

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

AGRADECIMENTO E COLABORADORES:



UFOPA

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

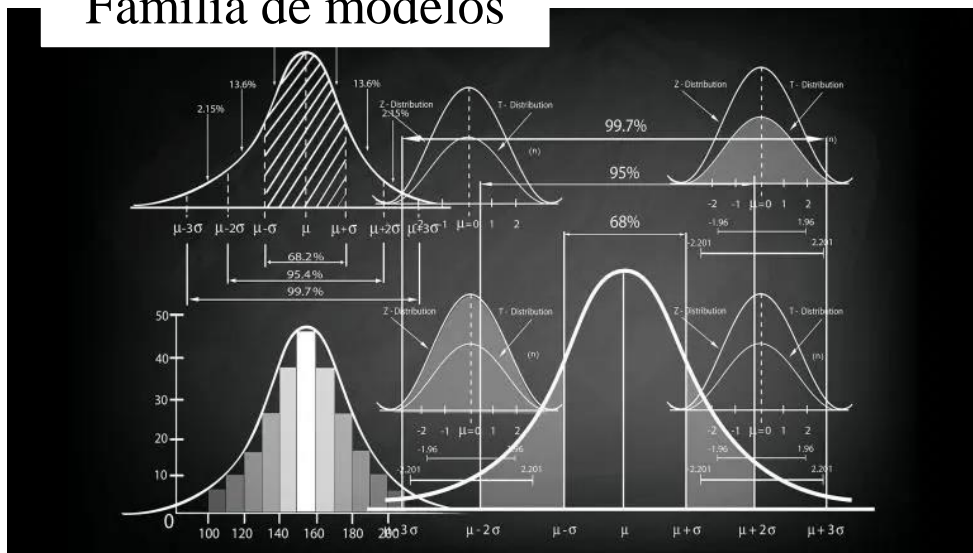
Data: 10/03/2022

Seleção e Diagnóstico de Modelos Bayesianos



- Contextualização modelos;
- Índices seleção modelos Bayesianos;
- Validação Cruzada (CV);
- Ilustrando CV Bayesianos;
- PSIS - Pareto smoothed importance sampling;
- Diagnóstico de Modelos Bayesianos;
- Exemplo;

Família de modelos



<https://www.cienciaedados.com/probabilidade-e-estatistica-os-fundamentos-para-cientistas-de-dados-parte-1/>

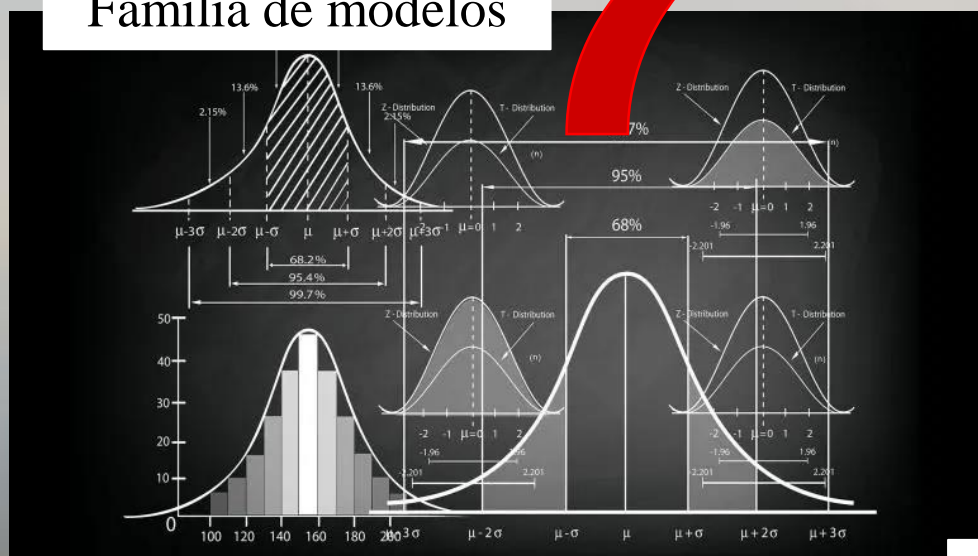
“Essencialmente, todos os modelos estão errados, mas alguns são úteis”

George Box (1976)

Journal of the American Statistical Association



Família de modelos



<https://www.cienciaedados.com/probabilidade-e-estatistica-os-fundamentos-para-cientistas-de-dados-parte-1/>



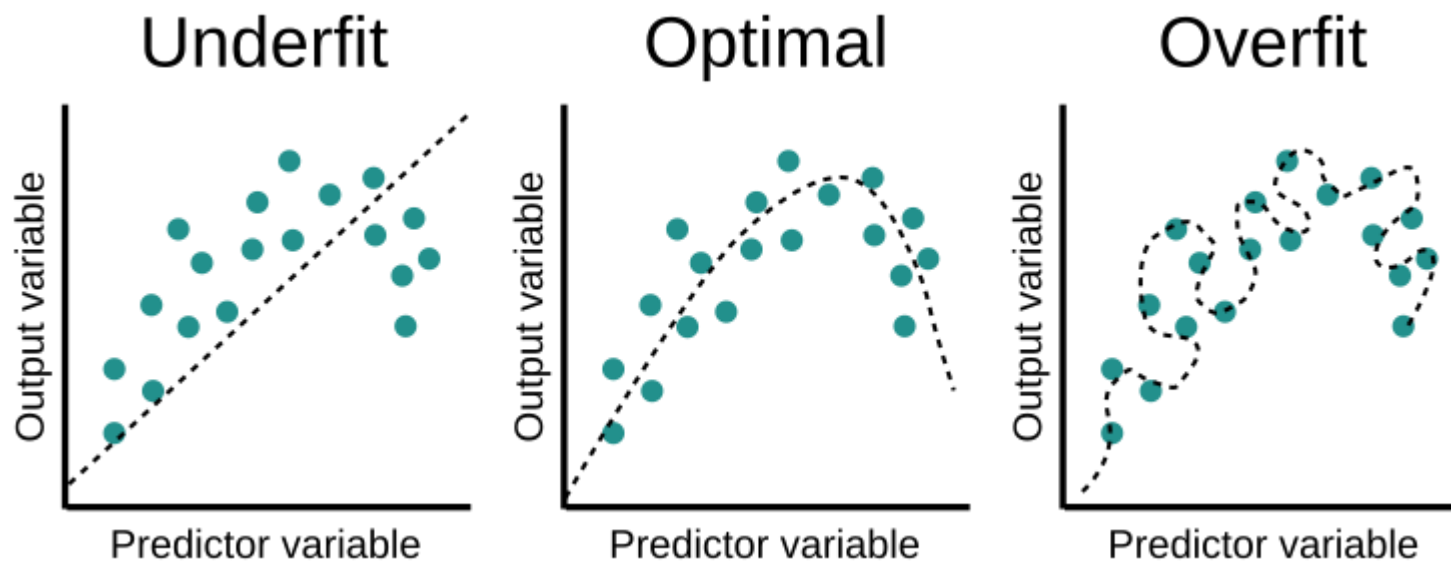
Vencedor
Ou um conjunto de
modelos vencedores

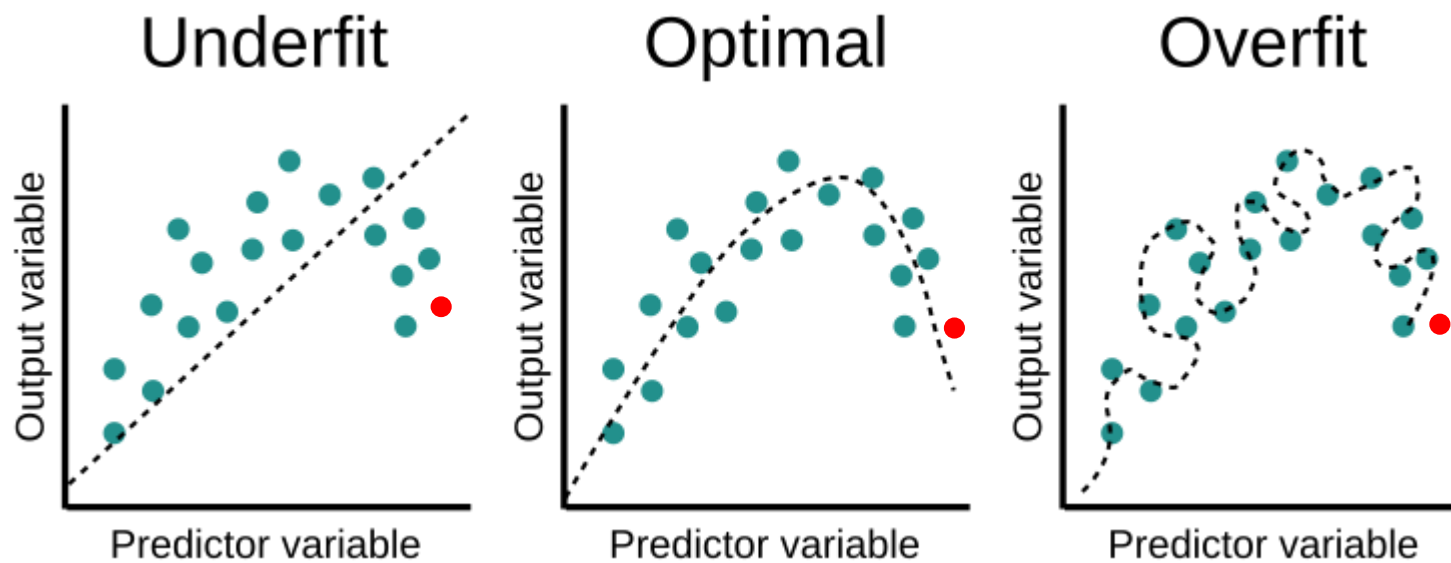
“Essencialmente, todos os
modelos estão errados, mas
alguns são úteis”

George Box (1976)

Journal of the American Statistical Association







Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC);
- Deviance information criterion (DIC);
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC);
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC);
- Deviance information criterion (DIC);
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC);
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

$$\text{AIC} = -2 \log p(y|\hat{\theta}_{\text{mle}}) + 2k$$

Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC); \longrightarrow
- Deviance information criterion (DIC);
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC);
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

$$\text{BIC} = -2 \log p(y|\hat{\theta}) + k \log n$$

Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC);
- Deviance information criterion (DIC); \longrightarrow
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC);
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

$$\text{DIC} = -2 \log p(y|\hat{\theta}_{\text{Bayes}}) + 2p_{\text{DIC}}.$$

$$p_{\text{DIC}} = 2 \left(\log p(y|\hat{\theta}_{\text{Bayes}}) - E_{\text{post}}(\log p(y|\theta)) \right)$$

Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC);
- Deviance information criterion (DIC);
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC); →
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

$$\widehat{\text{elppd}}_{\text{WAIC}} = \text{lppd} - p_{\text{WAIC}}.$$

$$p_{\text{WAIC}1} = 2 \sum_{i=1}^n \left(\log(E_{\text{post}} p(y_i | \theta)) - E_{\text{post}}(\log p(y_i | \theta)) \right)$$

Computada

$$p_{\text{WAIC}1} = 2 \sum_{i=1}^n \left(\log \left(\frac{1}{S} \sum_{s=1}^S p(y_i | \theta^s) \right) - \frac{1}{S} \sum_{s=1}^S \log p(y_i | \theta^s) \right)$$

$$p_{\text{WAIC}2} = \sum_{i=1}^n \text{var}_{\text{post}}(\log p(y_i | \theta))$$

Computada

$$p_{\text{WAIC}2} = \sum_{i=1}^n V_{s=1}^S (\log p(y_i | \theta^s))$$

Métodos de seleção de modelos com perspectiva Bayesiana:

- Akaike information criterion (AIC);
- Bayesian information criterion (BIC);
- Deviance information criterion (DIC);
- Teste de razão de verossimilhança
- Focused information criterion (FIC);
- Watanabe–Akaike information criterion (WAIC);
- Critério de informação de Fisher estendido (EFIC);
- *Mallows's Cp*;
- Fator de Bayes;
- Validação cruzada (CV);
- *Stepwise regression*;
 - R^2 ;
 - Mean Squared Error (MSE);
 - *Deviance*;

- Teorema de Bayes:

$$\text{a posteriori} \leftarrow \boxed{P(\theta | y)} = \frac{P(y | \theta) P(\theta)}{P(y)}$$

$\xleftarrow{\text{verossimilhança}}$ $\xrightarrow{\text{a priori}}$

$\xrightarrow{\text{Constante normalizadora}}$

- Inferência Bayesiana a posteriori;
- Verificação preditiva das posteriores do modelos;
- Ou seja, através de precisão de suas previsões (acurácia preditiva);
- Outros métodos como:
 - Verificação preditivas das distribuições a priori;
 - Verificações mistas para modelos hierárquicos (Gelman, Meng, Stern, 1996);
 - Método de expansão contínuo do modelo (Draper, 1999);

- Teorema de Bayes:

$$\begin{array}{c} \text{verossimilhança} \leftarrow \quad \quad \quad \rightarrow \text{a priori} \\ \text{a posteriori} \leftarrow \boxed{P(\theta | y)} = \frac{P(y | \theta) P(\theta)}{P(y)} \quad \rightarrow \text{Constante normalizadora} \end{array}$$

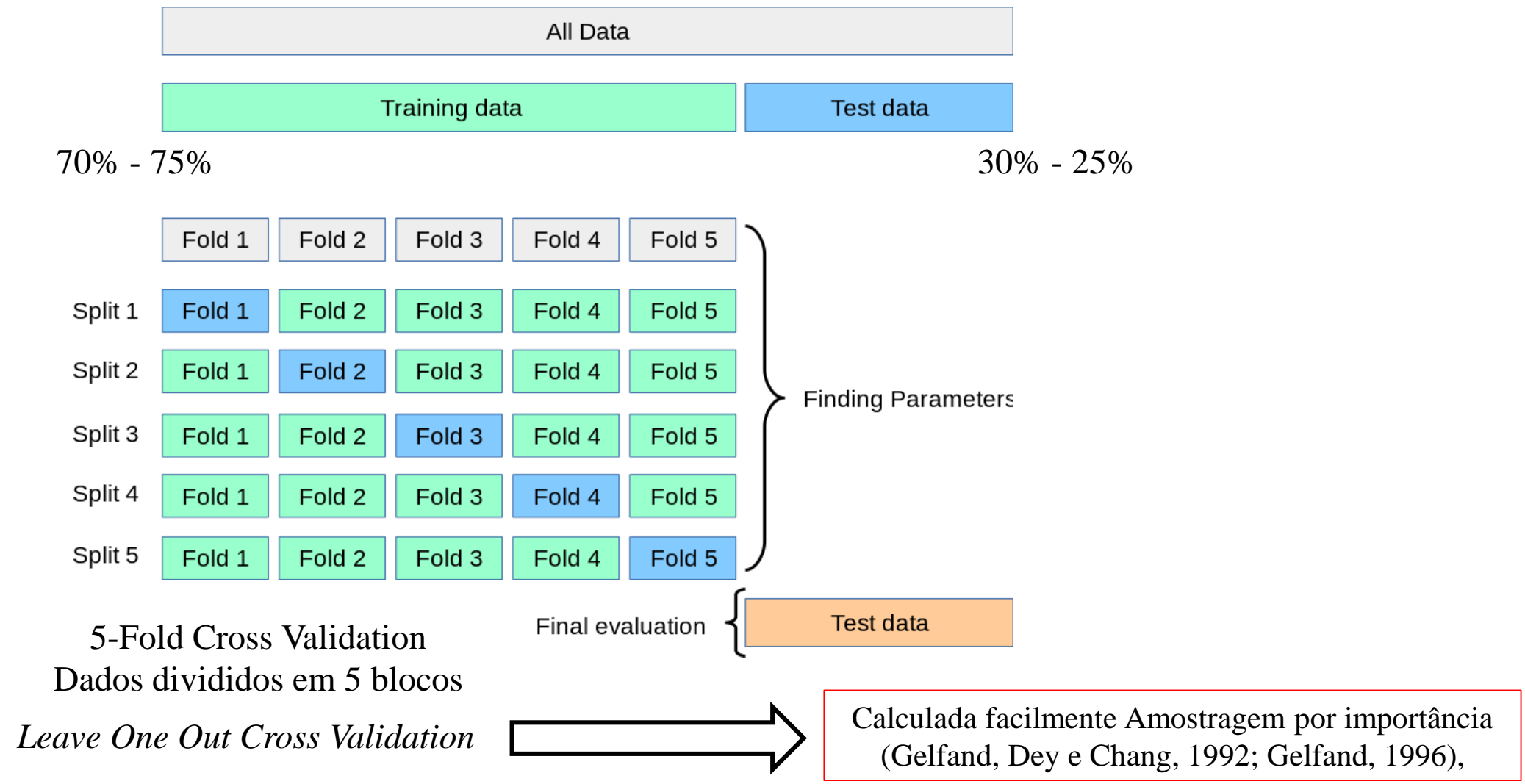
- Validação Cruzada (Cross Validation - CV);

- Inferência Bayesiana a posteriori;
- Verificação preditiva das posteriores do modelos;
- Ou seja, através de precisão de suas previsões (acurácia preditiva);

- Outros métodos como:
- Verificação preditivas das distribuições a priori;
- Verificações mistas para modelos hierárquicos (Gelman, Meng, Stern, 1996);
- Método de expansão contínuo do modelo (Draper, 1999);

Capacidade preditiva fora da amostra
utilizando ajuste dentro da amostra

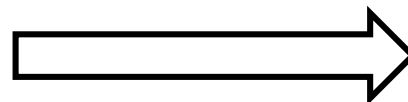
➤ Validação Cruzada (Cross Validation - CV);



➤ Validação Cruzada (Cross Validation - CV);

Importance Sampling: é uma método de aproximação que direciona amostragem (amostras importantes) de forma eficiente para estimar propriedades de uma distribuição alvo de interesse a partir de uma outra distribuição

5-Fold Cross Validation
Dados divididos em 5 blocos
Leave One Out Cross Validation



Calculada facilmente Amostragem por importância
(Gelfand, Dey e Chang, 1992; Gelfand, 1996),

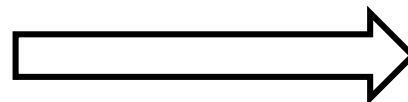


➤ Validação Cruzada (Cross Validation - CV);

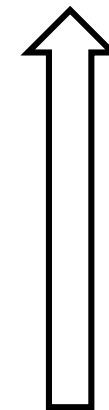
- A variância do estimador da amostragem por importância pode ser menor (ou maior) que a variância do estimador de Monte Carlo;
- O desafio é escolher essa função peso para conduzir a amostragem nas regiões de interesse;

Importance Sampling: é uma método de aproximação que direciona amostragem (amostras importantes) de forma eficiente para estimar propriedades de uma distribuição alvo de interesse a partir de uma outra distribuição

5-Fold Cross Validation
Dados divididos em 5 blocos
Leave One Out Cross Validation

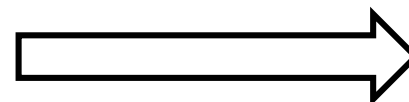


Calculada facilmente Amostragem por importância
(Gelfand, Dey e Chang, 1992; Gelfand, 1996),



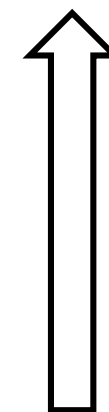
➤ Validação Cruzada (Cross Validation - CV);

5-Fold Cross Validation
Dados divididos em 5 blocos
Leave One Out Cross Validation



Calculada facilmente Amostragem por importância
(Gelfand, Dey e Chang, 1992; Gelfand, 1996),

Relacionado a LOO-CV
Problemas: fortes ruídos
Variância **Peso por importância**
grande ou infinita



➤ Validação Cruzada (Cross Validation - CV);

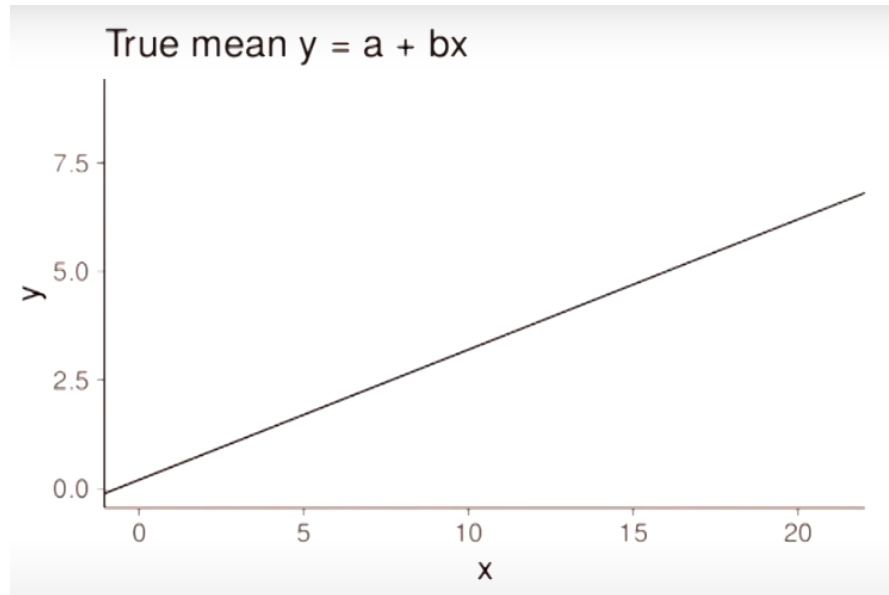
Vehtari, Gelman e Gabry (2017)
propuseram o uso da Amostragem por
Importância Suavizada de Pareto
(PSIS - Pareto smoothed importance sampling)

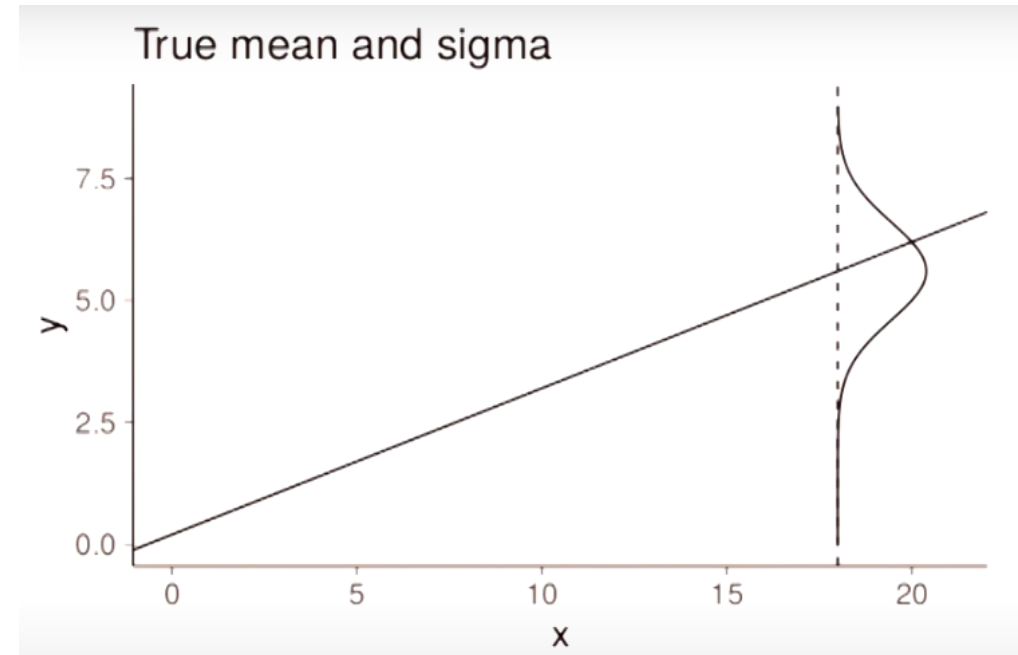
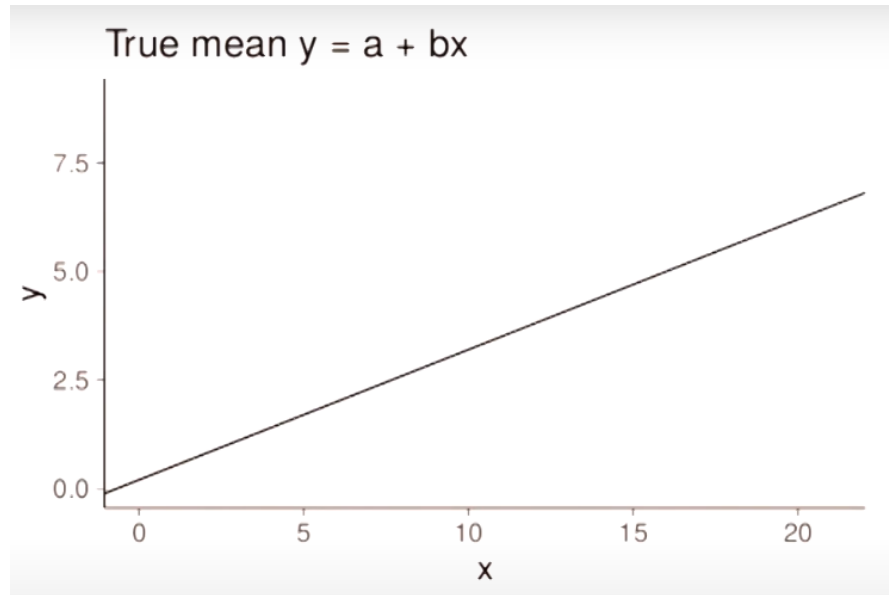
PSIS estabiliza a amostragem por importância

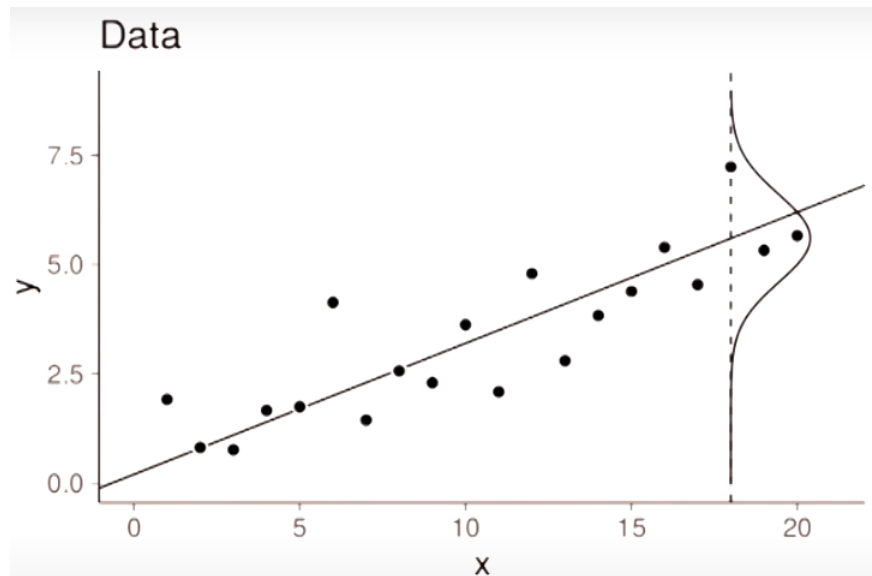
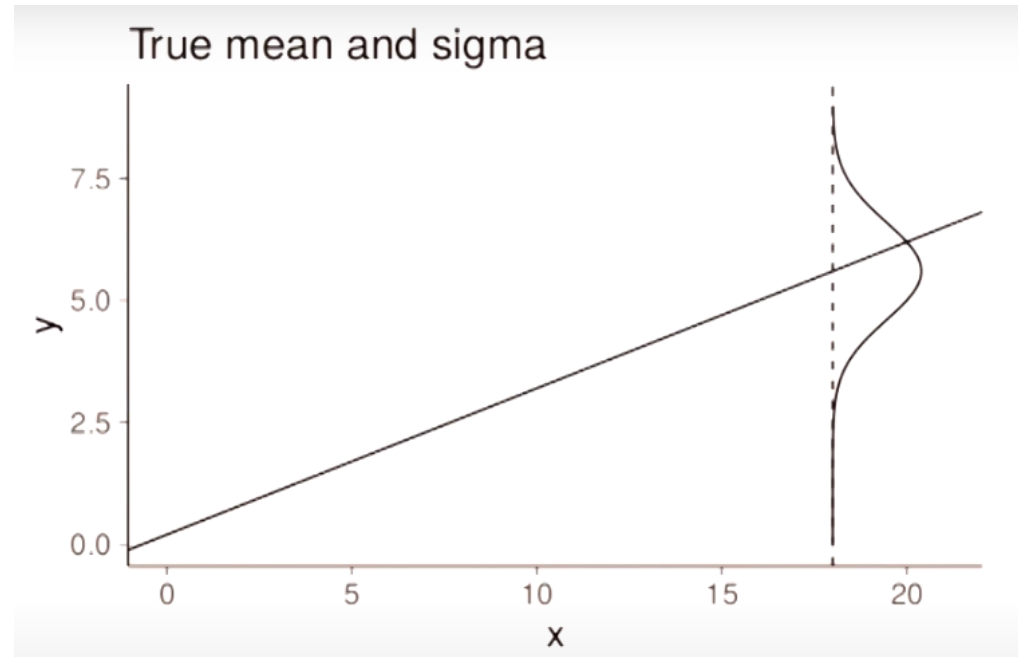
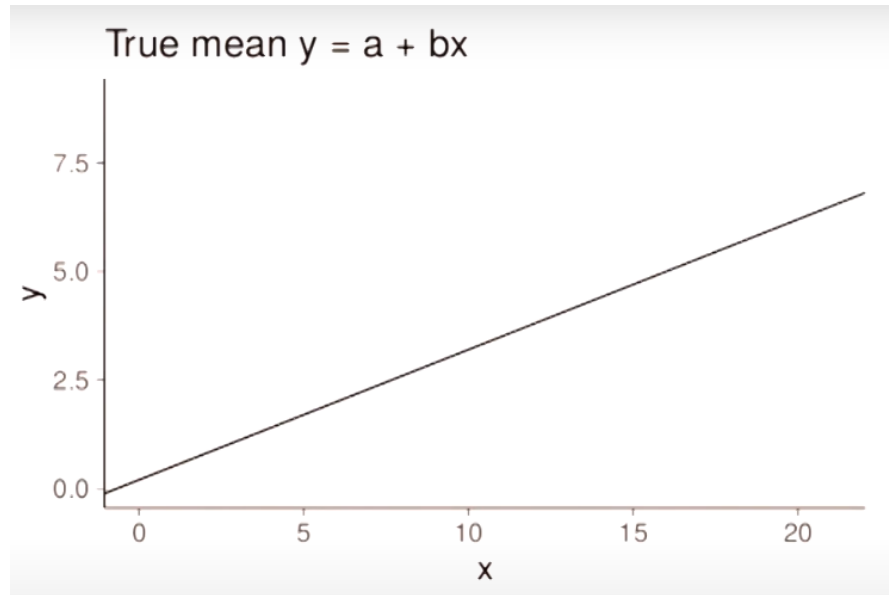
Relacionado a LOO-CV
Problemas: fortes ruídos
Variância **Peso por importância**
grande ou infinita

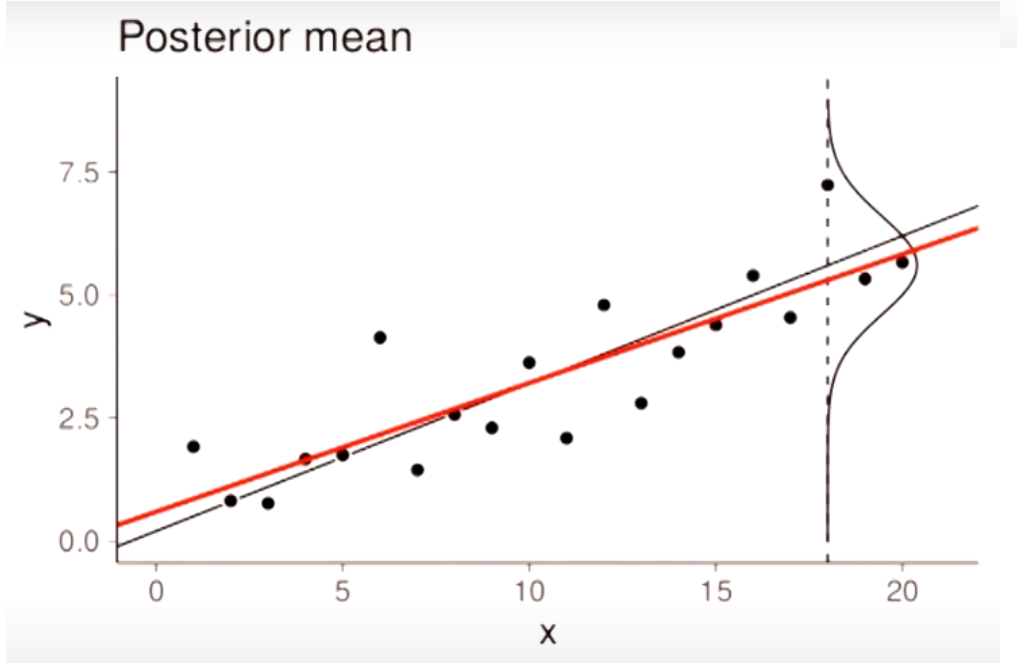
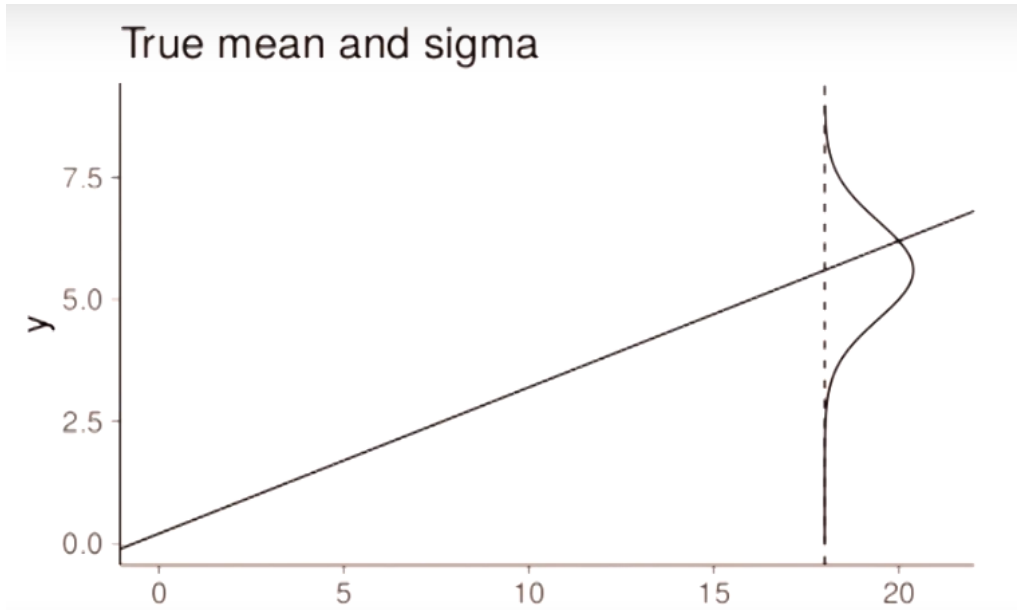
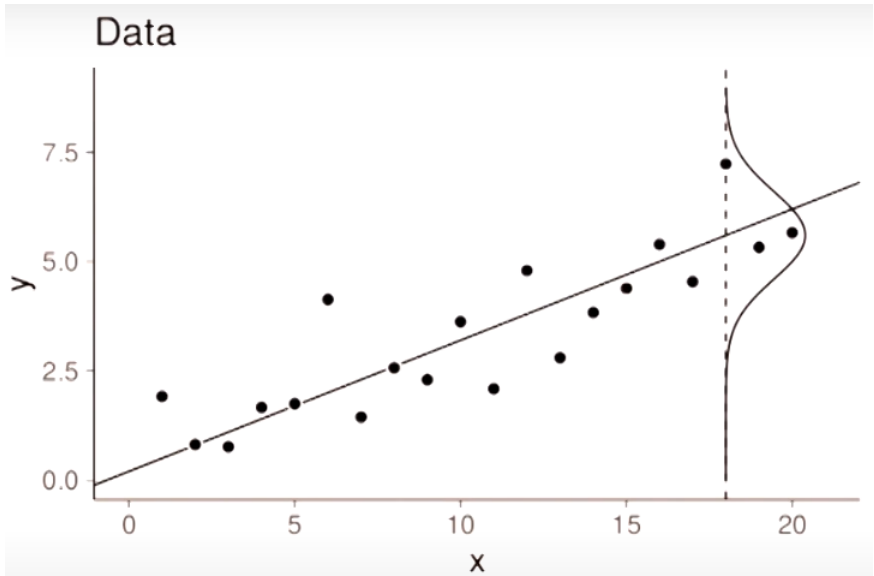
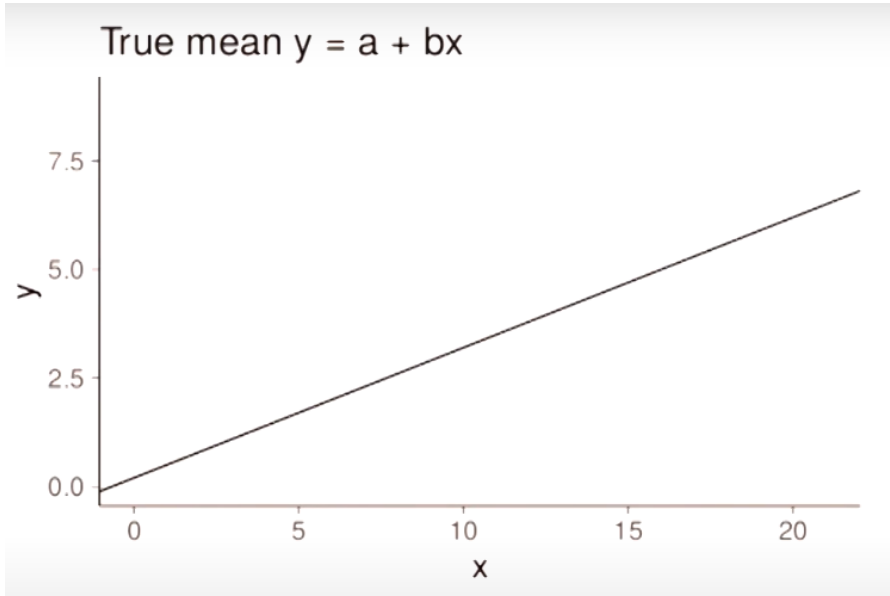
5-Fold Cross Validation
Dados divididos em 5 blocos
Leave One Out Cross Validation

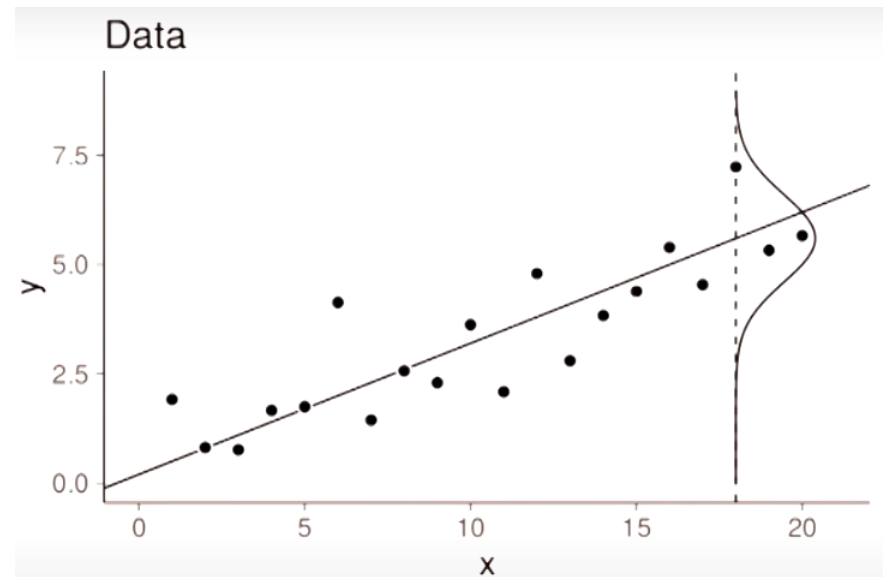
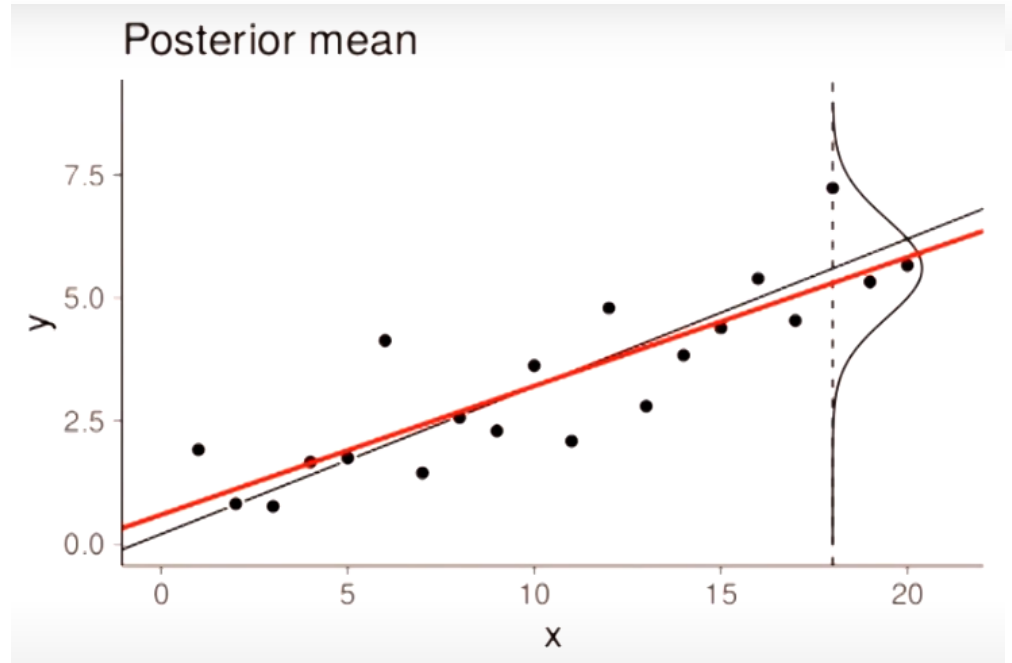
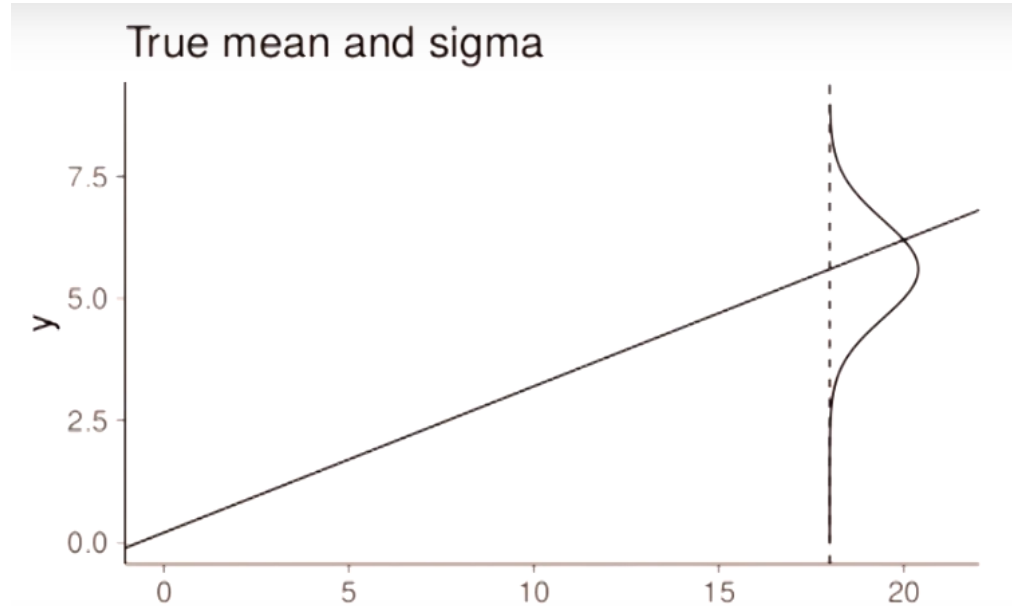
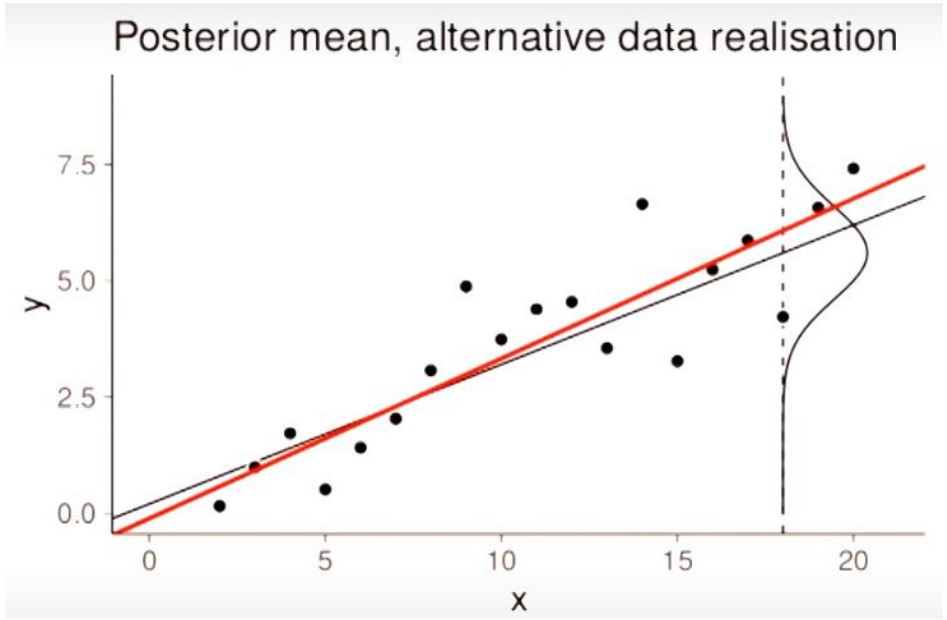
Calculada facilmente Amostragem por importância
(Gelfand, Dey e Chang, 1992; Gelfand, 1996),

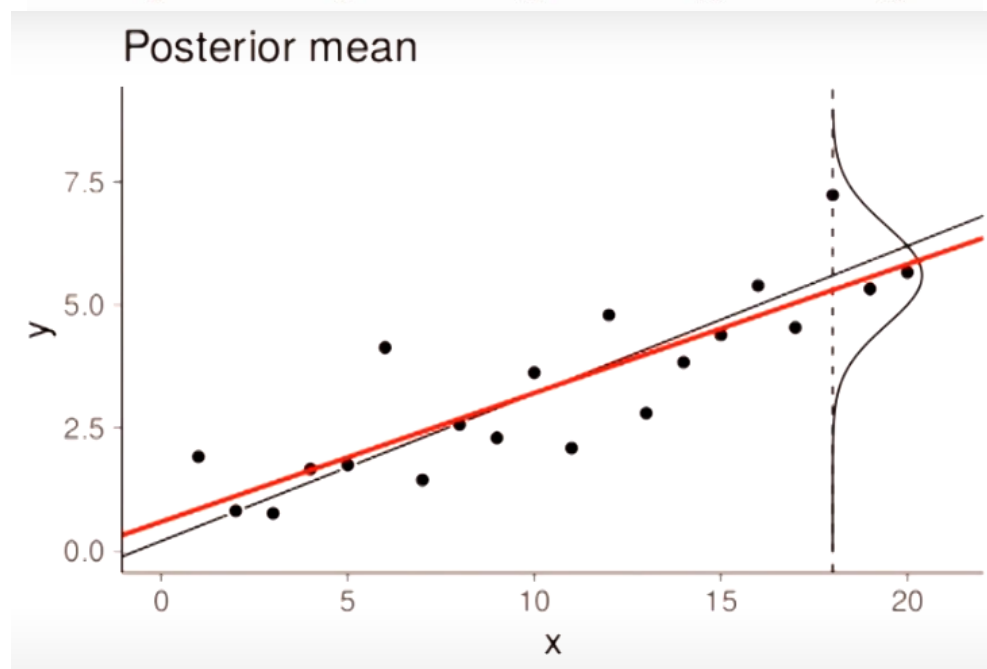
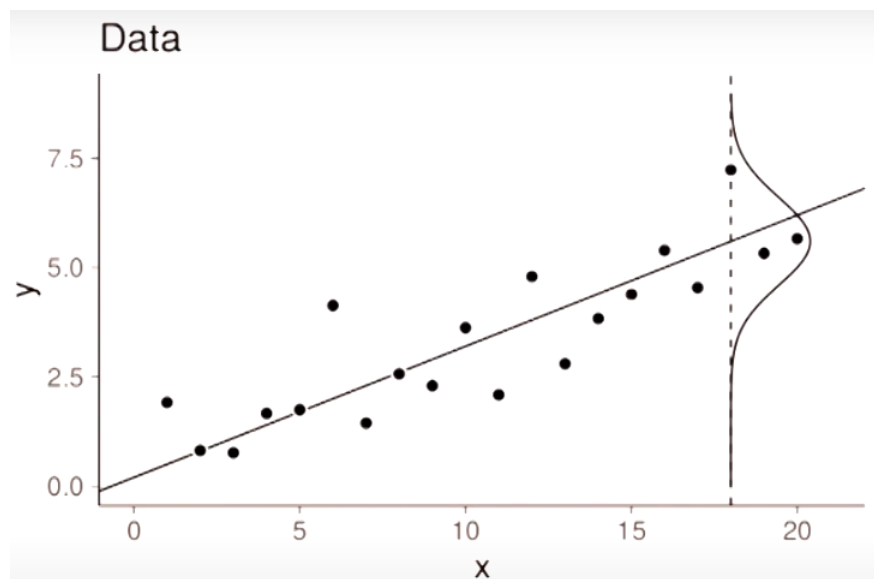
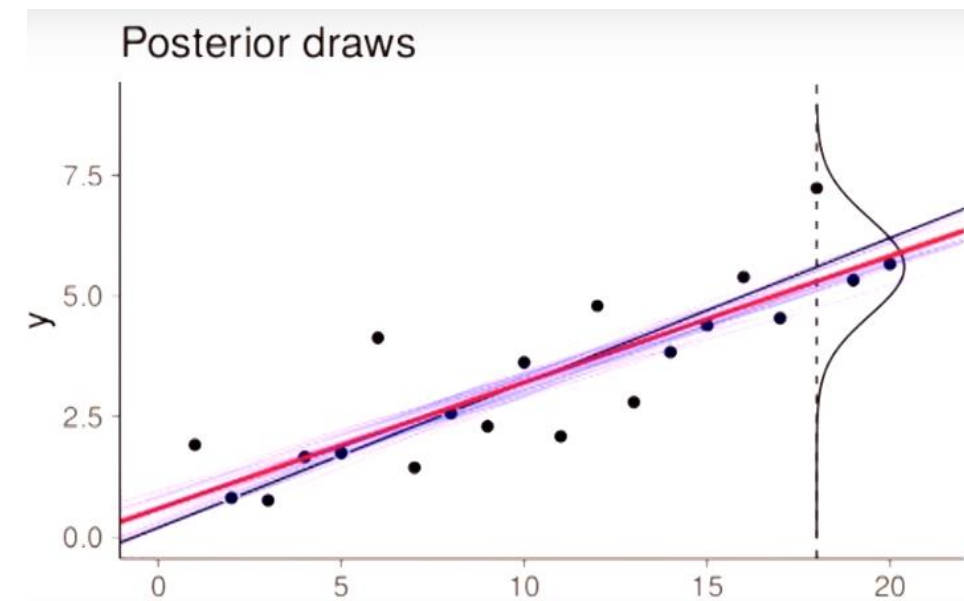
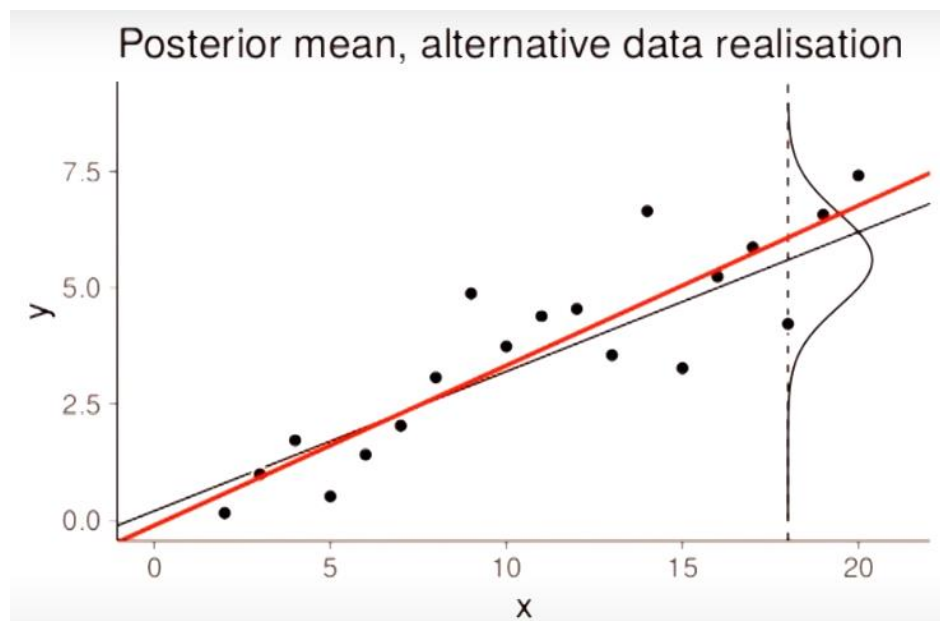


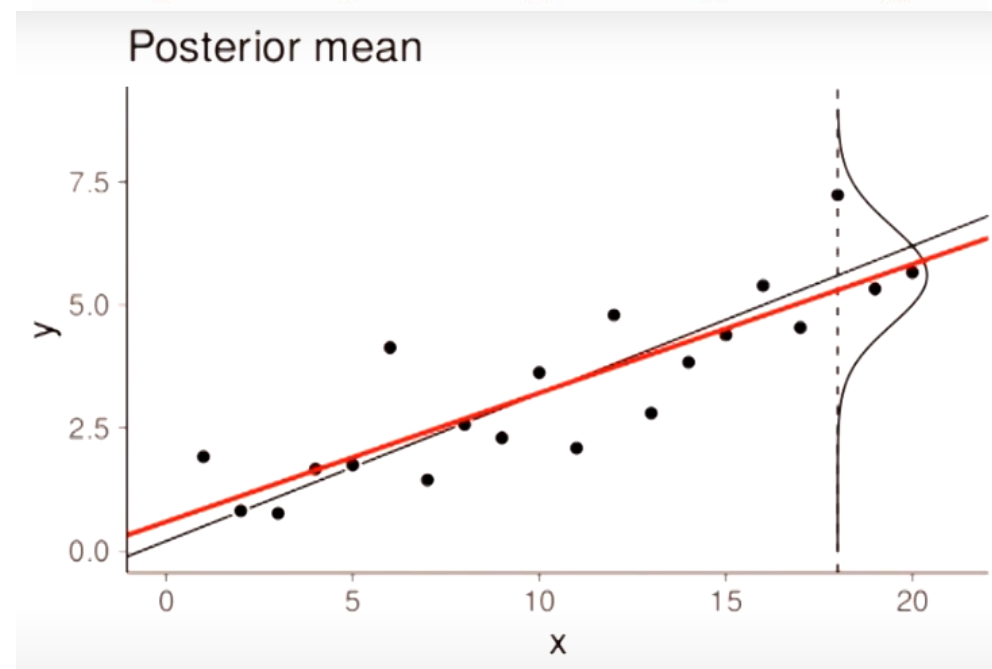
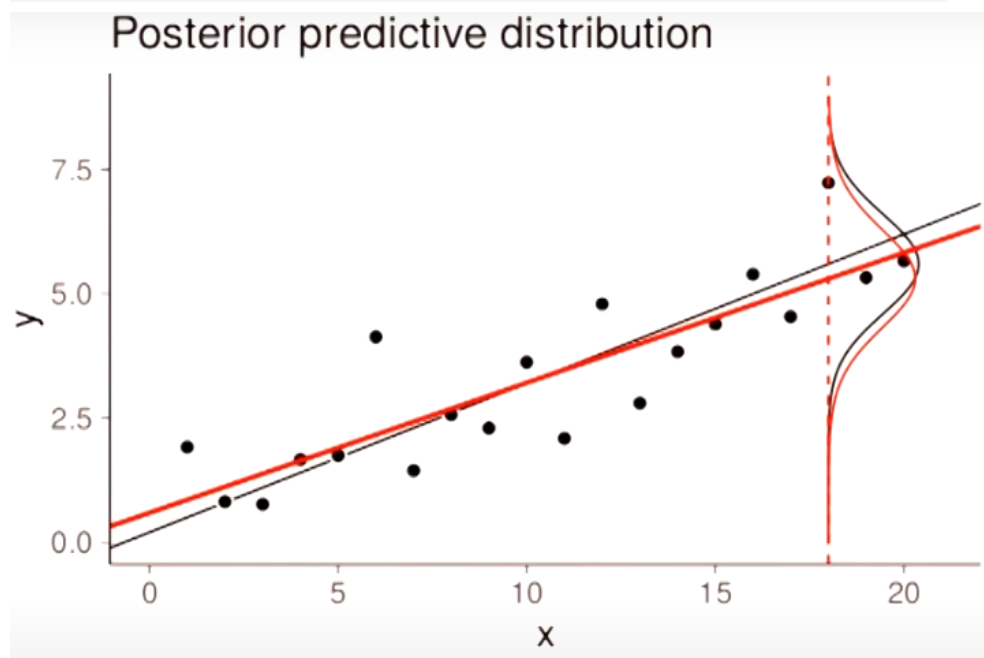
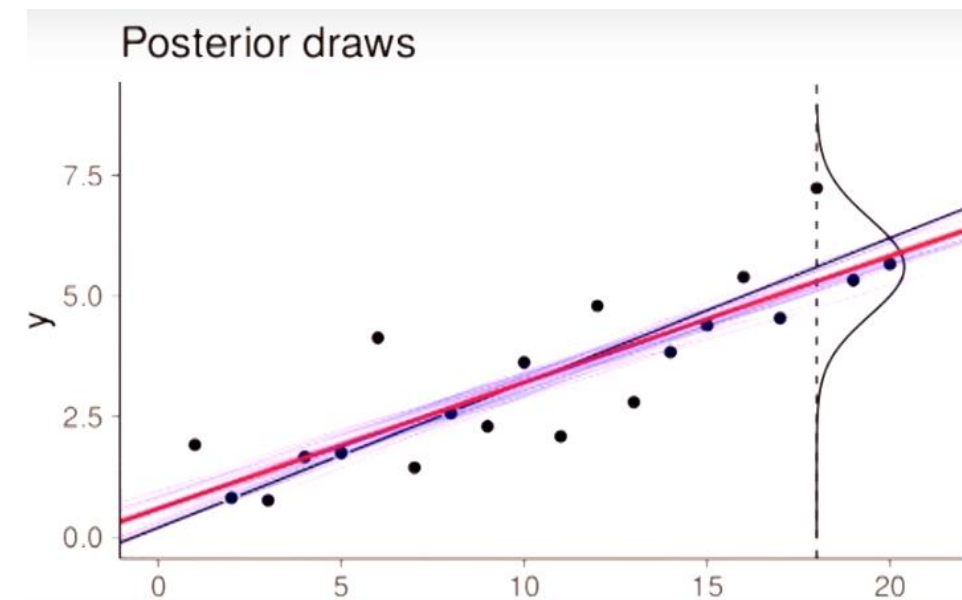
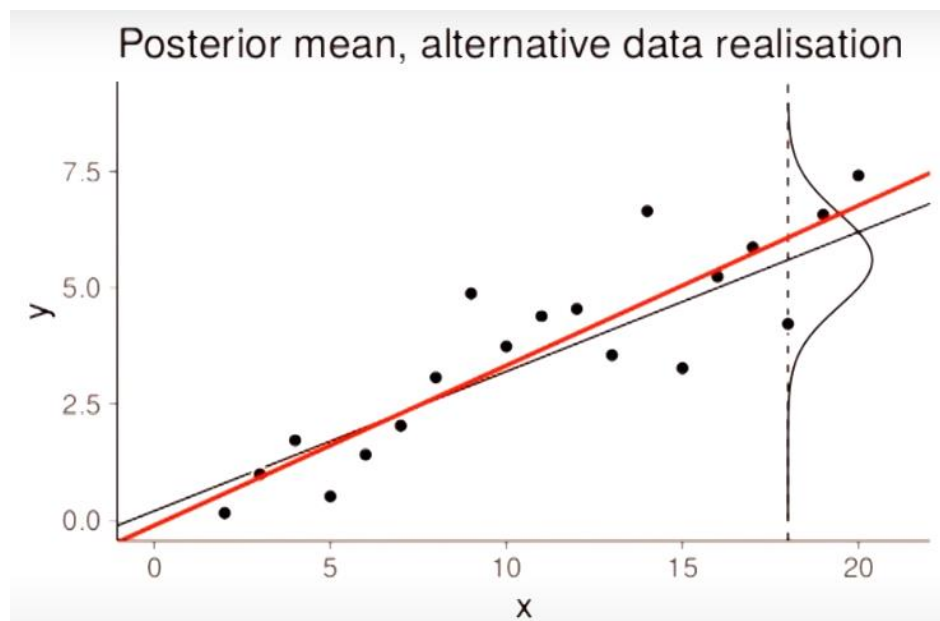








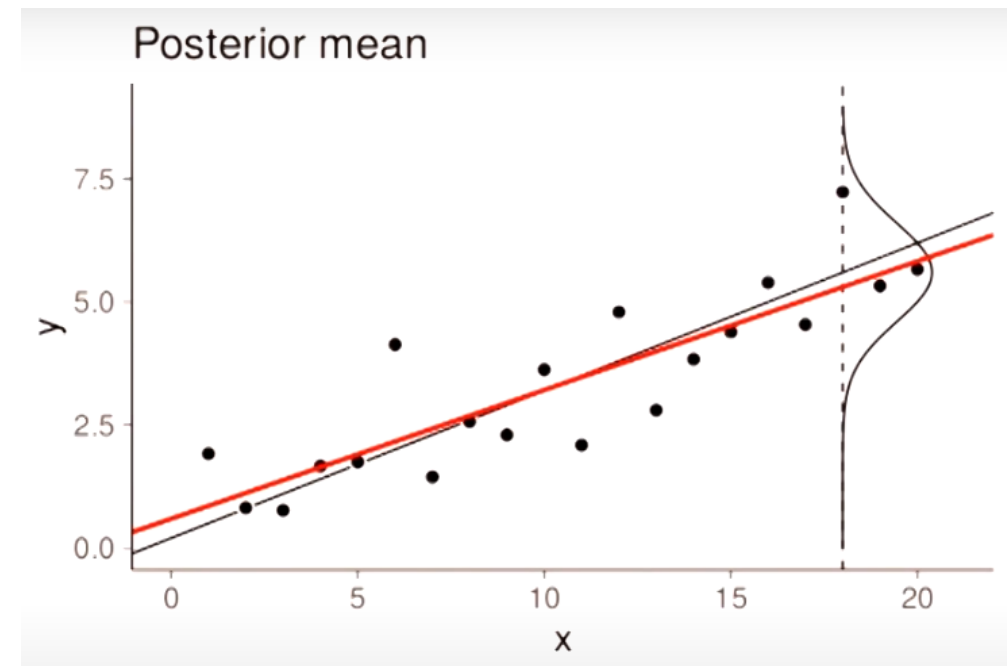
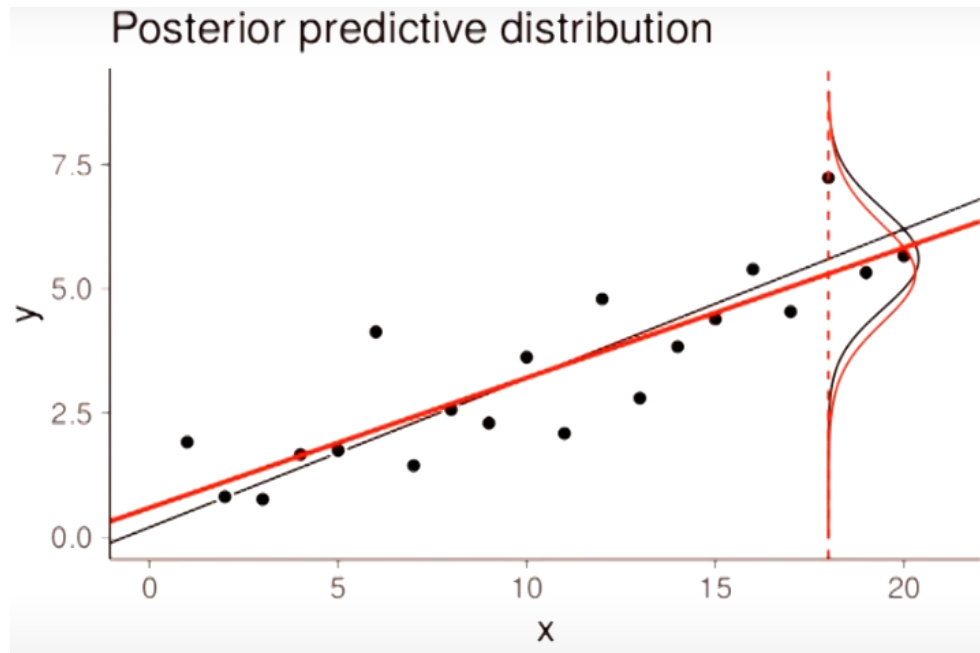




$$p(\tilde{y}|\tilde{x}, x, y) = \int p(\tilde{y}|\tilde{x}, \theta)p(\theta|x, y)d\theta$$

\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

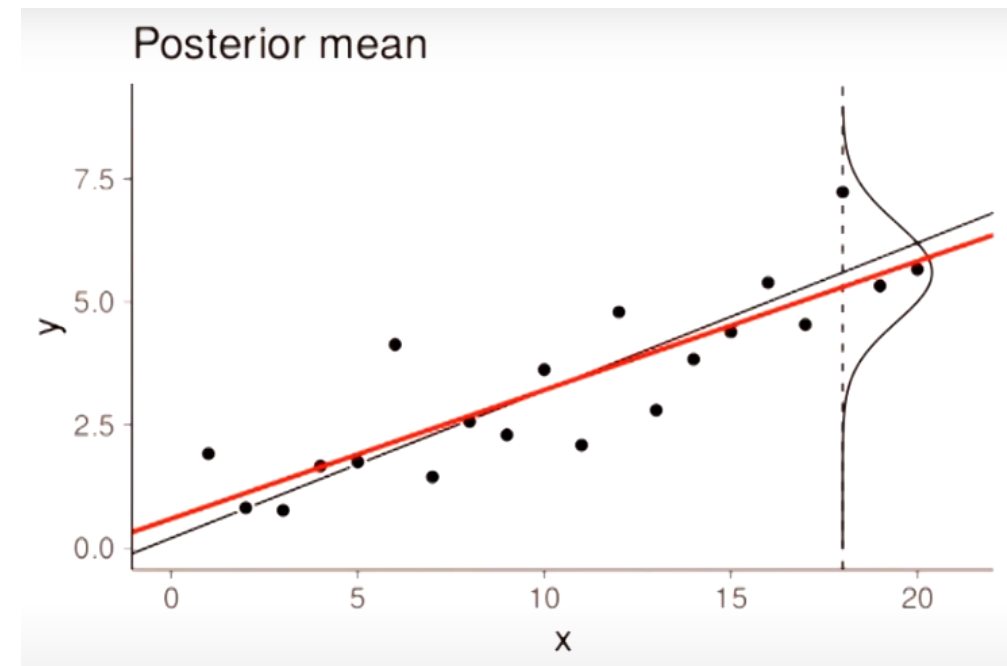
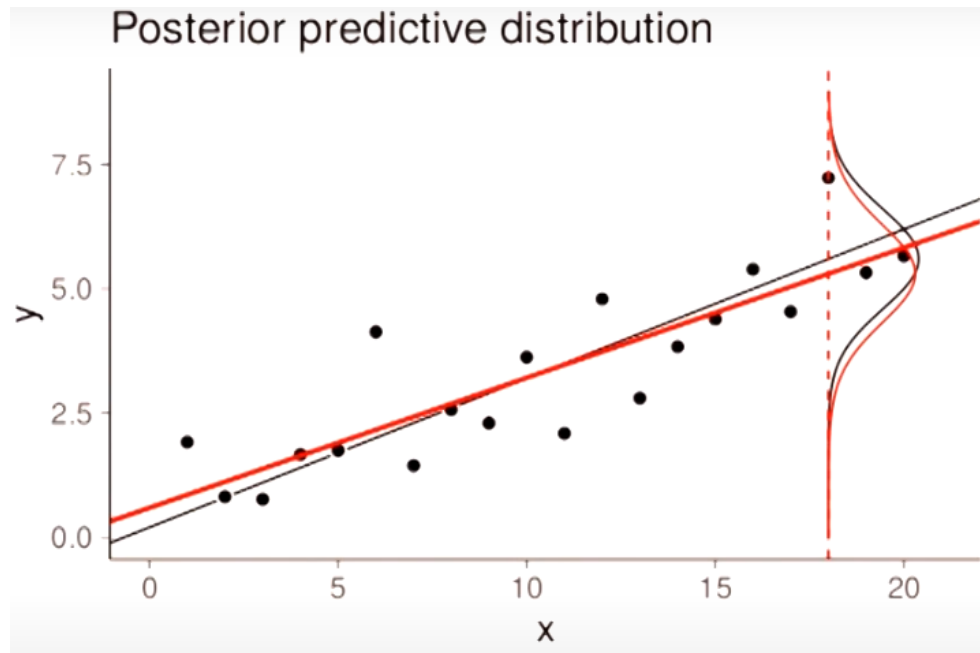


$$p(\tilde{y}|\tilde{x}, x, y) = \int p(\tilde{y}|\tilde{x}, \theta) p(\theta|x, y) d\theta$$

Distribuição a posteriori
do parâmetro

\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

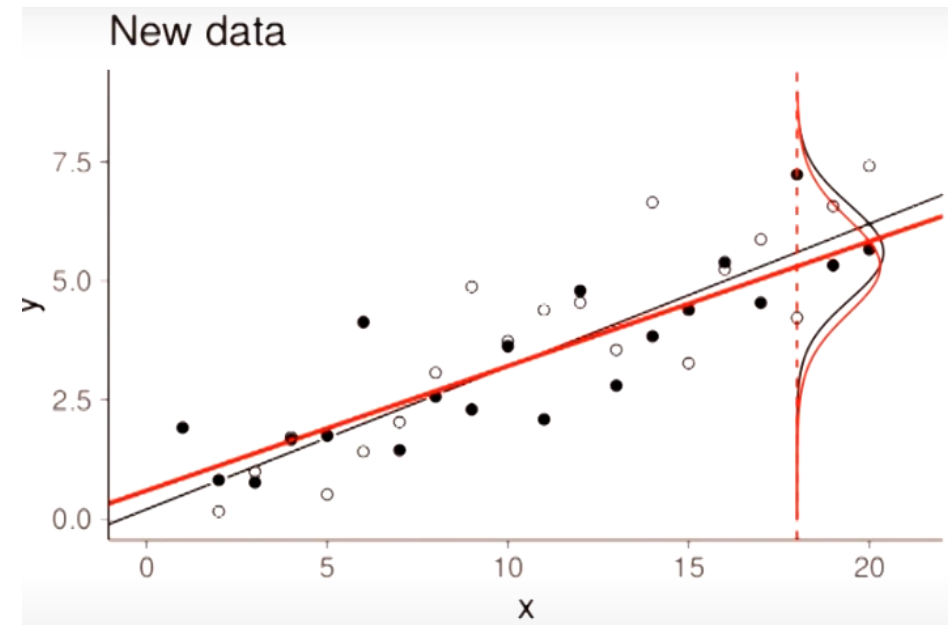
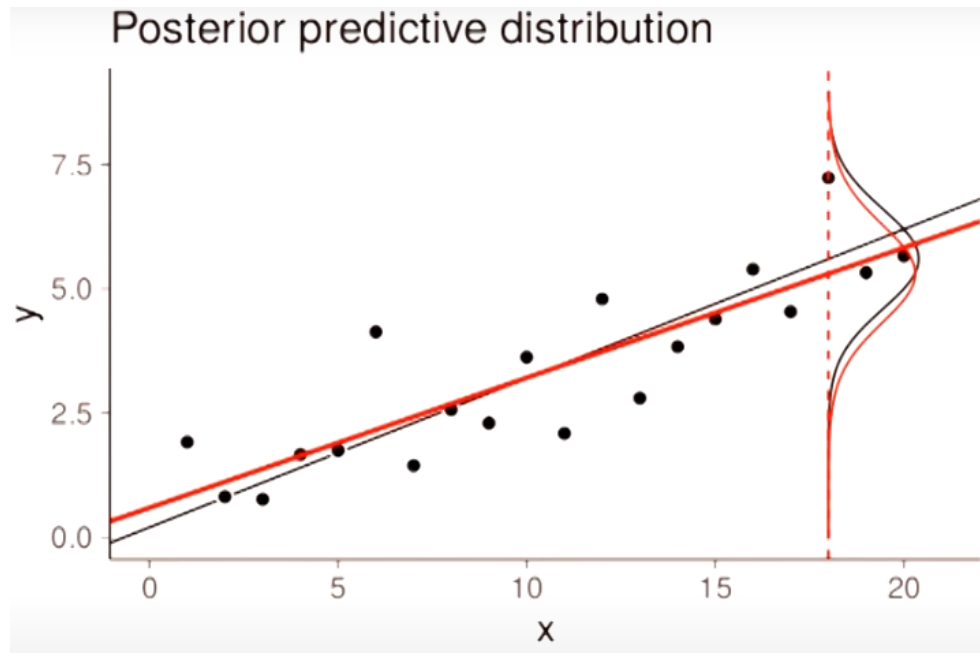


$$p(\tilde{y}|\tilde{x}, x, y) = \int p(\tilde{y}|\tilde{x}, \theta)p(\theta|x, y)d\theta$$

\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

Acurácia do modelo aos dados novos
Mas não temos os dados novos

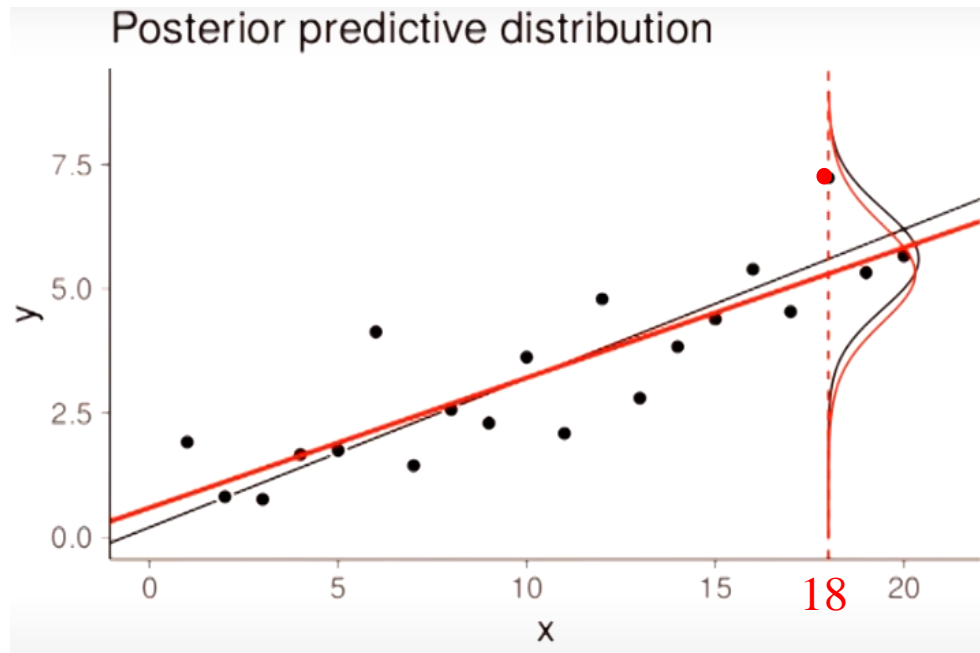


$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

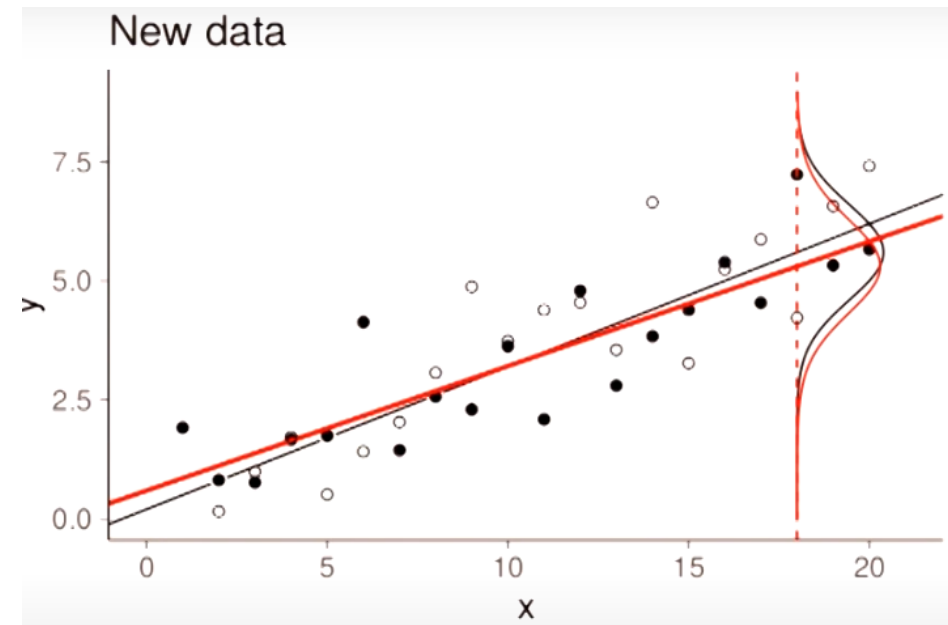
\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

Calcula a densidade de probabilidade preditiva a posteriori para $x=18$ com dados completos



Acurácia do modelo aos dados novos
Mas não temos os dados novos

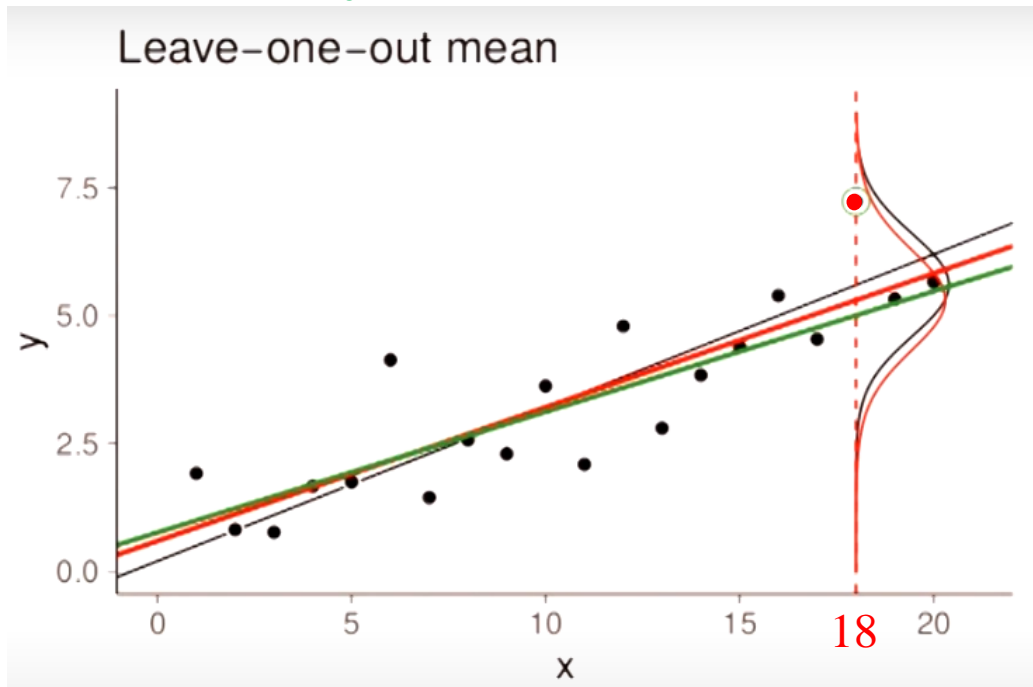


$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

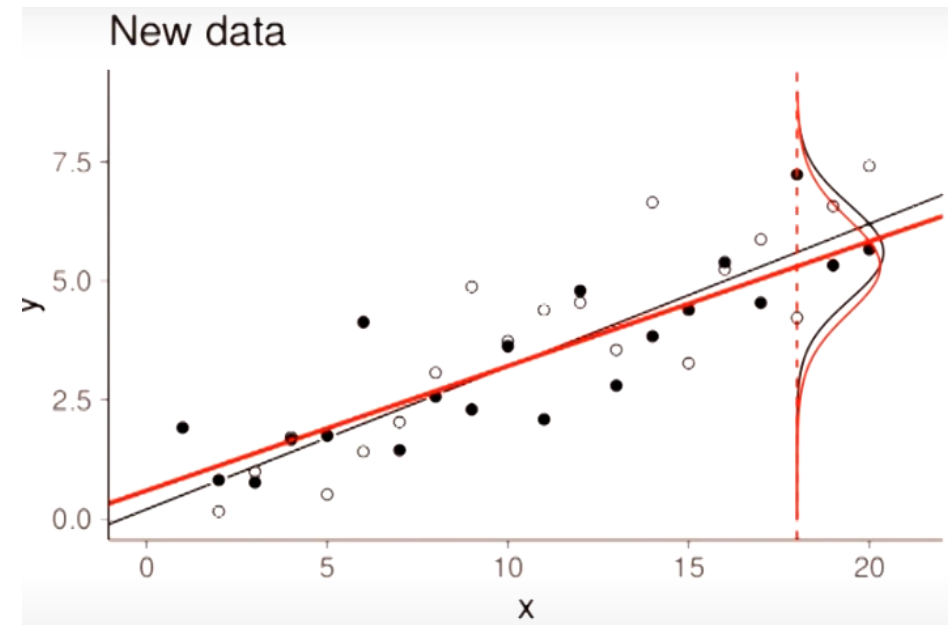
\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

Retirar uma observação
E ajustar um modelo LOO



Acurácia do modelo aos dados novos
Mas não temos os dados novos



$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

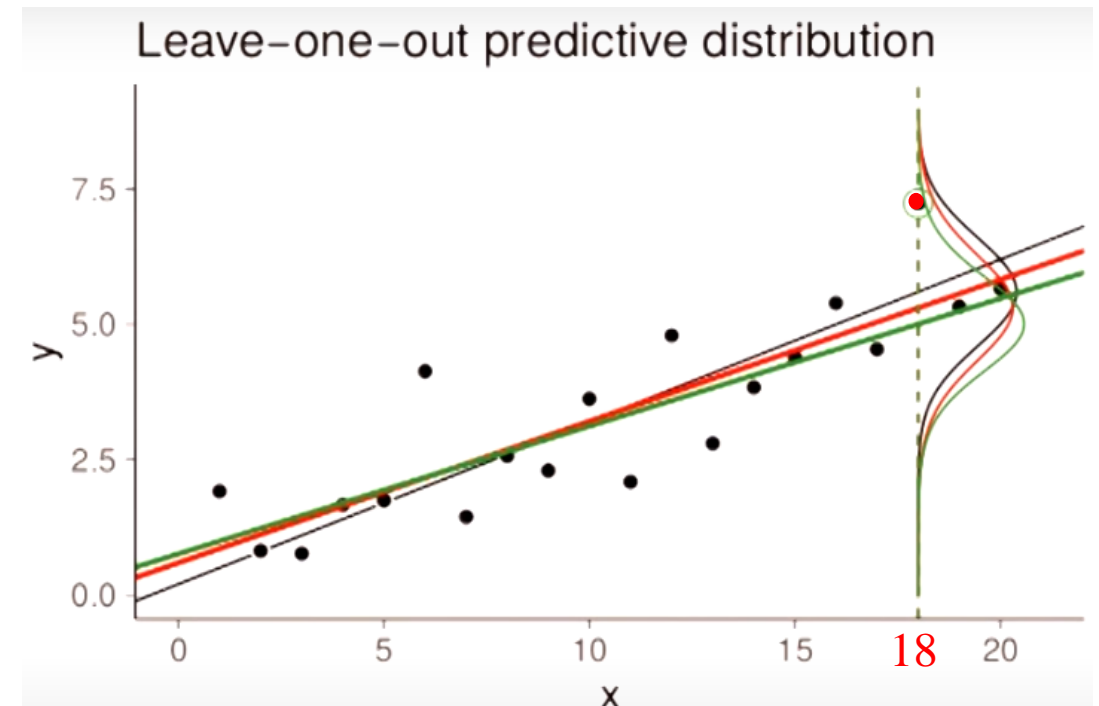
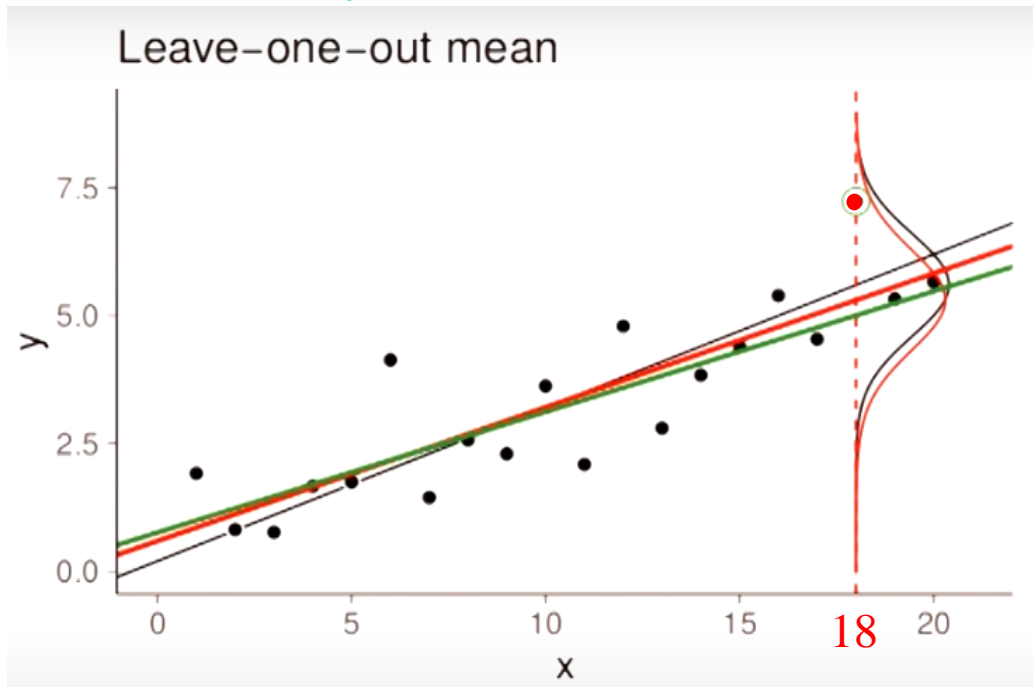
\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

Distribuição preditiva LOO

$$p(\tilde{y}|\tilde{x} = 18, x_{-18}, y_{-18}) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x_{-18}, y_{-18})d\theta$$

Retirar uma observação
E ajustar um modelo LOO



$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

\tilde{y} \tilde{x} → Dados novos

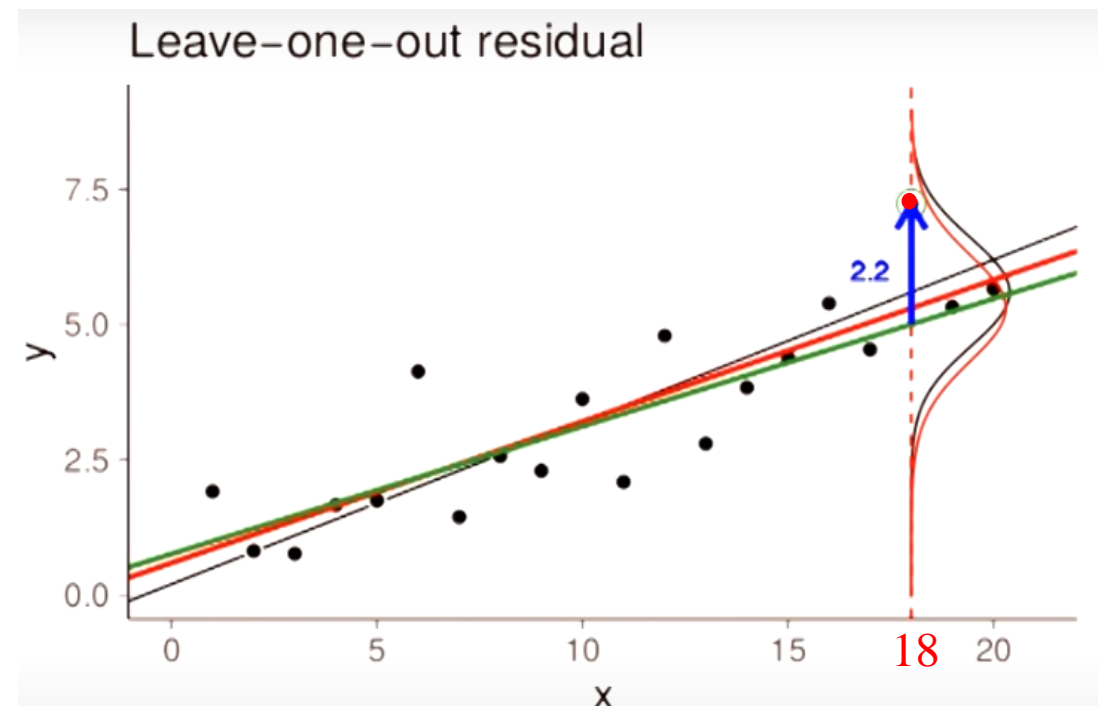
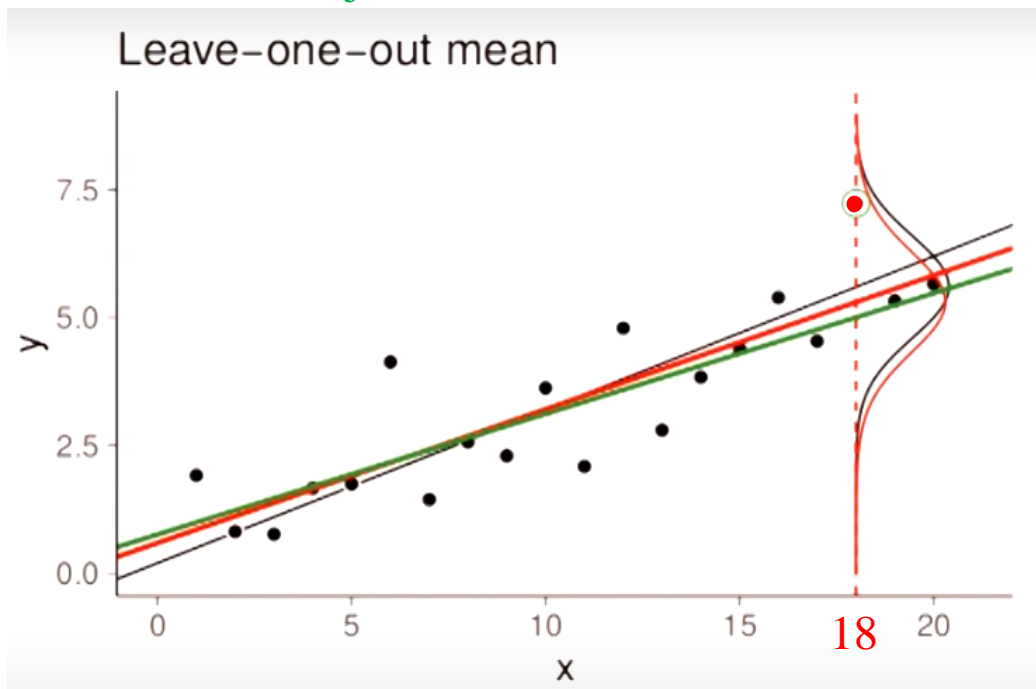
x, y → Dados antigos no qual o modelo foi ajustado

$$y_{18} - E[p(\tilde{y}|\tilde{x} = 18, x_{-18}, y_{-18})]$$

Retirar uma observação
E ajustar um modelo LOO

Podemos usar *essa distância*
(LOO posteriori preditiva) como
um resíduo de LOO

Estimativas pontuais:
Raiz quadrada do erro
médio (RMSE)
 R^2

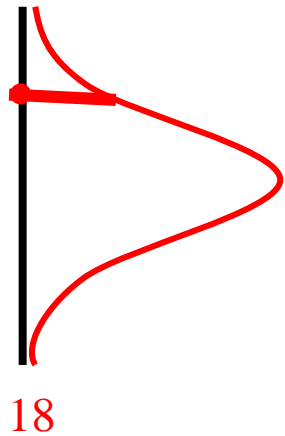


$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

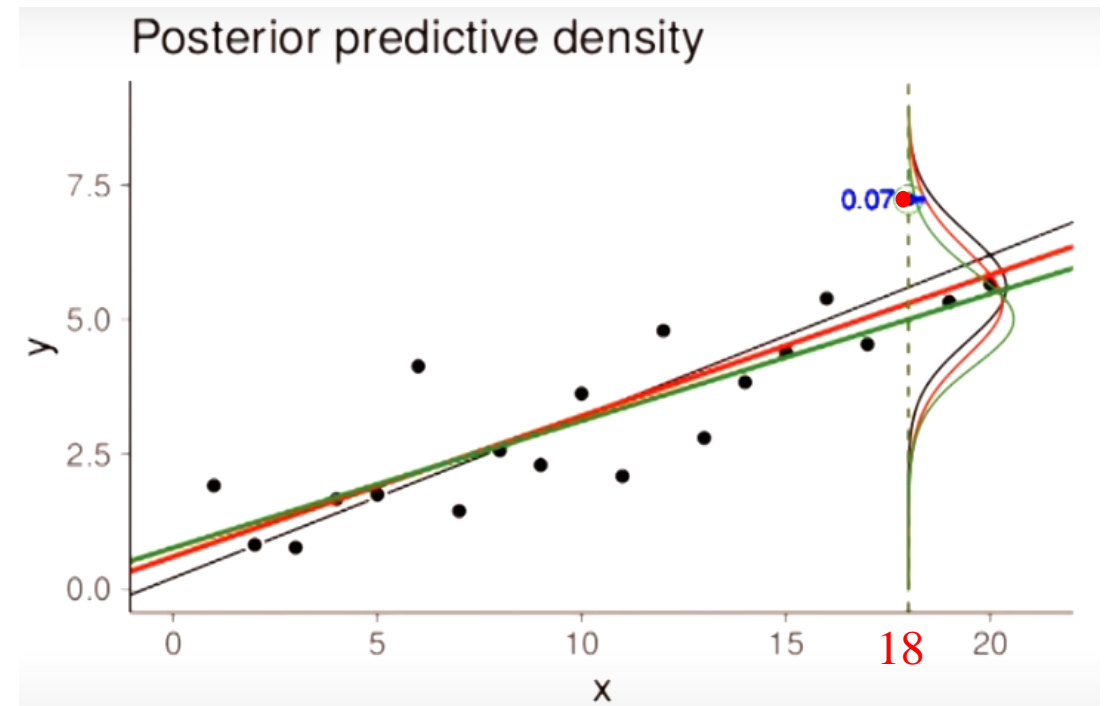
\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

$$p(\tilde{y}|\tilde{x} = 18, x_{-18}, y_{-18}) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x_{-18}, y_{-18})d\theta$$



$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x, y) \approx 0.07$$



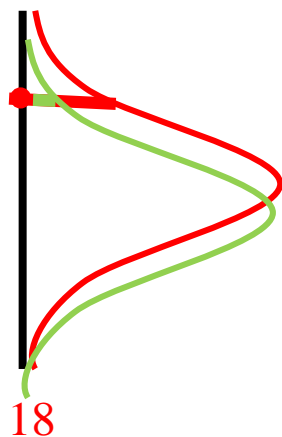


$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

\tilde{y} \tilde{x} → Dados novos

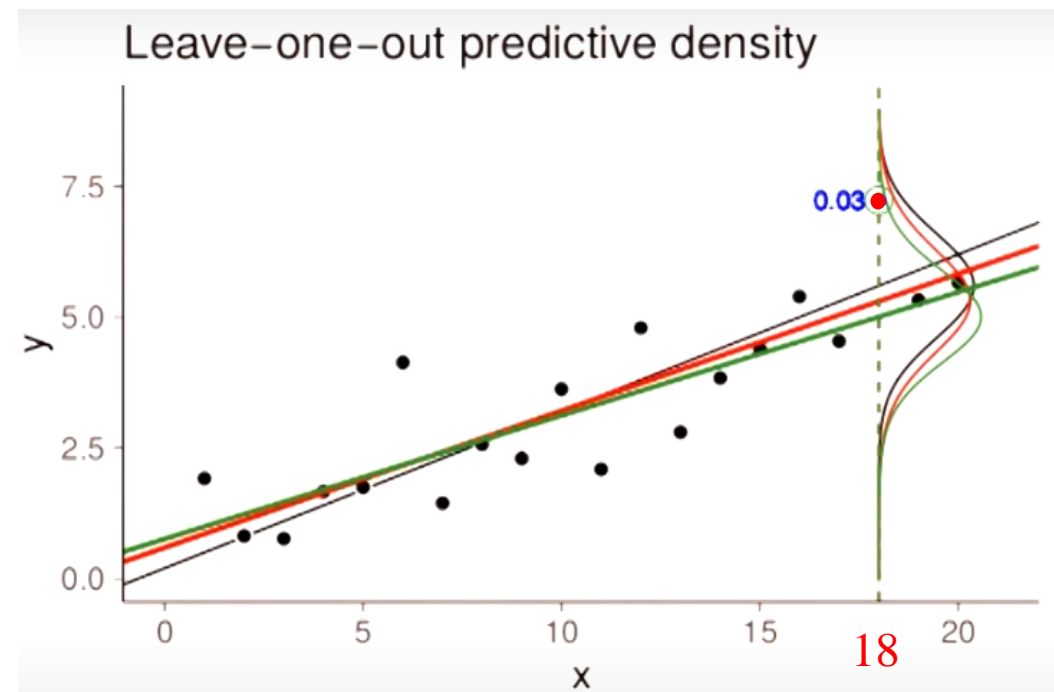
x, y → Dados antigos no qual o modelo foi ajustado

$$p(\tilde{y}|\tilde{x} = 18, x_{-18}, y_{-18}) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x_{-18}, y_{-18})d\theta$$



$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x, y) \approx 0.07$$

$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x_{-18}, y_{-18}) \approx 0.03$$



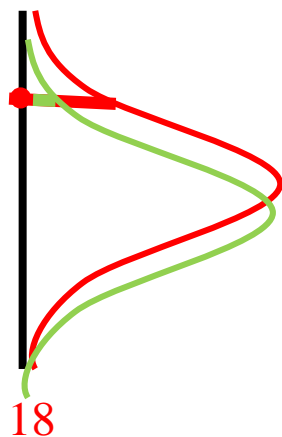


$$p(\tilde{y}|\tilde{x} = 18, x, y) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x, y)d\theta$$

\tilde{y} \tilde{x} → Dados novos

x, y → Dados antigos no qual o modelo foi ajustado

Dados completos



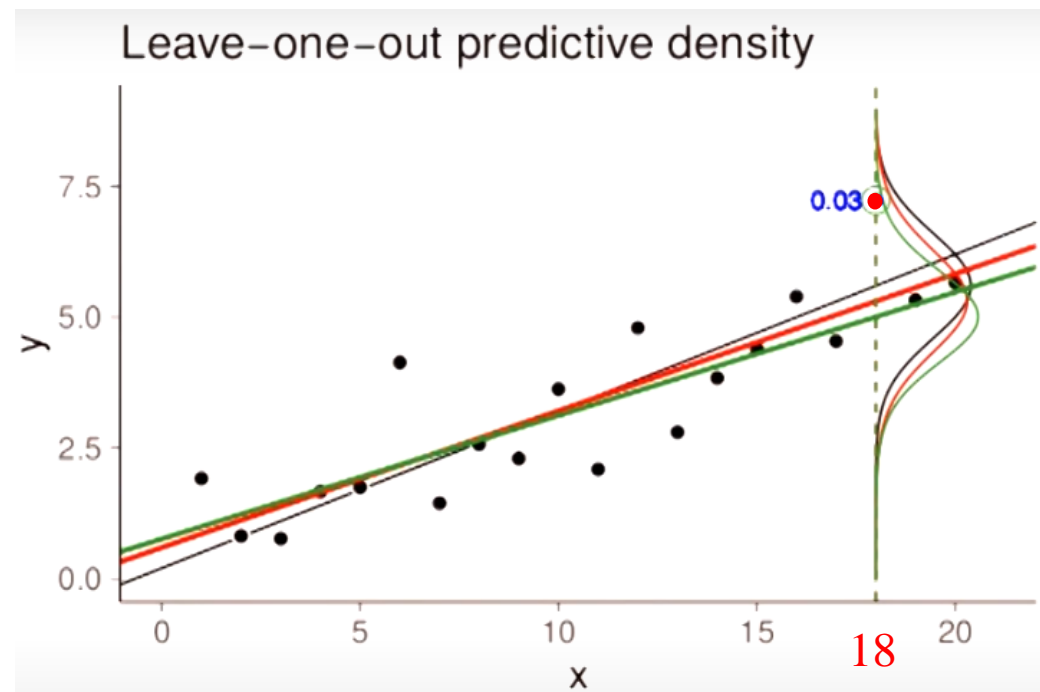
18

Dados $i = -18$ observação LOO

$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x, y) \approx 0.07$$

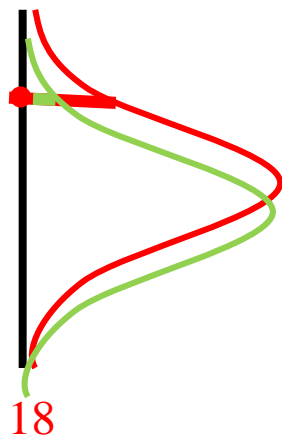
$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x_{-18}, y_{-18}) \approx 0.03$$

$$p(\tilde{y}|\tilde{x} = 18, x_{-18}, y_{-18}) = \int p(\tilde{y}|\tilde{x} = 18, \theta)p(\theta|x_{-18}, y_{-18})d\theta$$



$$p(y_i | x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

Dados completos

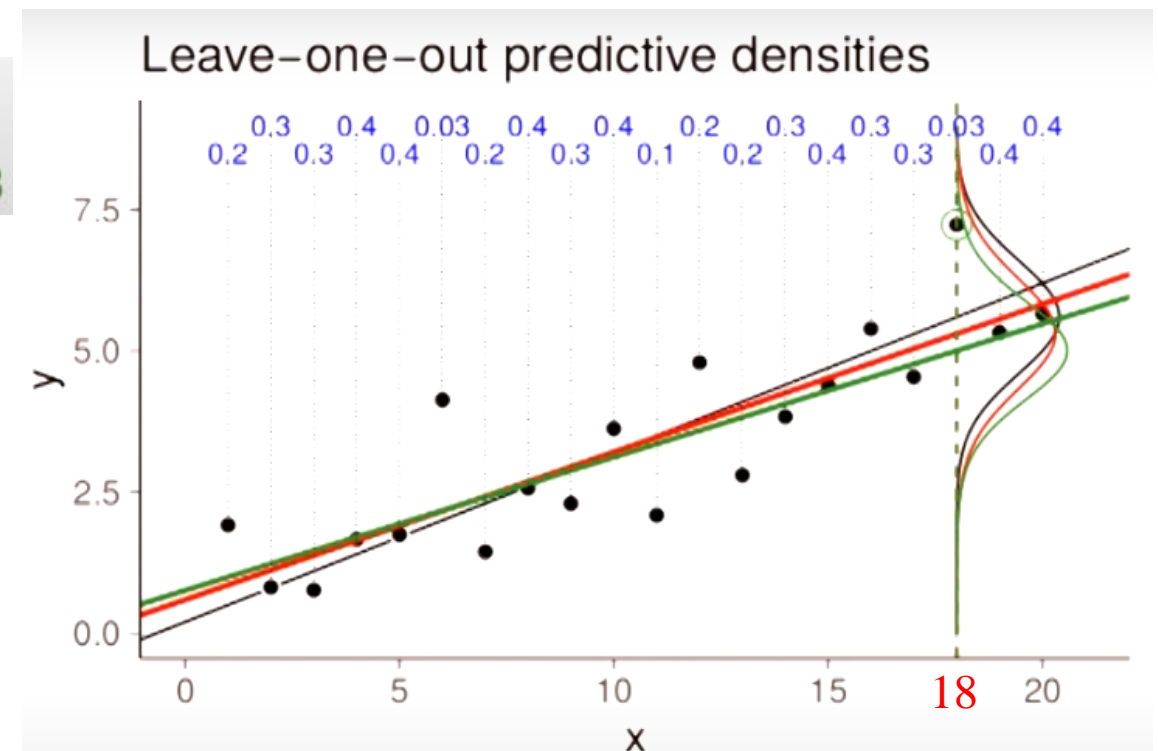


$$p(\tilde{y} = y_{18} | \tilde{x} = 18, x, y) \approx 0.07$$

$$p(\tilde{y} = y_{18} | \tilde{x} = 18, x_{-18}, y_{-18}) \approx 0.03$$

Dados $i = -18$ observação LOO

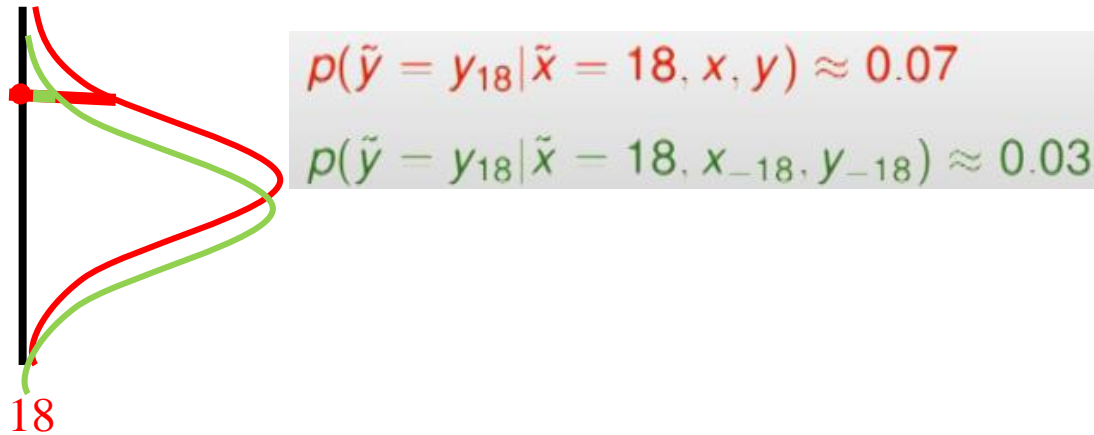
Repete o procedimento para todos os dados



$$p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

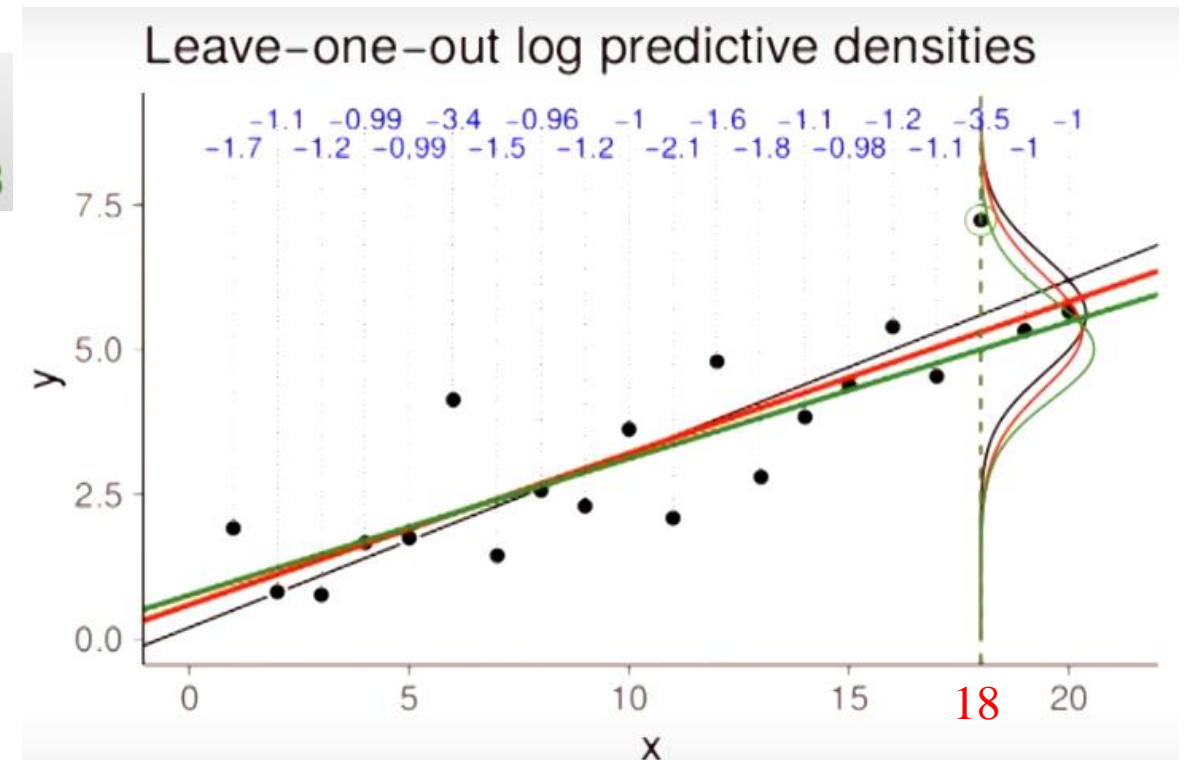
$$\log p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

Dados completos



Dados $i = -18$ observação LOO

Comum logaritmizar



$$p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

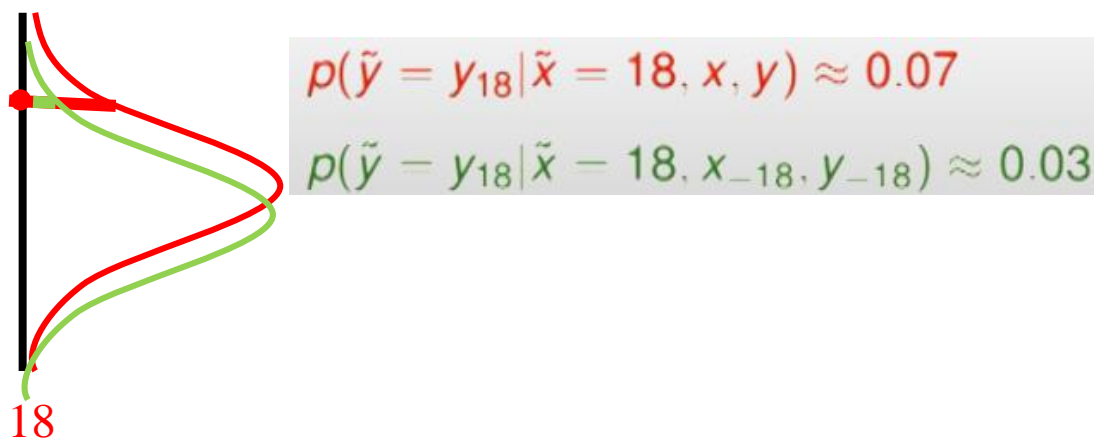
$$\log p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\text{elpd_loo} = \sum_{i=1}^{20} \log p(y_i|x_i, x_{-i}, y_{-i}) \approx -29.5$$

Estimativa não viesada da densidade
preditiva a posteriori produzida pelo novos
dados

Log da densidade preditiva esperada
Expected Log Predictive Densities
(ELPD_LOO)

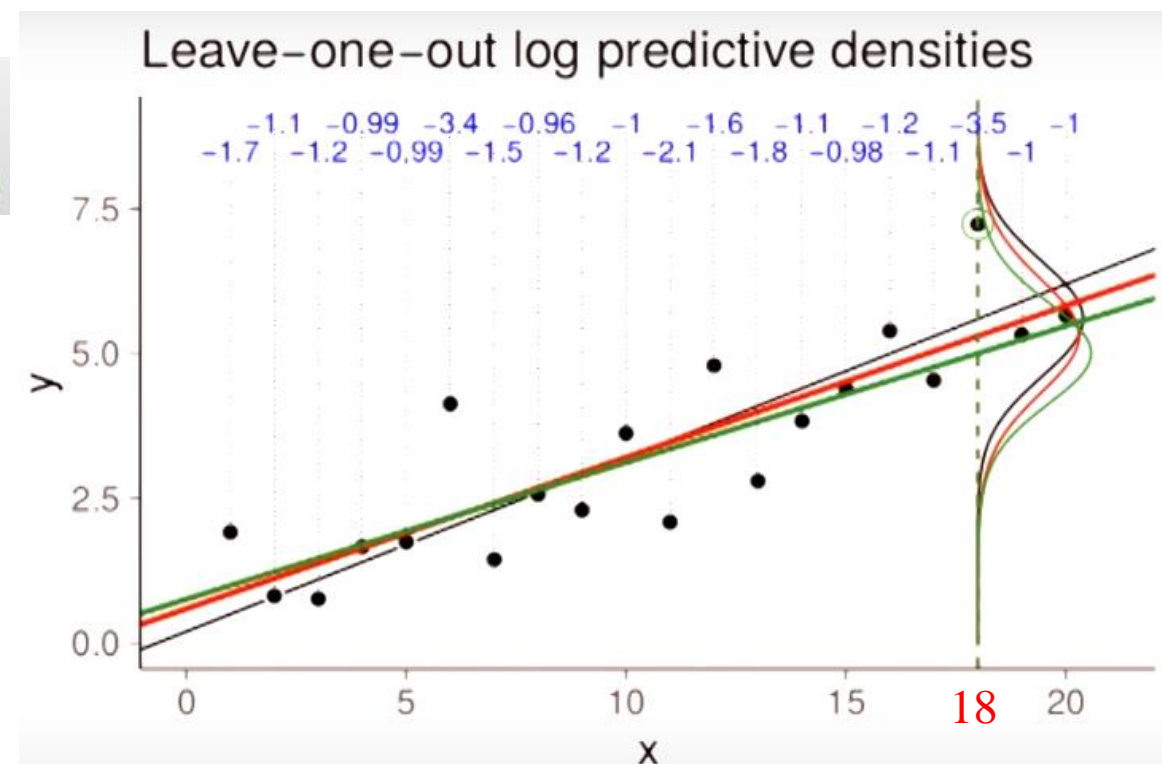
Dados completos



Dados $i = -18$ observação LOO

$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x, y) \approx 0.07$$

$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x_{-18}, y_{-18}) \approx 0.03$$



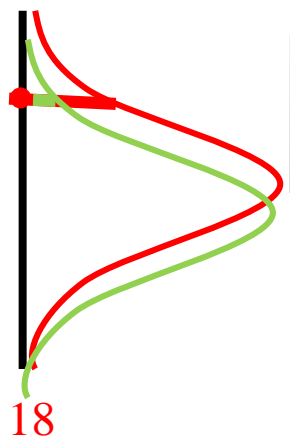
$$p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\log p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\text{elpd_loo} = \sum_{i=1}^{20} \log p(y_i|x_i, x_{-i}, y_{-i}) \approx -29.5$$

$$\text{lpd} = \sum_{i=1}^{20} \log p(y_i|x_i, x, y) \approx -26.8$$

Dados completos



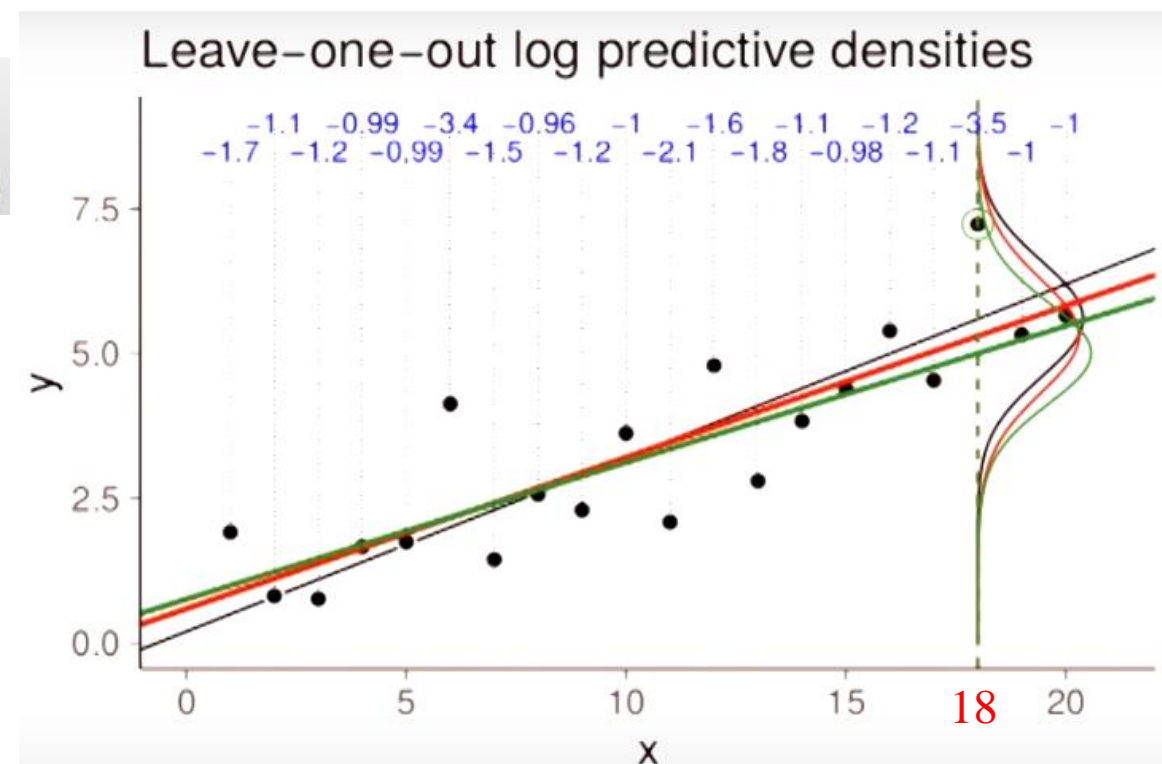
$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x, y) \approx 0.07$$

$$p(\tilde{y} = y_{18}|\tilde{x} = 18, x_{-18}, y_{-18}) \approx 0.03$$

Dados $i = -18$ observação LOO

Estimativa não viesada da densidade
preditiva a posteriori produzida pelo novos
dados

Log da densidade preditiva dos dados completos
Log Predictive Densities (LPD)



$$p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\log p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\text{elpd_loo} = \sum_{i=1}^{20} \log p(y_i|x_i, x_{-i}, y_{-i}) \approx -29.5$$

$$\text{lpd} = \sum_{i=1}^{20} \log p(y_i|x_i, x, y) \approx -26.8$$

$$\text{p_loo} = \text{lpd} - \text{elpd_loo} \approx 2.7$$

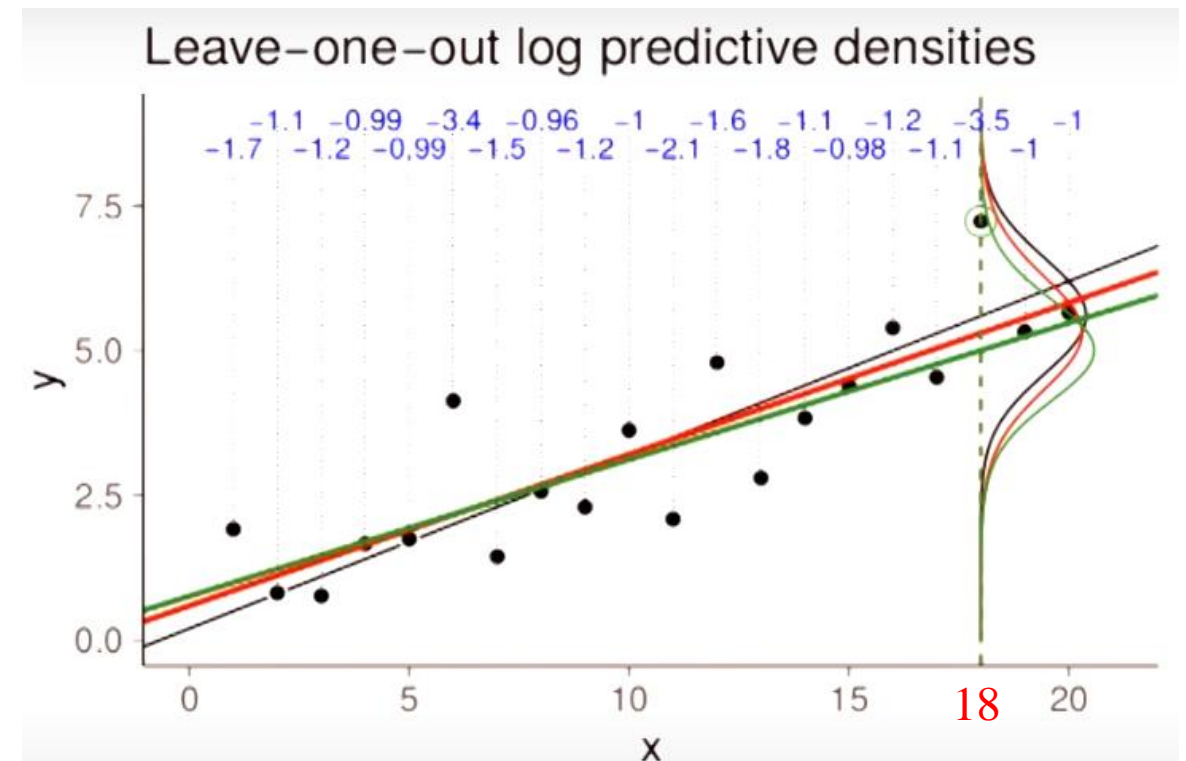
Número efetivo de parâmetros
Interpreta a complexidade do modelo

➤ Parâmetros:

- Intercepto
- Coeficiente angular
- sigma $\epsilon \sim N(0, \sigma^2)$

Estimativa não viesada da densidade
preditiva a posteriori produzida pelo novos
dados

Log da densidade preditiva esperada
Expected Log Predictive Densities
(ELPD_LOO)



$$p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\log p(y_i|x_i, x_{-i}, y_{-i}), \quad i = 1, \dots, 20$$

$$\text{elpd_loo} = \sum_{i=1}^{20} \log p(y_i|x_i, x_{-i}, y_{-i}) \approx -29.5$$

$$\text{lpd} = \sum_{i=1}^{20} \log p(y_i|x_i, x, y) \approx -26.8$$

$$\text{p_loo} = \text{lpd} - \text{elpd_loo} \approx 2.7$$

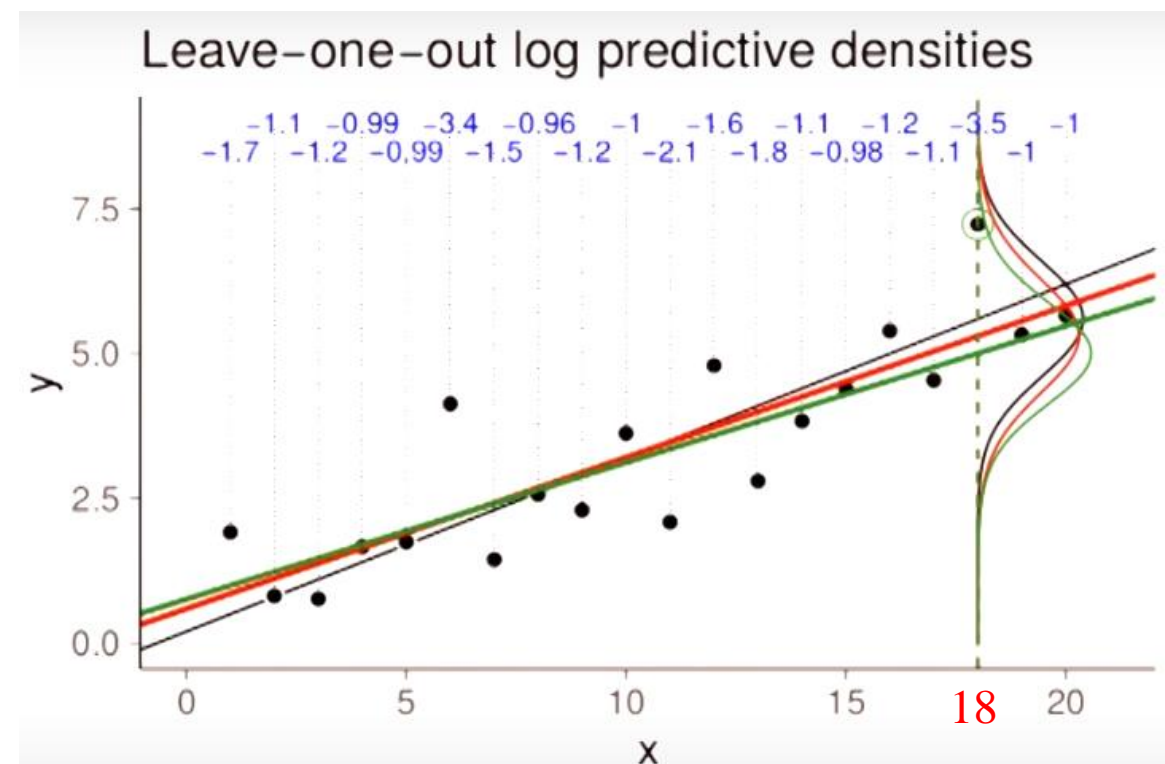
Número efetivo de parâmetros
Interpreta a complexidade do modelo

$$\text{SE} = \text{sd}(\log p(y_i|x_i, x_{-i}, y_{-i})) \cdot \sqrt{20} \approx 3.3$$

elpd_loo é uma estimativa de incerteza de $y|x$.
E portanto é interessante estimar o erro padrão
de elpd_loo .

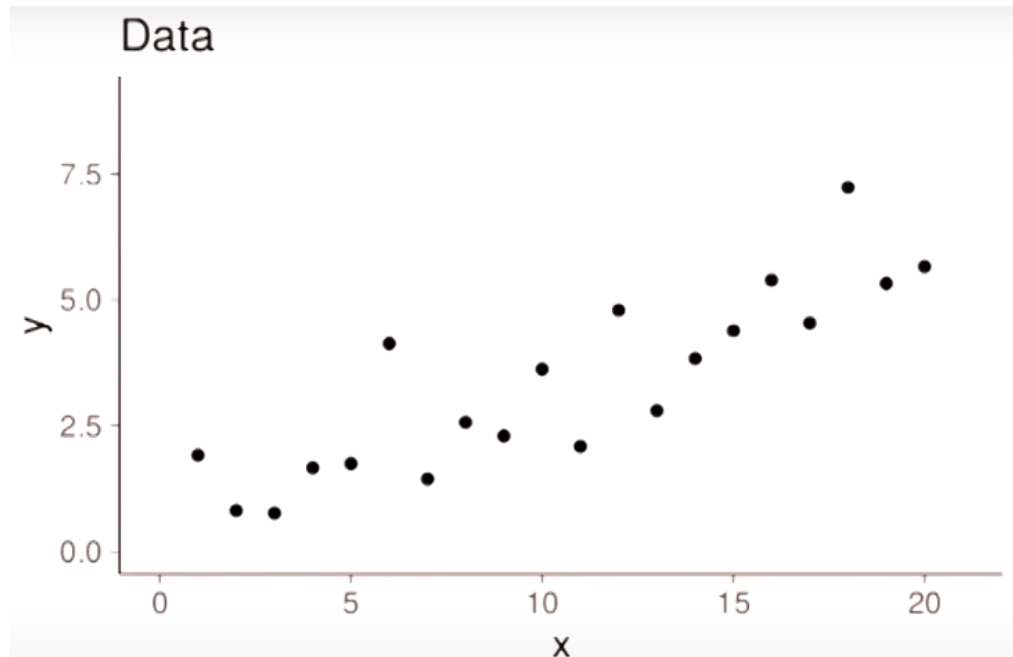
Estimativa não viesada da densidade
preditiva a posteriori produzida pelo novos
dados

Log da densidade preditiva esperada
Expected Log Predictive Densities
(ELPD_LOO)



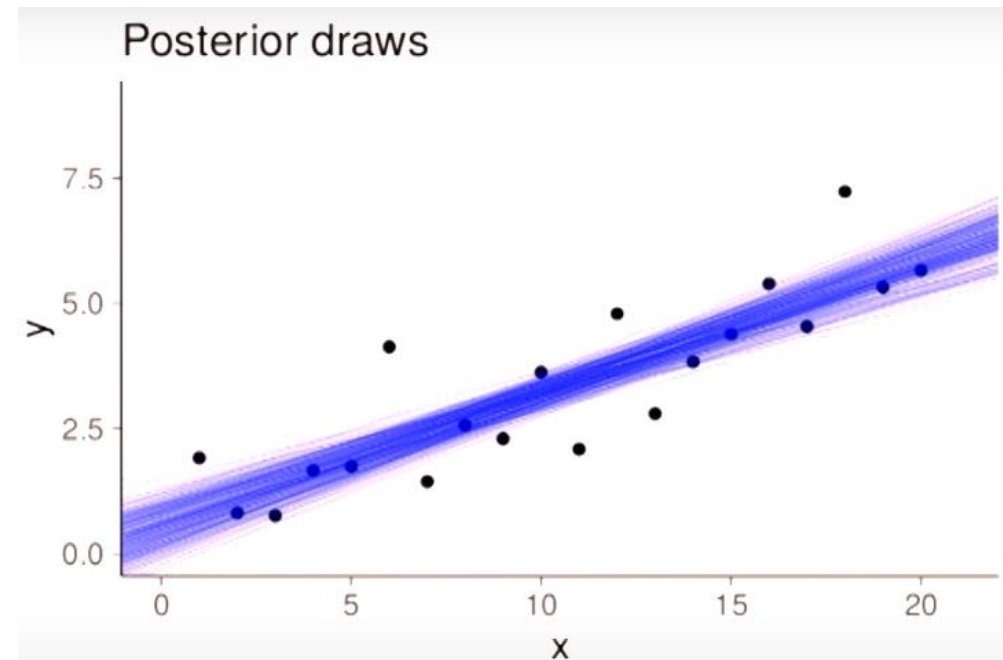
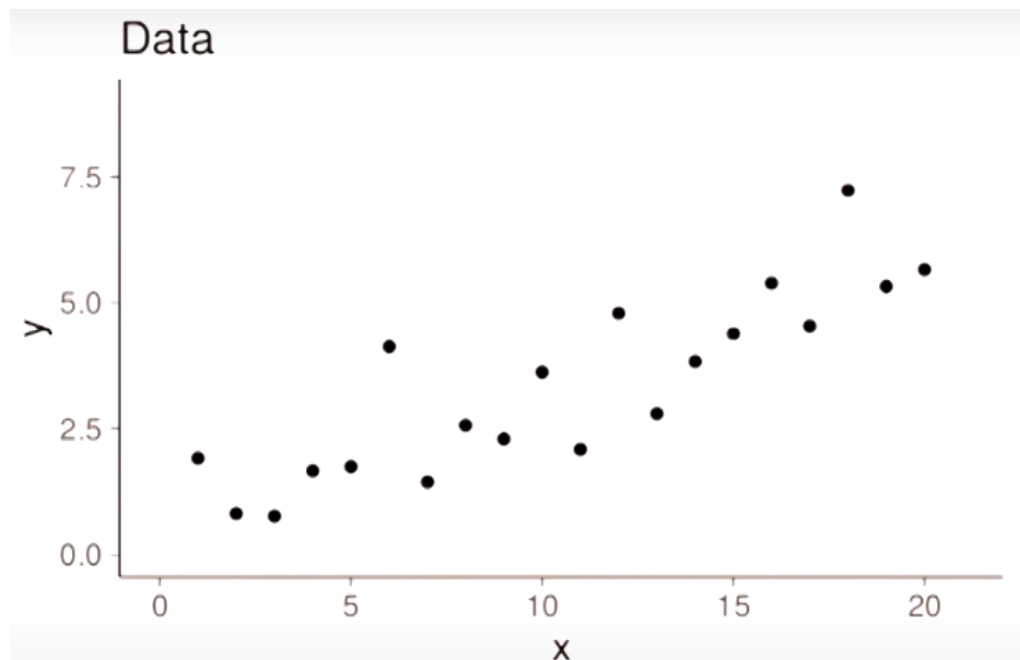
➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto



➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto



$$\theta^{(s)} \sim p(\theta|x, y)$$

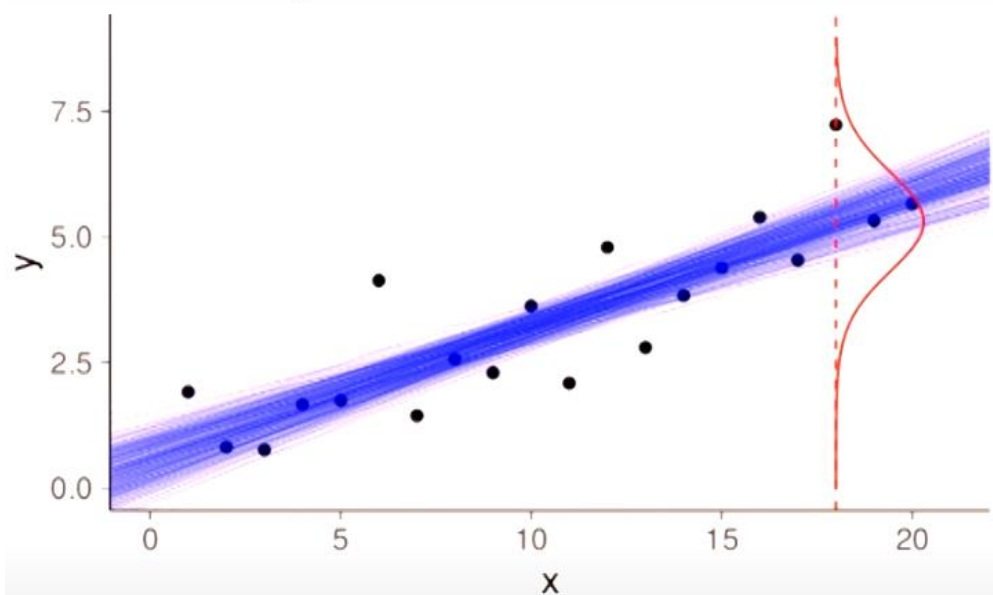
S - indexado das iterações do método iterativo (cálculo numérico e.g. MCMC)

$$\theta^s, s = 1, \dots, S$$

➤ LOO-CV Aproximado por PSIS

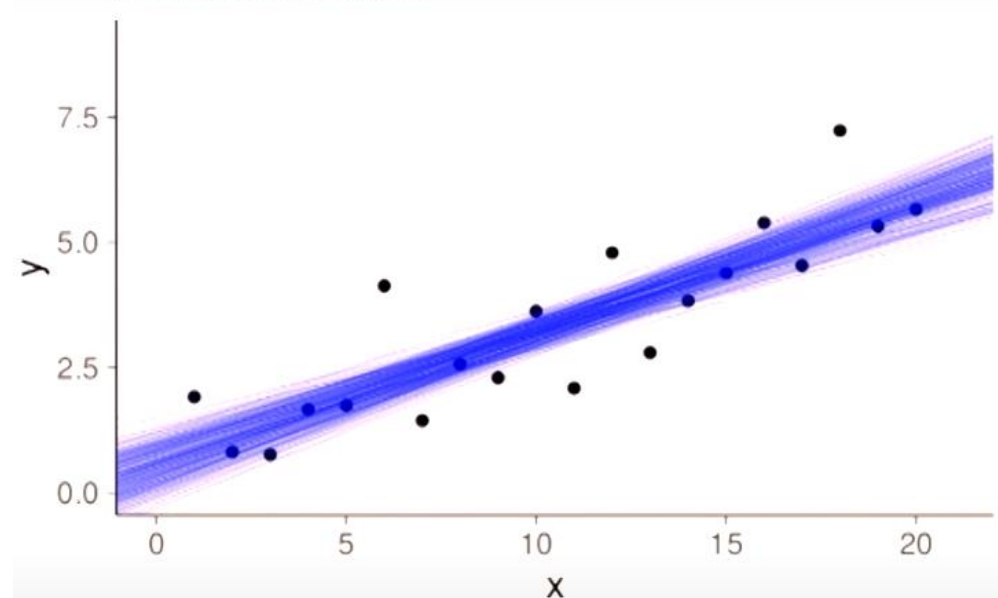
Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

Posterior predictive distribution



$$p(\tilde{y}|\tilde{x}, x, y) \approx \frac{1}{S} \sum_{s=1}^S p(\tilde{y}|\tilde{x}, \theta^{(s)})$$

Posterior draws

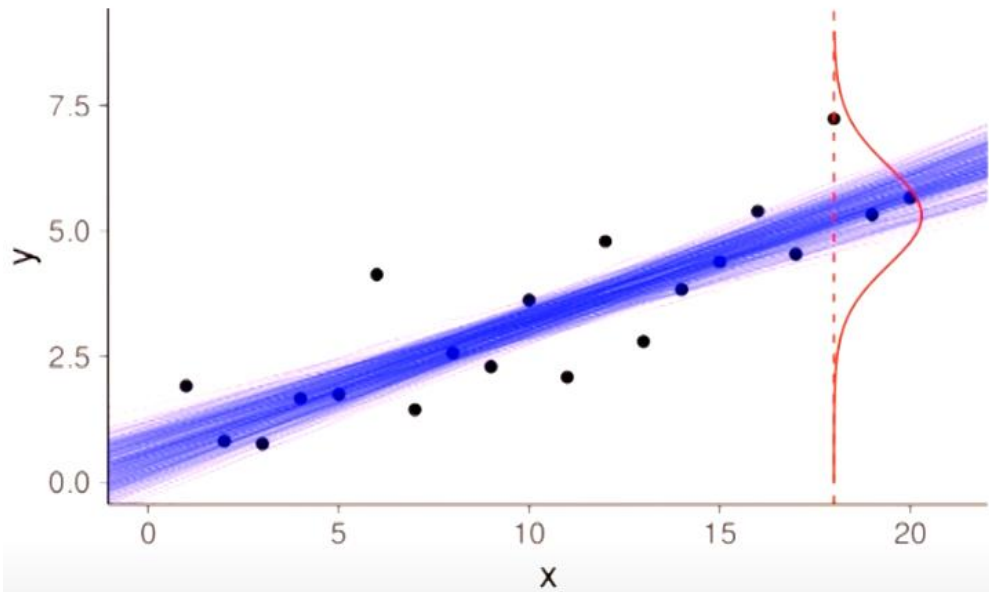


$$\theta^{(s)} \sim p(\theta|x, y)$$

➤ LOO-CV Aproximado por PSIS

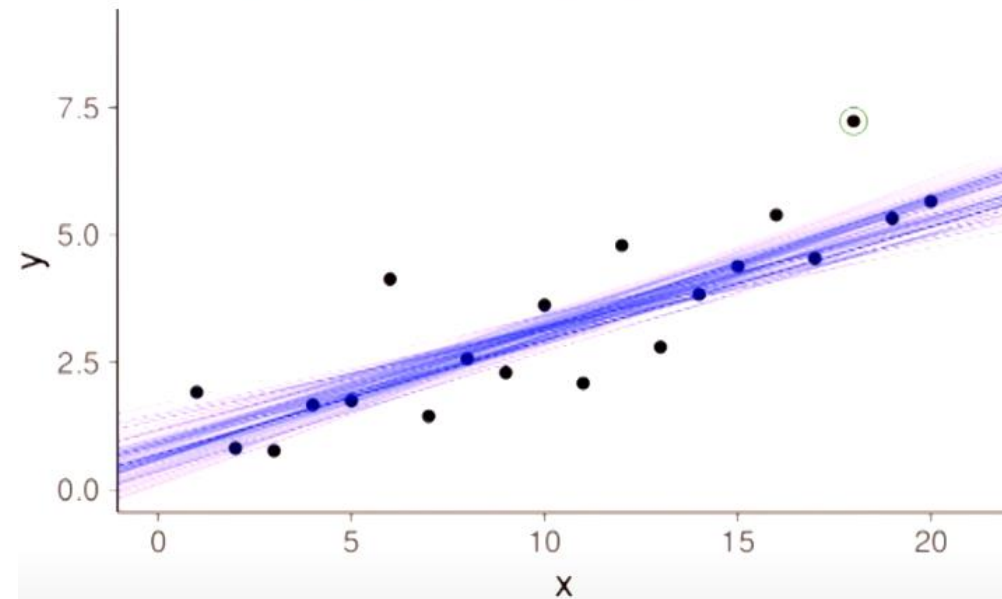
Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

Posterior predictive distribution



$$p(\tilde{y}|\tilde{x}, x, y) \approx \frac{1}{S} \sum_{s=1}^S p(\tilde{y}|\tilde{x}, \theta^{(s)})$$

PSIS-LOO weighted draws



$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

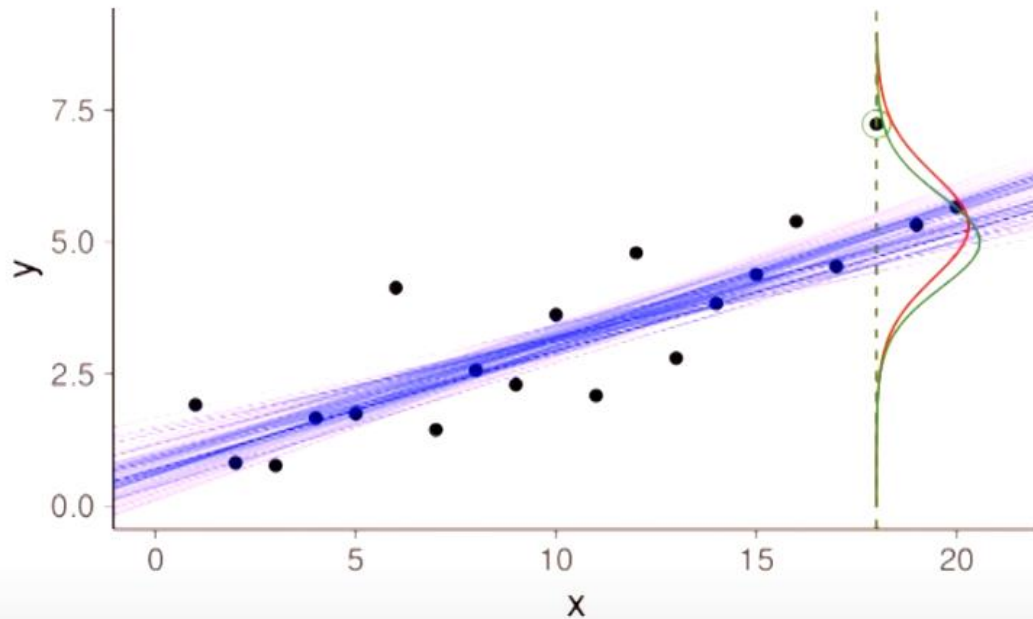
Verossimilhança avaliado nas iterações de $\theta^{(s)}$
a partir das posteriores

Computacionalmente fácil de se obter

➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

PSIS-LOO weighted predictive distribution

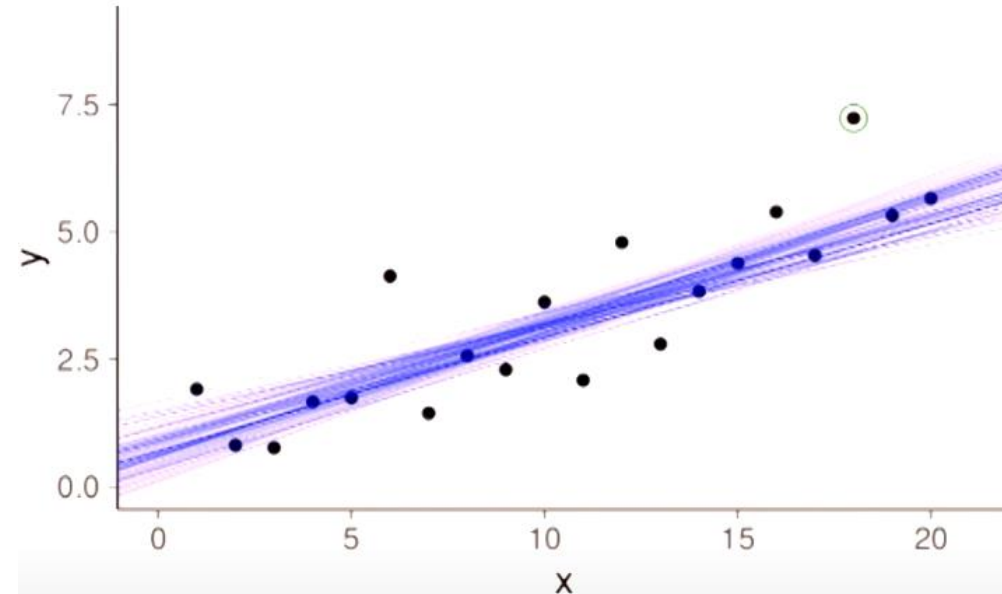


$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

$$p(y_i|x_i, x_{-i}, y_{-i}) \approx \sum_{s=1}^S [w_i^{(s)} p(y_i|x_i, \theta^{(s)})], \text{ where } w \leftarrow \text{PSIS}(r)$$

PSIS-LOO weighted draws



$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

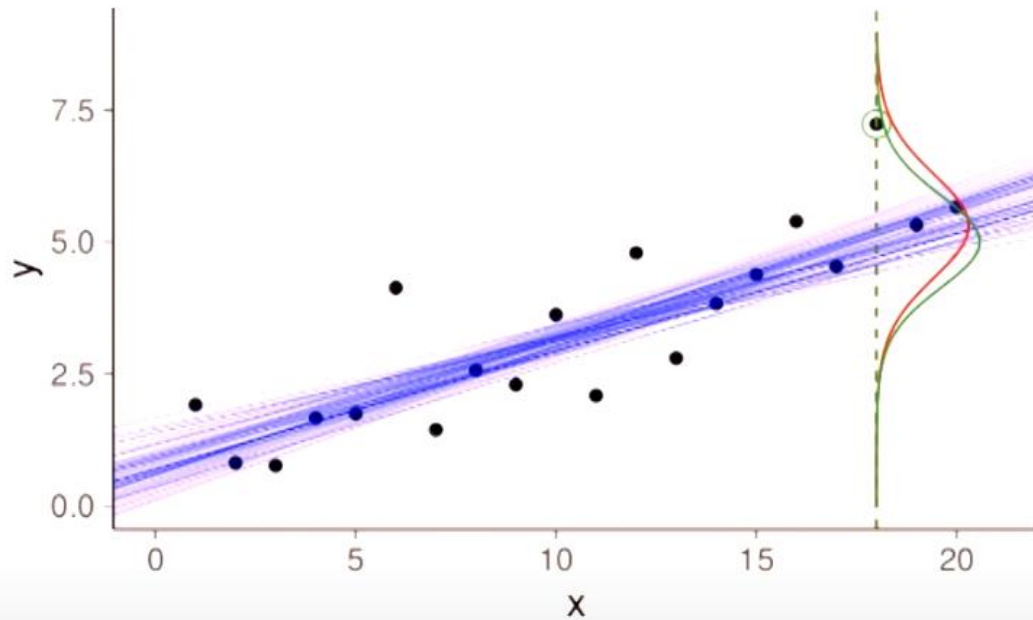
Densidade a posteriori preditiva
LOO-CV é proporcional ...

Custo computacional grande

➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

PSIS-LOO weighted predictive distribution

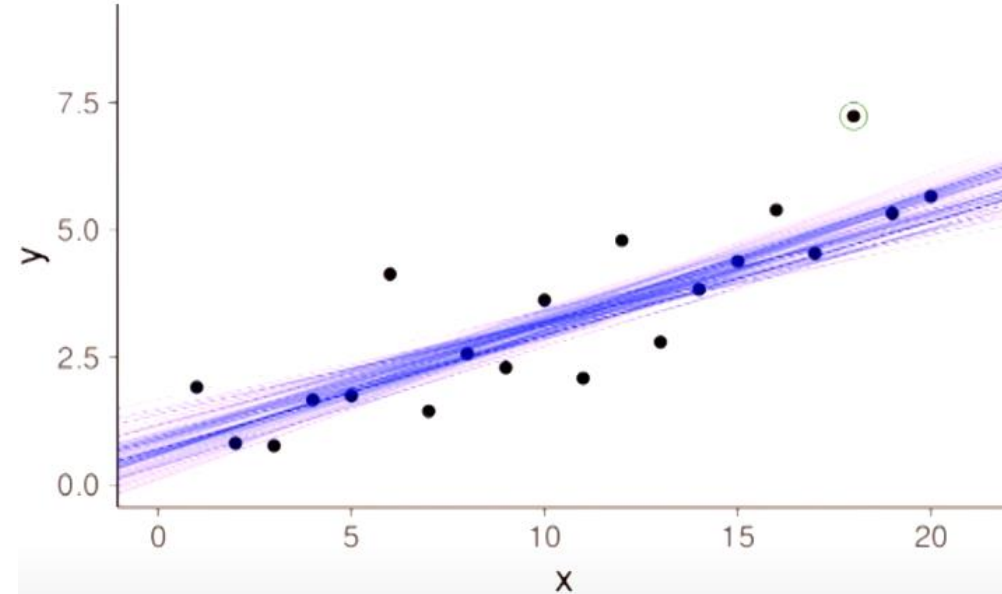


$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

$$p(y_i|x_i, x_{-i}, y_{-i}) \approx \sum_{s=1}^S [w_i^{(s)} p(y_i|x_i, \theta^{(s)})], \text{ where } w \leftarrow \text{PSIS}(r)$$

PSIS-LOO weighted draws



$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

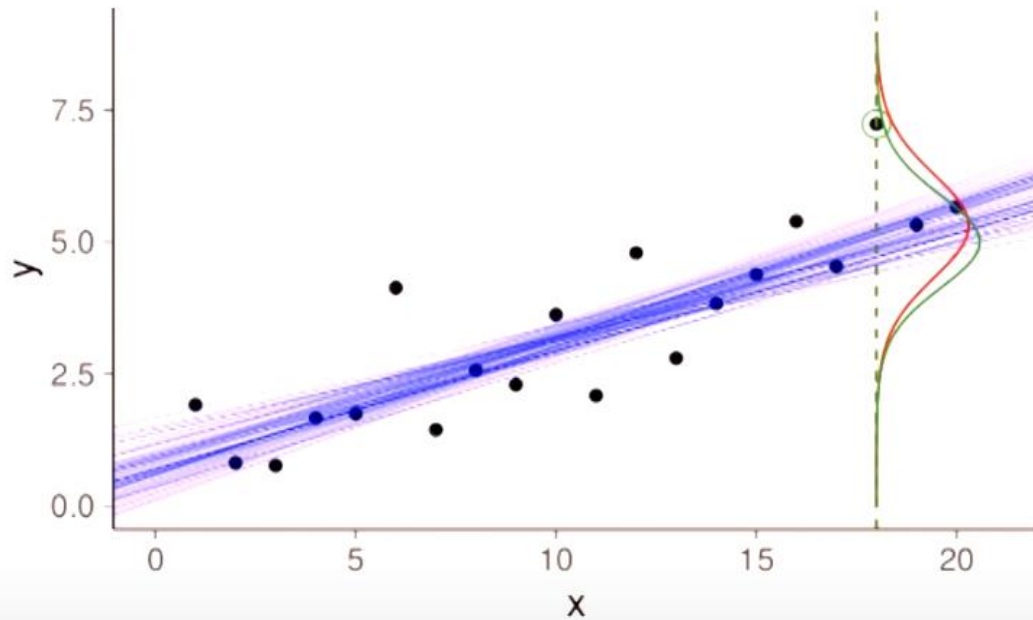
... Verossimilhança desenhada a partir das iterações de $\theta^{(s)}$ da posteriori do modelo e ...

Computacionalmente
fácil de se obter

➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

PSIS-LOO weighted predictive distribution

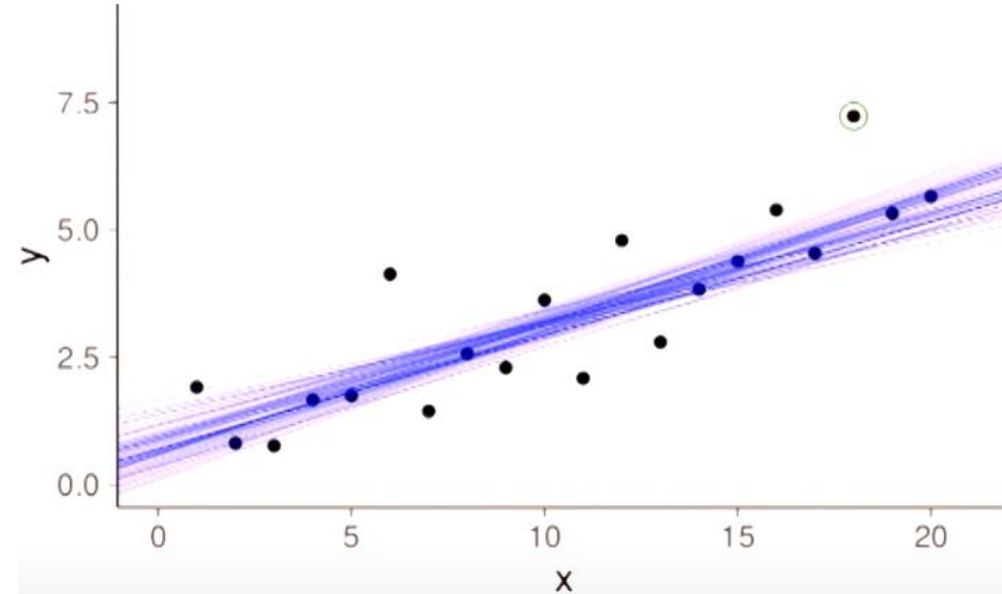


$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

$$p(y_i|x_i, x_{-i}, y_{-i}) \approx \sum_{s=1}^S [w_i^{(s)} p(y_i|x_i, \theta^{(s)})], \text{ where } w \leftarrow \text{PSIS}(r)$$

PSIS-LOO weighted draws



$$\theta^{(s)} \sim p(\theta|x, y)$$

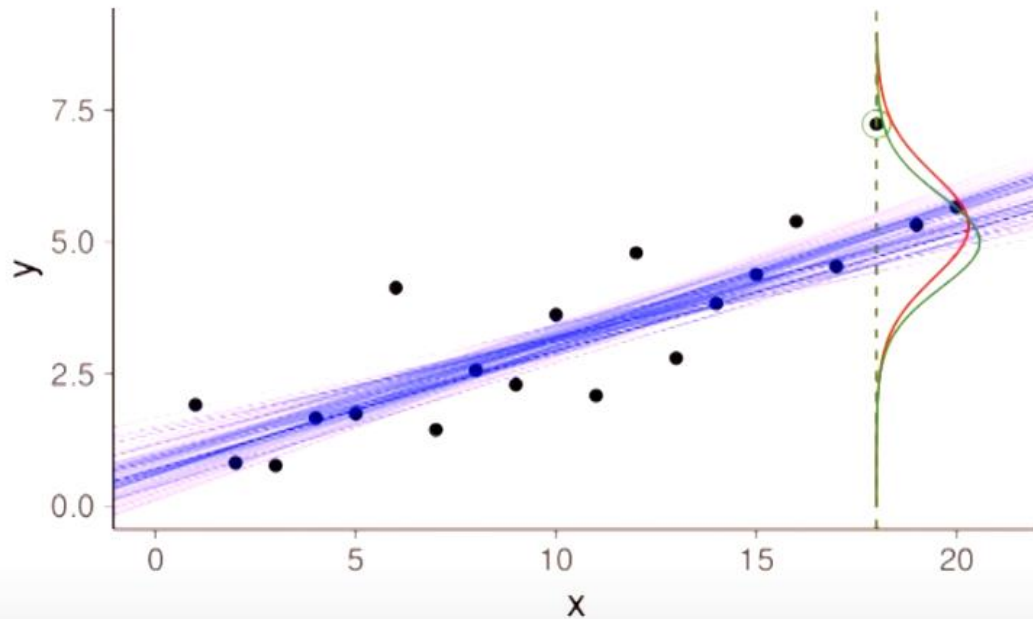
$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

... Verossimilhança ponderado pelo
Poso por importância

➤ LOO-CV Aproximado por PSIS

Leave One Out Cross Validation aproximado pelo método de Amostragem por Importância Suavizado de Pareto

PSIS-LOO weighted predictive distribution



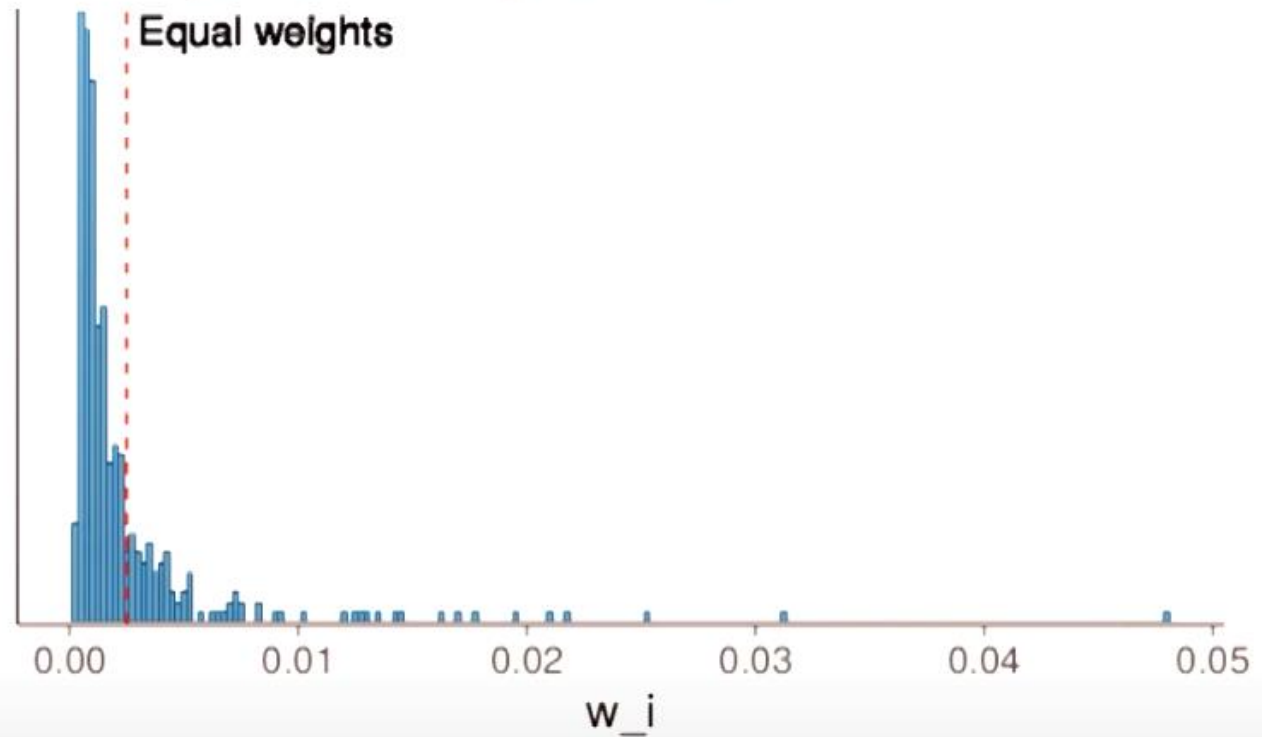
$$\theta^{(s)} \sim p(\theta|x, y)$$

$$r_i^{(s)} = p(\theta^{(s)}|x_{-i}, y_{-i})/p(\theta^{(s)}|x, y) \propto 1/p(y_i|x_i, \theta^{(s)})$$

$$p(y_i|x_i, x_{-i}, y_{-i}) \approx \sum_{s=1}^S [w_i^{(s)} p(y_i|x_i, \theta^{(s)})], \text{ where } w \leftarrow \text{PSIS}(r)$$

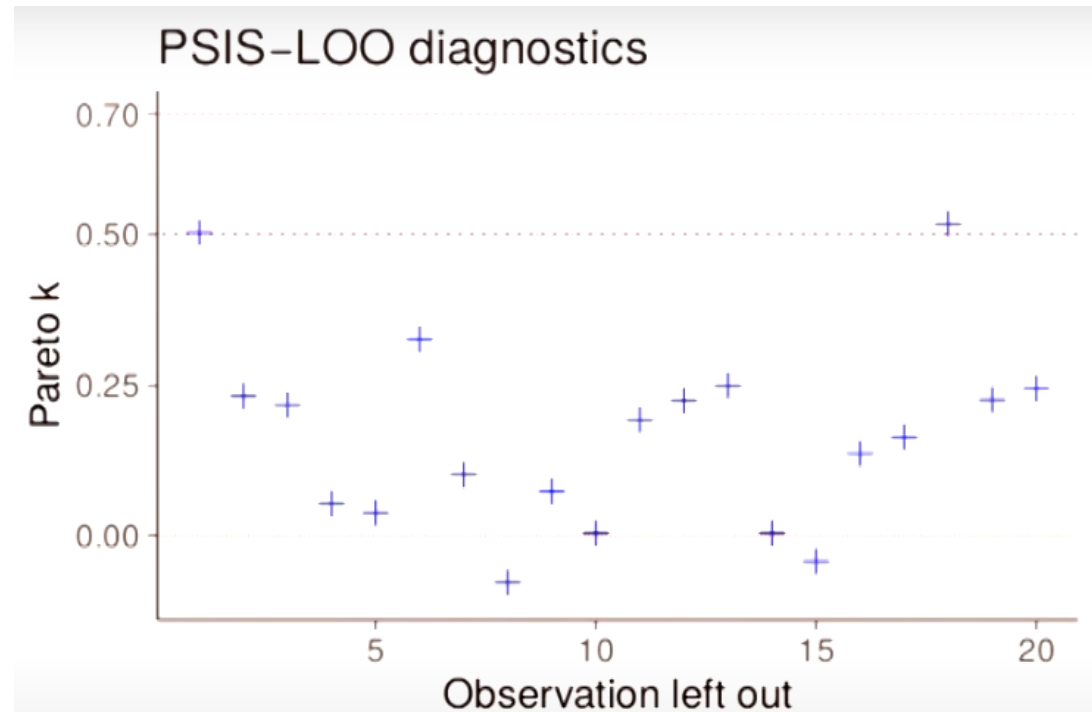
A cauda da distribuição do peso por importância que causava o problema da grande estimativa da variância

400 importance weights for leave-18th-out



Foi estabilizada pela amostragem por importância suavizada por Pareto

Já que o cálculo do tamanho efetivo de amostras pela amostragem por importância depende da estimativa dessa variância



Pareto k diagnostic values:

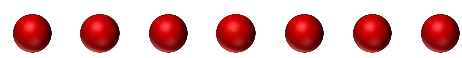
		Count	Pct.	Min.	n_eff
(-Inf, 0.5]	(good)	18	90.0%	899	
(0.5, 0.7]	(ok)	2	10.0%	459	
(0.7, 1]	(bad)	0	0.0%	<NA>	
(1, Inf)	(very bad)	0	0.0%	<NA>	

All Pareto k estimates are ok ($k < 0.7$).
See `help('pareto-k-diagnostic')` for details.

Stan code

$$\log(r_i^{(s)}) = \log(1/p(y_i|x_i, \theta^{(s)})) = -\text{log_lik}[i]$$

```
...  
model {  
  alpha ~ normal(pmualpha, psalpha);  
  beta ~ normal(pmubeta, psbeta);  
  y ~ normal(mu, sigma);  
}  
generated quantities {  
  vector[N] log_lik;  
  for (i in 1:N)  
    log_lik[i] = normal_lpdf(y[i] | mu[i], sigma);  
}
```



Saída do package loo mostrando os resultados do método de aproximação leave-one-out cross validation (LOO-CV) pelo método de amostragem por importância suavizada de Pareto (PSIS)

Este exemplo vem de uma pesquisa com residentes de uma pequena área em Bangladesh que foi afetada pelo arsênico na água potável. Os entrevistados com níveis elevados de arsênio em seus poços foram questionados se eles estavam interessados em obter água do poço de um vizinho, e uma série de regressões logísticas foram adequadas para prever essa resposta binária, dadas várias informações sobre as famílias (Gelman e Hill, 2007). Aqui, ajustamos um modelo para a resposta de troca de poço, dados dois preditores: o nível de arsênico da água na casa do residente e a distância da casa até o poço seguro mais próximo.

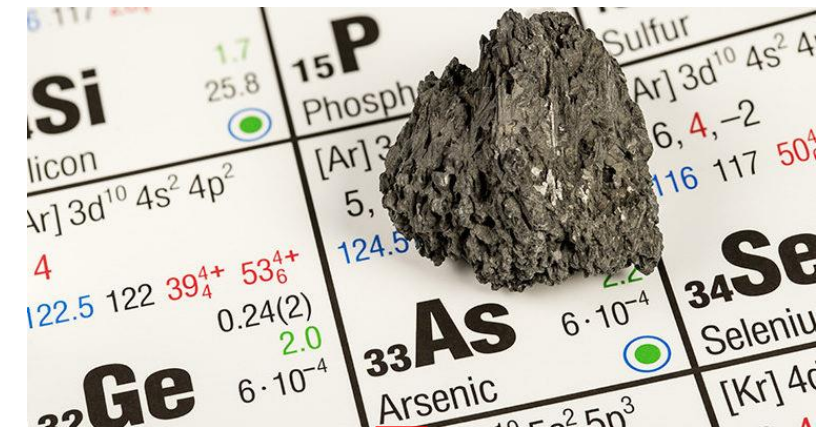
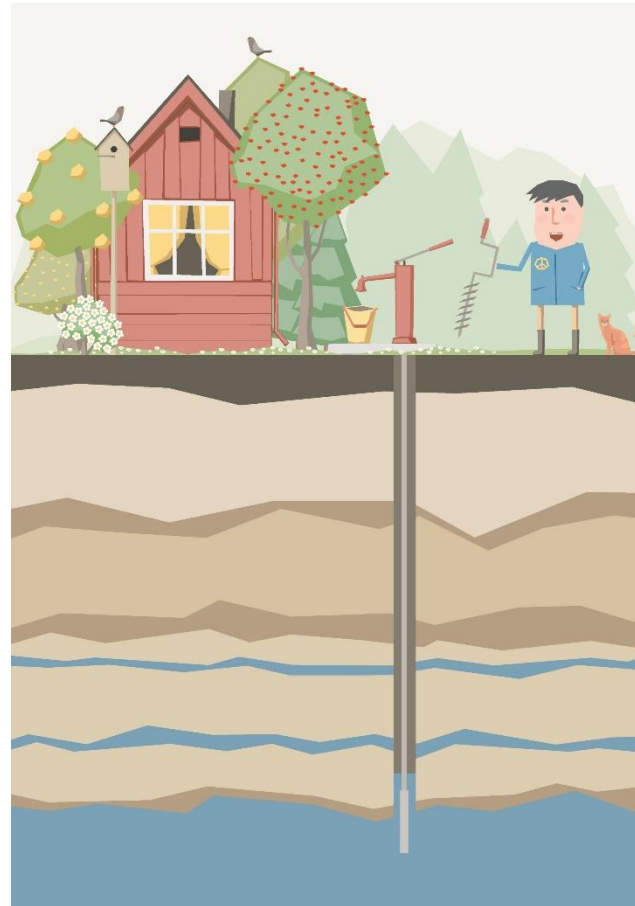
➤ Exemplo - 5 Segundo exemplo do modelo de regressão logístico

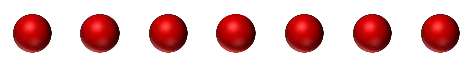
➤ Dados:

Devido a contaminação por arsênio na região de Bangladesh foi perguntado em uma entrevista com os moradores se eles estavam dispostos a obter água potável de outro poço.

- Variável dependente:
 - 1 – Sim interessado;
 - 0 – Não interessado;
- Variável independente:
 - Informações sobre a família;
 - Nível de arsênio na água dos moradores;
 - Distância da casa até o poço seguro mais próximo;

Ajustamos um modelo logístico para a resposta de troca de poço.





Saída do package loo mostrando os resultados do método de aproximação leave-one-out cross validation (LOO-CV) pelo método de amostragem por importância suavizada de Pareto (PSIS)

```
Computed from 4000 by 3020 log-likelihood matrix
```

	Estimate	SE
elpd_loo	-1968.5	15.6
p_loo	3.2	0.1
looic	3937.0	31.2

```
-----
```

```
Monte Carlo SE of elpd_loo is 0.0.
```

```
All Pareto k estimates are good (k < 0.5).  
See help('pareto-k-diagnostic') for details.
```

Saída do package loo mostrando os resultados do método de aproximação leave-one-out cross validation (LOO-CV) pelo método de amostragem por importância suavizada de Pareto (PSIS)

Critério de Informação de LOO

Número Efetivo de Parâmetros

Densidade Preditiva Logarítmica Esperada

Computed from 4000 by 3020 log-likelihood matrix

	Estimate	SE
elpd_loo	-1968.5	15.6
p_loo	3.2	0.1
looic	3937.0	31.2

Monte Carlo SE of elpd_loo is 0.0.

All Pareto k estimates are good (k < 0.5).
See help('pareto-k-diagnostic') for details.

Erro de Monte Carlo

Confiabilidade da
aproximação LOO

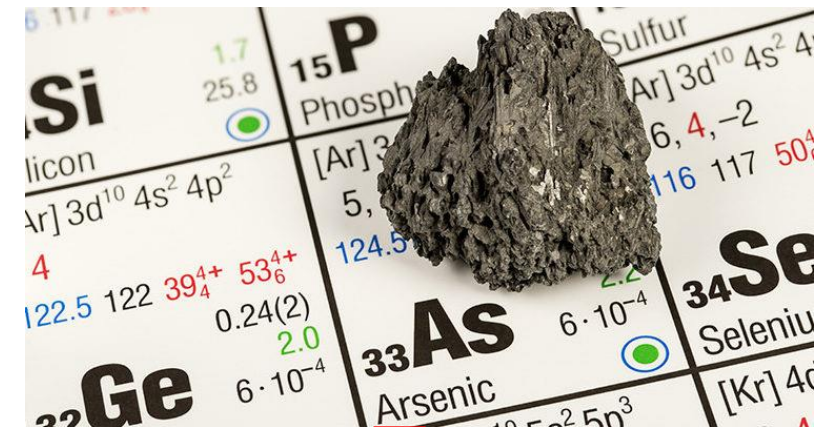
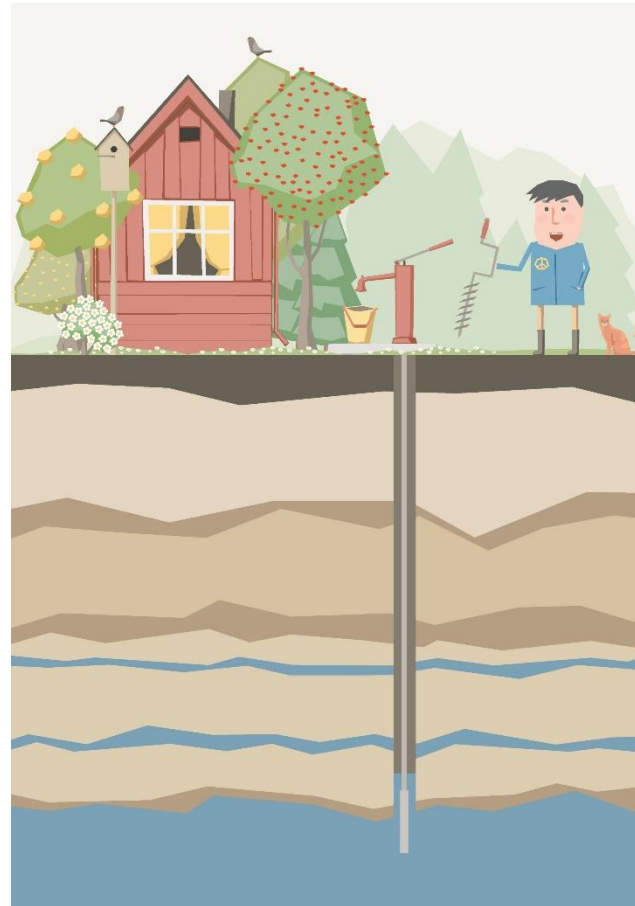
➤ Exemplo - 5 Segundo exemplo do modelo de regressão logístico

➤ Dados:

Devido a contaminação por arsênio na região de Bangladesh foi perguntado em uma entrevista com os moradores se eles estavam dispostos a obter água potável de outro poço.

- Variável dependente:
 - 1 – Sim interessado;
 - 0 – Não interessado;
- Variável independente:
 - Informações sobre a família;
 - Nível de **log(arsênio)** na água dos moradores;
 - Distância da casa até o poço seguro mais próximo;

Ajustamos um modelo logístico para a resposta de troca de poço.





Saída do package loo mostrando os resultados do método de aproximação leave-one-out cross validation (LOO-CV) pelo método de amostragem por importância suavizada de Pareto (PSIS)

Modelo 1

```
Computed from 4000 by 3020 log-likelihood matrix
```

	Estimate	SE
elpd_loo	-1968.5	15.6
p_loo	3.2	0.1
looic	3937.0	31.2

```
-----  
Monte Carlo SE of elpd_loo is 0.0.
```

```
All Pareto k estimates are good (k < 0.5).  
See help('pareto-k-diagnostic') for details.
```

Modelo 2

```
Computed from 4000 by 3020 log-likelihood matrix
```

	Estimate	SE
elpd_loo	-1952.2	16.2
p_loo	3.1	0.1
looic	3904.5	32.4

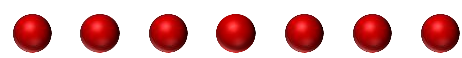
```
-----  
Monte Carlo SE of elpd_loo is 0.0.
```

```
All Pareto k estimates are good (k < 0.5).  
See help('pareto-k-diagnostic') for details.
```

	elpd_diff	se_diff
model2	0.0	0.0
model1	-16.2	4.4

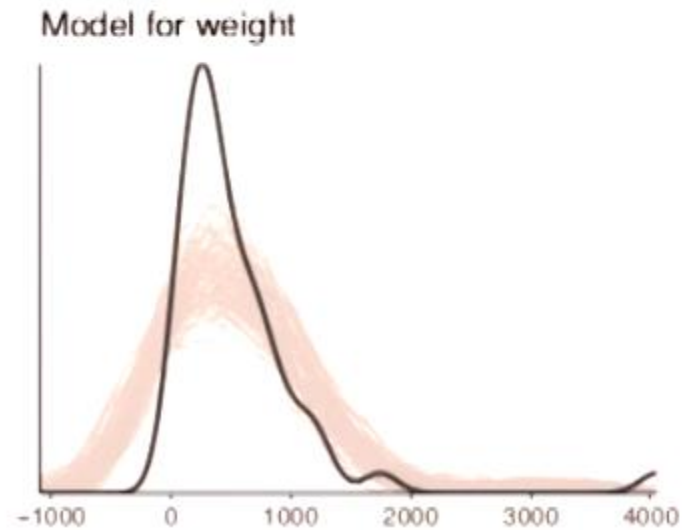
A primeira coluna mostra a diferença no ELPD em relação ao modelo com o maior ELPD

Nesse caso, a diferença em elpd e sua escala com relação ao erro padrão aproximado da diferença) indica uma preferência pelo segundo modelo

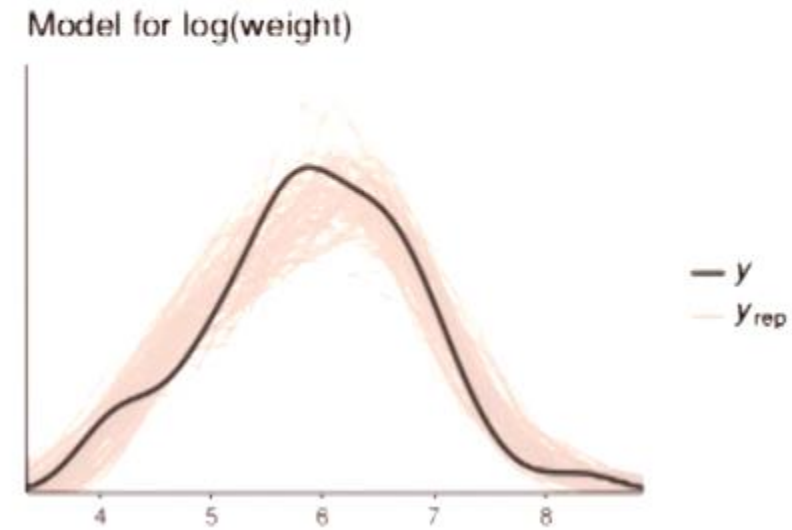


Sometimes cross-validation is not needed

- Posterior predictive checking is often sufficient



Modelo 1



Modelo 2

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: 10/03/2022

AGRADECIMENTO E COLABORADORES:



UFOPA