Introdução a linguagem Stan (*rstan*), um software para modelos bayesianos.

Universidade Federal do Oeste do Pará (UFOPA)
Campus de Monte Alegre – Engenharia de Aquicultura

Universidade Federal de Lavras (UFLA)
PPG – Estatística e Experimentação Agropecuária

Professor: Carlos Antônio Zarzar
E-mail: carloszarzar_@hotmail.com
carlos.zarzar@ufopa.edu.br

Data: 09/03/2022

Agradecimento e Colaboradores:

UNIVERSIDADE FEDERAL DE LAVRAS

NÚCLEO DE ESTUDOS EM REGRESSÃO NÃO LINEAR APLICADA

UFOPA

➢ Sumário

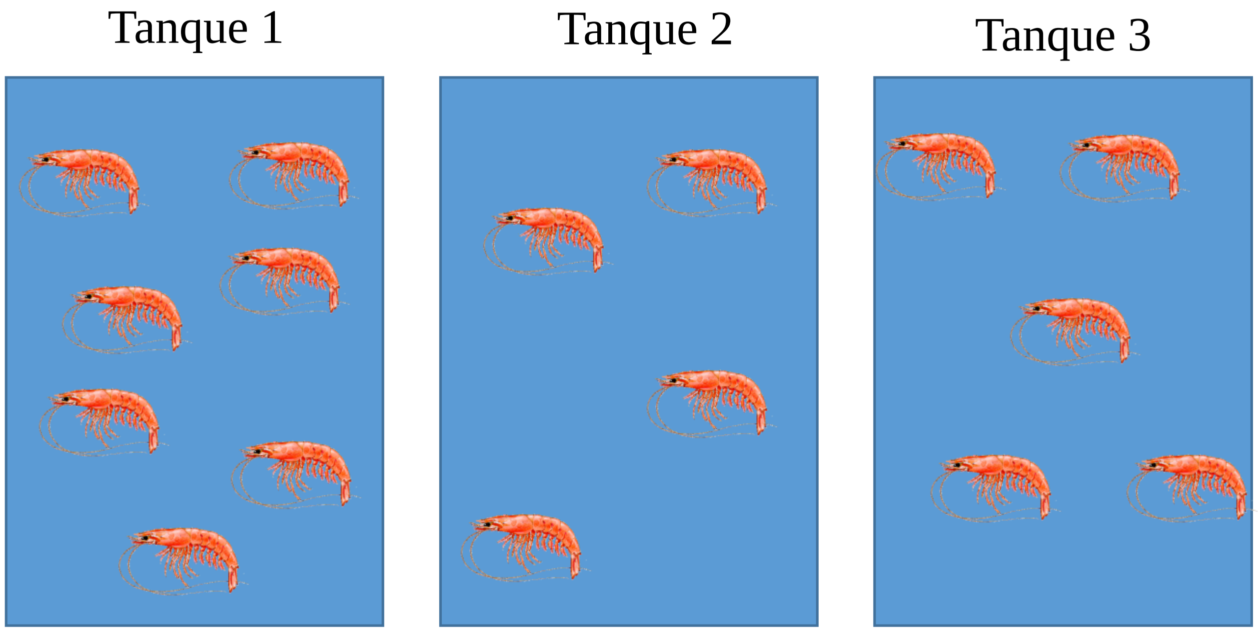
- Estatística Bayesiana;
- **Modelos Hierárquicos;**
- Método computacional de integração MCMC;
- Algoritmo de amostragem;
  - Gibbs;
  - Metropolis-Hasting;
  - Hamiltoniano Monte Carlo;
  - NUTS;

Modelos Hierárquicos

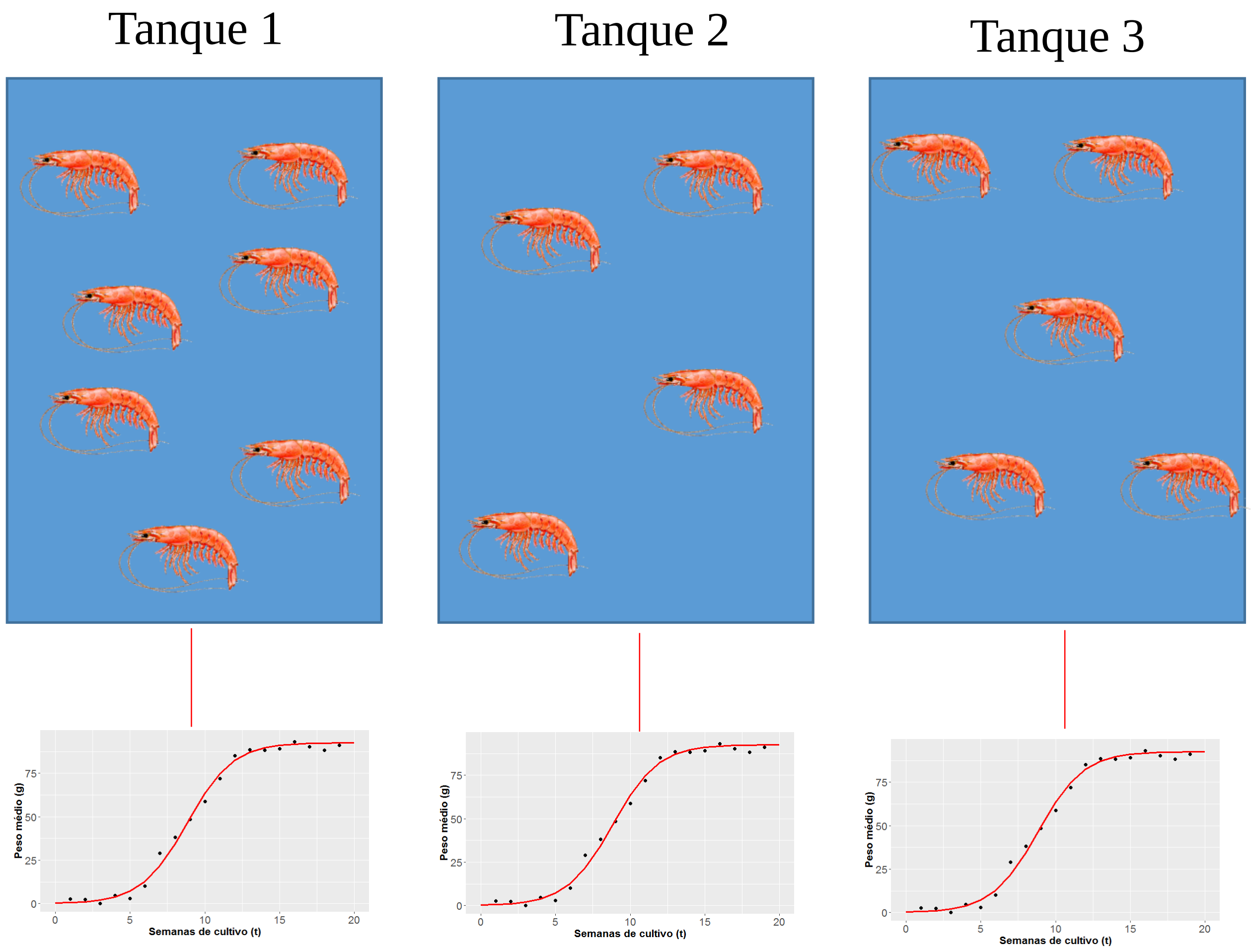
- Ilustração dos modelos Hierárquicos;
- Conceito e definição;
- Modelos Hierárquicos Bayesianos.

- Várias nomes representam os Modelos Hierárquicos ou casos especiais dele:
  - Modelos hierárquicos;
  - Modelo multinível;
  - Modelo de efeitos aleatórios;
  - Modelo de efeitos mistos;
  - Bayes empírico;
  - Regressão regularizada / penalizada / reduzida;
  - Determinação automática de relevância (ARD) em processos gaussianos (GP);
  - Adaptação de domínio;
  - Modelo de componentes de variância;
  - Modelo transversal;
  - Modelo de dados aninhados, design de gráfico dividido, coeficiente aleatório.

- <u>Modelo Hierárquico:</u>

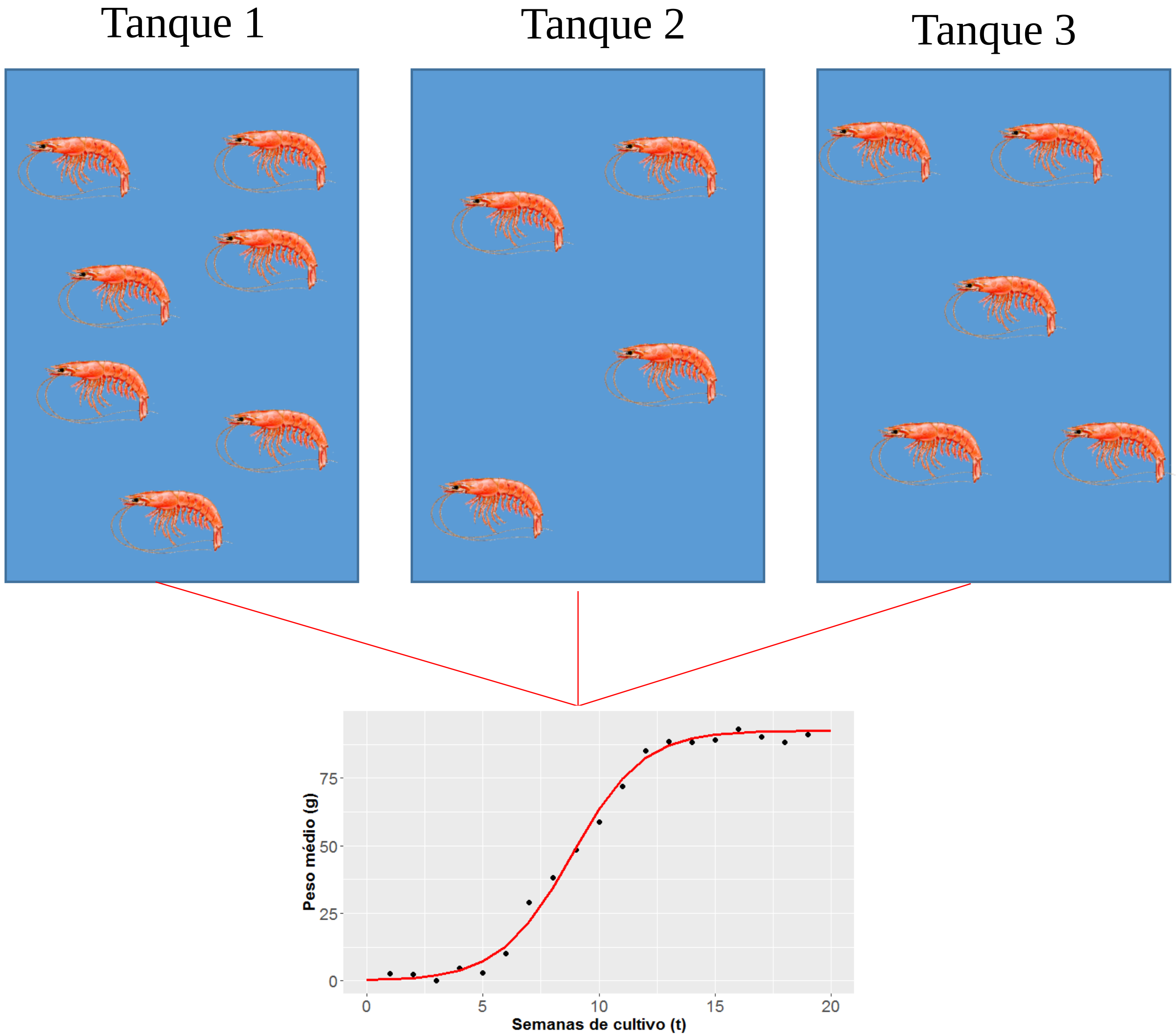


- <u>Modelo Hierárquico:</u>



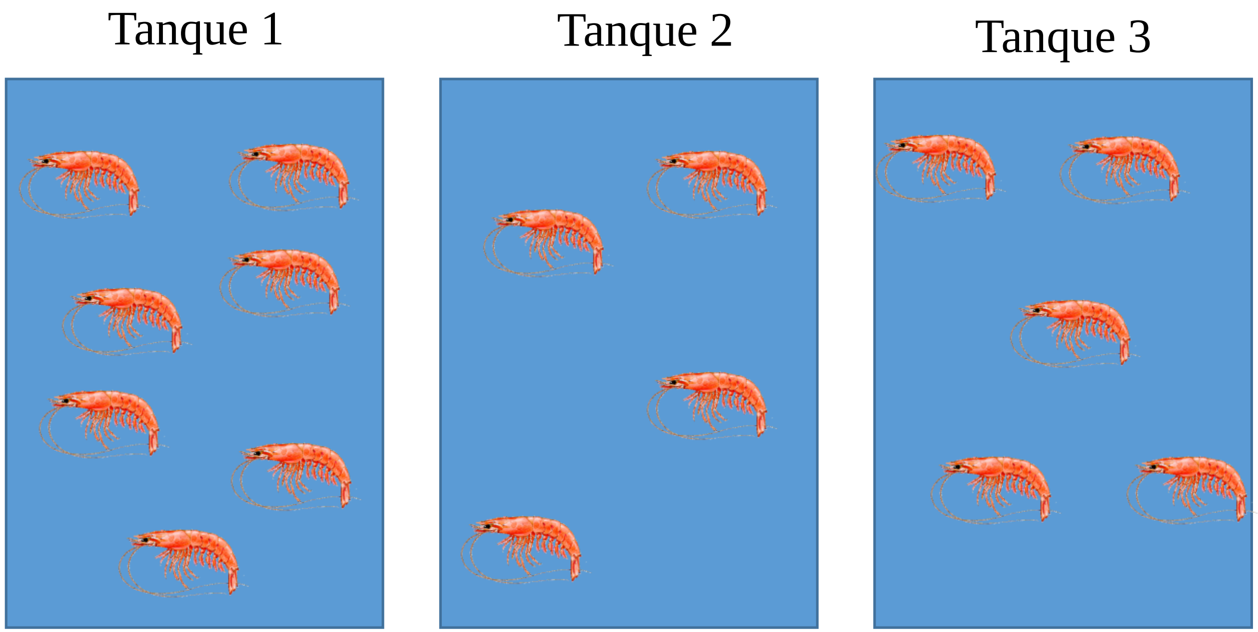
Modelos Hierárquicos

- <u>Modelo Hierárquico:</u>



Tanque 1 Tanque 2 Tanque 3

- <u>Modelo Hierárquico:</u>



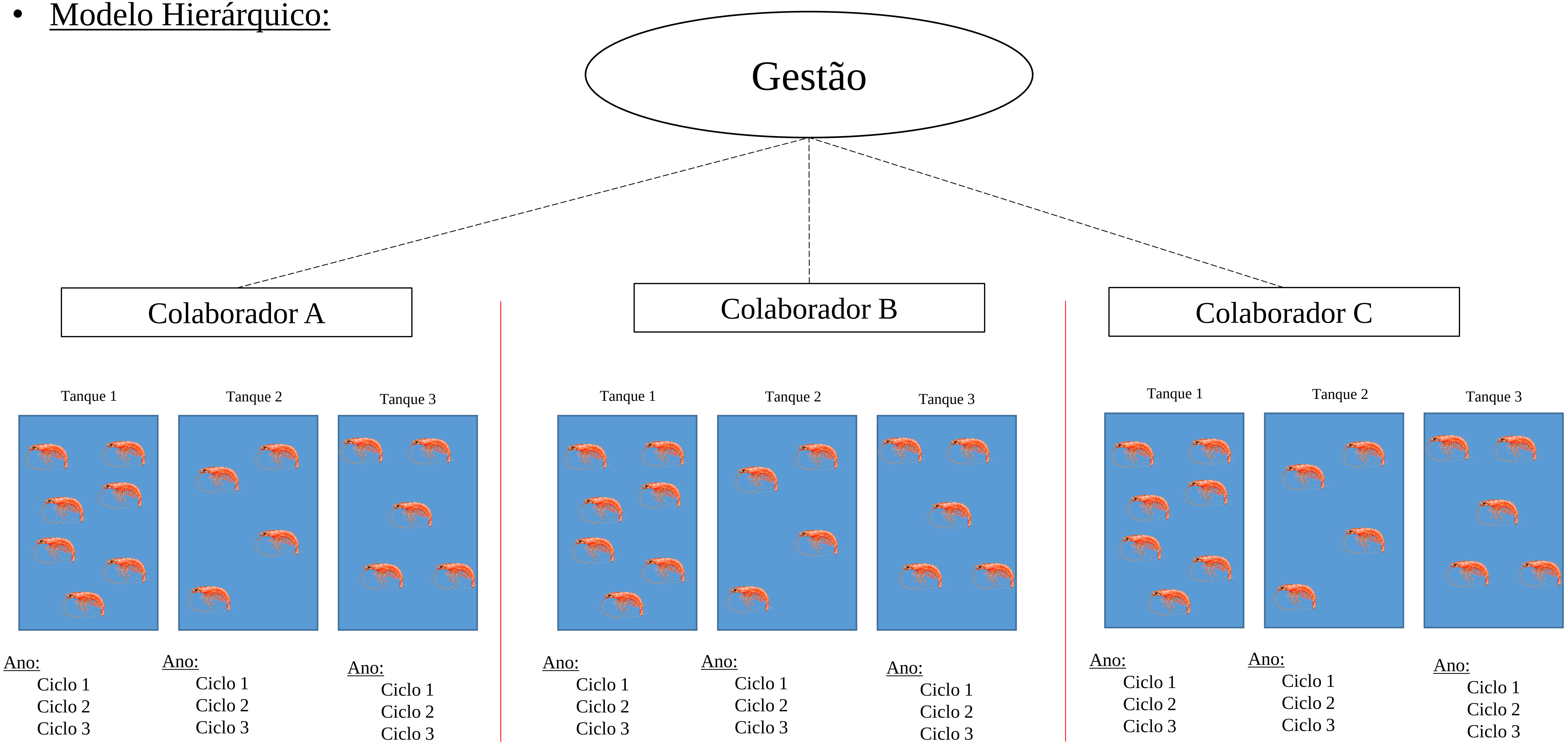
Tanque 1 Tanque 2 Tanque 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

- <u>Modelo Hierárquico:</u>



Gestão

Colaborador A

Tanque 1
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 2
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 3
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Colaborador B

Tanque 1
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 2
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 3
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

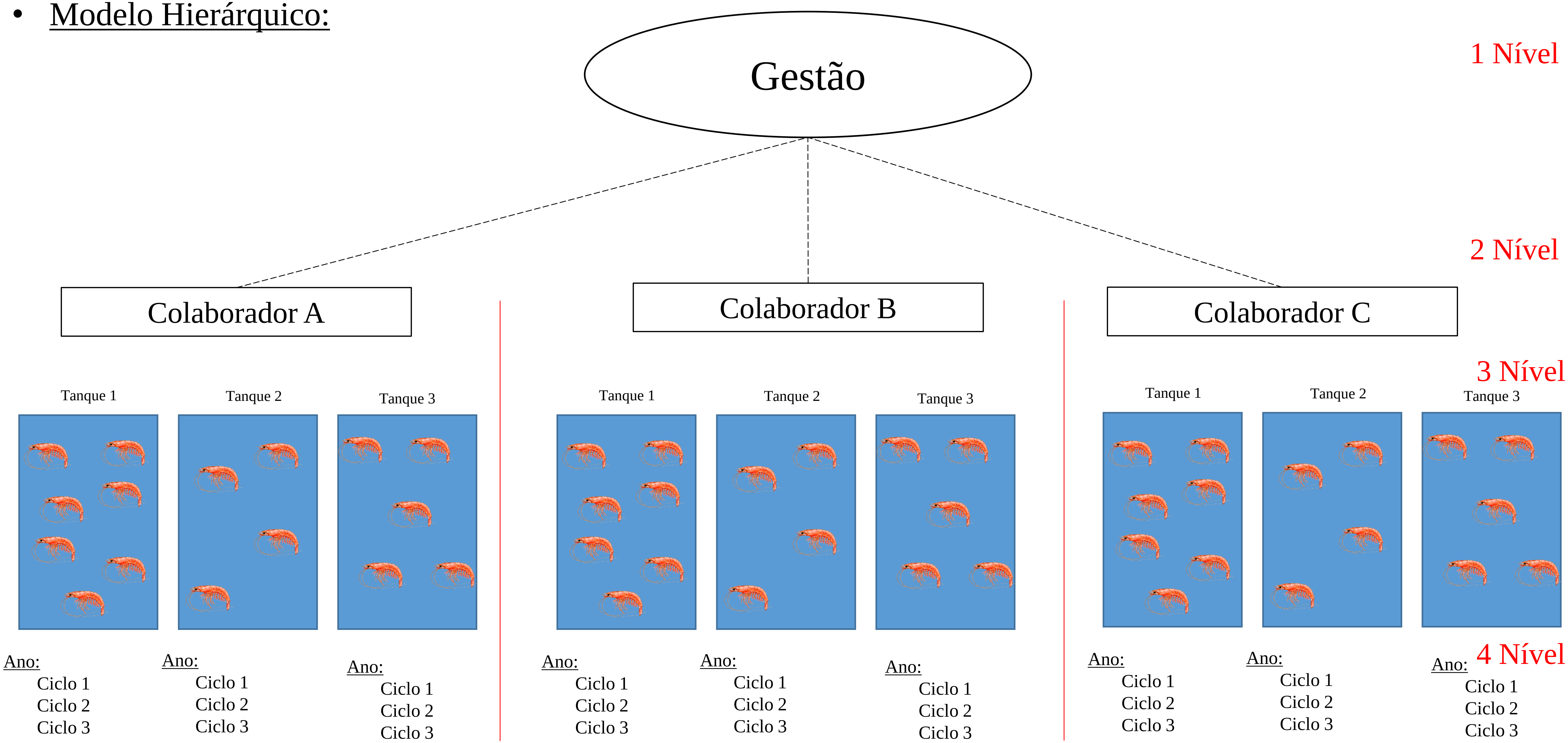
Colaborador C

Tanque 1
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 2
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

Tanque 3
Ano:
Ciclo 1
Ciclo 2
Ciclo 3

- <u>Modelo Hierárquico:</u>



Gestão

1 Nível

2 Nível

Colaborador A

Colaborador B

Colaborador C

3 Nível

Tanque 1
Tanque 2
Tanque 3

Tanque 1
Tanque 2
Tanque 3

Tanque 1
Tanque 2
Tanque 3

4 Nível

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

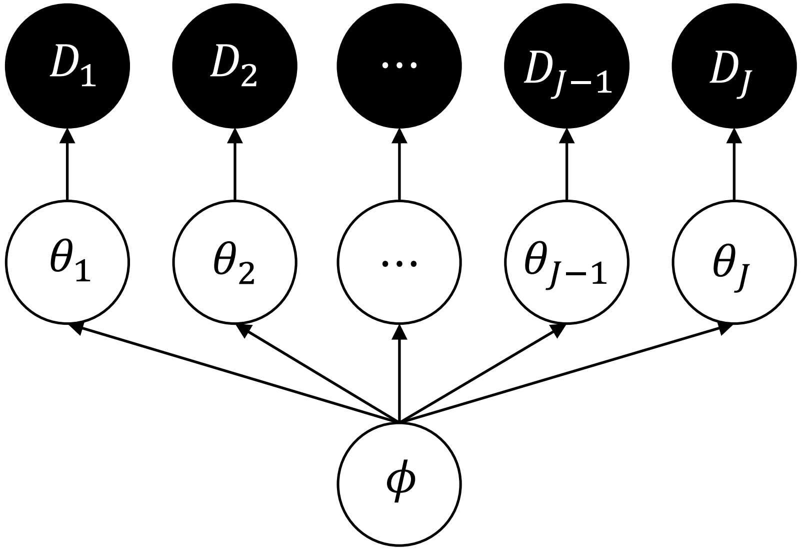
<u>Ano:</u>
Ciclo 1
Ciclo 2
Ciclo 3

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \textbf{para todo } i = 1, \ldots, n_j$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \textbf{para todo } j = 1, \ldots, J$$



- Modelo Hierárquica

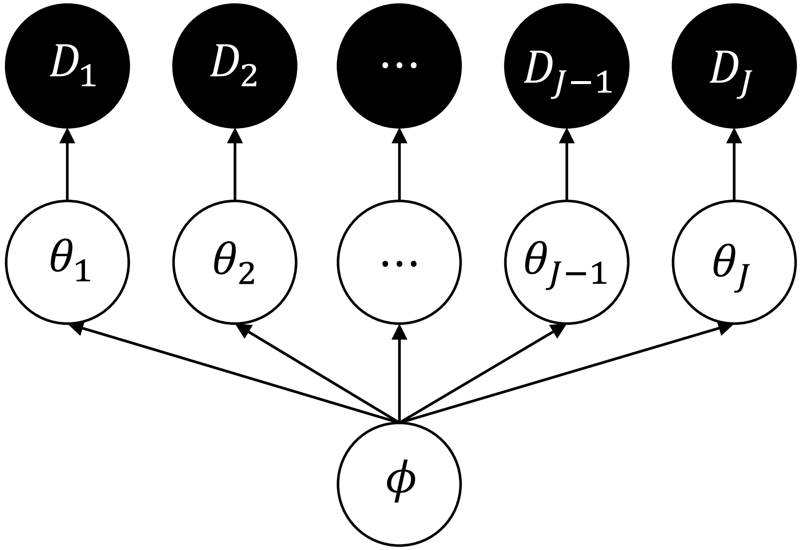
J grupos diferentes e $n_1,\ldots,n_J$ observações

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1,\ldots,n_j$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1,\ldots,J$$

$$Y_{11},\ldots,Y_{n_1 1},\ldots,Y_{1J},\ldots,Y_{n_J J} \perp\!\!\!\perp \mid \theta$$

$$\theta_1,\ldots,\theta_J \perp\!\!\!\perp \mid \phi,$$



- Modelo Hierárquica

J grupos diferentes e $n_1,\ldots,n_J$ observações

$$Y_{ij} \mid \theta_j \sim \boxed{p(y_{ij} \mid \theta_j)} \text{ para todo } i = 1,\ldots,n_j$$

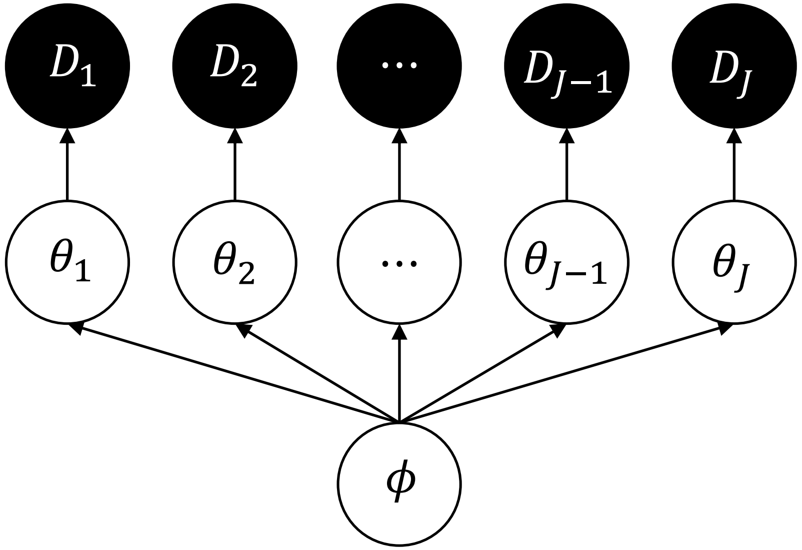
$$\theta_j \mid \phi \sim \boxed{p(\theta_j \mid \phi)} \text{ para todo } j = 1,\ldots,J$$

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$Y_{11},\ldots,Y_{n_1 1},\ldots,Y_{1J},\ldots,Y_{n_J J} \perp\!\!\!\perp \mid \theta$$

$$\theta_1,\ldots,\theta_J \perp\!\!\!\perp \mid \phi,$$



- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

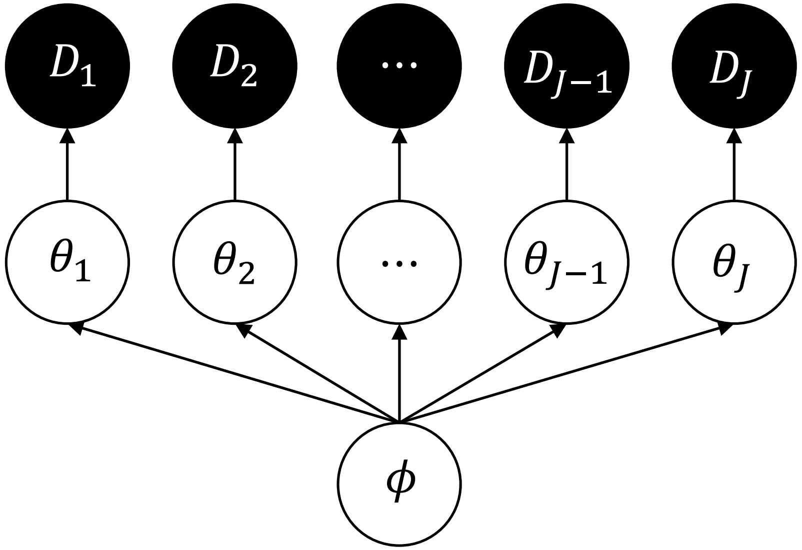
$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \textbf{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \textbf{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$



- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

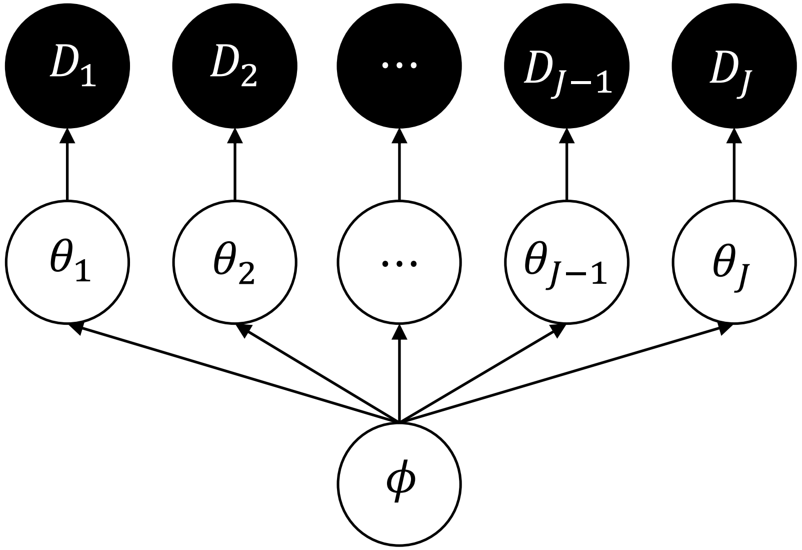
$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \textbf{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \textbf{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

1. Fixa-los a alguns valores constantes;



- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

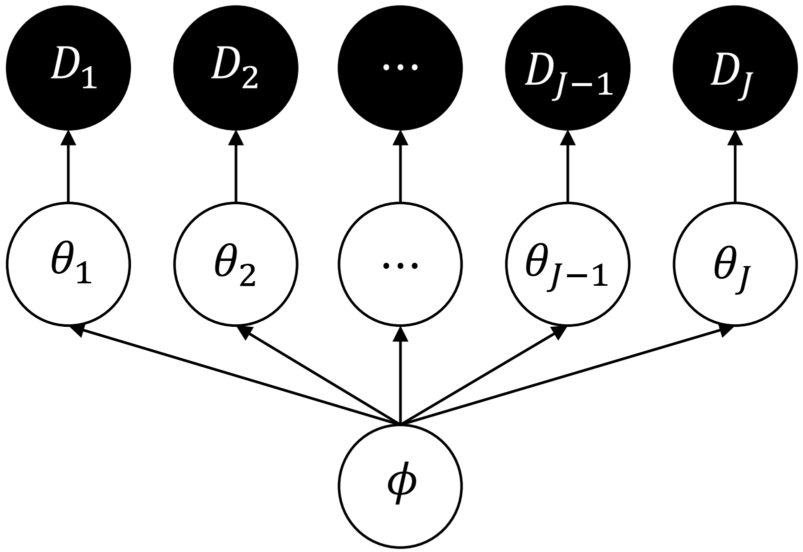
$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j)$**para todo** $i = 1, \ldots, n_j$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$\theta_j \mid \phi \sim p(\theta_j \mid \phi)$**para todo** $j = 1, \ldots, J$

$\phi \sim p(\phi)$

1. Fixa-los a alguns valores constantes;
2. Usar estimativas pontuais estimadas a partir dos dados;



- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

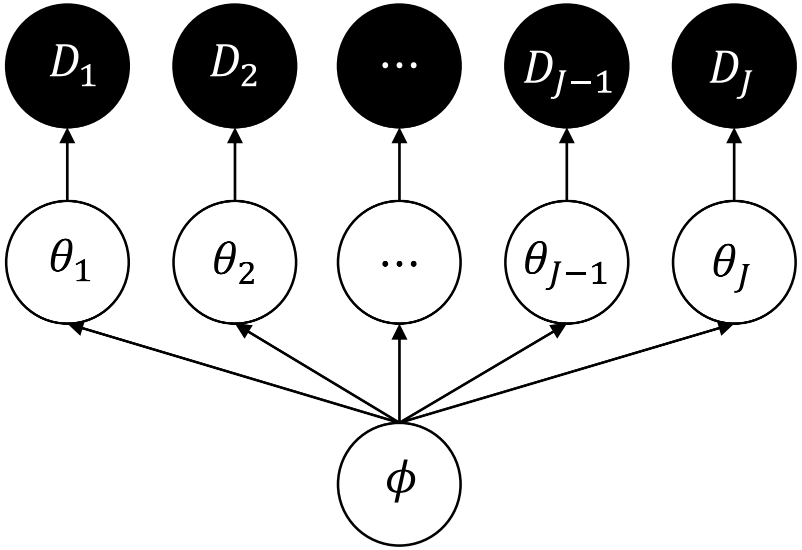
$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \textbf{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \textbf{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

1. Fixa-los a alguns valores constantes;
2. Usar estimativas pontuais estimadas a partir dos dados;
3. Definir uma distribuição de probabilidade sobre eles.



- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j)\textbf{para todo } i = 1, \ldots, n_j$$

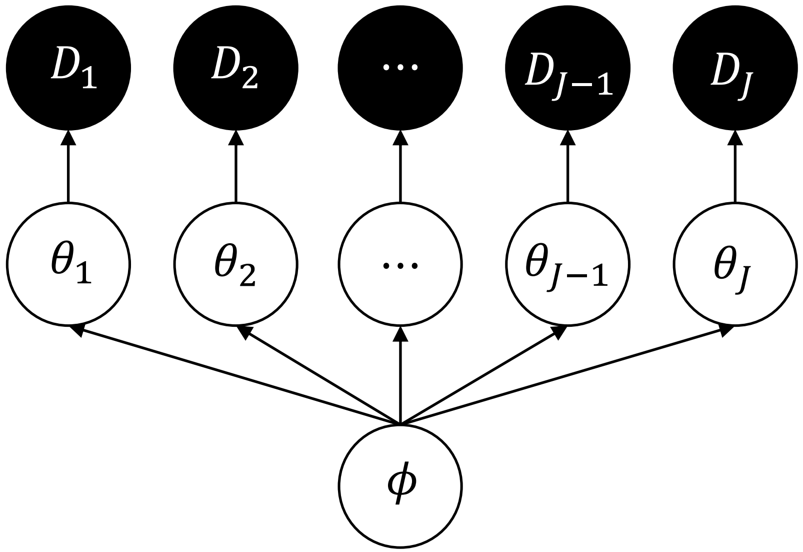
$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi)\textbf{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

1. Fixa-los a alguns valores constantes;
2. Usar estimativas pontuais estimadas a partir dos dados;
3. Definir uma distribuição de probabilidade sobre eles.

Modelo Hierárquico
totalmente Bayesiano



Modelos Hierárquicos

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

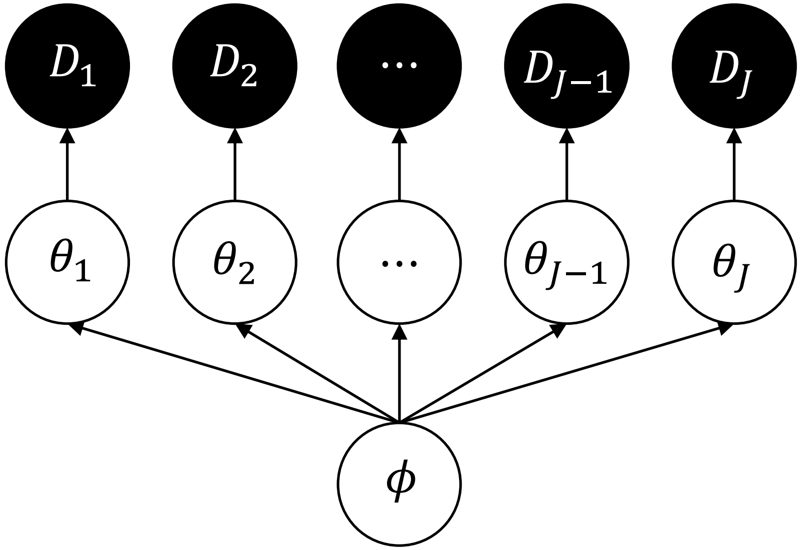
$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$



Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$\boxed{\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta}$$

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

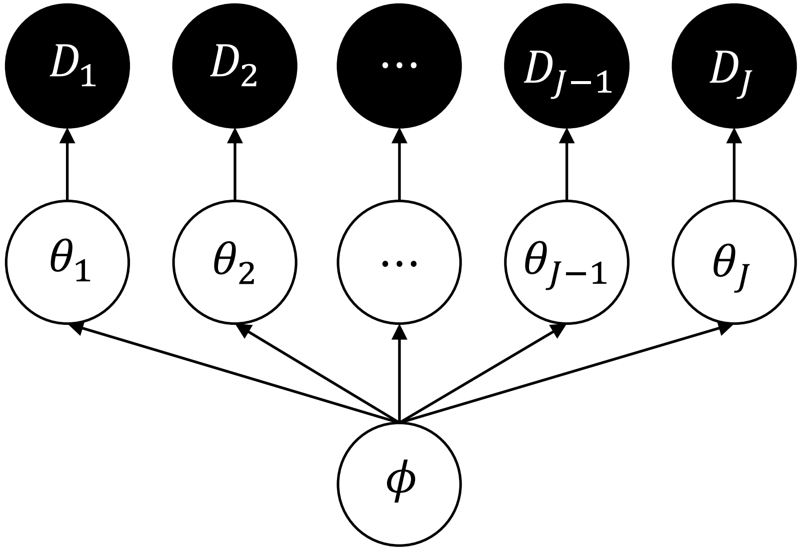
$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$



Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow \boxed{p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),}$$

Modelos Hierárquicos

- Modelo Hierárquica

J grupos diferentes e $n_1, \dots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j)\text{para todo } i = 1, \dots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi)\text{para todo } j = 1, \dots, J$$

$$\phi \sim p(\phi)$$

Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto \; p(\mathbf{y}|\boldsymbol{\theta}, \phi) \; p(\boldsymbol{\theta}, \phi)$$

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),$$

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

Distribuição condicional

Verossimilhança

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, |\mathbf{y}) \propto \boxed{p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\phi})}\ p(\boldsymbol{\theta}, \boldsymbol{\phi})$$

$$p(\boldsymbol{\theta}, \boldsymbol{\phi}, |\mathbf{y}) \propto \underline{p(\mathbf{y}|\boldsymbol{\theta})}\ p(\boldsymbol{\theta}|\boldsymbol{\phi})\ p(\boldsymbol{\phi})$$

Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow \underline{p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),}$$

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

Distribuição conjunta

A priori

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto \ p(\mathbf{y}|\boldsymbol{\theta}, \phi) \boxed{p(\boldsymbol{\theta}, \phi)}$$

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) \ \underline{p(\boldsymbol{\theta}|\phi) \ p(\phi)}$$

Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),$$

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j) \text{para todo } i = 1, \ldots, n_j$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi) \text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),$$

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \phi)\ p(\boldsymbol{\theta}, \phi)$$

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto \underline{p(\mathbf{y}|\boldsymbol{\theta})}\ \underline{p(\boldsymbol{\theta}|\phi)}\ p(\phi)$$

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto p(\phi) \prod_{j=1}^{J} \underline{p(\mathbf{y}_j|\boldsymbol{\theta}_j)\ p(\boldsymbol{\theta}_j|\phi).}$$

Modelos Hierárquicos

- Modelo Hierárquica

J grupos diferentes e $n_1, \ldots, n_J$ observações

$$p(\mathbf{y}_j \mid \theta_j) = \prod_{i=1}^{n_j} p(y_{ij} \mid \theta_j).$$

$$Y_{ij} \mid \theta_j \sim p(y_{ij} \mid \theta_j)\text{para todo } i = 1, \ldots, n_j$$

$$p(\theta \mid \phi) = \prod_{j=1}^{J} p(\theta_j \mid \phi).$$

$$\theta_j \mid \phi \sim p(\theta_j \mid \phi)\text{para todo } j = 1, \ldots, J$$

$$\phi \sim p(\phi)$$

Dados dependem do hiperparâmetro apenas através dos parâmetros a nível do seu grupo.

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto \; p(\mathbf{y}|\boldsymbol{\theta}, \phi) \; p(\boldsymbol{\theta}, \phi)$$

$$\mathbf{Y} \perp\!\!\!\perp \phi \mid \theta \longrightarrow p(\mathbf{y} \mid \theta, \phi) = p(\mathbf{y} \mid \theta),$$

$$p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) \; p(\boldsymbol{\theta}|\phi) \; p(\phi)$$

$$\boxed{p(\boldsymbol{\theta}, \phi, |\mathbf{y}) \propto p(\phi) \prod_{j=1}^{J} p(\mathbf{y}_j|\boldsymbol{\theta}_j) \; p(\boldsymbol{\theta}_j|\phi).} \longrightarrow$$

A posteriori

Modelos Hierárquicos Bayesianos;

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int p(\boldsymbol{\theta},\phi|\mathbf{y})\ \mathrm{d}\phi = \int p(\boldsymbol{\theta}|\phi,\mathbf{y})\ p(\phi|\mathbf{y})\ \mathrm{d}\phi$$

Distribuição marginal
a posteriori dos $\boldsymbol{\theta}$
a nível de grupo

$$p(\boldsymbol{\theta},\phi,|\mathbf{y}) \propto\ p(\mathbf{y}|\boldsymbol{\theta},\phi)\ p(\boldsymbol{\theta},\phi)$$

$$p(\boldsymbol{\theta},\phi,|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta})\ p(\boldsymbol{\theta}|\phi)\ p(\phi)$$

$$\boxed{p(\boldsymbol{\theta},\phi,|\mathbf{y}) \propto p(\phi)\prod_{j=1}^{J} p(\mathbf{y}_j|\boldsymbol{\theta}_j)\ p(\boldsymbol{\theta}_j|\phi).}$$

A posteriori
Modelos Hierárquicos Bayesianos;

Introdução a linguagem Stan (*rstan*), um software para modelos bayesianos.

Universidade Federal do Oeste do Pará (UFOPA)
Campus de Monte Alegre – Engenharia de Aquicultura

Obrigado !!!

Professor: Carlos Antônio Zarzar
E-mail: carloszarzar_@hotmail.com
carlos.zarzar@ufopa.edu.br

Data: 09/03/2022

AGRADECIMENTO E COLABORADORES:

UNIVERSIDADE FEDERAL DE LAVRAS

NÚCLEO DE ESTUDOS EM REGRESSÃO NÃO LINEAR APLICADA

UFOPA