



**Universidade de Brasília
Departamento de Estatística**

Estimação da cobertura de mortalidade de óbitos no Peru
via modelos hierárquicos bayesianos

Irenaldo Ferreira Lima Júnior

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2023**

Irenaldo Ferreira Lima Júnior

Estimação da cobertura de mortalidade de óbitos no Peru
via modelos hierárquicos bayesianos

Orientador(a): Prof. Dr. Leandro Tavares Correia

Projeto apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2023

Resumo

A qualidade de dados públicos é essencial para criação de políticas públicas eficientes, que priorizem quem realmente necessita e utilize os recursos disponíveis da melhor forma possível. Existem diferentes formas e modelos para estimar o número real de óbitos ocorridos em uma população, mas muitos ignoram o agrupamento natural que os dados apresentam. A estrutura hierárquica é a divisão de um grupo de dados em subgrupos, no caso de um país como o Peru, os dados se subdividem em departamentos, que se subdividem em províncias, que por sua vez, se subdividem em distritos. O objetivo deste trabalho é criar um modelo hierárquico que considera esses agrupamentos de dados utilizando a abordagem bayesiana, comparar com resultados obtidos analiticamente e discutir os resultados obtidos. O modelo se mostrou eficiente em estimar os óbitos registrados e, conseqüentemente, a cobertura de mortalidade dos bancos do Sistema Informático Nacional de Óbitos (SINADEF). Comparando com os resultados apresentados pela metodologia clássica, o modelo apresentou um menor número de zeros possibilitando o cálculo de cobertura para regiões menores, além disso, mais de 99% dos distritos tiveram a estimação do modelo clássico contida no intervalo de credibilidade construído com base no modelo bayesiano. Além disso, é claro o reflexo do esforço por parte do Peru para melhorar a qualidade de seus dados, que tinham aproximadamente 30% dos registros feitos de forma online em 2017 e tinha quase 80% destes mesmos registros realizados em sistemas online em 2020.

Palavras-chave: Cobertura de Mortalidade, Demografia, Modelos Hierárquicos, Regressão Bayesiana

Lista de Tabelas

1	Cobertura estimada de óbitos registrados no SINADEF por sexo segundo localizações - Método GGB. Peru, 2017.	20
2	Tábua de vida abreviada de mortalidade para homens, Peru, 2015-2020 . . .	22
3	Óbitos esperados por grupos de idade. Província de Lima, 2017 - 2020 . . .	23
4	Óbitos esperados por grupos de idade. Distrito San Juan de Lurigancho, 2017 - 2020	24
5	Óbitos esperados por grupos de idade. Distrito El parco, 2017 - 2020 . . .	24
6	Óbitos esperados por grupos de idade. Distrito El parco, 2017	29
7	Óbitos esperados por grupos de idade. Distrito El parco, 2018	32
8	Óbitos esperados por grupos de idade. Distrito El parco, 2019	36
9	Óbitos esperados por grupos de idade. Distrito El parco, 2020	40

Lista de Figuras

1	Representação da estrutura hierárquica	9
2	Representação da estrutura hierárquica do Peru	16
3	Pirâmide etária do Peru - 2007 e 2017	18
4	Média de registros de óbitos realizados de forma online no Peru	18
5	Convergência dos parâmetros do modelo bayesiano, 2017	25
6	Convergência dos parâmetros do modelo bayesiano, 2017	26
7	Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2017	27
8	Checagem da distribuição preditiva a <i>posteriori</i> do modelo bayesiano, 2017	27
9	Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2017	28
10	Convergência dos parâmetros do modelo bayesiano, 2018	29
11	Convergência dos parâmetros do modelo bayesiano, 2018	30
12	Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2018	31
13	Checagem da distribuição preditiva a <i>posteriori</i> do modelo bayesiano, 2018	31
14	Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2018	32
15	Convergência dos parâmetros do modelo bayesiano, 2019	33
16	Convergência dos parâmetros do modelo bayesiano, 2019	33
17	Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2019	34
18	Checagem da distribuição preditiva a <i>posteriori</i> do modelo bayesiano, 2019	35
19	Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2019	35
20	Convergência dos parâmetros do modelo bayesiano, 2020	37
21	Convergência dos parâmetros do modelo bayesiano, 2020	37

22	Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2020	38
23	Checagem da distribuição preditiva a <i>posteriori</i> do modelo bayesiano, 2020	39
24	Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2020	39

Sumário

1 Introdução	8
2 Referencial Teórico	10
2.1 Inferência Bayesiana	10
2.1.1 Teorema de Bayes	10
2.1.2 Modelos Hierárquicos Bayesianos	11
2.2 Algoritmos MCMC	13
2.2.1 Algoritmo Metropolis-Hastings	13
2.2.2 Amostrador de Gibbs	13
2.3 Métodos de Distribuição de Óbitos (DDM)	14
2.3.1 General Growth Balance (GGB)	14
2.3.2 Synthetic Extinct Generation (SEG)	15
3 Metodologia	16
3.1 Conjunto de dados	16
3.2 Modelo	16
3.3 Métodos Clássicos	17
4 Resultados	21
4.1 Metodologia Analítica	21
4.2 Métodos Bayesianos	25
4.2.1 Modelo 2017	25
4.2.2 Modelo 2018	29
4.2.3 Modelo 2019	33
4.2.4 Modelo 2020	36
5 Conclusão	41
Referências	42
Apêndice	43

1 Introdução

A eficácia e eficiência de políticas públicas passa por dados confiáveis e fidedignos a realidade. De acordo com AbouZahr e Boerma (2005) as decisões corretas para a melhora da saúde pública só podem ser tomadas se tiverem como suporte informações de boa qualidade sobre a saúde da população alvo como mortalidade, morbidade e causas de morte, que, por sua vez, dependem de um sistema que registre esses eventos com boa cobertura e legitimidade. Dados lícitos aumentam o conhecimento da dinâmica demográfica e auxiliam no planejamento fiscal e na criação de políticas sociais que utilizem os recursos públicos de forma mais eficiente.

Estudos que tenham como objetivo examinar mortalidade ainda enfrentam dificuldades devido à falta de qualidade dos dados, sendo ainda mais agravados quando se examina os grupos em níveis sub-nacionais, principalmente em áreas mais pobres e de difícil acesso (QUEIROZ et al., 2017). Em 2010 o Peru tinha uma das piores qualidades de informação relacionadas a mortalidade do mundo, com mais de 45% das causas de morte sendo classificadas como pouco úteis ou mal definidas (NAGHAVI et al., 2010) e em 2015 atingiu um nível baixo no Índice de desempenho de estatísticas Vitais (VSPI), um indicador que avalia a cobertura, qualidade, nível de detalhe das causas de morte, a consistência interna, a qualidade dos registros por idade e sexo e a disponibilidade dos dados (MIKKELSEN et al., 2015). Além disso, a estimação de mortalidade em pequenas áreas é um desafio devido obviamente a pequena população observada em alguns pontos e alta variabilidade amostral nas mortes registradas (PLETCHER, 1999).

Os dados de mortalidade apresentam naturalmente uma estrutura hierárquica, como representado na Figura 2, onde podemos observar uma relação de grupos e sub-grupos, onde os óbitos observados em um nível podem ser classificados em um subgrupo de outro nível e assim por diante. Um exemplo seria um país, como o Peru, com níveis hierárquicos em sua divisão geográfica, sendo eles os departamentos, províncias e distritos. Os distritos são subgrupos de uma província, que por sua vez, são sub-grupos de um departamento. A Figura 2 exemplifica essa situação.

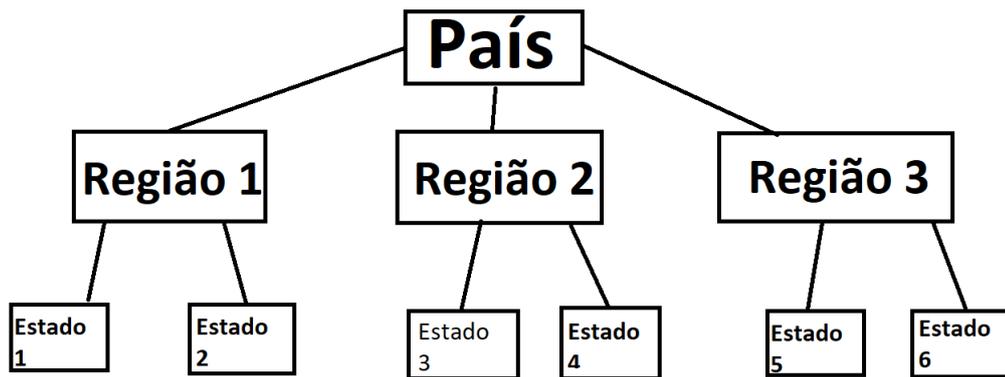


Figura 1: Representação da estrutura hierárquica

Modelos de regressão linear desconsideram esses agrupamentos e geram estimativas para o todo, levando a uma perda das informações individuais de cada nível da estrutura hierárquica. É nesse contexto que surgem os modelos hierárquicos que gera estimativas para cada um dos níveis presentes, diminuindo a variação nas estimativas pela existência de semelhanças entre os parâmetros dos modelos individuais.

Novas abordagens utilizando modelos Bayesianos para estimação da cobertura de mortalidade em pequenas áreas mostraram-se muito eficientes, pois esse modelo permite utilizar de informações previamente conhecidas como idade, tempo, sexo para aumentar a precisão das estimativas de mortalidade e deixar claro a incerteza associada a essas estimativas (SCHMERTMANN; GONZAGA, 2018).

O presente trabalho tem em vista comparar resultados analíticos de estimação de cobertura de mortalidade com estimações realizadas por modelos hierárquicos bayesianos. Para isso serão apresentados na introdução a motivação do tema e uma breve explicação do que é estrutura hierárquica e por que essa técnica será utilizada para estimação dos registros de óbitos e, conseqüentemente, da cobertura de mortalidade. O referencial teórico expõe a revisão bibliográfica, esclarecendo as técnicas estatísticas utilizadas no decorrer deste trabalho. A Metodologia explicita o conjunto de dados utilizado nas análises e os detalhes do modelo proposto, como variáveis consideradas, sintaxe e forma dos modelos. Os resultados apresentam a aplicação das técnicas expostas anteriormente e os resultados obtidos. A conclusão descreve as principais percepções geradas pelos resultados e discorre brevemente sobre o desempenho do modelo. Ao final, estão disponíveis as referências utilizadas para construção deste trabalho e os códigos empregados nas modelagens.

2 Referencial Teórico

2.1 Inferência Bayesiana

A inferência Bayesiana tem como propósito a estimação de parâmetros de um modelo assumindo uma distribuição de probabilidade que contenha a informação dos dados da amostra.

2.1.1 Teorema de Bayes

A inferência Bayesiana se constrói entorno do teorema de Bayes, um corolário da lei de probabilidade total, que relaciona as probabilidades marginais e condicionais. O teorema de Bayes é dado por:

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{j=1}^n P(A_j)P(B/A_j)}. \quad (2.1.1)$$

O teorema de Bayes representa a probabilidade de um evento a partir da informação a *priori*. No contexto inferencial, o teorema de Bayes será a quantidade conhecida θ e pelo conjunto que contém a informação a *priori* sobre θ .

Definindo $p(\theta|H)$ como a informação sobre o parâmetro dado o conjunto de informações sobre ele e $p(y|\theta, H)$ e $p(y|H)$ como a nova informação obtida através da amostra, tem-se a forma de obtenção de informação do parâmetro através da distribuição inicial e da informação dos dados dada por:

$$p(\theta|y, H) = \frac{p(y|\theta, H)p(\theta|H)}{p(y|H)}, \quad (2.1.2)$$

e sendo $p(y|H)$ a distribuição preditiva de y com base na informação de H , dada por:

$$p(y|H) = \int_{\Theta} p(y, \theta|H)d\theta = \int_{\Theta} p(y|\theta, H)p(\theta|H)d\theta. \quad (2.1.3)$$

Tem-se, portanto, que $p(\theta|H)$ será a distribuição a *priori* de θ , pois representa a distribuição anterior a investigação dos dados y . Já $p(y|\theta, H)$ é a função de verossimilhança, pois representa o parâmetro a partir das observações. E $p(\theta|y, H)$ é a distribuição a posteriori, pois representa o resultado após a obtenção dos dados. A partir disso a equação (2.1.3) pode ser simplificada por:

$$p(\theta|y, H) \propto p(y|\theta, H)p(\theta, H) \quad (2.1.4)$$

Portanto, na inferência Bayesiana a estimação é o resultado da junção da distribuição a *priori* e da observada nos dados, resultando por fim na distribuição a *posteriori* de θ . Ao contrário da inferência Bayesiana que utiliza a função de verossimilhança e a informação os dados observados, a inferência clássica utiliza somente a função de verossimilhança.

2.1.2 Modelos Hierárquicos Bayesianos

Como visto, a inferência Bayesiana ajusta a informação da estrutura dos dados com a informação obtida da distribuição a *priori*. Na modelagem, a estrutura de linearidade se ajusta bem com a estrutura de modelagem hierárquica para modelos normais, apresentando resultados similares.

Serão apresentados os três modelos mais comuns com a ótica Bayesiana. A formulação das prioris é igual para os três assim como sua equação, diferindo-se apenas na matrix $X_1\beta_1$ que especificará o tipo de modelo misto: efeitos fixos, interceptos aleatórios ou interceptos e coeficientes aleatórios.

As suposições de linearidade e normalidade são adicionadas a estrutura dos dados, sendo o modelo ajustado para todos os estágios do modelo. Um modelo hierárquico normal Bayesiano de 2 níveis, terá as seguintes prioris para os dois níveis:

$$\begin{aligned} Y|\beta_1, \phi &\sim N(X_1\beta_1, \phi^{-1}I_n), \\ \beta_1|\beta_2, \phi &\sim N(X_2\beta_2, \phi^{-1}C_1^{-1}), \\ \beta_2|\phi &\sim N(\mu, \phi^{-1}C_2^{-1}), \\ n_0\sigma_0^2\phi &\sim X_{n_0}^2, \end{aligned}$$

com X_1 sendo a matriz de variáveis do primeiro nível e X_2 a de variáveis do segundo nível. β_1 e β_2 são os coeficientes de cada nível.

Os próximos passos são obter as distribuições a *posteriori* de β_1, β_2 e ϕ . Observe-se que a inferência de β_1 tem sua distribuição a *posteriori* condicionada a β_2 e ϕ sendo substituídos por seus estimadores de ligação quando integrado pela *posteriori* marginal de β_1 .

Para o modelo de efeitos fixos temos:

$$y_i \sim N(\mu_i, \sigma^2),$$

$$\mu_i = \alpha_i + \beta_i x_i,$$

$$\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix} \sim N \left[\begin{pmatrix} \alpha \\ \beta_p \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_{\beta_p}^2 \end{pmatrix} \right].$$

Para o modelos de interceptos aleatórios, temos a seguinte determinação *a priori*:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2),$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^P \beta_p x_{pij},$$

$$\alpha_{0j} \sim N(\alpha_0 + \omega_{0j}, \sigma^2).$$

E para o modelo com interceptos e coeficientes aleatórios temos:

$$y_{ij} \sim N(\mu_{ij}, \sigma^2),$$

$$\mu_{ij} = \alpha_{0j} + \sum_{p=1}^P \beta_p x_{pi},$$

$$\alpha_{0j} \sim N(\alpha_0 + \omega_{0j}, \sigma_{0j}^2),$$

$$\beta_{pj} \sim N(\beta_p + \omega_{pj}, \sigma_j^2).$$

2.2 Algoritmos MCMC

É muito comum que a distribuição *a posteriori* não possa ser obtida analiticamente e para contornar esse problema pode-se utilizar algoritmos MCMC (Monte Carlo via Cadeias de Markov), que permitem simular variáveis aleatórias (iterativamente) a partir das distribuições condicionais completas para se obter uma distribuição *a posteriori* numericamente. Através de uma cadeia de Markov homogênea, ergódica e irredutível simula-se uma densidade $p(\cdot)$, sendo a distribuição estacionária a distribuição de $p(\cdot)$. Seus métodos mais comuns são o Metropolis-Hastings e o Amostrador de Gibbs.

2.2.1 Algoritmo Metropolis-Hastings

O algoritmo de Metropolis-Hastings (HASTINGS, 1970) propõe que, seja $q(\theta, \cdot)$ o núcleo de uma distribuição em que se consegue gerar valores de θ e sendo θ^k o estado da cadeia de Markov na k -ésima iteração, a posição $k+1$ pode ser gerada da seguinte forma:

- Gerar um θ^* com base em $q(\theta^k, \cdot)$
- Aceitar o valor gerado com probabilidade:

$$\alpha(\theta^k, \theta^*) = \min\left(1, \frac{\frac{p(\theta^*)}{q(\theta^k, \theta^*)}}{\frac{p(\theta^k)}{q(\theta^*, \theta^k)}}\right)$$

Sendo assim, $\theta^{k+1} = \theta^*$, caso se aceite, e $\theta^{k+1} = \theta^k$ com probabilidade $1 - \alpha(\theta^k, \theta^*)$, caso se rejeite.

2.2.2 Amostrador de Gibbs

Já o algoritmo do Amostrador de Gibbs simula valores utilizando como núcleo da distribuição o produto das distribuições condicionais completas. Executa-se da seguinte forma:

- Define-se um "chute inicial" para θ^0
- Gera-se $\theta_1^{(k)}$ de $f(\theta_1 | \theta_{-1}^{(k-1)}, x)$, que é a distribuição condicional completa.
- Gera-se $\theta_2^{(k)}$ de $f(\theta_2 | \theta_1^{(k)}, \theta_3^{(k-1)}, \dots, \theta_j^{(k-1)}, x)$.

⋮

- Gera-se $\theta_{j-1}^{(k)}$ de $f(\theta_{j-1}|\theta_1^{(k)}, \dots, \theta_{j-2}^{(k)}, \theta_j^{(k-1)}, x)$.
- Gera-se $\theta_j^{(k)}$ de $f(\theta_j|\theta_{-j}^{(k)}, x)$

2.3 Métodos de Distribuição de Óbitos (DDM)

Os métodos de distribuição de óbitos são utilizados para estimar a cobertura de mortalidade adulta, que comparam a distribuição de óbitos por idade com a distribuição etária da população, provendo assim o padrão etário de mortalidade para um dado período. Os dois principais métodos são o General Growth Balance (GGB), proposto por Hill (1987), e o Synthetic Extinct Generation (SEG), proposto por Bennett e Horiuchi (1981). Seguem os pressupostos adotados relacionados a dinâmica demográfica:

- A população não está sujeita a migração;
- A cobertura dos óbitos por idade é constante;
- A cobertura da contagem populacional por idade é constante; e
- As idades de vivos e dos óbitos não contêm erros.

2.3.1 General Growth Balance (GGB)

O método GGB foi desenvolvido por Hill (1987), tomando como base a equação básica de equilíbrio demográfico, que define a taxa de entrada em uma idade $a+$, $b(a+)$, como igual à taxa de crescimento na idade $a+$, $r(a+)$, mais taxa de saída (morte) na idade $a+$, $d(a+)$, como evidenciado abaixo:

$$b(a+) = r(a+) + d(a+)$$

Em populações fechadas, a taxa de crescimento é constante para todas as idades, fazendo com que a taxa de entrada e a taxa de mortes sejam linearmente relacionadas. Ou seja, através da diferença entre a taxa de entrada e a taxa de crescimento em relação à taxa de mortalidade em cada idade, pode-se estimar um intercepto que evidencia qualquer alteração na cobertura de óbitos entre dois censos.

2.3.2 Synthetic Extinct Generation (SEG)

O método SEG criado por Bennett e Horiuchi (1981), utiliza-se de taxas específicas de crescimento para cada idade para transformar uma distribuição de óbitos em uma estimativa da distribuição etária daquela população, ou seja, tomando os óbitos de uma idade $x+$, podemos estimar a população com idade x . O grau de cobertura dos óbitos é então calculado com base na razão entre óbitos estimados pela população acima da idade x e a população observada, de fato.

3 Metodologia

3.1 Conjunto de dados

Serão utilizados dados do Sistema Informático Nacional de Óbitos (SINADEF) de 2017 até 2020 com o intuito de avaliar a qualidade de cobertura dos registros de óbitos nas regiões do Peru e compará-las também no nível de Província, além de dados do Censo de 2007 e 2017 e as projeções populacionais até 2020 por sexo e idade para regiões, províncias e distritos realizadas no projeto Vasconcelos et al. (2022).

3.2 Modelo

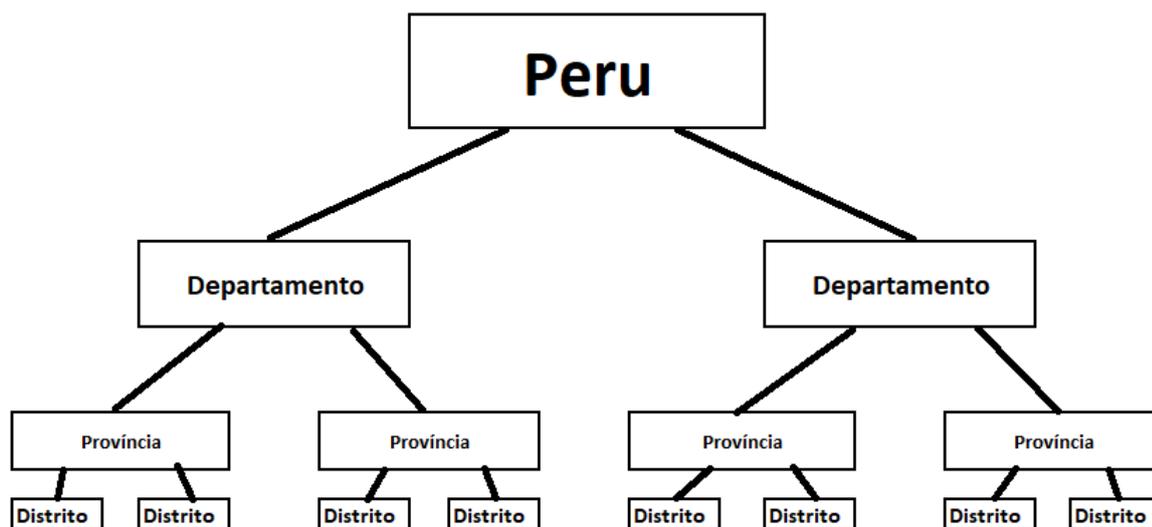


Figura 2: Representação da estrutura hierárquica do Peru

As variáveis consideradas para criação do modelo são a idade, dividida em 3 classes, de $[0,5)$, $[5,65)$ e $[65,+)$ com o intuito de diminuir a chance de ocorrências de 0 nas estimativas. Serão consideradas ainda a População e número de mortes. Cada uma dessas variáveis estará disposta dentro de 3 níveis hierárquicos, como já citado anteriormente, os departamentos, as províncias e os distritos. Dado a escala da população e a grande variação é proposto um modelo que considera o logaritmo da população. A sintaxe e forma do modelo é apresentada abaixo:

$$\text{Modelo} = \text{Óbitos} \sim \text{Log}(\text{População}) + \text{Idades} + (1|\text{Departamento}) + (1|\text{Província}) + (1|\text{Distrito})$$

$$\text{Óbitos} \sim \text{Poisson}(\lambda_i)$$

$$\text{Log}(\lambda_i) = \alpha + \alpha_{\text{Departamento}_i} + \alpha_{\text{Província}_i} + \alpha_{\text{Distrito}_i} + \beta_1 \text{Log}(\text{População}) + \beta_j \text{Idades}$$

$$\alpha_{\text{Departamento}} \sim \text{Normal}(0, \sigma_{\text{Departamento}})$$

$$\alpha_{\text{Província}} \sim \text{Normal}(0, \sigma_{\text{Província}})$$

$$\alpha_{\text{Distrito}} \sim \text{Normal}(0, \sigma_{\text{Distrito}})$$

$$\alpha \sim \text{T-Student}(3, 2.2, 2.5)$$

$$\beta_1 \sim \text{Normal}(0, 100)$$

$$\beta_j \sim \text{Normal}(0, 100)$$

$$\sigma_{\text{Departamento}} \sim \text{T-Student}(3, 0, 2.5)$$

$$\sigma_{\text{Província}} \sim \text{T-Student}(3, 0, 2.5)$$

$$\sigma_{\text{Distrito}} \sim \text{T-Student}(3, 0, 2.5)$$

As distribuições dos parâmetros β_1 e β_j , onde j representa os 3 grupos de idade existentes, foram escolhidas para que as distribuições *a priori* fossem não-informativas. As distribuições dos demais parâmetros foram definidas por *default* no pacote computacional “BRMS” do software R.

3.3 Métodos Clássicos

A população do Peru é de 31.237.884 pessoas, segundo o Censo realizado em 2017. Observa-se na Figura 3, que a base da pirâmide etária peruana foi reduzida ao longo das últimas décadas com um aumento no centro, que reflete menos nascimentos e maior população em idade ativa. A divisão regional do Peru apresenta em seu primeiro nível de subdivisão administrativa as regiões, também conhecidas como departamentos, que, se subdividem em seu segundo nível de subdivisão administrativa, as províncias, que, por sua vez, se dividem em distritos. Essa estrutura apresenta um desafio para a estimação de óbitos, principalmente nos dois últimos níveis administrativos, pois ainda que existam províncias e distritos com populações consideráveis, como a província de Lima com 8.512.046 pessoas e o distrito San Juan de Lurigancho com 1.038.495, existem outros com populações menores, como a província de Purus com 2.860 pessoas e o distrito de Huampara com apenas 149 pessoas.

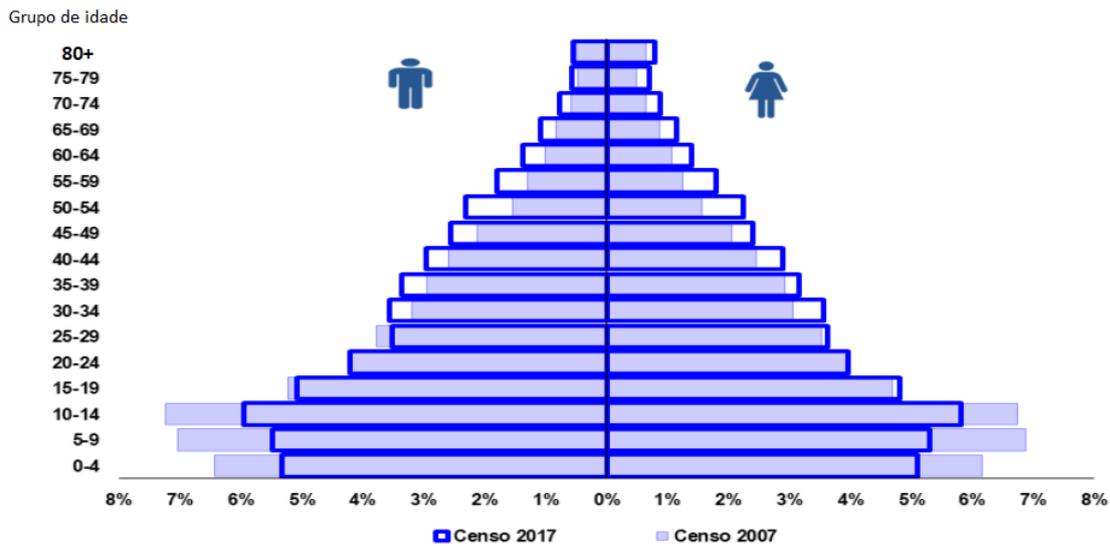


Figura 3: Pirâmide etária do Peru - 2007 e 2017

A qualidade das informações relativas ao registro de óbitos de um país são crucias para possibilitar que o governo e outros agentes, nacionais ou internacionais, tenham uma boa utilização dos recursos e criem ações, com maior efetividade, ações que variam desde políticas públicas, econômicas, previdenciárias e etc. Nos últimos anos o Peru vêm despendendo esforços para melhorar a qualidade dos seus registros, como mostra a Figura 4.

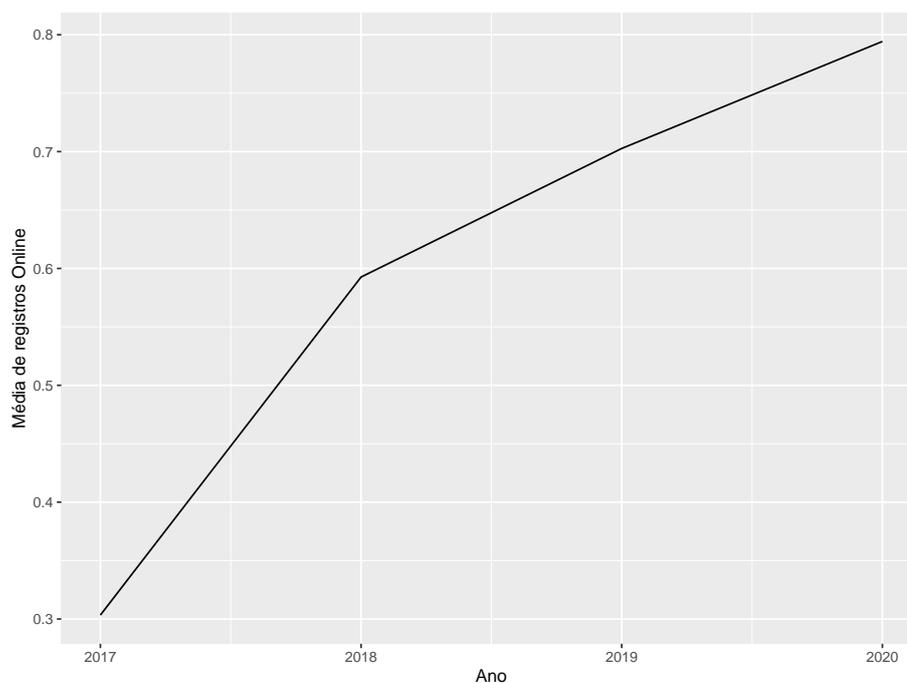


Figura 4: Média de registros de óbitos realizados de forma online no Peru

Como mencionado anteriormente, uma das formas de estimar a mortalidade de uma população é através dos Métodos de Distribuição de Óbitos (DDM), sendo possível destacar o General Growth Balance, proposto por Hill (1987), e o Synthetic Extinct Generation (SEG), proposto por Bennett e Horiuchi (1981). Ademais, para aplicação do SEG são necessários dados da população em dois pontos distintos de tempo e os respectivos registros de óbitos deste período, porém o presente estudo se propõe a estimar as coberturas de mortalidade entre os anos 2017 à 2020, estando disponível somente o censo realizado em 2017, impossibilitando a utilização deste método.

As coberturas estimadas para o Peru e suas regiões pelo método GGB no ano de 2017 são apresentados na Tabela 1. Foram utilizados os dados do Censo de 2017, do SINADEF do mesmo ano e as planilhas de cálculo disponíveis on-line pela União Internacional para o Estudo Científico da População (IUSSP) (<http://demographicestimation.iussp.org/>).

É importante lembrar que o Método GGB tem pressupostos muito rigorosos, como população fechada e com taxas de mortalidade constantes, podendo, portanto, subestimar as coberturas de mortalidade das regiões. Outrossim, pode-se observar uma grande heterogeneidade entre os valores estimados entre as regiões, como Lambayeque com coberturas acima 85% para ambos os sexos, enquanto outras regiões como Amazonas com coberturas inferiores a 40% para ambos os sexos. O Peru tem taxas de cobertura de óbitos de 57,9% para homens e 54,9% para mulheres, valores ainda baixos quando comparados, por exemplo, com o Brasil que apresenta coberturas de 98% para homens e 96% para mulheres, segundo Queiroz et al. (2017).

Outra forma de se estimar a cobertura de registro de óbitos é analiticamente, comparando os registros do SINADEF e comparando com os números esperados para a região no período de interesse. Os números de óbitos esperados podem ser estimado de diferentes formas, uma delas sendo, por uma tábua de vida modelo que represente de forma satisfatória os níveis e estruturas de mortalidade da população em questão. O Instituto Nacional de Estatística e Informação (INEI) disponibilizou, em 2010, as tábuas de vida para o Peru e suas regiões através da publicação “Situación y perspectivas de la mortalidad por sexo y grupos de edad, nacional y departamentos, 1990-2025”, sendo essa a tábua utilizada para as estimações apresentadas neste estudo. É importante salientar que não foram apresentadas tábuas para as províncias, sendo considerada as funções apresentadas para a região, uma boa estimativa da estrutura das províncias.

A partir da tábua de vida por regiões para os anos de 2015 a 2020, serão utilizados a função taxa de mortalidade entre as idades x e $x + n$, $m_{s,d}(x; x + n)$ na região d .

Tabela 1: Cobertura estimada de óbitos registrados no SINADEF por sexo segundo localizações - Método GGB. Peru, 2017.

Localizações	Cobertura (%)	
	Homens	Mulheres
Peru	57,9	54,9
Amazonas	33,2	35,4
Ancash	59,6	59,4
Apurimac	55,9	57,7
Arequipa	63,6	58,4
Ayacucho	28,5	29,7
Cajamarca	36,5	37,7
Prov. Const. Del Callao	63,8	50,2
Cusco	59,2	57,2
Huancavelica	55,5	63,0
Huanuco	59,3	68,9
Ica	79,7	79,5
Junin	58,7	54,6
La Libertad	68,9	65,6
Lambayeque	91,6	85,2
Lima	56,7	50,3
Loreto	36,7	36,7
Madre de Dios	53,7	42,6
Moquegua	72,2	68,1
Pasco	34,1	39,6
Piura	68,4	64,0
Puno	43,2	46,1
San Martin	40,8	43,4
Tacna	52,3	56,6
Tumbes	53,3	55,6
Ucayali	42,3	47,1

Fonte: INEI, Censo demográfico, 2017; SINADEF, 2017.

4 Resultados

4.1 Metodologia Analítica

Uma das alternativas para se calcular a cobertura de mortalidade por grupos de idade até o nível de distritos no Peru, é utilizar um método analítico, que usa tábuas de vida para definir a intensidade e estrutura da mortalidade na população e assim estimar as mortes esperadas. A tábua de vida utilizado neste trabalho foi “Situación y perspectivas de la mortalidad por sexo y grupos de edad, nacional y departamentos, 1990-2025”, fornecida pelo INEI, apresentada na Tabela 2, que tem intervalos de idade de 5 anos, mas para se evitar estimativas de óbitos iguais a 0, os intervalos utilizados neste trabalho serão mais amplos, se dividindo entre as seguintes categorias: 1) menores de 5 anos; 2) entre 5 e 64 anos; e 3) 65 anos ou mais.

Os óbitos esperados são calculados na forma a seguir:

$$D_{e,s,l}(x; x+n) = m_{s,r}(x; x+n) * P_{s,l}(x; x+n),$$

Onde:

$D_{e,s,l}(x; x+n)$: Óbitos esperados entre as idades x e $x+n$, por sexo e localização;

$m_{s,r}(x; x+n)$: função “taxa de mortalidade” da Tábua de vida entre as idades x e $x+n$, por sexo e região;

$P_{s,l}(x; x+n)$: População estimada por idade, segundo sexo e localização.

Tabela 2: Tábua de vida abreviada de mortalidade para homens, Peru, 2015-2020

Idade (x)	n	m(x,n)	q(x,n)	l(x)	d(x,n)	L(x,n)	P(x,x+5)	T(x)	e(x)
0	1	0,0142	0,0140	100.000	1400	98.716	0,9844	7.371.155	73,7
1	4	0,0010	0,0038	98.600	375	393.505	0,9963	7.272.439	73,8
5	5	0,0006	0,0029	98.225	289	490.405	0,9976	6.878.934	70,0
10	5	0,0004	0,0019	97.937	182	489.227	0,9966	6.388.530	65,2
15	5	0,0011	0,0057	97.754	558	487.554	0,9927	5.899.303	60,4
20	5	0,0017	0,0086	97.196	837	483.997	0,9902	5.411.749	55,7
25	5	0,0022	0,0108	96.359	1040	479.244	0,9891	4.927.752	51,1
30	5	0,0022	0,0109	95.320	1038	474.031	0,9884	4.448.508	46,7
35	5	0,0025	0,0125	94.282	1180	468.549	0,9858	3.974.477	42,2
40	5	0,0033	0,0162	93.101	1505	461.898	0,9815	3.595.929	37,7
45	5	0,0043	0,0212	91.596	1939	453.369	0,9747	3.044.030	33,2
50	5	0,0061	0,0302	89.657	2708	441.881	0,9640	2.590.662	28,9
55	5	0,0087	0,0427	86.949	3711	425.975	0,9482	2.148.781	24,7
60	5	0,0129	0,0625	83.238	5206	403.902	0,9241	1.722.806	20,7
65	5	0,0194	0,0926	78.032	7227	373.255	0,8783	1.318.903	16,9
70	5	0,0338	0,1567	70.805	11094	327.845	0,8070	945.648	13,4
75	5	0,0532	0,2359	59.711	14086	264.561	0,7088	617.804	10,4
80	5	0,0874	0,3593	45.625	16391	187.522	0,5682	353.243	7,7
85	5	0,1424	0,5189	29.234	15171	106.547	0,4144	165.721	5,7
90	5	0,2123	0,6664	14.063	9372	44.156	0,2842	59.174	4,2
95	5	0,2937	0,7855	4.691	3685	12.547	0,1645	15.017	3,2
100	w	0,4075	1,0000	1.007	1007	2.471	0,0000	2.471	2,5

Fonte: INEI, Situación y perspectivas de la mortalidad por sexo y grupos de edad, nacional y departamentos, 1990-2025, 2010.

As fórmulas utilizadas para se adequar as taxas de mortalidade para as novas categorias são apresentadas abaixo:

1. Menores de 5 anos (5 anos exatos)

$$m_{s,d}(0;5) = \frac{\sum_0^5 d_{s,d}(x; x+n)}{\sum_0^5 L_{s,d}(x; x+n)};$$

2. Entre 5 e 65 anos (65 anos exatos)

$$m_{s,d}(5;65) = \frac{\sum_5^{65} d_{s,d}(x; x+n)}{\sum_5^{65} L_{s,d}(x; x+n)};$$

3. 65 anos ou mais (65 anos até o grupo aberto final)

$$m_{s,d}(65+) = \frac{\sum_{65}^{\infty} d_{s,d}(x; x+n)}{\sum_{65}^{\infty} L_{s,d}(x; x+n)}.$$

Onde:

$m_{s,r}(x; x+n)$: função “taxa de mortalidade” da Tábua de vida entre as idades x e $x+n$, por sexo e região;

$d_{s,r}(x; x+n)$: função “óbitos entre as idades x e $x+n$ ” da Tábua de vida, por sexo e região;

$L_{s,r}(x; x+n)$: função “Tempo vivido pelos sobreviventes entre as idades x e $x+n$ ” da Tábua de vida, segundo sexo e região.

O cálculo da cobertura de óbitos será dado pela seguinte expressão:

$$\text{Cobertura}_{s,l}(x; x+n) = \frac{D_{obs,s,l}(x; x+n)}{D_{e,s,l}(x; x+n)},$$

sendo:

$\text{Cobertura}_{s,l}(x; x+n)$: Cobertura entre as idades x e $x+n$, por sexo e localização;

$D_{obs,s,l}(x; x+n)$: Óbitos observados entre as idades x e $x+n$, por sexo e localização;

$D_{e,s,l}(x; x+n)$: Óbitos esperados entre as idades x e $x+n$, por sexo e localização.

Essa forma de cálculo apresenta uma boa estimativa de óbitos esperados para províncias e distritos com grandes populações, como para Província de Lima e o distrito de San Juan de Lurigancho, apresentados nas Tabelas 3 e 4, respectivamente.

Tabela 3: Óbitos esperados por grupos de idade. Província de Lima, 2017 - 2020

Província	Idade	Anos			
		2017	2018	2019	2020
Província de Lima	[0,5)	1557	1771	1903	2028
	[5,65)	16815	18004	18293	18539
	[65,+)	42543	47067	49197	51177

Tabela 4: Óbitos esperados por grupos de idade. Distrito San Juan de Lurigancho, 2017 - 2020

Distrito	Idade	Anos			
		2017	2018	2019	2020
San Juan de Lurigancho	[0,5)	210	233	245	257
	[5,65)	2065	2220	2262	2298
	[65,+)	3665	4134	4400	4652

Já para províncias e distritos com populações menores como o distrito de Huampara os cálculos de óbitos esperados são menores que 1 e por vezes 0, como demonstrado na Tabela 5. Nesses casos a cobertura de mortalidade não pode ser calculada ou tende ao infinito, dado que seu cálculo é definido como a divisão de óbitos observados pelos óbitos esperados.

Tabela 5: Óbitos esperados por grupos de idade. Distrito El parco, 2017 - 2020

Distrito	Idade	Anos			
		2017	2018	2019	2020
El parco	[0,5)	< 1	< 1	< 1	< 1
	[5,65)	4	4	4	4
	[65,+)	10	11	11	12

4.2 Métodos Bayesianos

As análises realizadas a seguir utilizam o pacote “BRMS” (BürkNER, 2017) do software R para criação de modelos hierárquicos bayesianos que tentam prever o número de óbitos esperados nos diversos níveis geográficos do Peru. Para maior simplicidade e facilidade de interpretação dos dados foram criados modelos para cada ano. Os principais resultados estão expostos a seguir.

4.2.1 Modelo 2017

Serão apresentados agora os resultados para o modelo hierárquico bayesiano com a forma e distribuição dos parâmetros apresentadas na seção anterior que utilizou os dados de óbitos do Peru no ano de 2017.

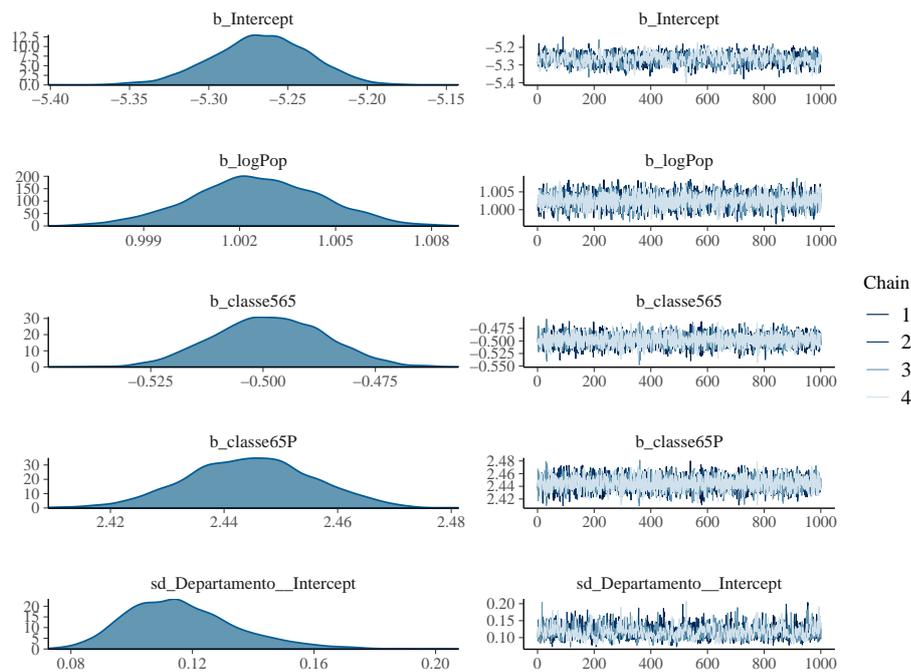


Figura 5: Convergência dos parâmetros do modelo bayesiano, 2017

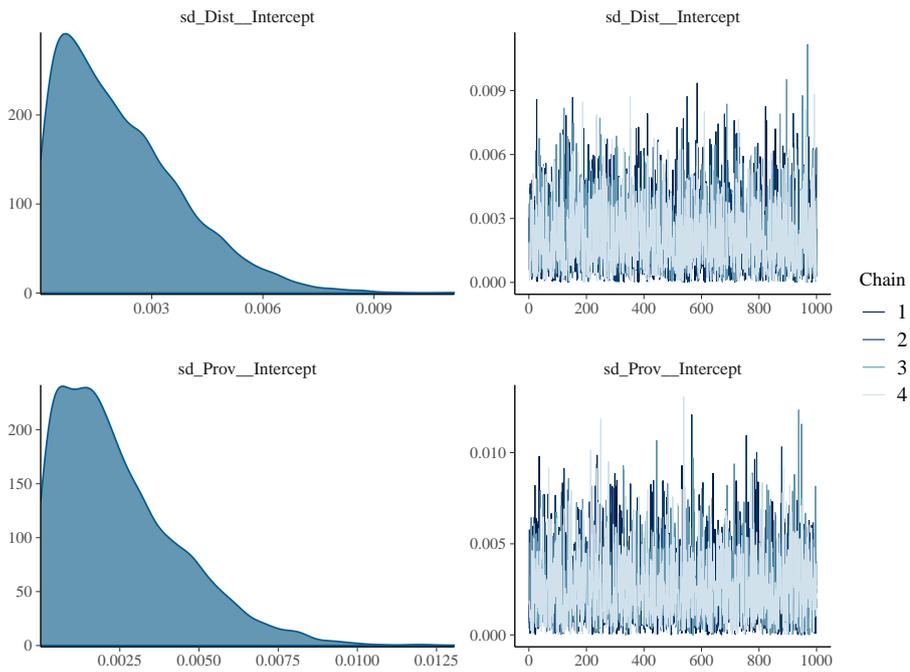


Figura 6: Convergência dos parâmetros do modelo bayesiano, 2017

Nota-se pelos gráficos acima, que o traço apresentado na imagem à direita da Figura 5 e 6 se assemelham muito ao ruído branco, indicando que não existem problemas de convergência para o modelo em nenhum dos parâmetros amostrados. Os valores estimados de órbitos quando comparados com os valores reais, mostrados na Figura 7, são sempre bem próximos, excluindo somente alguns valores que destoam dos demais e poderiam ser considerados outliers.

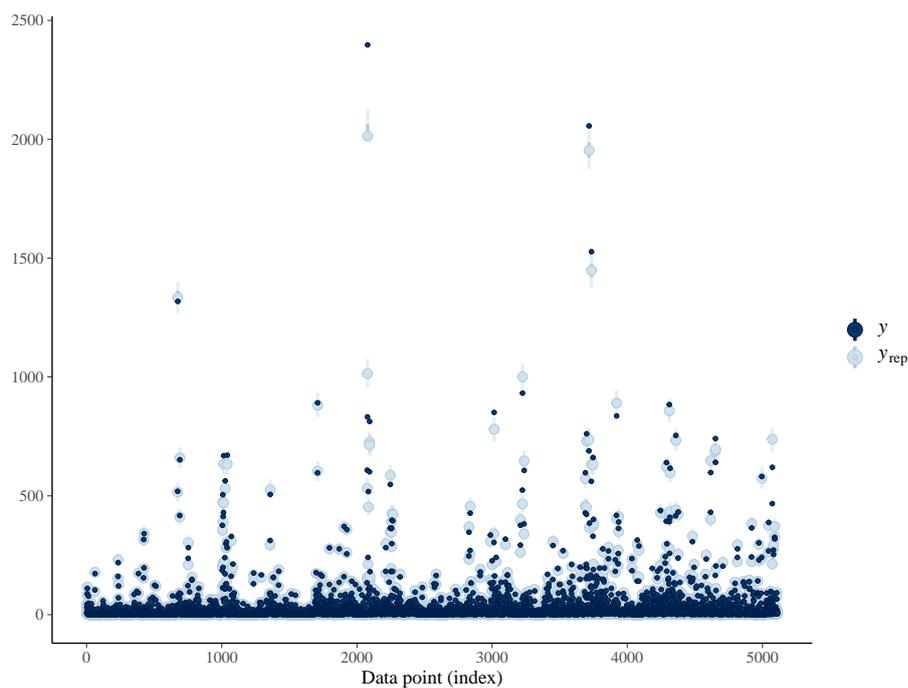


Figura 7: Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2017

Dos 1.704 distritos existentes na base de dados, mais de 99% têm os valores de óbitos contidos no intervalo de credibilidade de 95% gerado para as estimativas. A Figura 8 corrobora com os dados apresentados anteriormente, com as linhas da distribuição preditiva *a posteriori* quase que coincidindo com os dados reais.

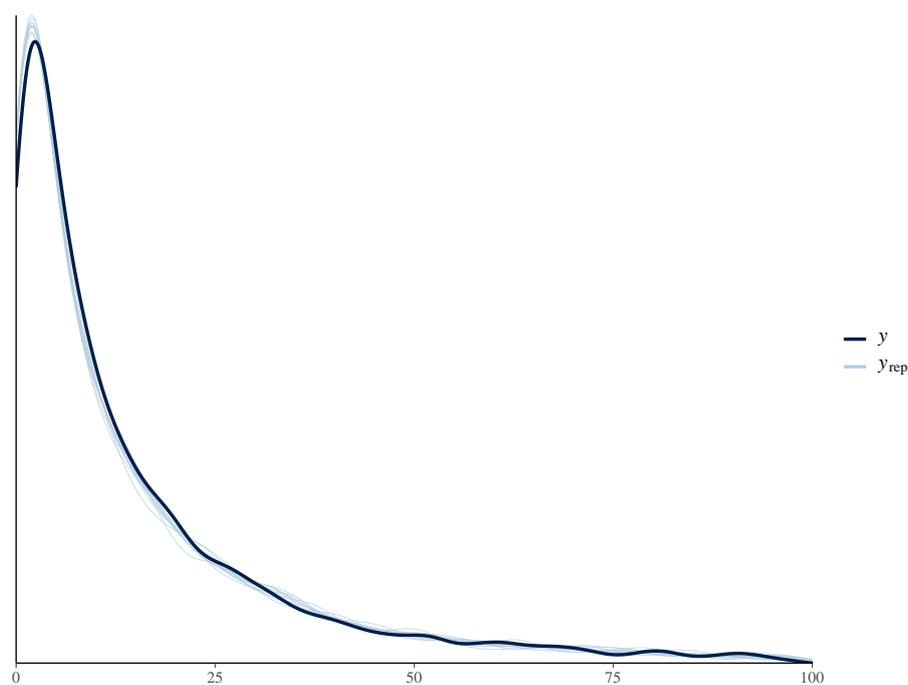


Figura 8: Checagem da distribuição preditiva *a posteriori* do modelo bayesiano, 2017

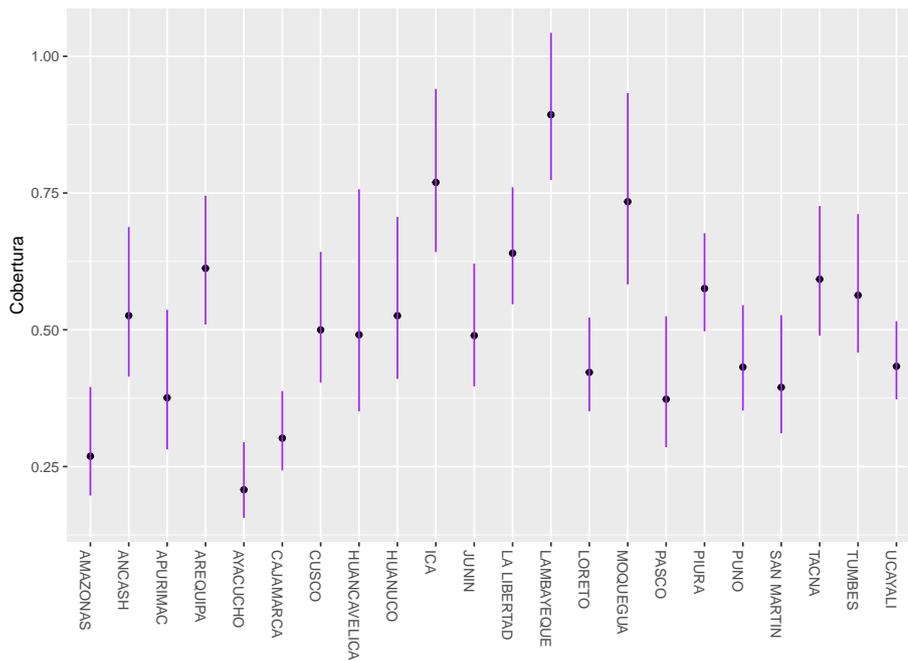


Figura 9: Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2017

Em todos os departamentos a estimativa de cobertura de óbitos, o número de óbitos observados divididos pelos óbitos estimados, do modelo clássico está contido no intervalo de 95% da cobertura de óbitos do modelo bayesiano como exemplificado na Figura 9, corroborando mais uma vez para qualidade do ajuste do modelo proposto.

O desafio da estimação de áreas menores, com menos população e óbitos, consequentemente, apresenta muitos zeros e dificulta a estimação da cobertura de registros de morte nessas regiões, como já discutido anteriormente. O modelo bayesiano proposto tem menor incidência de zeros, como, por exemplo, o distrito de El parco, que nos modelos analíticos apresenta valores estimados < 1 para o ano de 2017 para idades entre 0 e 5 anos, mas no modelo bayesiano apresenta estimativa pontual igual a 1, possibilitando, por sua vez, o cálculo da taxa de cobertura como mostrado na Tabela 6.

Tabela 6: Óbitos esperados por grupos de idade. Distrito El parco, 2017

Distrito	Idade	Óbitos Observados	Óbitos Esperados (IC 95%)		Cobertura	
			Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
El parco	[0,5)	0	0	2	-	0%
	[5,65)	5	1	7	71%	500%
	[65,+)	5	5	15	33%	100%

4.2.2 Modelo 2018

Serão apresentados agora os resultados para o modelo hierárquico bayesiano com a forma e distribuição dos parâmetros apresentadas na seção anterior que utilizou os dados de óbitos do Peru no ano de 2018.

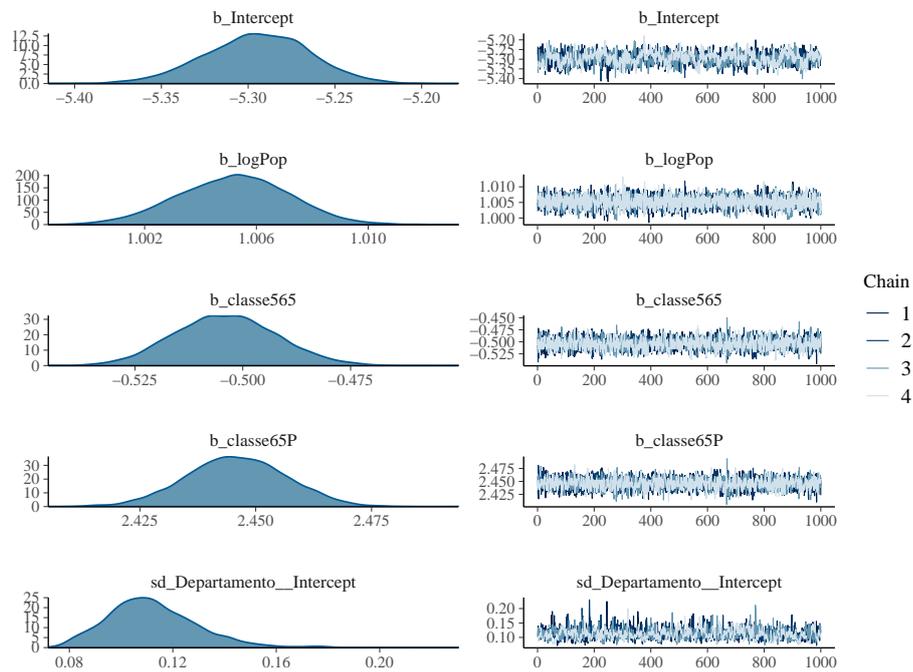


Figura 10: Convergência dos parâmetros do modelo bayesiano, 2018

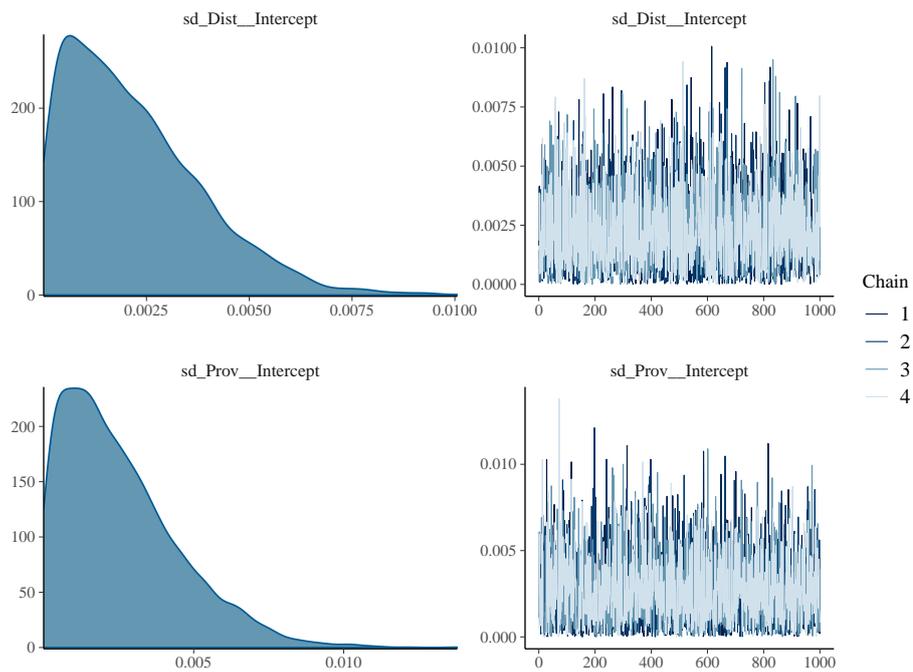


Figura 11: Convergência dos parâmetros do modelo bayesiano, 2018

Nota-se pelos gráficos acima, que o traço apresentado na imagem à direita da Figura 10 e 11 se assemelham muito ao ruído branco, indicando que não existem problemas de convergência para o modelo em nenhum dos parâmetros amostrados. Os valores estimados de órbitas quando comparados com os valores reais, mostrados na Figura 12, são sempre bem próximos, excluindo somente alguns valores que destoam dos demais e poderiam ser considerados outliers.

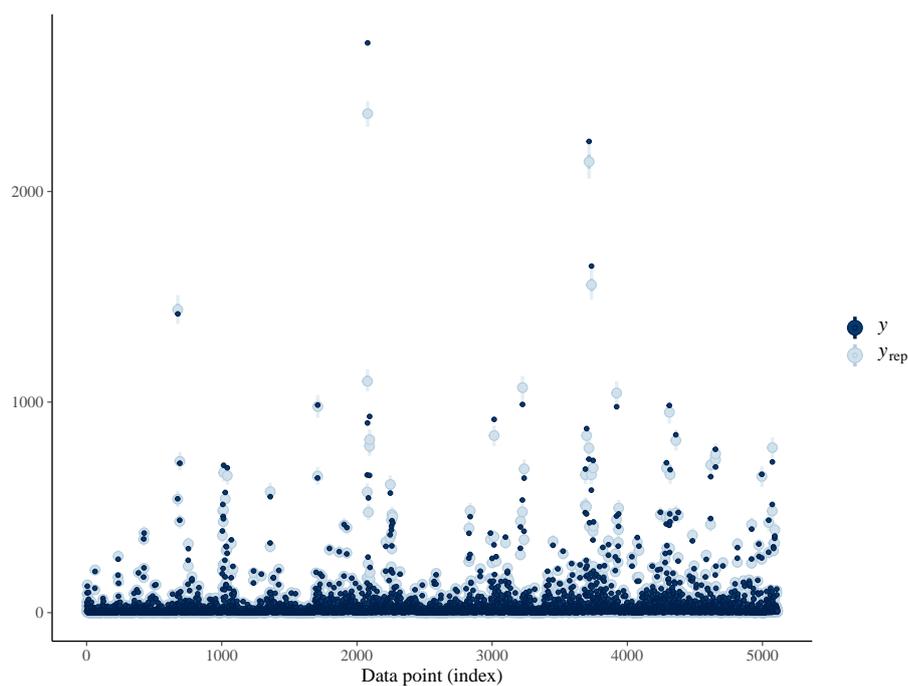


Figura 12: Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2018

Dos 1.704 distritos existentes na base de dados, mais de 99% têm os valores de óbitos contidos no intervalo de credibilidade de 95% gerado para as estimativas. A Figura 13 corrobora com os dados apresentados anteriormente, com as linhas da distribuição preditiva *a posteriori* quase que coincidindo com os dados reais.

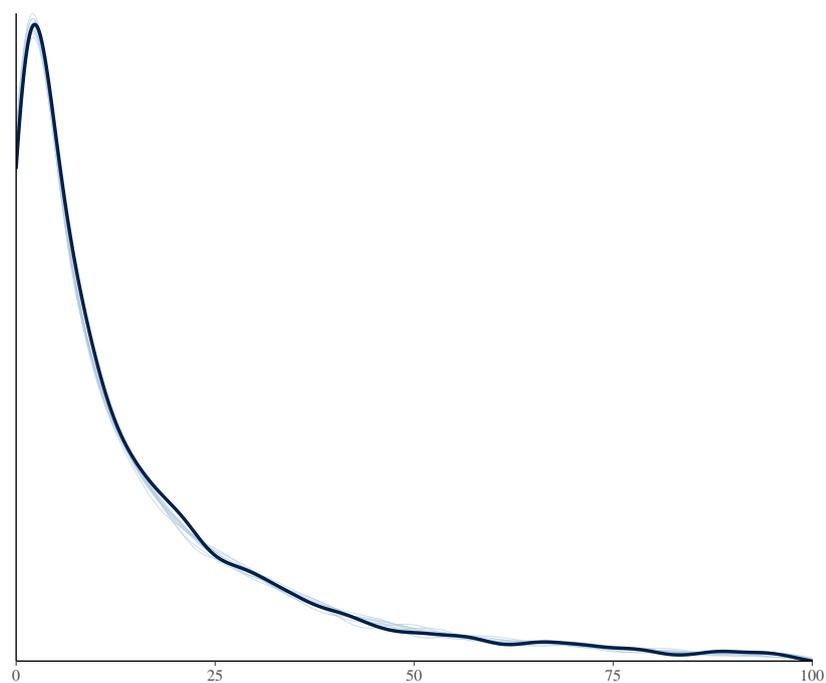
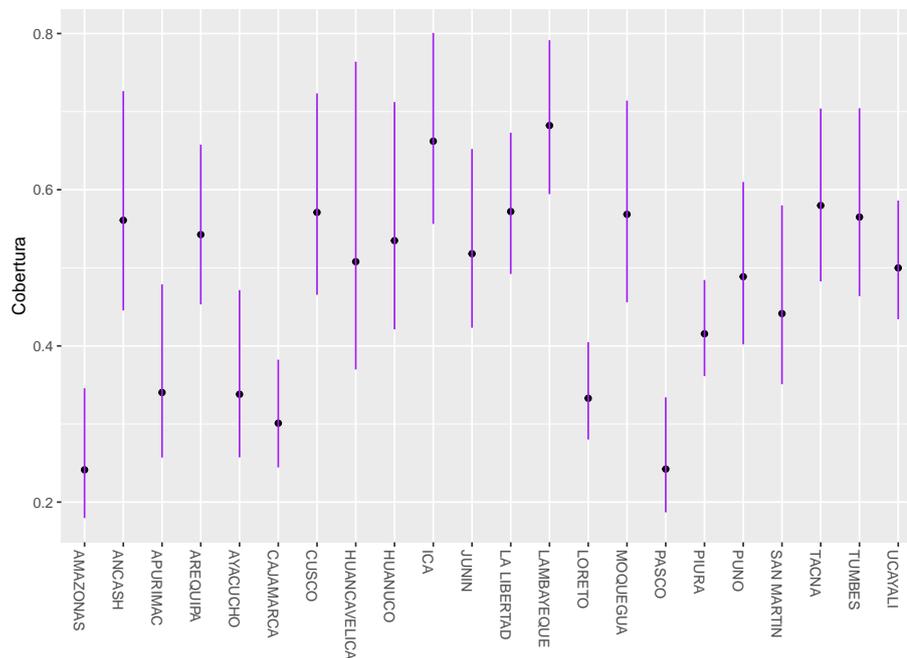


Figura 13: Checagem da distribuição preditiva *a posteriori* do modelo bayesiano, 2018

Figura 14: Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2018



Em todos os departamentos a estimativa de cobertura de óbitos, o número de óbitos observados divididos pelos óbitos estimados, do modelo clássico está contido no intervalo de 95% da cobertura de óbitos do modelo bayesiano como exemplificado na Figura 14, corroborando mais uma vez para qualidade do ajuste do modelo proposto.

Novamente o modelo bayesiano proposto para o ano de 2018 tem menor incidência de zeros, como, por exemplo, o distrito de El parco, que nos modelos analíticos apresenta valores estimados < 1 para o ano de 2018 para idades entre 0 e 5 anos, mas no modelo bayesiano apresenta estimativa pontual igual a 1, possibilitando, por sua vez, o cálculo da taxa de cobertura como mostrado na Tabela 7.

Tabela 7: Óbitos esperados por grupos de idade. Distrito El parco, 2018

Distrito	Idade	Óbitos Observados	Óbitos Esperados (IC 95%)		Cobertura	
			Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
El parco	[0,5)	0	0	2	-	0%
	[5,65)	1	1	7	14%	100%
	[65,+)	1	6	16	6%	17%

4.2.3 Modelo 2019

Serão apresentados agora os resultados para o modelo hierárquico bayesiano com a forma e distribuição dos parâmetros apresentadas na seção anterior que utilizou os dados de órbitos do Peru no ano de 2019.

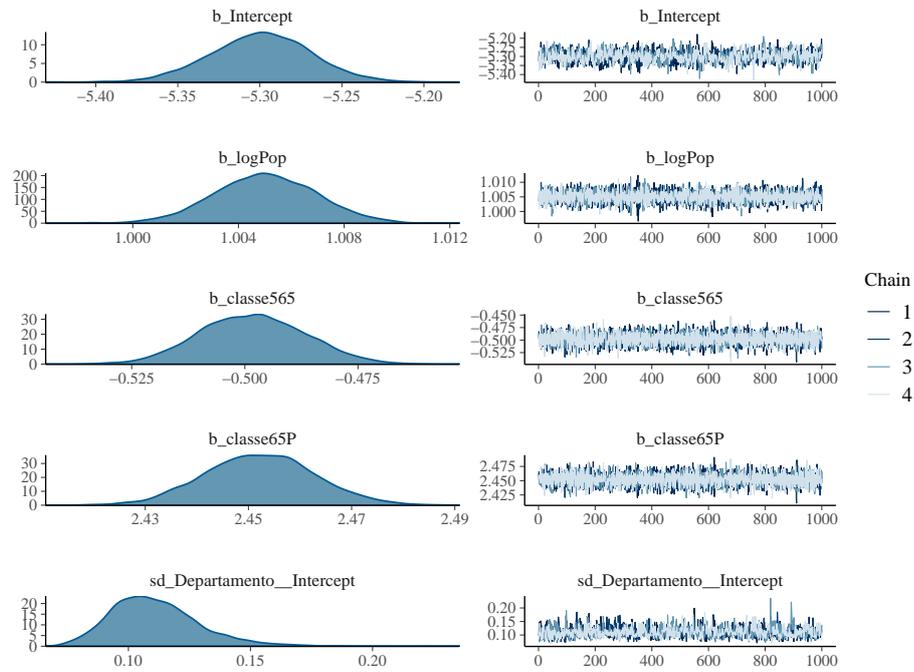


Figura 15: Convergência dos parâmetros do modelo bayesiano, 2019

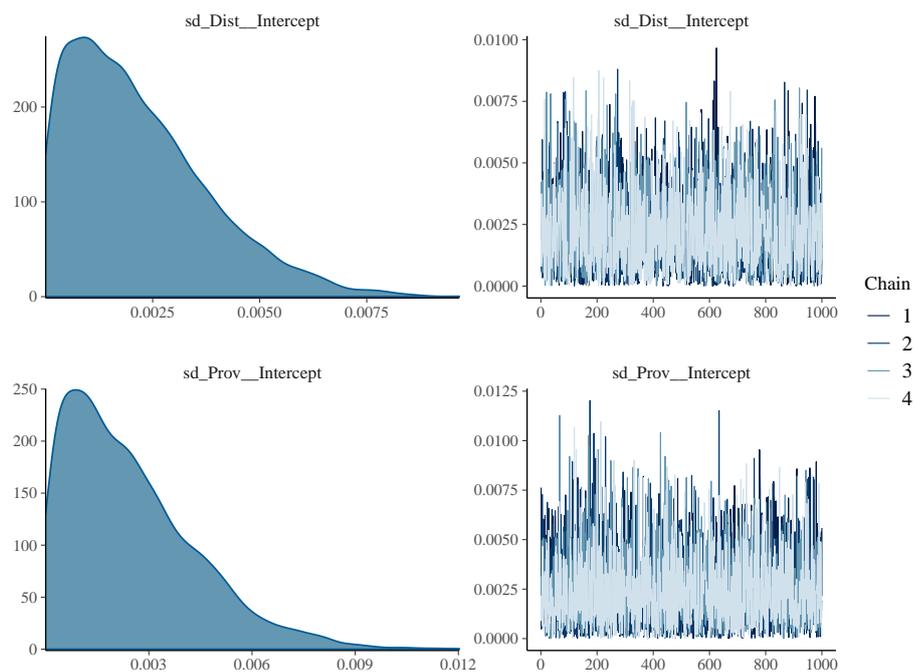


Figura 16: Convergência dos parâmetros do modelo bayesiano, 2019

Nota-se pelos gráficos acima, que o traço apresentado na imagem à direita da Figura 15 e 16 se assemelham muito ao ruído branco, indicando que não existem problemas de convergência para o modelo em nenhum dos parâmetros amostrados. Os valores estimados de óbitos quando comparados com os valores reais, mostrados na Figura 17, são sempre bem próximos, excluindo somente alguns valores que destoam dos demais e poderiam ser considerados outliers.

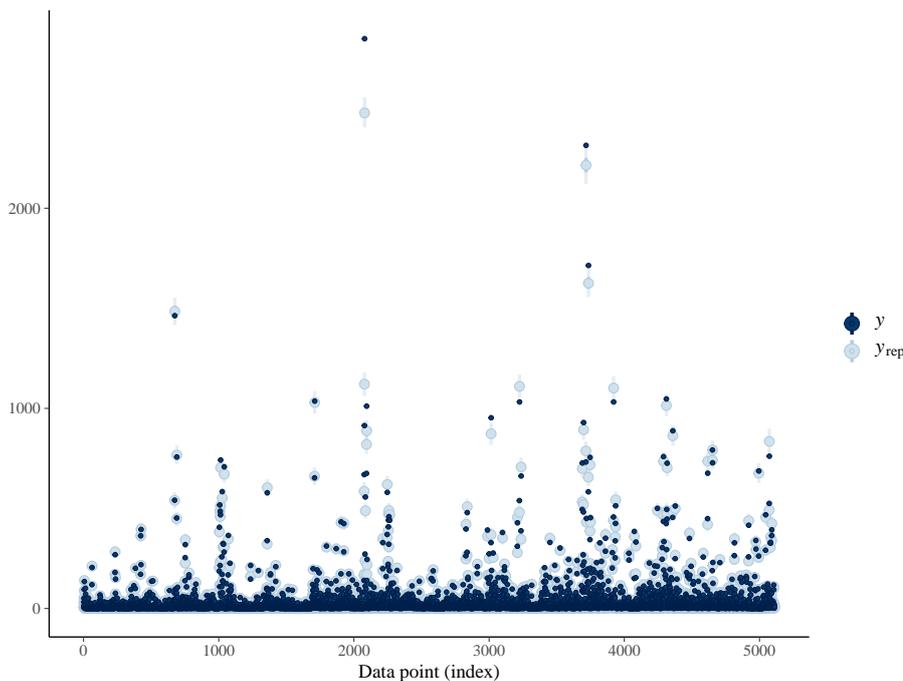


Figura 17: Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2019

Dos 1.704 distritos existentes na base de dados, mais de 99% têm os valores de óbitos contidos no intervalo de credibilidade de 95% gerado para as estimativas. A Figura 18 corrobora com os dados apresentados anteriormente, com as linhas da distribuição preditiva *a posteriori* quase que coincidindo com os dados reais.

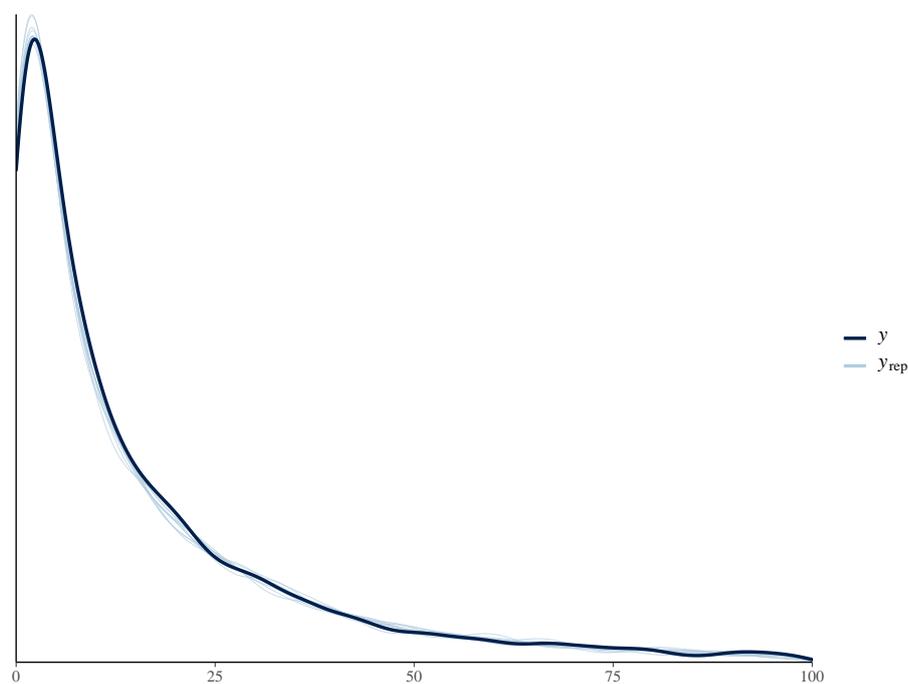


Figura 18: Checagem da distribuição preditiva a *posteriori* do modelo bayesiano, 2019

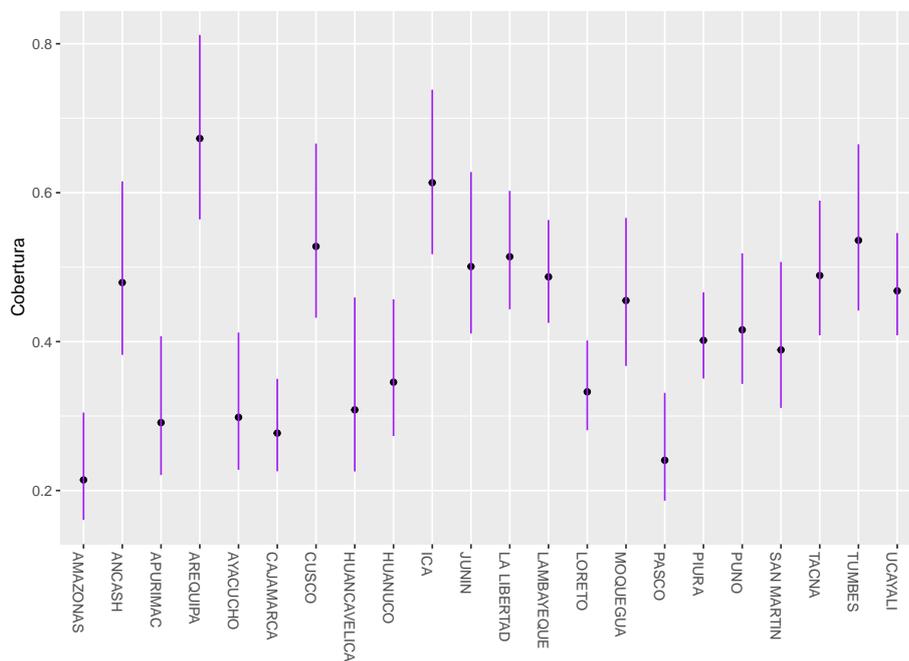


Figura 19: Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2019

Em todos os departamentos a estimativa de cobertura de óbitos, o número de óbitos observados divididos pelos óbitos estimados, do modelo clássico está contido no intervalo de 95% da cobertura de óbitos do modelo bayesiano como exemplificado na

Figura 19, corroborando mais uma vez para qualidade do ajuste do modelo proposto.

O modelo bayesiano proposto para o ano de 2019 tem menor incidência de zeros assim como os anteriores. O distrito de El parco, que nos modelos analíticos apresenta valores estimados < 1 para o ano de 2019 para idades entre 0 e 5 anos, mas no modelo bayesiano apresenta estimativa pontual igual a 1, possibilitando, por sua vez, o cálculo da taxa de cobertura como mostrado na Tabela 8.

Tabela 8: Óbitos esperados por grupos de idade. Distrito El parco, 2019

Distrito	Idade	Óbitos Observados	Óbitos Esperados (IC 95%)		Cobertura	
			Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
El parco	[0,5)	0	0	2	0%	0%
	[5,65)	0	1	7	0%	0%
	[65,+)	2	6	17	12%	33%

4.2.4 Modelo 2020

Serão apresentados agora os resultados para o modelo hierárquico bayesiano com a forma e distribuição dos parâmetros apresentadas na seção anterior que utilizou os dados de óbitos do Peru no ano de 2020.

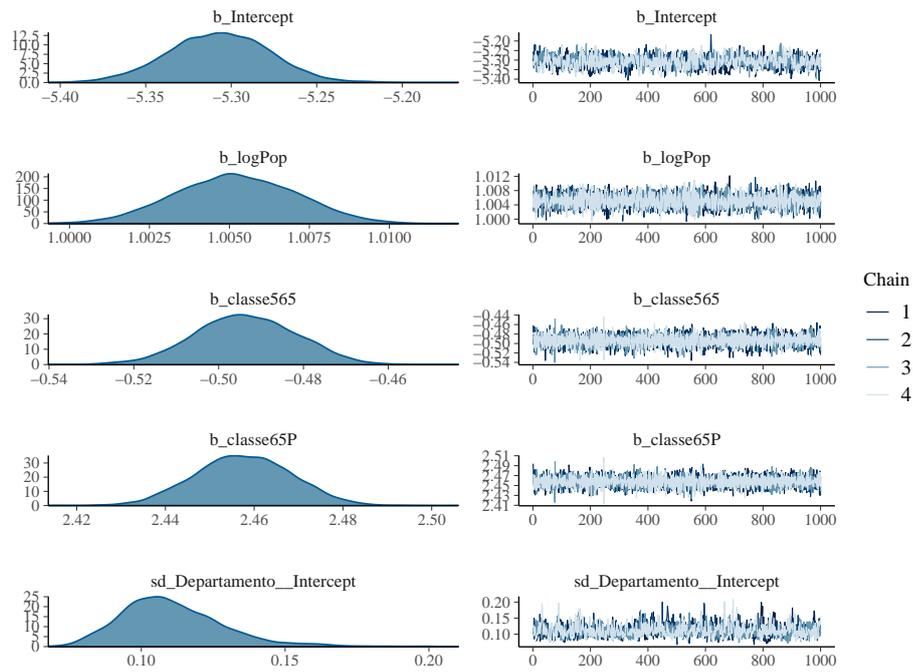


Figura 20: Convergência dos parâmetros do modelo bayesiano, 2020

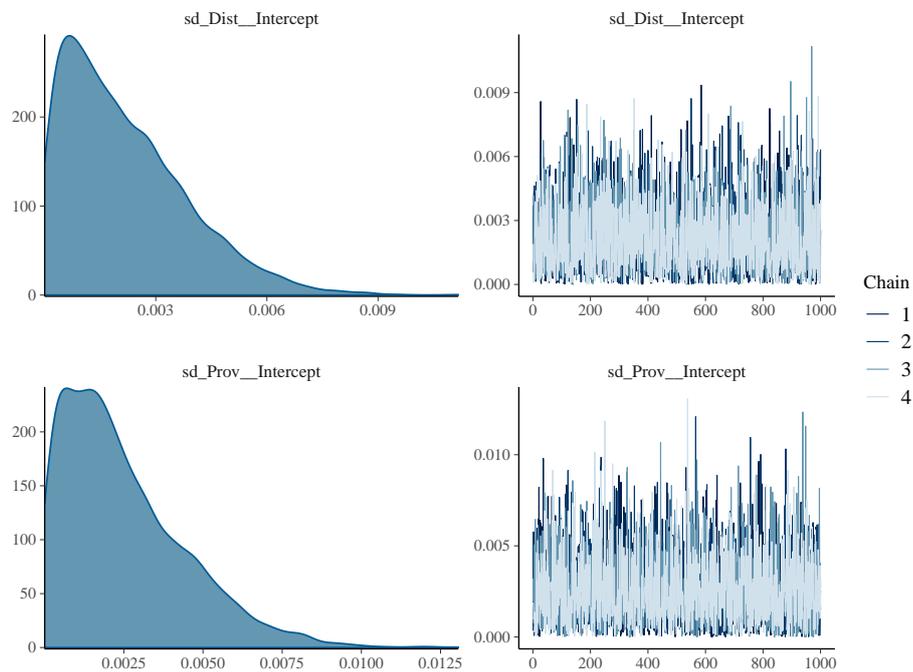


Figura 21: Convergência dos parâmetros do modelo bayesiano, 2020

Nota-se pelos gráficos acima, que o traço apresentado na imagem à direita da Figura 20 e 21 se assemelham muito ao ruído branco, indicando que não existem problemas de convergência para o modelo em nenhum dos parâmetros amostrados. Os valores estimados de órbitas quando comparados com os valores reais, mostrados na Figura 22,

são sempre bem próximos, excluindo somente alguns valores que destoam dos demais e poderiam ser considerados outliers.

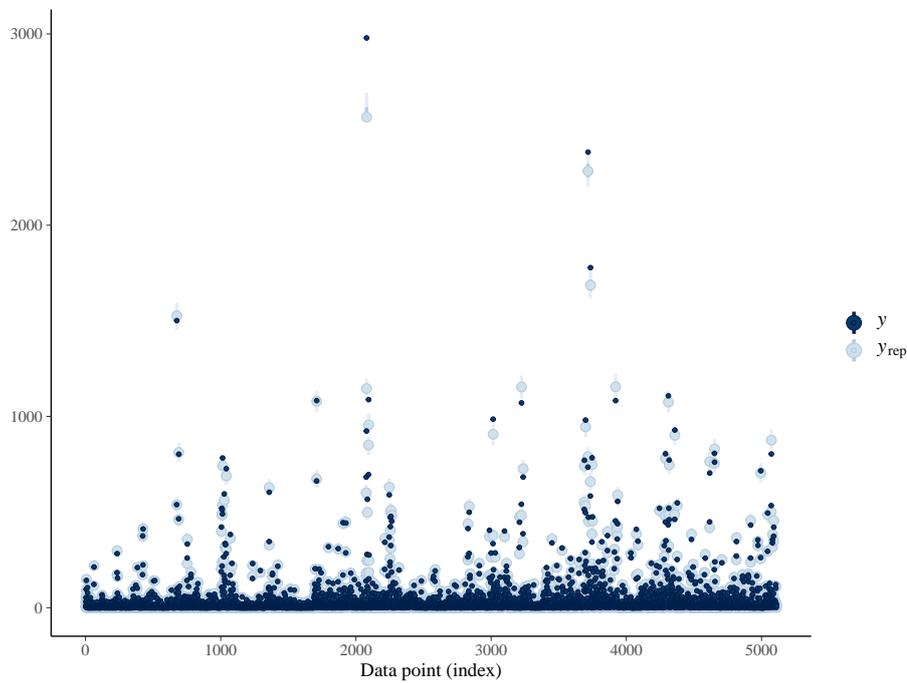


Figura 22: Comparação entre os valores estimados pelo modelo bayesiano e valores reais para os distritos, 2020

Dos 1.704 distritos existentes na base de dados, mais de 99% têm os valores de óbitos contidos no intervalo de credibilidade de 95% gerado para as estimativas. A Figura 23 corrobora com os dados apresentados anteriormente, com as linhas da distribuição preditiva *a posteriori* quase que coincidindo com os dados reais.

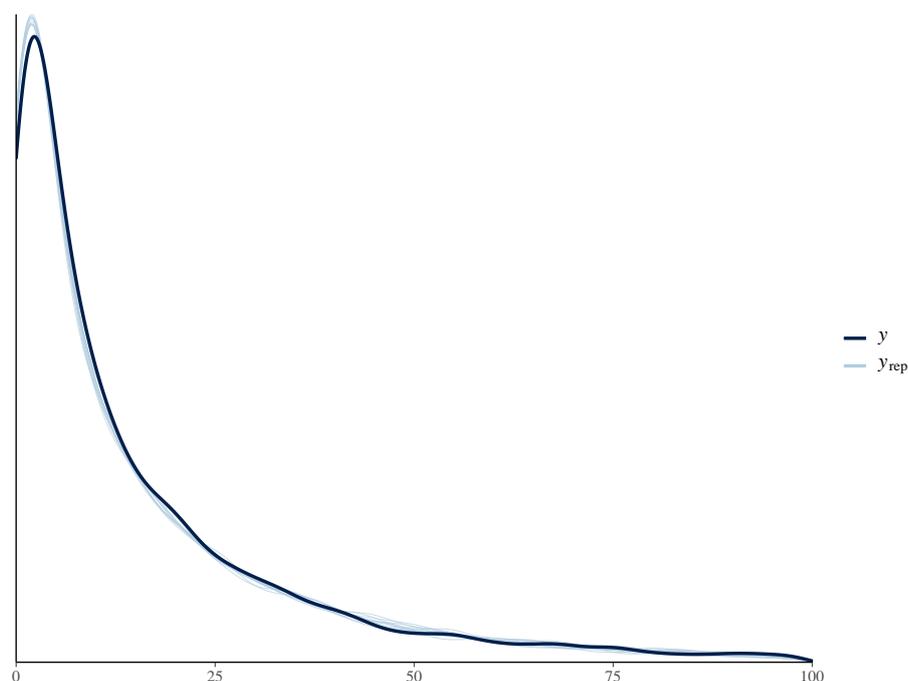


Figura 23: Checagem da distribuição preditiva a *posteriori* do modelo bayesiano, 2020

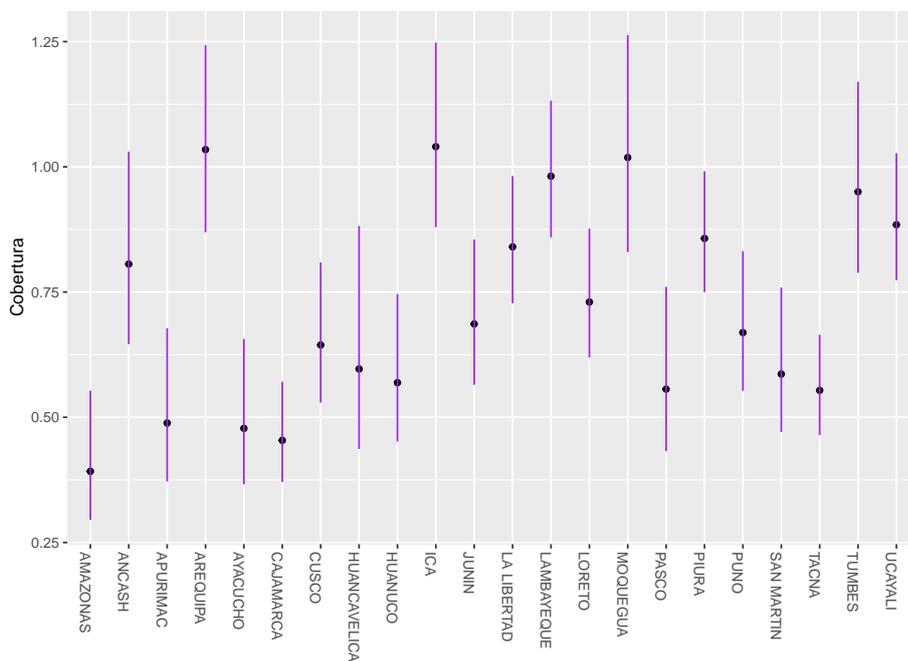


Figura 24: Cobertura de registro de óbitos da metodologia clássica e intervalos de credibilidade de 95% do modelo bayesiano por departamento, 2020

Em todos os departamentos a estimativa de cobertura de óbitos, o número de óbitos observados divididos pelos óbitos estimados, do modelo clássico está contido no intervalo de 95% da cobertura de óbitos do modelo bayesiano como exemplificado na

Figura 24, corroborando mais uma vez para qualidade do ajuste do modelo proposto.

Tabela 9: Óbitos esperados por grupos de idade. Distrito El parco, 2020

Distrito	Idade	Óbitos Observados	Óbitos Esperados (IC 95%)		Cobertura	
			Limite Inferior	Limite Superior	Limite Inferior	Limite Superior
	[0,5)	0	0	2	-	0%
El parco	[5,65)	2	1	7	14%	200%
	[65,+)	7	7	18	58%	100%

O limite inferior é sempre maior ou igual a 1.

Para o ano de 2020 o modelo bayesiano proposto também apresenta menor volume de zeros na estimação em relação ao modelo clássico. O distrito de El parco, que nos modelos analíticos apresenta valores estimados < 1 para o ano de 2020 para idades entre 0 e 5 anos, mas no modelo bayesiano apresenta estimativa pontual igual a 1, possibilitando, por sua vez, o cálculo da taxa de cobertura como mostrado na Tabela 9.

5 Conclusão

O Peru está empreendendo esforços contínuos para melhoria da qualidade de suas bases de registros de óbitos. No ano de 2017 mais de 90% dos departamentos tinham mais de 50% dos seus registros de óbitos feitos manualmente, o que prejudica a qualidade da informação. O resultado dos esforços empreendidos para melhora do registro reflete nos anos seguintes. Em 2018 apenas 27% dos departamentos tinham mais de 50% dos registros realizados manualmente, este valor caiu para 14% em 2019 e chegando a 5% em 2020, sendo que neste ano quase 70% dos departamentos tinham mais de 70% dos registros de óbitos feitos em sistemas on-line. Essa melhora é percebida também nas estimativas de cobertura de mortalidade estimadas pelo modelo. Tomando como exemplo o distrito de El parco, tinha uma cobertura de 71% em 2017, muito por causa de uma superestimação nas idades [5, 65) de 167%, que caiu para 13% em 2018, mantendo-se em 13% em 2019, mas com a cobertura sendo igual a 0% nas idades entre 0 e 65 anos, e finalmente em 2020 subindo para 53% cobertura dos registros de morte.

O modelo bayesiano proposto para os anos estimou com qualidade os valores de óbitos e, conseqüentemente, as coberturas de mortalidade do Peru. Para todos os anos, os parâmetros dos modelos convergiram sem grandes problemas e os intervalos de credibilidade de 95% propostos para as coberturas continham o valor do modelo clássico em todos os departamentos e, novamente, em todos os anos, os intervalos de credibilidade de 95% construídos a partir dos modelos continham os valores estimados pelo modelo clássico em 99% dos distritos, apresentando um número menor de zeros que impossibilitam o cálculo da cobertura de mortalidade e mostrando uma ótima eficiência no ponto mais desafiador que é a estimação de pequenas regiões devido a sua pequena população observada em alguns pontos e alta variabilidade amostral nas mortes registradas (PLETCHER, 1999).

Referências

- ABOUZAHAR, C.; BOERMA, T. Health information systems: the foundations of public health. *Bulletin of the World Health Organization*, SciELO Public Health, v. 83, p. 578–583, 2005.
- BENNETT, N. G.; HORIUCHI, S. Estimating the completeness of death registration in a closed population. *Population index*, JSTOR, p. 207–221, 1981.
- BÜRKNER, P.-C. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, v. 80, n. 1, p. 1–28, 2017.
- HASTINGS, W. K. Monte carlo sampling methods using markov chains and their applications. Oxford University Press, 1970.
- HILL, K. Estimating census and death registration completeness. In: THE ASIAN & PACIFIC POPULATION FORUM. *Asian and Pacific population forum/East-West Population Institute, East-West Center*. [S.l.], 1987. v. 1, n. 3, p. 8–13.
- MIKKELSEN, L. et al. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, Elsevier, v. 386, n. 10001, p. 1395–1406, 2015.
- NAGHAVI, M. et al. Algorithms for enhancing public health utility of national causes-of-death data. *Population health metrics*, Springer, v. 8, n. 1, p. 1–14, 2010.
- PLETCHER, S. D. Model fitting and hypothesis testing for age-specific mortality data. *Journal of Evolutionary Biology*, v. 12, n. 3, p. 430–439, 1999.
- QUEIROZ, B. L. et al. Estimativas do grau de cobertura e da mortalidade adulta (45q15) para as unidades da federação no brasil entre 1980 e 2010. *Revista Brasileira de Epidemiologia*, SciELO Brasil, v. 20, p. 21–33, 2017.
- SCHMERTMANN, C. P.; GONZAGA, M. R. Bayesian estimation of age-specific mortality and life expectancy for small areas with defective vital records. *Demography*, Duke University Press, v. 55, n. 4, p. 1363–1388, 2018.
- VASCONCELOS, A. M. N. et al. *Estudio Sobre La Cobertura Del Sistema Nacional De Defunciones En Perú a Nivel Distrital Entre 2017 y 2021*. [S.l.], 2022. Disponível em: <https://drive.google.com/file/d/1opqHJvVZdQ8FyYyW6wP6FBC5WYjHq7gu/view?usp=sharing>.

Apêndice

Neste apêndice são apresentados os códigos utilizados no software R que geraram os resultados apresentados anteriormente. O pacote BRMS utiliza outro pacote chamado Rstan, que usa um módulo integrado no R para criar inferências em modelos bayesianos a partir de algoritmos MCMC. Segue os códigos gerados:

```
1
2
3 library(brms)
4
5 ##### Dados de 2017 #####
6
7
8 # Criando-se 3 modelos, um com log(Inc_pobreza), outro com Inc_pobreza e
9   um terceiro sem a variável Inc_pobreza
10 mod1.7 <- brm(Def_esp ~ log(Pop) + classe + (1|Departamento) + (1|Prov)
11   + (1|Dist),data = data17,family = poisson(),prior = c(prior(normal
12   (0,10),class=b)),cores = 4)
13
14
15
16 # Resumo dos modelos
17 summary(mod1.7)
18
19
20 # Convergência dos parâmetros
21 plot(mod1.7)
22
23
24 # Comparando a distribuição a posteriori e os pontos dos dados com a
25   predições intervalares do modelo
26 pp_check(mod1.7)
27
28
29 pp_check(mod1.7,type = 'loo_intervals')
30
31
32 # Gerando intervalos de credibilidade
33 predict1.7 <- predictive_interval(mod1.7,prob = 0.95)
34
35
36 # Tabela com quantos valores estão contidos no intervalo de
37   credibilidade
38 table(cbind( predict1.7[,1] <= data17$Def_esp & predict1.7[,2] >=
39   data17$Def_esp ))
```

```
34
35
36 # Modelo com intervalo de confiança
37 plot(conditional_effects(mod1.7),points=T)
38
39
40
41 # Tabela com quantos valores de cobertura estão contidos no intervalo
    de credibilidade
42 cobert_int.17 <- predict1.7/data17$Def_Obs
43 table(cbind( cobert_int.17[,1] <= data17$Def_esp/data17$Def_Obs &
    cobert_int.17[,2] >= data17$Def_esp/data17$Def_Obs ))
44
45
46                                     ##### Dados de 2018 #####
47
48
49 # Criando-se 3 modelos, um com log(Inc_pobreza), outro com Inc_pobreza e
    um terceiro sem a variável Inc_pobreza
50 mod1.8 <- brm(Def_esp ~ log(Pop) + classe + (1|Dist) + (1|Prov) + (1|
    Departamento),data = data18,family = poisson(),prior = c(prior(normal
    (0,10),class=b)),cores = 4)
51
52 # Resumo dos modelos
53 summary(mod1.8)
54
55
56 # Convergência dos parâmetros
57 plot(mod1.8)
58
59
60 # Comparando a distribuição a posteriori e os pontos dos dados com a
    predições intervalares do modelo
61 pp_check(mod1.8)
62
63
64 pp_check(mod1.8,type = 'loo_intervals')
65
66
67 # Gerando intervalos de credibilidade
68 predict1.8 <- predictive_interval(mod1.8,prob = 0.95)
69
70
71 # Tabela com quantos valores estão contidos no intervalo de
    credibilidade
72 table(cbind( predict1.8[,1] <= data18$Def_esp & predict1.8[,2] >=
    data18$Def_esp ))
```

```
73
74
75 # Modelo com intervalo de confian a
76 plot(conditional_effects(mod1.8),points=T)
77
78 # Tabela com quantos valores de cobertura est o contidos no intervalo
  de credibilidade
79 cobert_int.18 <- predict1.8/data18$Def_Obs
80 table(cbind( cobert_int.18[,1] <= data18$Def_esp/data18$Def_Obs &
  cobert_int.18[,2] >= data18$Def_esp/data18$Def_Obs ))
81
82
83 ##### Dados de 2019 #####
84
85
86 # Criando-se 3 modelos, um com log(Inc_pobreza), outro com Inc_pobreza e
  um terceiro sem a vari vel Inc_pobreza
87 mod1.9 <- brm(Def_esp ~ log(Pop) + classe + (1|Departamento) + (1|Prov)
  + (1|Dist),data = data19,family = poisson(),prior = c(prior(normal
  (0,10),class=b)),cores = 4)
88
89
90
91 # Resumo dos modelos
92 summary(mod1.9)
93
94
95 # Converg ncia dos par metros
96 plot(mod1.9)
97
98
99 # Comparando a distribui o a posteriori e os pontos dos dados com a
  predi es intervalares do modelo
100 pp_check(mod1.9)
101
102
103 pp_check(mod1.9,type = 'loo_intervals')
104
105
106 # Gerando intervalos de credibilidade
107 predict1.9 <- predictive_interval(mod1.9,prob = 0.95)
108
109 # Tabela com quantos valores est o contidos no intervalo de
  credibilidade
110 table(cbind( predict1.9[,1] <= data19$Def_esp & predict1.9[,2] >=
  data19$Def_esp ))
111
```

```
112
113
114 # Modelo com intervalo de confiança
115 plot(conditional_effects(mod1.9),points=T)
116
117
118 # Tabela com quantos valores de cobertura estão contidos no intervalo
    de credibilidade
119 cobert_int.19 <- predict1.9/data19$Def_Obs
120 table(cbind( cobert_int.19[,1] <= data19$Def_esp/data19$Def_Obs &
    cobert_int.19[,2] >= data19$Def_esp/data19$Def_Obs ))
121
122
123 ##### Dados de 2020 #####
124
125
126 # Criando-se 3 modelos, um com log(Inc_pobreza), outro com Inc_pobreza e
    um terceiro sem a variável Inc_pobreza
127 mod1.20 <- brm(Def_esp ~ log(Pop) + classe + (1|Departamento) + (1|Prov
    ) + (1|Dist),data = data20,family = poisson(),prior = c(prior(normal
    (0,10),class=b)),cores = 4)
128
129
130
131
132 # Resumo dos modelos
133 summary(mod1.20)
134
135
136 # Convergência dos parâmetros
137 plot(mod1.20)
138
139
140 # Comparando a distribuição a posteriori e os pontos dos dados com a
    predição intervalares do modelo
141 pp_check(mod1.20)
142
143
144 pp_check(mod1.20,type = 'loo_intervals')
145
146
147
148 # Gerando intervalos de credibilidade
149 predict1.20 <- predictive_interval(mod1.20,prob = 0.95)
150
151
```

```
152 # Tabela com quantos valores est o contidos no intervalo de
      credibilidade
153 table(cbind( predict1.20[,1] <= data20$Def_esp & predict1.20[,2] >=
      data20$Def_esp ))
154
155
156
157 # Modelo com intervalo de confian a
158 plot(conditional_effects(mod1.20),points=T)
159
160
161 # Tabela com quantos valores de cobertura est o contidos no intervalo
      de credibilidade
162 cobert_int.20 <- predict1.20/data20$Def_Obs
163 table(cbind( cobert_int.20[,1] <= data20$Def_esp/data20$Def_Obs &
      cobert_int.20[,2] >= data20$Def_esp/data20$Def_Obs ))
```