre o  $\mathcal{R}$ .

**FIM**

Responsável pelo desenvolvimento do exercício:

Magnus Dehli Vigeland

[magnusdv@gmail.com](mailto:magnusdv@gmail.com)

Tradução:

Cíntia Alves ([cintialves70@gmail.com](mailto:cintialves70@gmail.com))

Comité Executivo do GHEP-ISFG

[info@ghep-isfg.org](mailto:info@ghep-isfg.org)