



A dynamic model to analyze genetic variations of host traits that affect disease transmission

Daiane A. Zuanetti

Milena Nascimento Lima

Oswaldo Anacleto Jr

Programa Interinstitucional de Pós-Graduação em Estatística - UFSCar & USP

Ciclo de Palestras do Programa de Pós Graduação em Estatística / IM-UFRJ

Como reduzir a disseminação de doenças e prevenir epidemias?



O problema:

As atuais estratégias de controle de doenças nem sempre são eficientes ou disponíveis.

O melhoramento genético oferece uma alternativa viável ao controle de doenças

- Criação seletiva de animais;
- Genética quantitativa;
- Estatística



Características do hospedeiro que afetam a propagação de doenças infecciosas

Susceptibilidade

Definição:

É a capacidade do hospedeiro bloquear a entrada do patógeno ou restringir sua replicação.

Aspectos:

- Relação inversamente proporcional à resistência à doenças.
- Efeito Direto.

Importância:

Auxilia no diagnóstico e controle de doenças infecciosas.

É o alvo dos atuais estudos genéticos de doenças!

Infectividade

Definição:

É a capacidade do hospedeiro causar a infecção por unidade de contato.

Aspectos:

- Efeito Indireto (social)
- Relação diretamente proporcional ao risco de transmissão.

Importância:

Contribui para a previsão e prevenção de surtos futuros de doenças.

Alta infectividade leva ao superespalhamento!

Objetivos

Propor modelos estatísticos que possam capturar a **dinâmica da transmissão de doenças** e consequentemente aprimorar a estimação de efeitos genéticos na **susceptibilidade e infectividade**.

Analisar o efeito de **diferentes desenhos experimentais** na qualidade das estimativas do modelo *dnIGE*;

Estender o modelo, implementando a **correlação genética** entre as características epidêmicas;

Conceitos de Genética Quantitativa

**Modelo Básico da Genética
quantitativa**

$$P = A + E$$

Componentes de variação

$$\sigma_P^2 = \sigma_A^2 + \sigma_E^2$$

Herdabilidade

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}$$

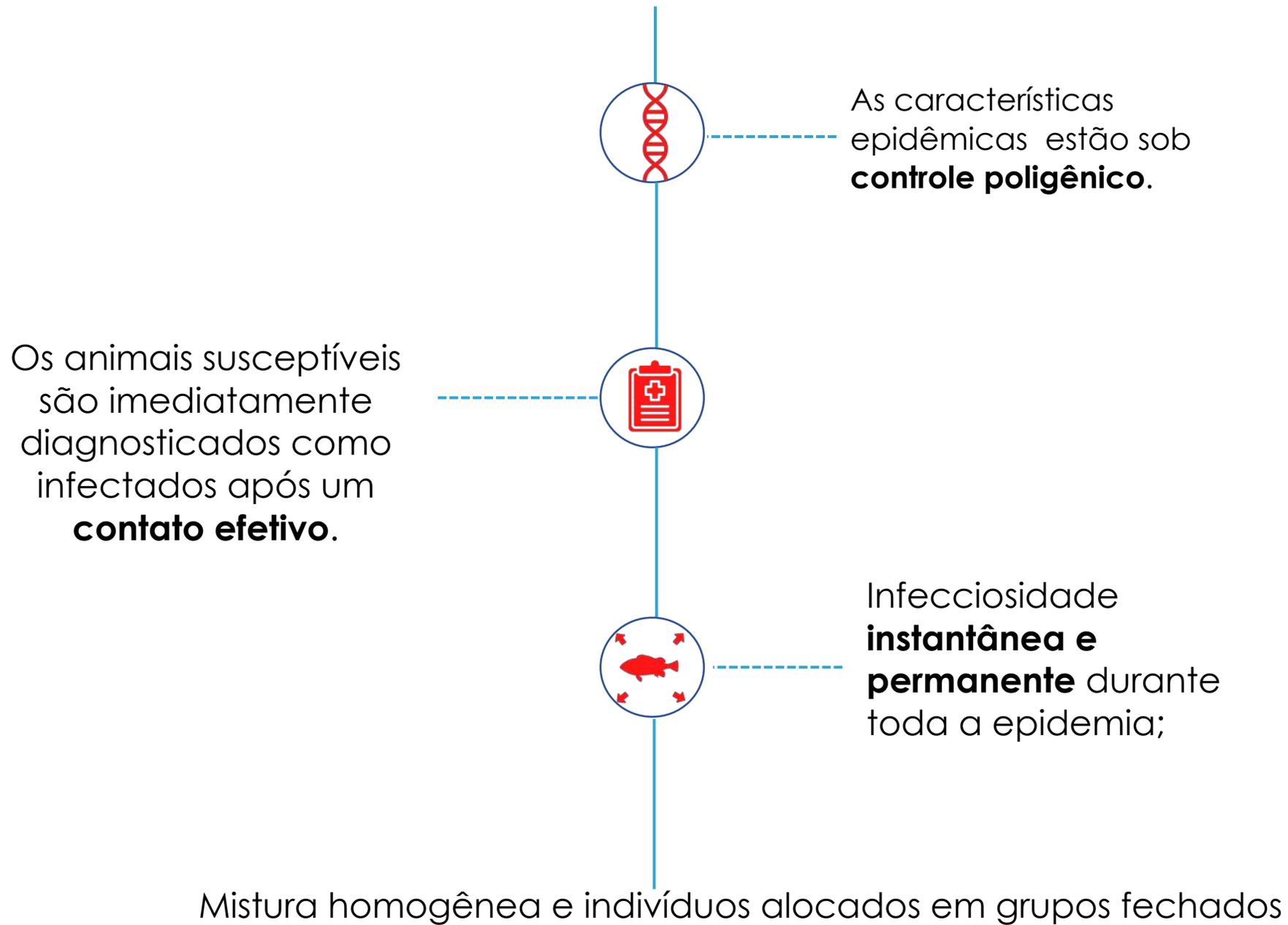
Modelos Epidêmicos

- Modelos compartimentais *SIR*

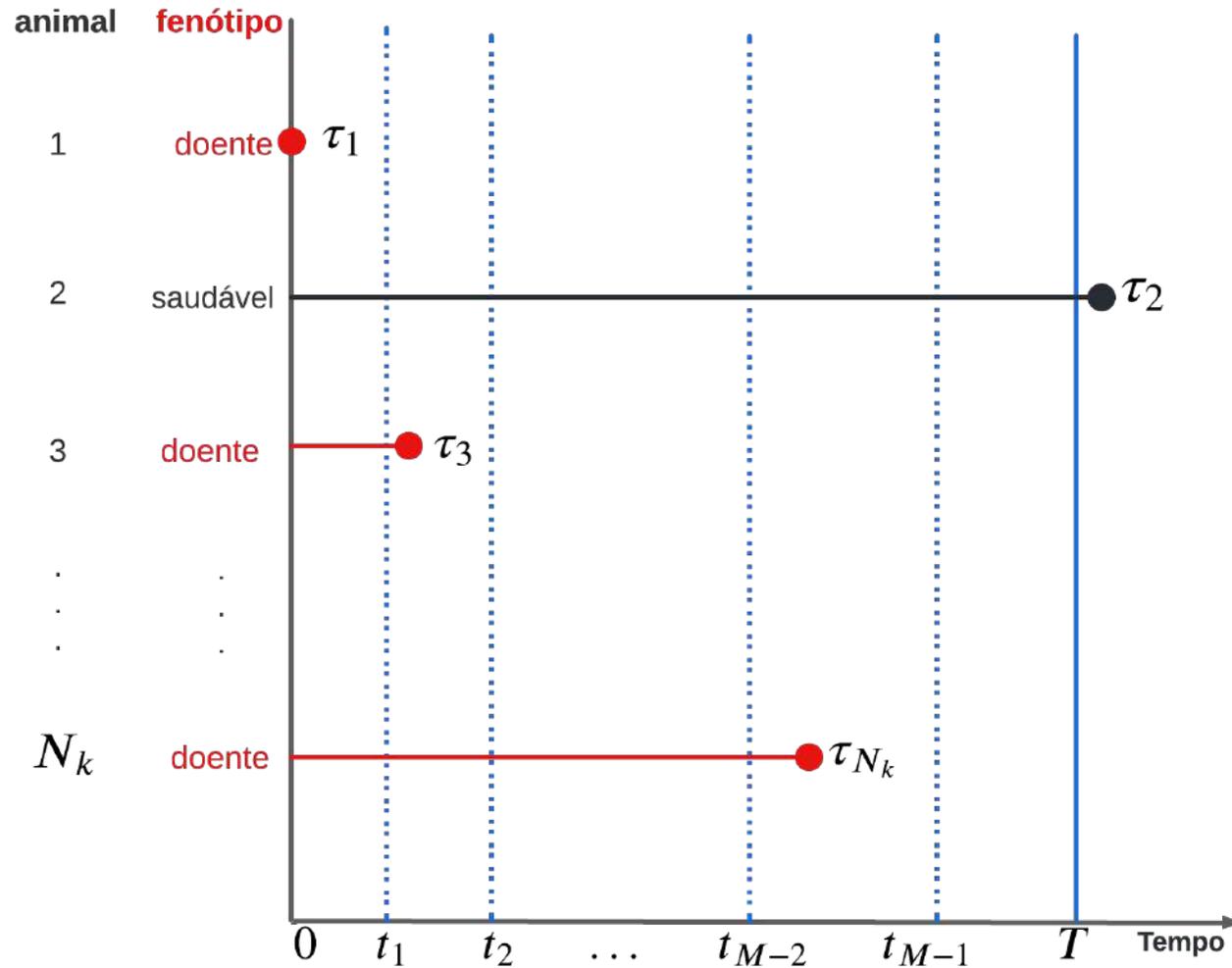
Suposições:

- *População de tamanho fixo*
- *Mistura homogênea*
- *Caso índice*
- *Divisão em compartimentos (**S-I-R**)*

Suposições do modelo proposto (dnIGE)



Estrutura de dados de infecção



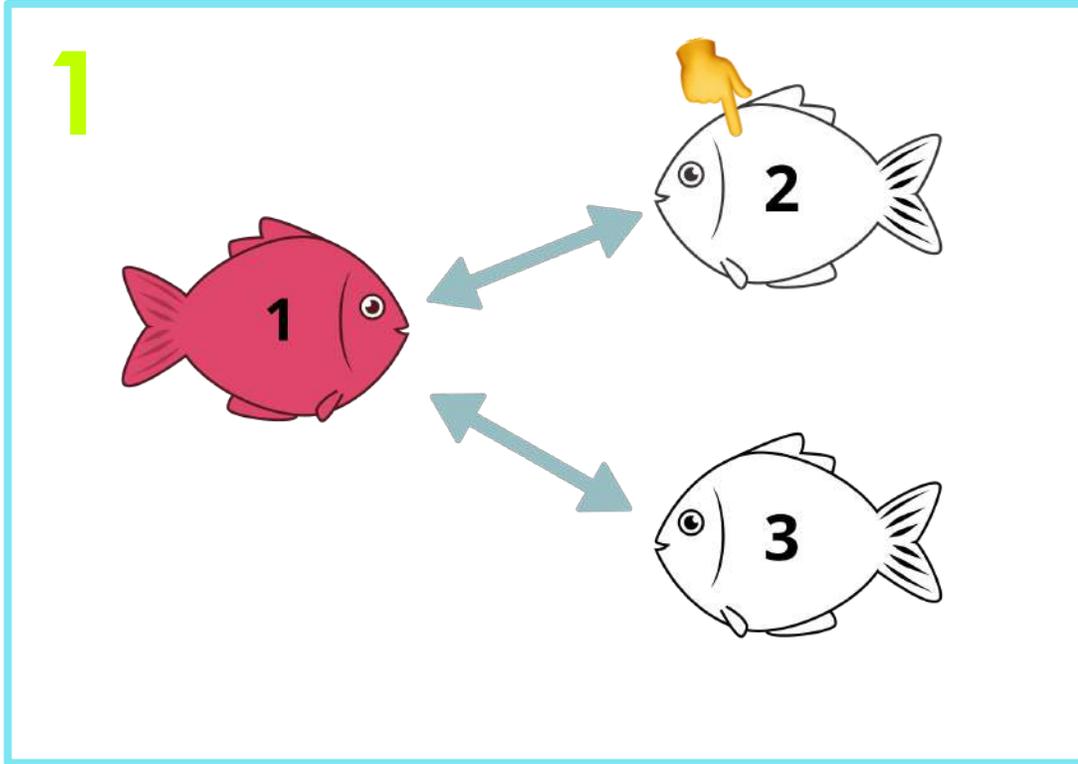
Taxa de infecção do indivíduo j no tempo t

$$\lambda_j(t) = g_j \beta \sum_{k:p_k=p_j} f_k \ell_k(t)$$

Em que:

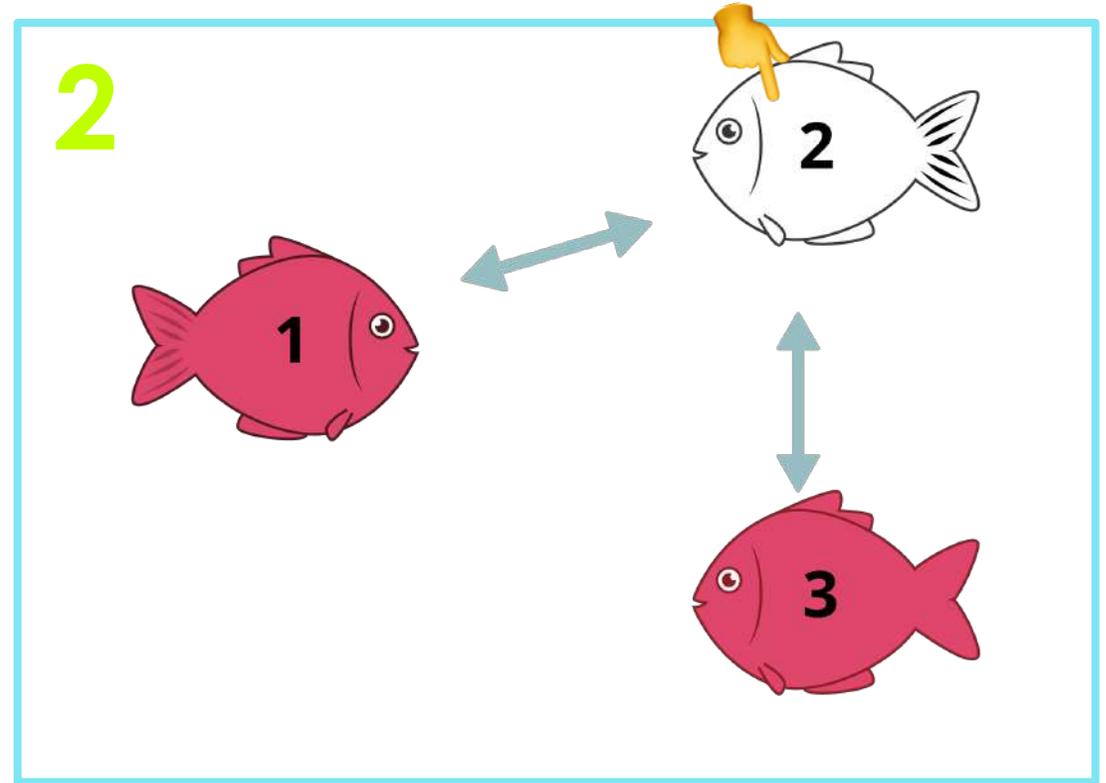
- g_j é a susceptibilidade do animal j ;
- β é a taxa de transmissão (taxa de contato efetivo);
- $\sum f_k \ell_k(t)$ é a infectividade dos infectados antes de j ;
- $\ell_k(t)$ é o indicador de infecção: $\ell_k(t) = 1$ se $\tau_k < t$ e $\ell_k(t) = 0$ caso contrário;
- p_j indica o grupo do animal j .

Ilustração da taxa de infecção



A taxa de infecção do animal 2 é dada por

$$\lambda_2(t) = g_2 \beta f_1$$



A taxa de infecção do animal 2 é dada por

$$\lambda_2(t) = g_2 \beta (f_1 + f_3)$$

Componente de variação para a susceptibilidade

$$\log(g_j) = a_{g,j} + e_{g,j}$$

\mathbf{a}_g

É um vetor de efeitos aleatórios genéticos (correlacionados) de dimensão $(N - n_0)$.

\mathbf{e}_g

É um vetor de efeitos aleatórios ambientais (independentes) de dimensão $(N - n_0)$.

$$\begin{aligned}\mathbf{a}_g | \sigma_{A,g}^2 &\sim N(\mathbf{0}, \sigma_{A,g}^2 \mathbf{A}_g) \\ \mathbf{e}_g | \sigma_{E,g}^2 &\sim N(\mathbf{0}, \sigma_{E,g}^2 \mathbf{I}_g)\end{aligned}$$

Assumindo susceptibilidade e infectividade independentes

Componente de variação para a infectividade

$$\log(f_j) = a_{f,j} + e_{f,j}$$

a_f

É um vetor de efeitos aleatórios genéticos (correlacionados) de dimensão (n de inf.)

e_f

É um vetor de efeitos aleatórios ambientais (independentes) de dimensão (n de inf.)

$$\begin{aligned} a_f | \sigma_{A,f}^2 &\sim N(\mathbf{0}, \sigma_{A,f}^2 \mathbf{A}_f) \\ e_f | \sigma_{E,f}^2 &\sim N(\mathbf{0}, \sigma_{E,f}^2 \mathbf{I}_f) \end{aligned}$$

Assumindo susceptibilidade e inventividade independentes

Componentes de variação com correlação

Se assumirmos que susceptibilidade e infectividade são correlacionados e sejam

$$\mathbf{a} = [\mathbf{a}_g^\top, \mathbf{a}_f^\top]^\top \text{ e}$$
$$\mathbf{e} = [\mathbf{e}_g^\top, \mathbf{e}_f^\top]^\top .$$

Então,

$$\mathbf{a} \mid \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A \sim N_{2S} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{S,g}^2 & \rho_A \sigma_{S,g} \sigma_{S,f} \\ \rho_A \sigma_{S,g} \sigma_{S,f} & \sigma_{S,f}^2 \end{bmatrix} \otimes \mathbf{I}_S \right)$$
$$\mathbf{e} \mid \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E \sim N_{2N} \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma_{E,g}^2 & \rho_E \sigma_{E,g} \sigma_{E,f} \\ \rho_E \sigma_{E,g} \sigma_{E,f} & \sigma_{E,f}^2 \end{bmatrix} \otimes \mathbf{I}_N \right) .$$

Função de verossimilhança

Seja $Y_j(w)$ o número de contatos efetivos do j -ésimo indivíduo até o momento w e $Y_j(t)$ a taxa de infecção deste indivíduo j no momento t , para $j=1, \dots, N$,

$$Y_j(w) \sim \text{Poisson} \left(\int_0^w \lambda_j(t) dt \right).$$

Assim, a probabilidade de que o tempo de infecção do j -ésimo indivíduo seja exatamente τ_j é dada por

$$\begin{aligned} P(Y_j(\tau_j) = 0) \lambda_j(\tau_j) &= \exp \left\{ - \int_0^{\tau_j} \lambda_j(t) dt \right\} \lambda_j(\tau_j) \\ &= \exp \left\{ - \int_0^{\tau_j} \exp\{a_{g,j} + e_{g,j}\} \beta \sum_{k:p_k=p_j} \exp\{a_{f,j} + e_{f,k}\} \ell_k(t) dt \right\} \\ &\quad \times \exp\{a_{g,j} + e_{g,j}\} \beta \sum_{k:p_k=p_j} \exp\{a_{f,j} + e_{f,k}\} \ell_k(\tau_j) \end{aligned}$$

Função de verossimilhança

A probabilidade de que o tempo de infecção do j -ésimo indivíduo seja maior que o tempo final T do experimento é

$$P(Y_j(T) = 0) = \exp \left\{ - \int_0^T \lambda_j(t) dt \right\} .$$

A função de verossimilhança é dada pela multiplicação da última ou penúltima probabilidade (de acordo com a situação) para cada indivíduo da amostra.

Distribuições a priori e a posteriori

$$P(\boldsymbol{\theta} | \boldsymbol{\tau}) \propto \mathbb{L}(\boldsymbol{\theta} | \boldsymbol{\tau})P(\mathbf{a} | \sigma_{S,g}^2, \sigma_{S,f}^2, \rho_A)P(\mathbf{e} | \sigma_{E,g}^2, \sigma_{E,f}^2, \rho_E) \\ \times P(\sigma_{A,g}^2)P(\sigma_{A,f}^2)P(\sigma_{E,g}^2)P(\sigma_{E,f}^2)P(\rho_A)P(\rho_E)P(\beta)$$

Priori da taxa de transmissão

$$\beta \sim \text{Gamma} (.001, .001)$$

Priori das correlações genéticas e ambientais

$$\rho_A \sim \text{Lewandowski} - \text{Kurowicka} - \text{Joe} (1)$$

$$\rho_E \sim \text{Lewandowski} - \text{Kurowicka} - \text{Joe} (1)$$

Prioris dos desvios padrões

$$\sigma_{Sg} \sim \text{Half} - \text{Cauchy} (0,5)$$

$$\sigma_{Sf} \sim \text{Half} - \text{Cauchy} (0,5)$$

$$\sigma_{Eg} \sim \text{Half} - \text{Cauchy} (0,5)$$

$$\sigma_{Ef} \sim \text{Half} - \text{Cauchy} (0,5)$$

Amostras MCMC da distribuição a posteriori conjunta foram simuladas via Stan usando rstan.

Estudo de simulação

Simulamos estruturas de meio-irmão

- Geração parental: 100 genitores masculinos e 20 mães/pai;
- Prole: 2000 animais;
- Herdabilidade: 0.4 e 0.8;
- Tamanhos dos grupos: 2, 10 e 20;
- Métodos de alocação: *random* e *2FAM*;
- 20 replicações foram consideradas em cada simulação;
- Coeficiente de correlação: 0, +-0.2, +-0.4, +-0.6, +-0.8

Os efeitos genéticos aditivos dos pais de suscetibilidade e infectividade foram obtidos através da distribuição normal bivariada como comentado antes;

Estudo de simulação

Os efeitos genéticos aditivos dos indivíduos prole são obtidos somando-se os efeitos genéticos aditivos médios dos genitores e um termo mendeliano, simulado a partir de uma distribuição normal com média zero e variância correspondente à metade da variância genética aditiva da característica de interesse.

Para cada grupo fechado, um indivíduo foi escolhido aleatoriamente para ser o caso índice responsável pelo início do surto epidêmico e usamos o método direto de Gillespie para espalhar a epidemia em cada grupo fechado.

Métodos de alocação

Random

Consiste em selecionar, **ao acaso**, os animais constituintes da amostra, que serão alocados em cada um dos grupos independentes.

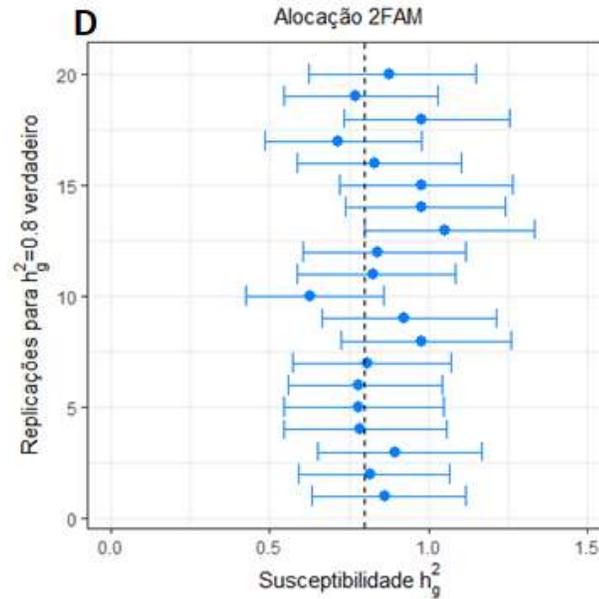
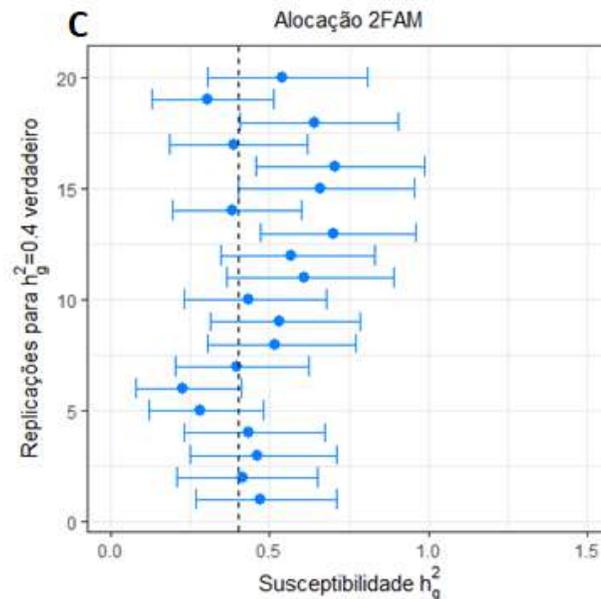
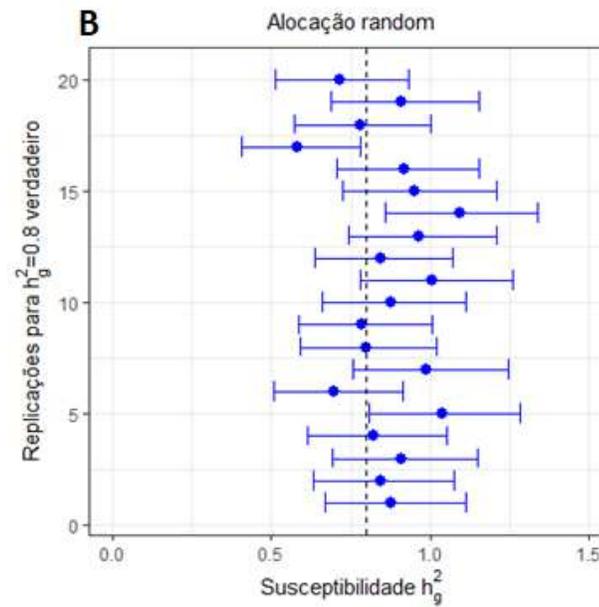
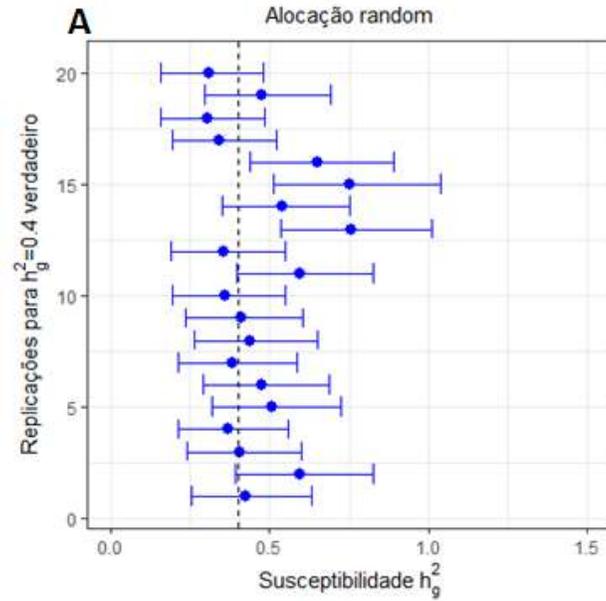
2FAM

Consiste em compor os grupos **por membros de duas famílias** distintas (selecionadas na estrutura de blocos), de maneira que cada família contribua com uma quantidade de membros equivalente à metade do tamanho do grupo.

FAMÍLIAS

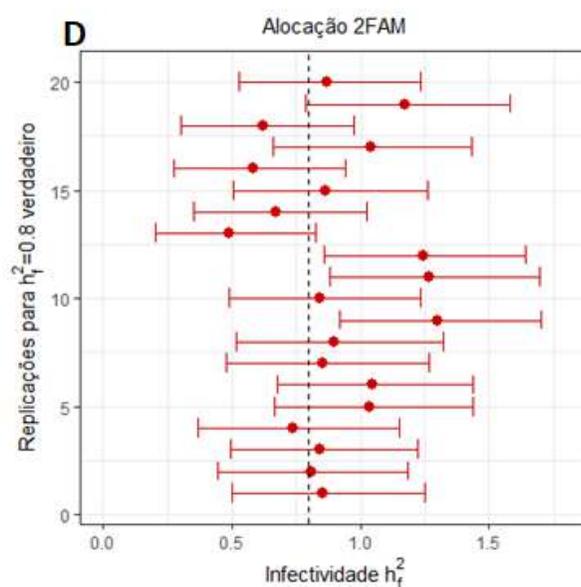
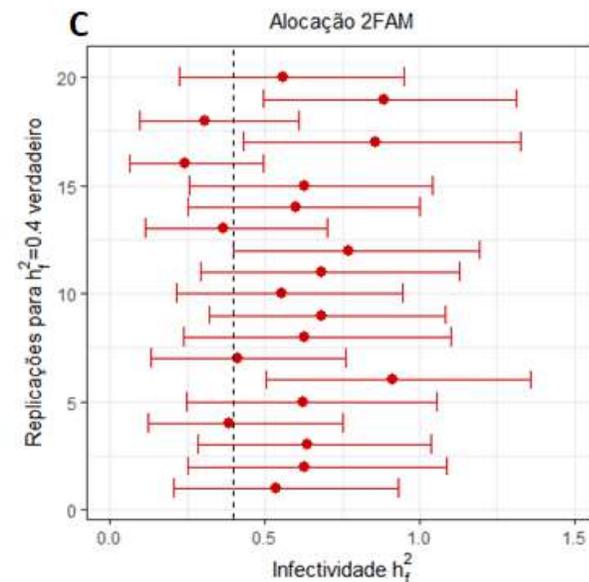
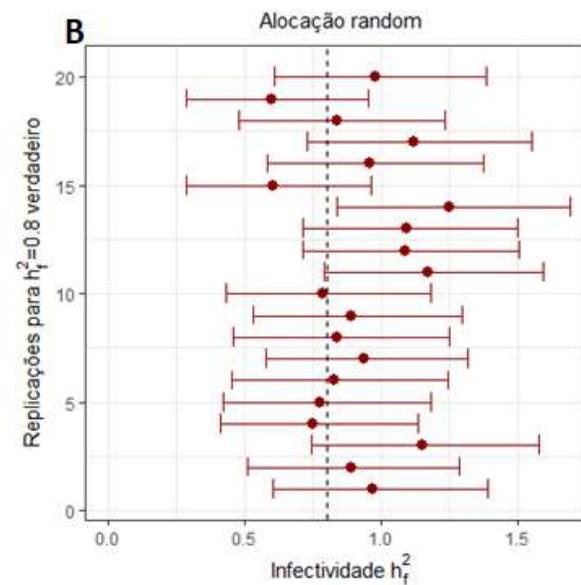
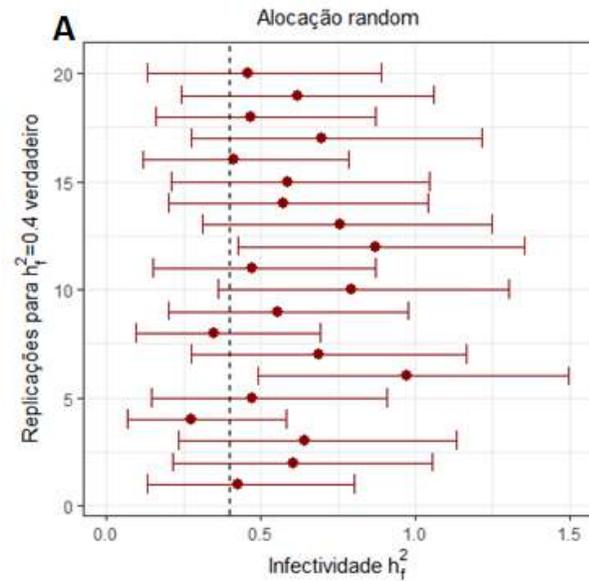
São agrupamentos de meios-irmãos.

Herdabilidade de susceptibilidade - IC90%



Considerando independência entre susceptibilidade e infectividade e grupo de tamanho 10.

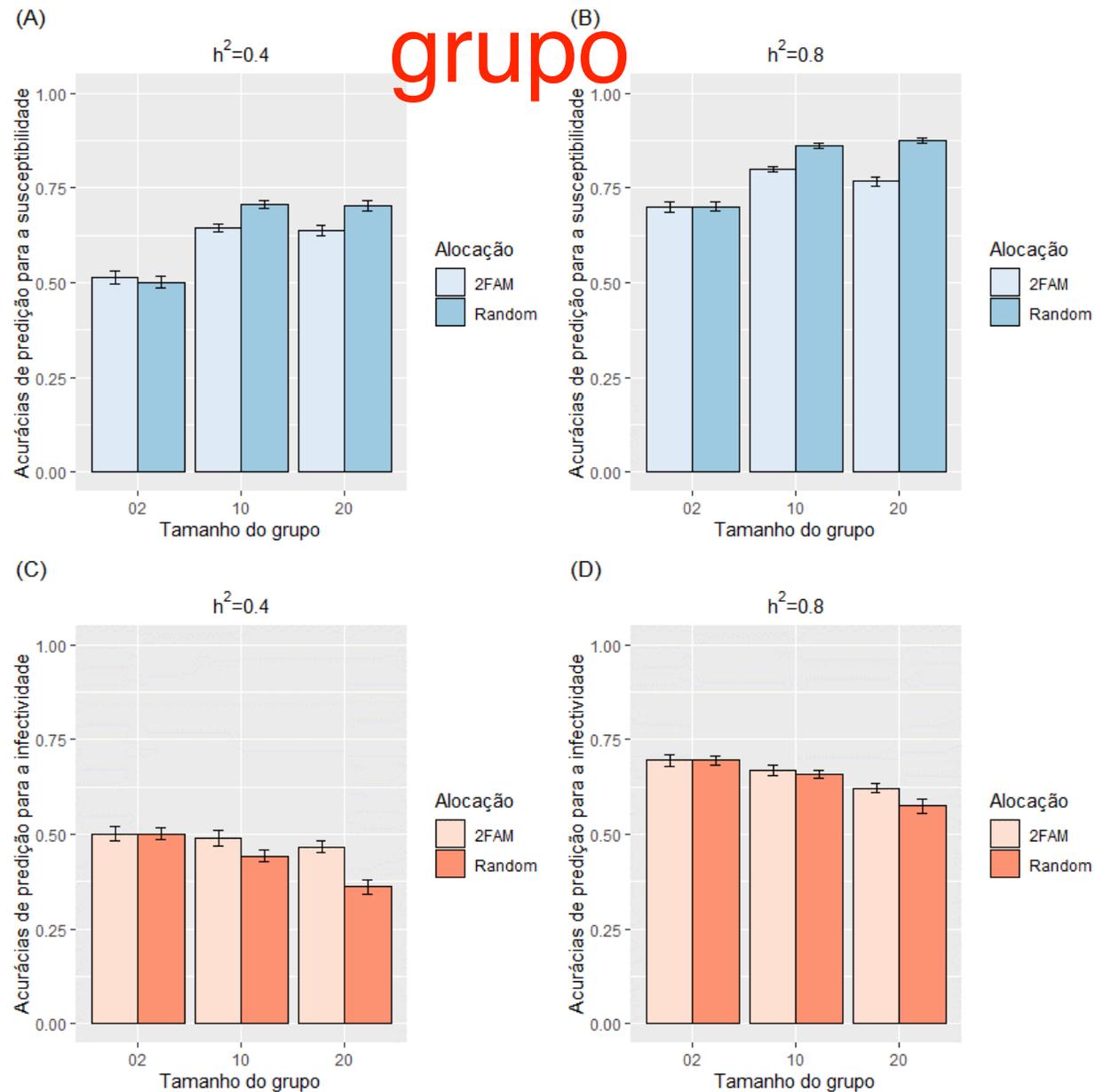
Herdabilidade da infectividade - IC90%



Considerando independência entre susceptibilidade e infectividade e grupo de tamanho 10.

A herdabilidade da susceptibilidade e da infectividade é apropriadamente estimada em ambos os métodos de alocação.

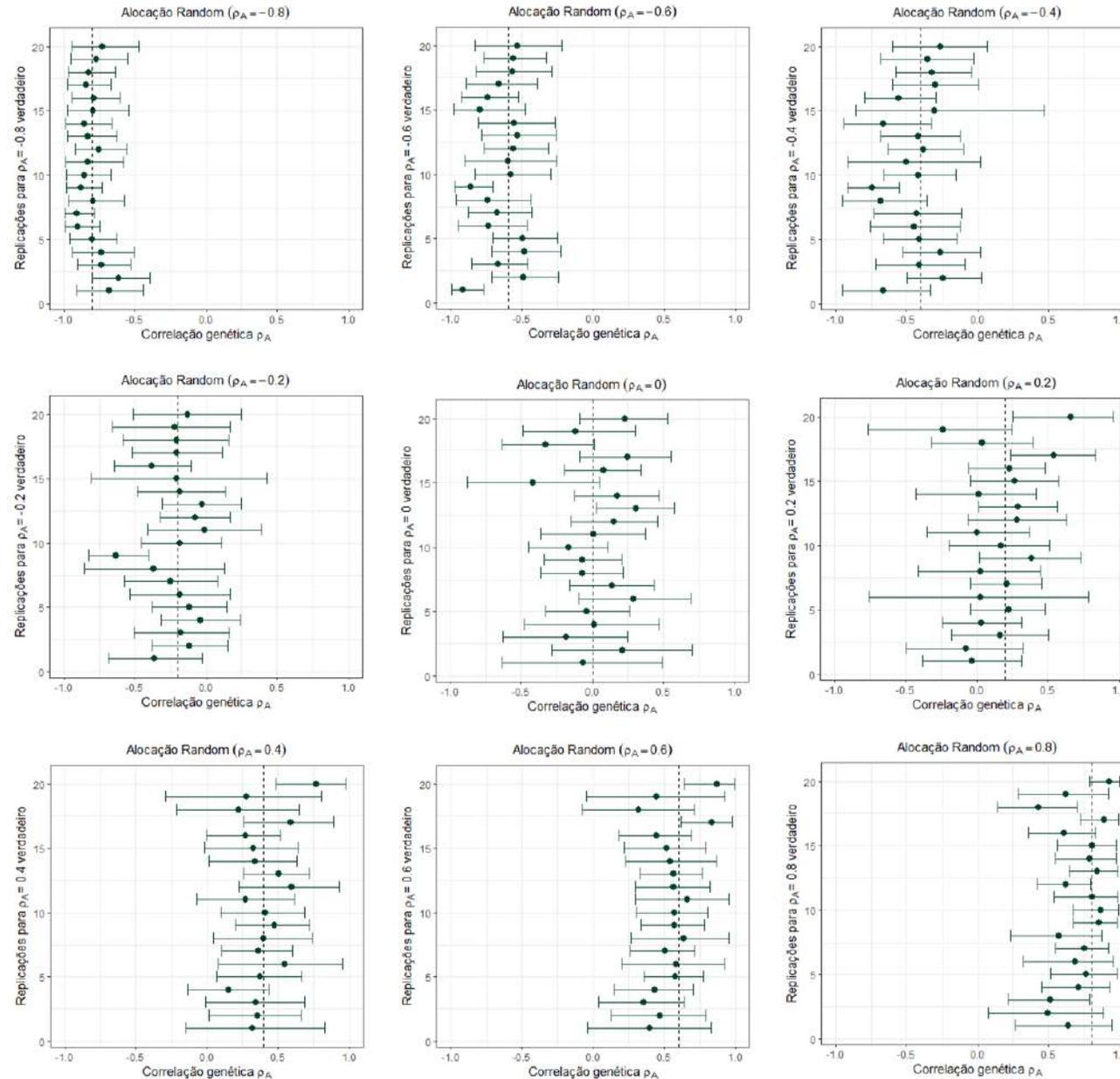
Efeito dos desenhos experimentais e tamanho de grupo



Efeito dos desenhos experimentais e tamanho de grupo

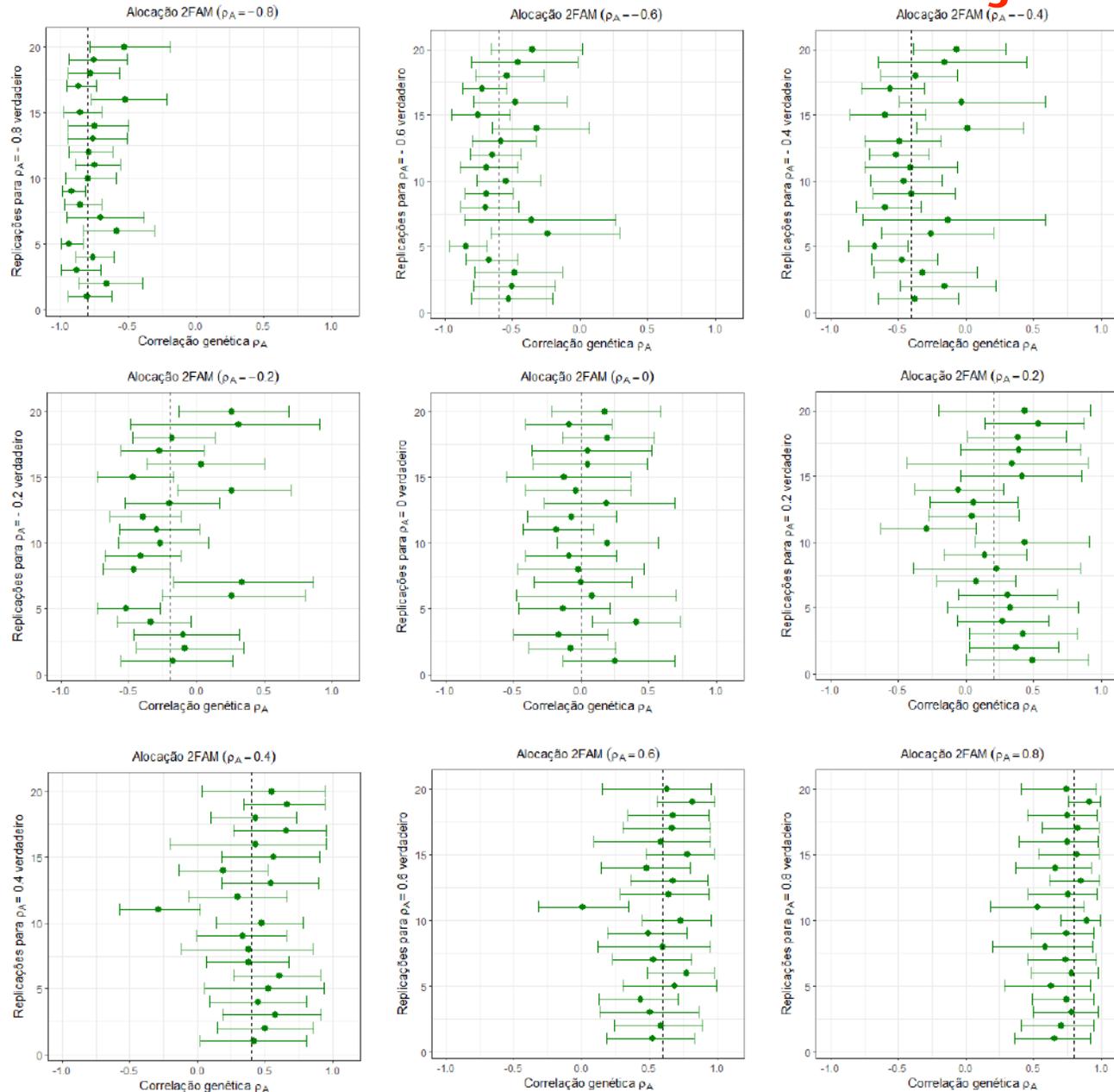
- O aumento da diversidade genética resulta em melhores acurácias.
- Grupos maiores favorecem a susceptibilidade.
- Grupos menores produzem as melhores estimativas da infectividade.
- **dnIGE com alocação random não é o mais eficiente para predizer os efeitos genéticos da infectividade.**
- O modelo estima mesmo não sabendo quem infectou quem.

Coeficiente de correlação



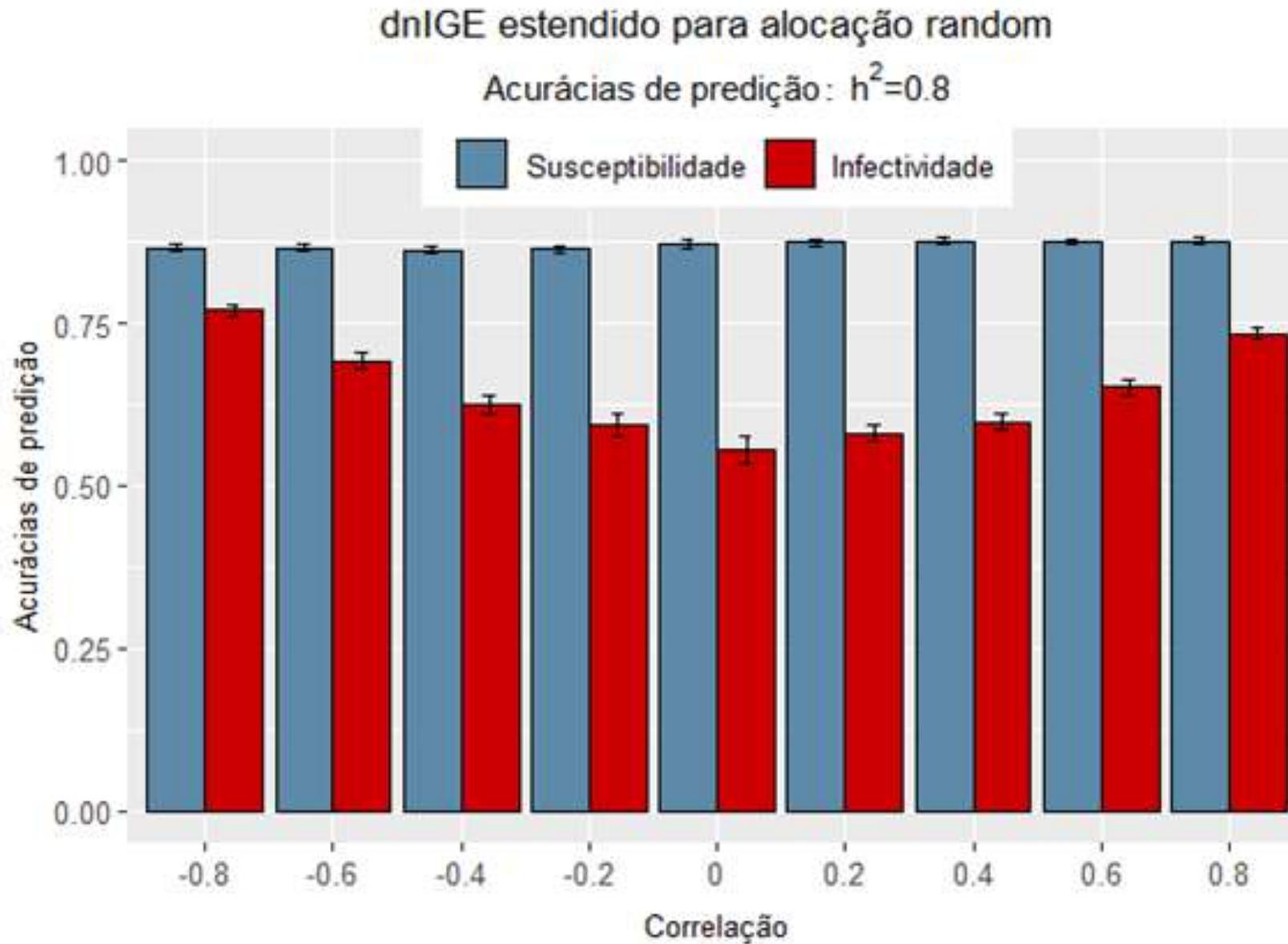
Considerando alocação random e grupo de tamanho 10.

Coeficiente de correlação



Considerando alocação 2FAM e grupo de tamanho 10.

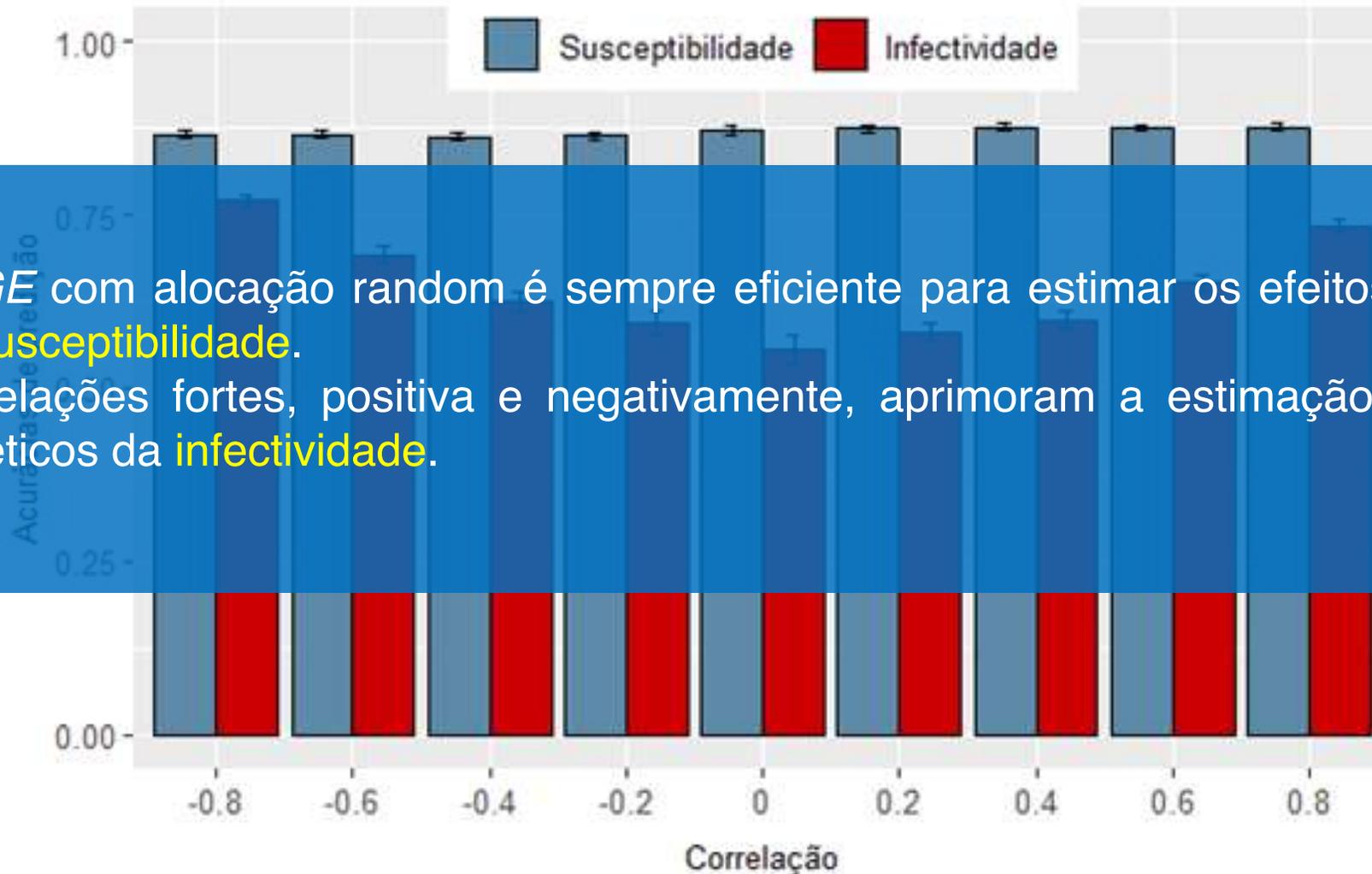
Efeito da correlação na predição



Efeito da correlação na predição

dnIGE estendido para alocação random

Acurácias de predição: $h^2=0.8$

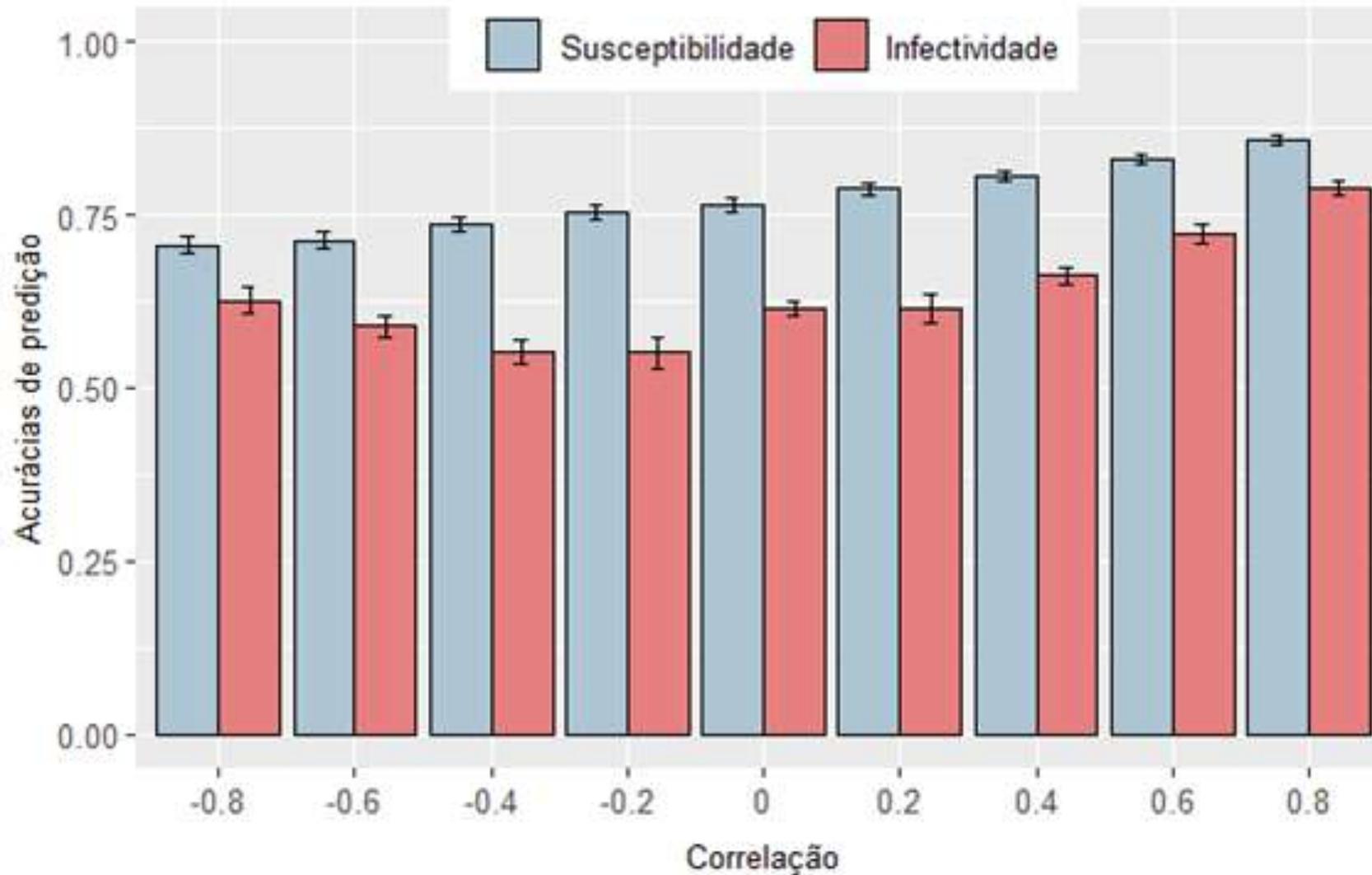


- *dnIGE* com alocação random é sempre eficiente para estimar os efeitos genéticos da **susceptibilidade**.
- Correlações fortes, positiva e negativamente, aprimoram a estimativa de efeitos genéticos da **infectividade**.

Efeito da correlação na predição

dnIGE estendido para alocação familiar

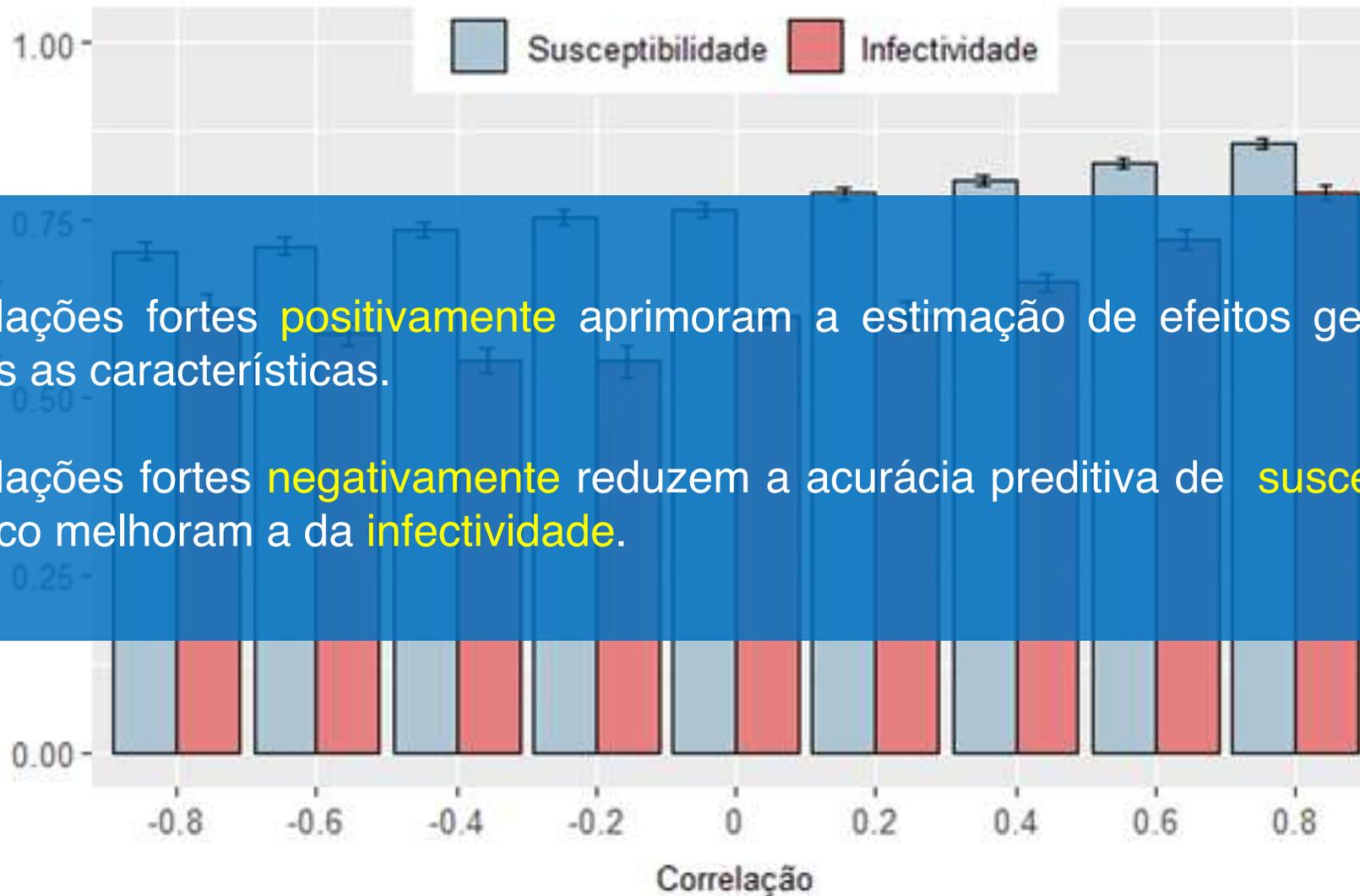
Acurácias de predição: $h^2=0.8$ 🖱️



Efeito da correlação na predição

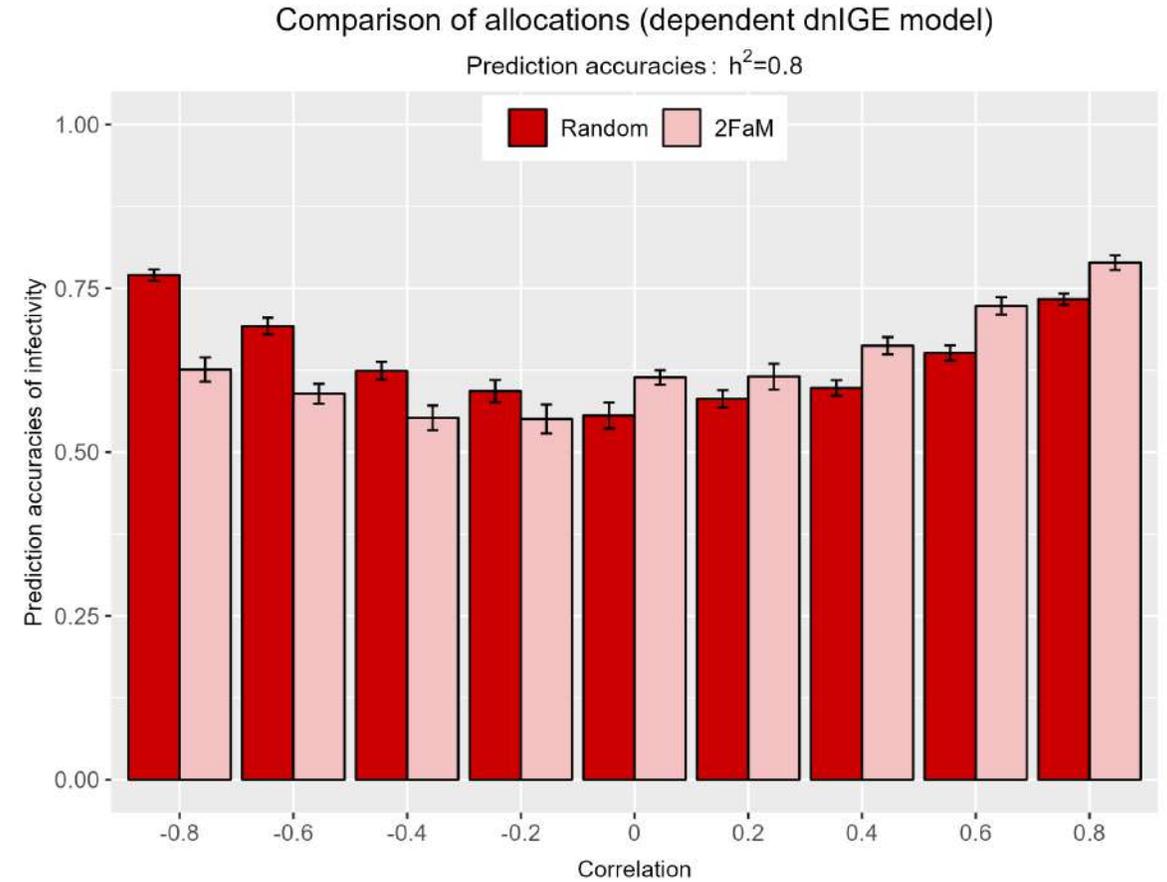
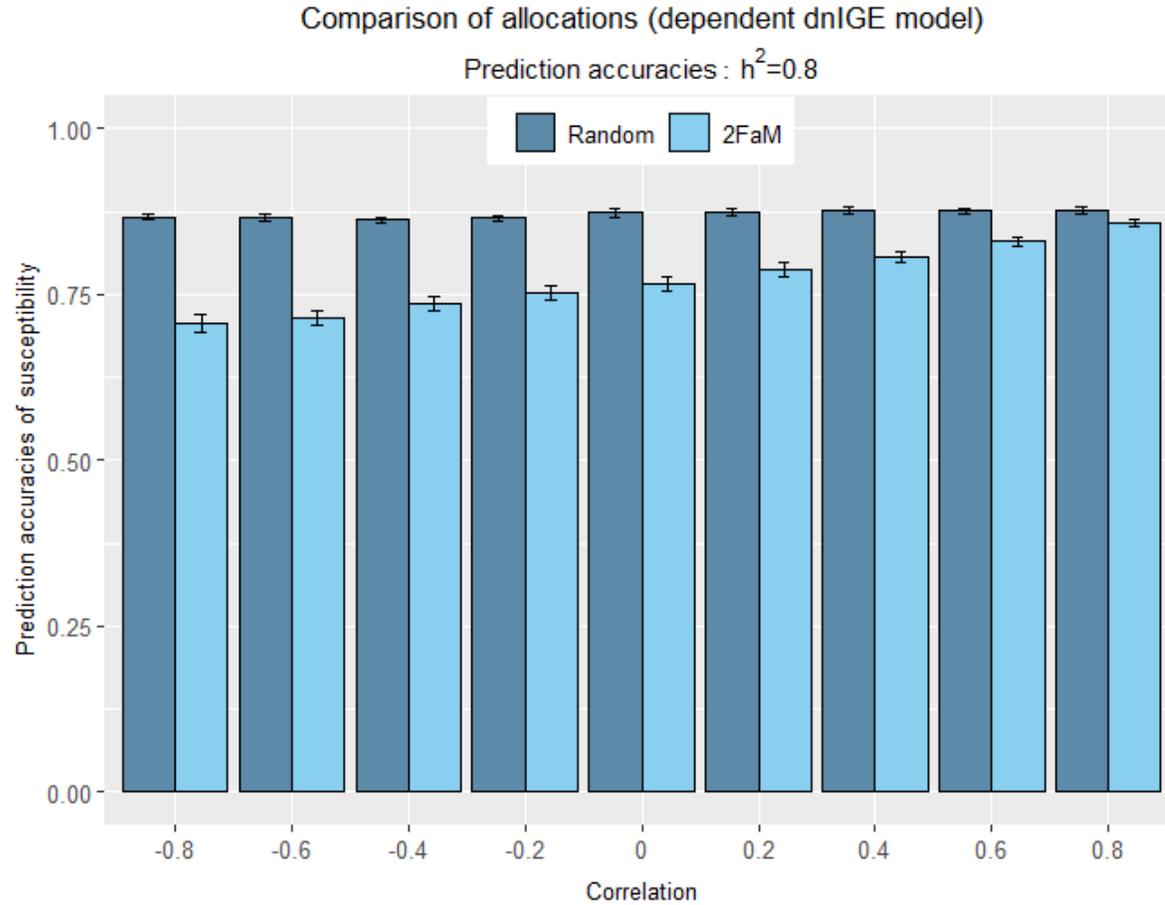
dnIGE estendido para alocação familiar

Acurácias de predição: $h^2=0.8$ 🖱️



- Correlações fortes **positivamente** aprimoram a estimativa de efeitos genéticos de ambas as características.
- Correlações fortes **negativamente** reduzem a acurácia preditiva de **susceptibilidade** e pouco melhoram a da **infectividade**.

Efeito da correlação na predição



Aplicação em dados reais

Seria realizada em dados de peixes considerando uma doença cutânea, de um experimento conduzido na Espanha. Mas novas regras de segurança da informação proibiram a divulgação do conjunto de dados.

A obtenção de outros dados reais para estimar os efeitos genéticos na susceptibilidade e infectividade requer experimentos em larga escala. Por isso, nos limitamos aqui a trazer conclusões em dados simulados.

Conclusões e estudos futuros

O modelo proposto prediz de maneira adequada os efeitos da susceptibilidade e infectividade, assim como a correlação entre eles;

Logo, pode ser uma poderosa ferramenta em estudos de melhoramento genético para identificar os indivíduos menos susceptíveis e menos infecciosos a uma determinada doença;

A alocação aleatória dos indivíduos nos grupos fechados parece ser mais indicada na previsão de valores de susceptibilidade, enquanto a alocação em família prevê melhor os valores de infectividade quando a correlação entre essas duas características é nula ou positiva;

O modelo pode ser aplicado a doenças que permitem recuperação ou infecções recorrentes, estendendo-o ao modelo compartimental SIR;

Se for sabido que a infecção ocorreu em um determinado intervalo de tempo, é possível prever em que ponto do intervalo cada indivíduo fica infectado.

A metodologia proposta pode ser estendida para estudar características complexas em populações humanas, incluindo estrutura de comunidades ou redes de contato entre os indivíduos.

Muito obrigada!

Contato: dzuanetti@ufscar.br