

Introdução

Em ictiologia (estudo dos peixes), é comum que se queira medir e registrar as características físicas de um animal em questão. Entretanto, em certas situações, não é possível ou prático capturá-lo para que seja analisado; por exemplo, caso ele já esteja extinto — sobrando apenas seus fósseis — ou, devido a sua raridade de ser encontrado, câmeras fotográficas tenham que ser usadas. Em ambos os casos, os pesquisadores são capazes de medir suas características relacionadas ao comprimento, mas não à massa. Nesse sentido, nosso trabalho tem como objetivo fornecer a esses pesquisadores um método no qual é possível que se infira o peso de um peixe dado que saibam suas medidas, sua espécie e tenham mais observações sobre indivíduos dessa espécie ou espécies parecidas.

Métodos e Materias

Para atingir esse objetivo, foram amostrados 159 peixes comumente encontrados em mercados, e juntamente com a sua espécie, foram anotadas as medidas de 5 partes de seus corpos: o peso (Weight), a altura (Height), o comprimento da boca ao começo do rabo (Length 1), o comprimento da diagonal da boca até o começo do rabo (Length 2), o comprimento da boca até o final do rabo (Length 3) e a espessura (Width); o peso está em gramas e as outras medidas, em centímetros. As espécies contidas na amostra são: Bream, Roach, Whitefish, Parkki, Perch, Pike e Smelt.

Com os dados em mãos, primeiro, analisamos o nível de correlação entre as variáveis preditoras e decidimos quais seriam descartadas por agregar pouca informação. Também, com intuito de melhorar o ajuste do modelo, fizemos uma transformação logarítmica na variável Weight, e adicionamos 4 interações.

Em seguida, criamos um modelo de regressão linear múltipla, onde seus parâmetros foram estimados por uma técnica bayesiana denominada amostrador de Gibbs, aplicando burn-in na cadeia gerada — mas não thinning, tendo em vista que essa técnica é aconselhada apenas em casos muito específicos (veja <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2041-210X.2011.00131.x>). Assumimos distribuição Normal centrada em 0 para todos os parâmetros do modelo, com exceção da variância, que assumimos ser Gama Inversa. No mais, criamos os gráficos e tabelas necessários para se verificar as suposições utilizadas para realizar todos esses procedimentos: convergência da cadeia de Markov à distribuição estacionária; comparação da variância intra-cadeia com a entre-cadeia por meio da estatística R-chapéu; e distribuição dos resíduos pela variável resposta.

Por fim, analisamos os resultados obtidos por meio dos gráficos e intervalos de credibilidade da distribuição à posteriori dos parâmetros.

Vale dizer que, para manter o trabalho conciso, foram omitidos os gráficos da evolução da cadeia de Markov, da distribuição a posteriori e da distribuição preditiva para diversos dos parâmetros. Dito isso, esses gráficos se comportaram como era de se esperar, e de maneira muito semelhante aos gráficos que foram plotados.

Resultados

Uma primeira análise descritiva dos dados é dada pelas Figuras 1 e 2. Nelas, é possível observar que as variáveis Height, Length 1, Length 2 e Length 3 estão na mesma ordem de grandeza, enquanto Weight e Width, em uma ordem de grandeza menor. Nota-se, também, que existe uma observação na qual o peixe tem peso 0g; essa observação foi retirada dos dados.

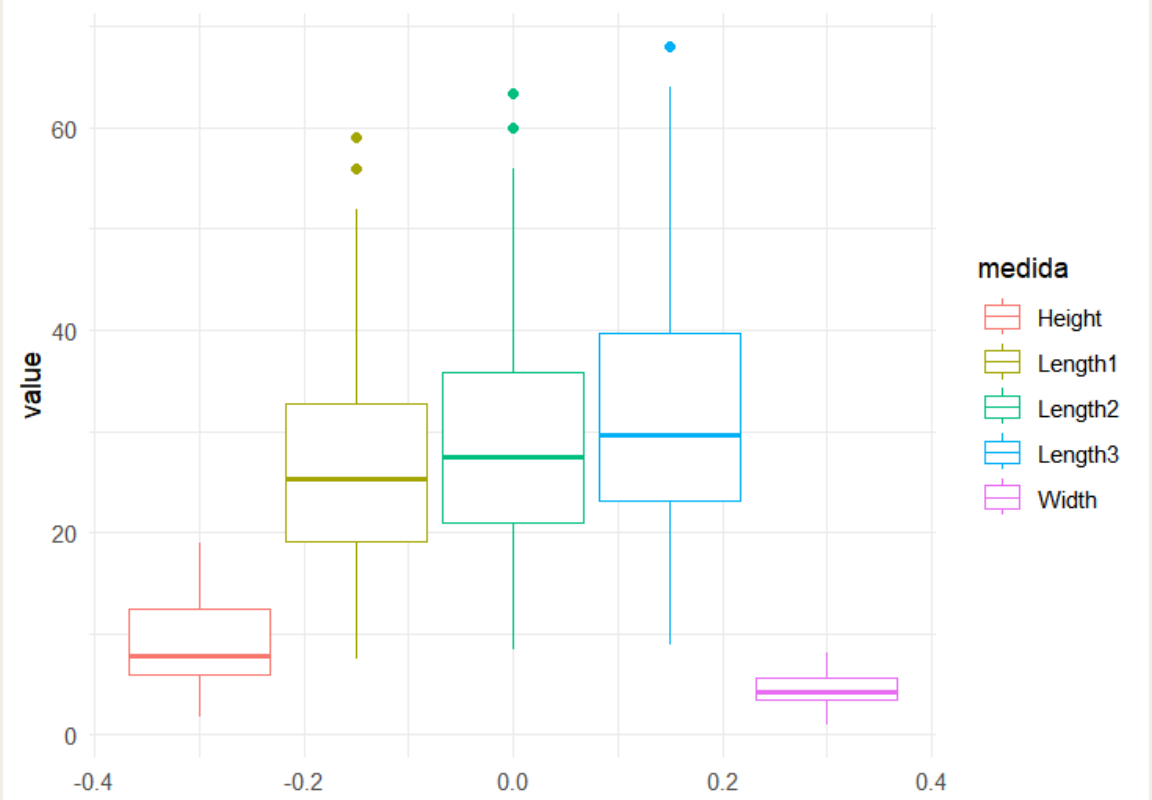


Figura 1

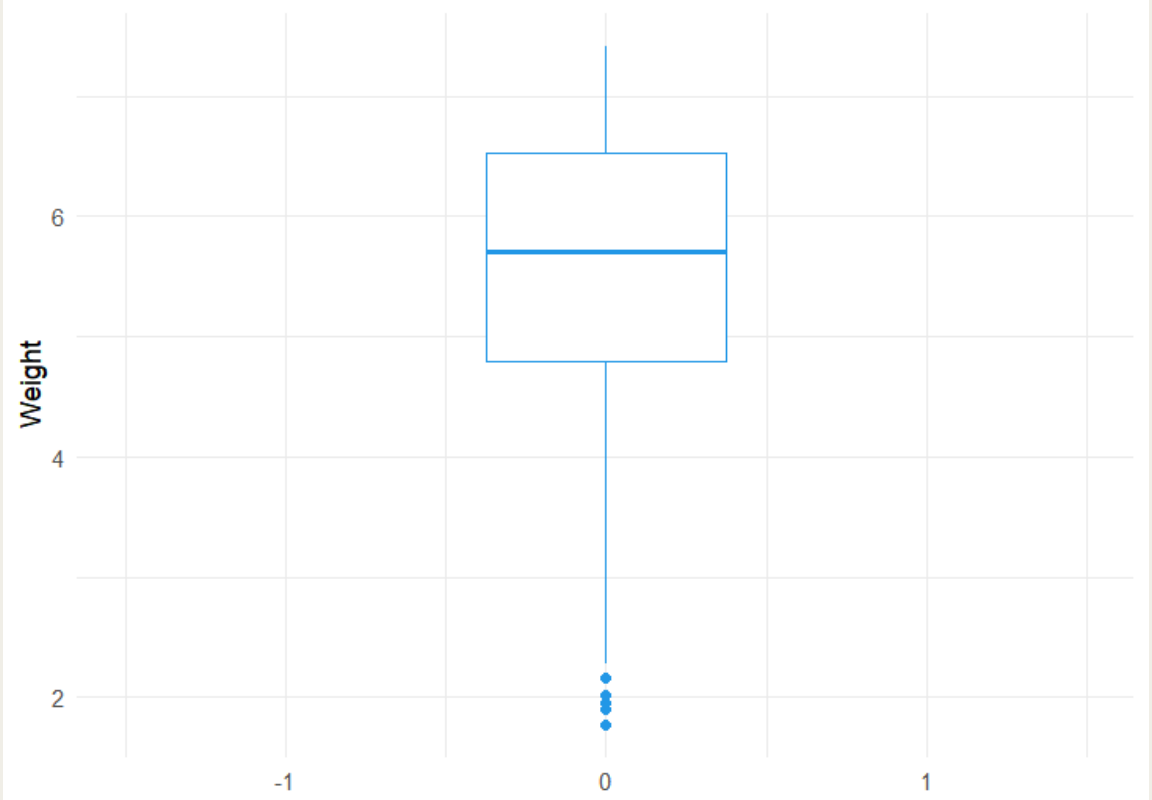
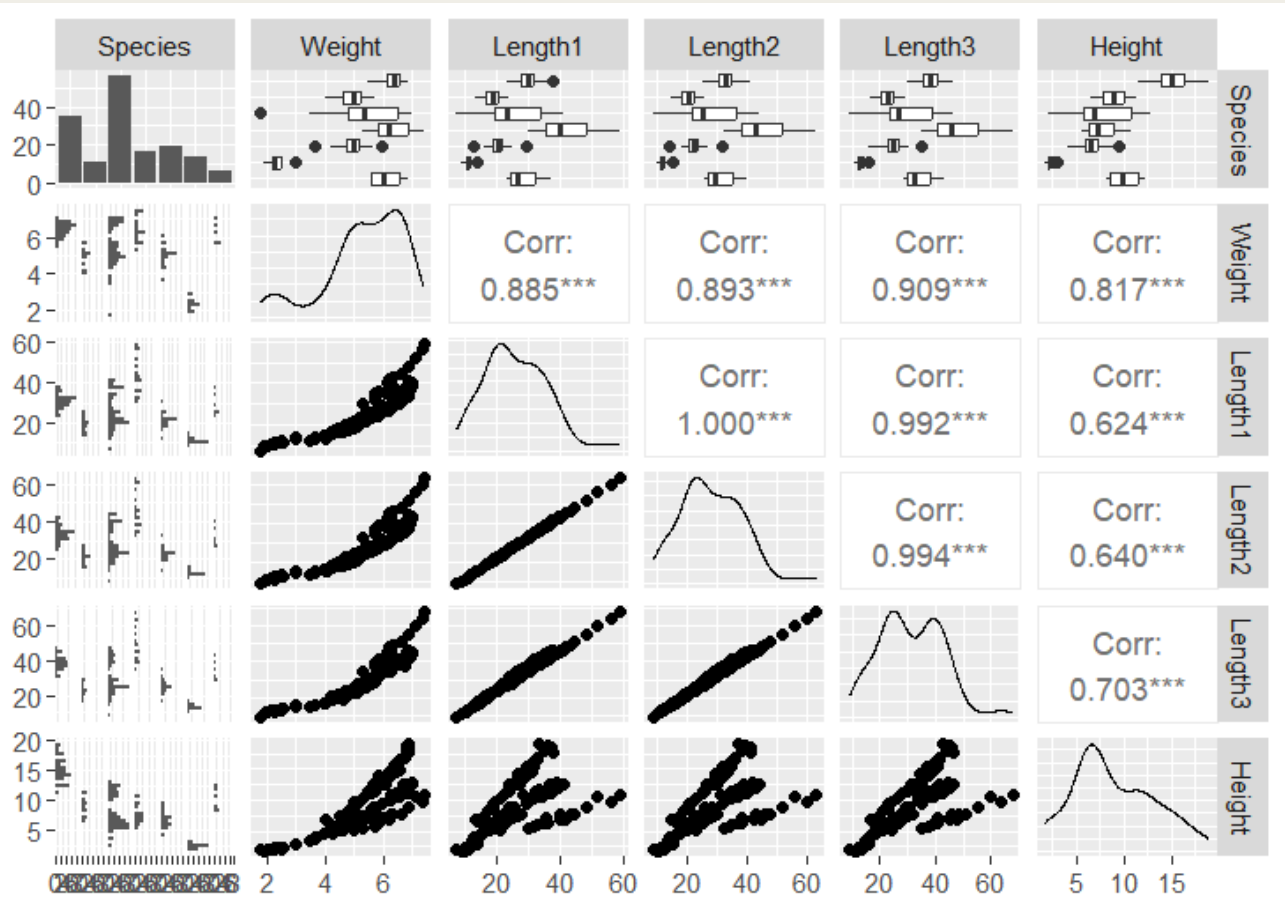


Figura 2

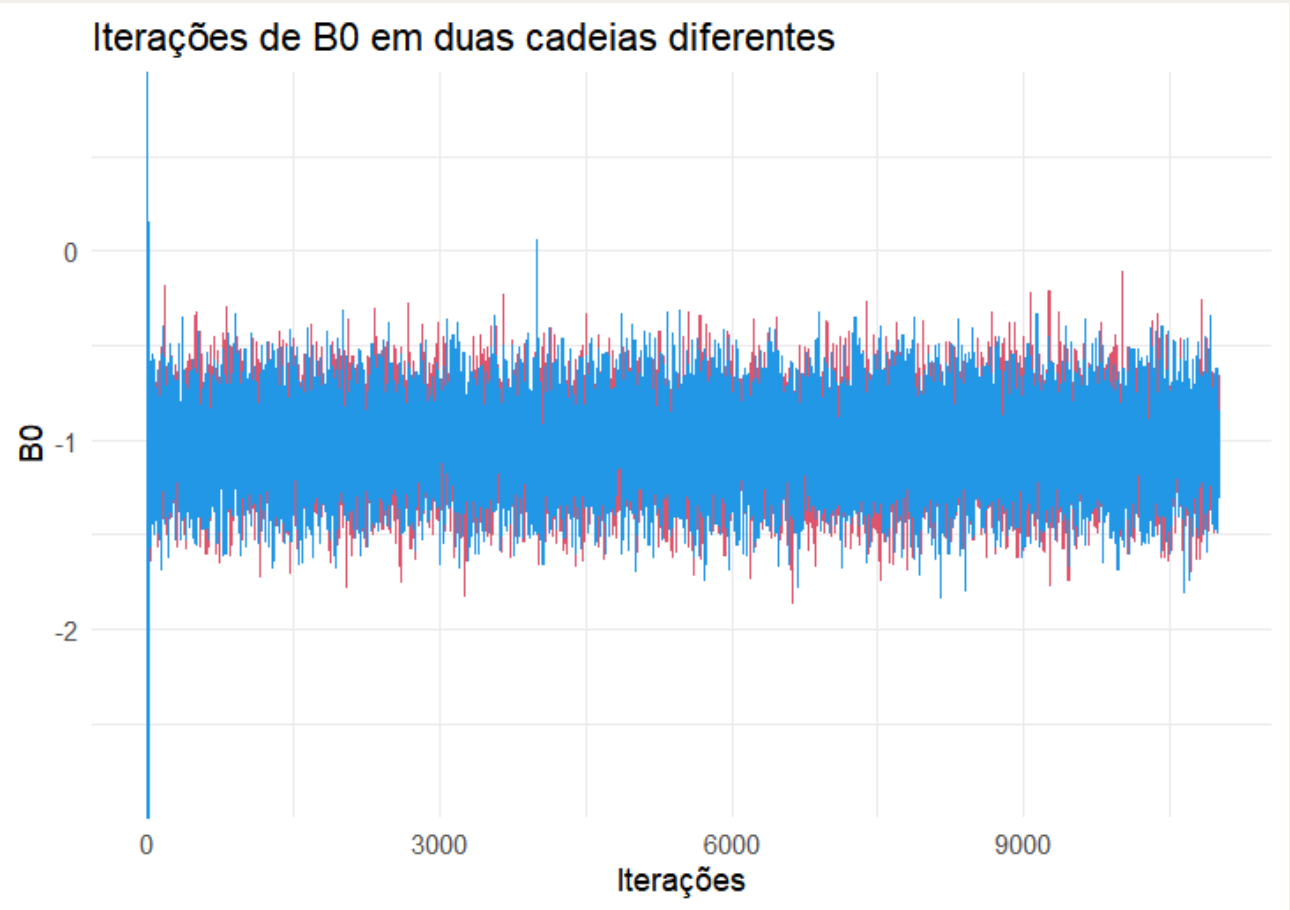
A Figura 3 nos mostra a correlação entre cada par de variáveis dos dados. Por conta da altíssima correlação entre Length 1 com Length 2 (1) e Length 1 com Length 3 (0.99), decidimos retirá-las da análise. Os parâmetros Betas representam as seguintes variáveis:

- B1: Length 1
- B2: Height
- B3: Width
- B4: Espécie Parkki
- B5: Espécie Perch
- B6: Espécie Pike
- B7: Espécie Roach
- B8: Espécie Smelt
- B9: Espécie Whitefish
- B10: Length 1:Height
- B11: Length 1:Width
- B12: Height:Width
- B13: Length 1:Height:Width

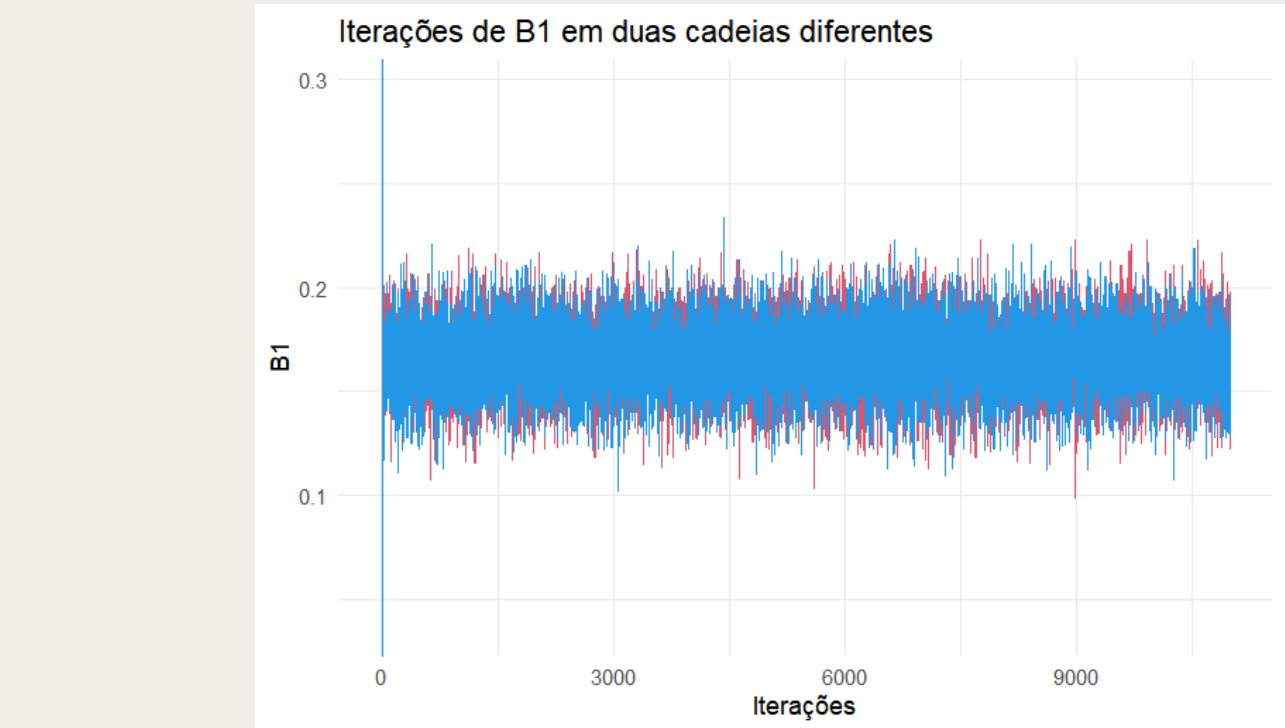
A evolução de ambas as cadeias de Markov para cada um dos parâmetros do modelo é dada pelas Figuras 4, 5 e 6. Analisando-as, fica claro que as cadeias convergiram para a distribuição estacionária.



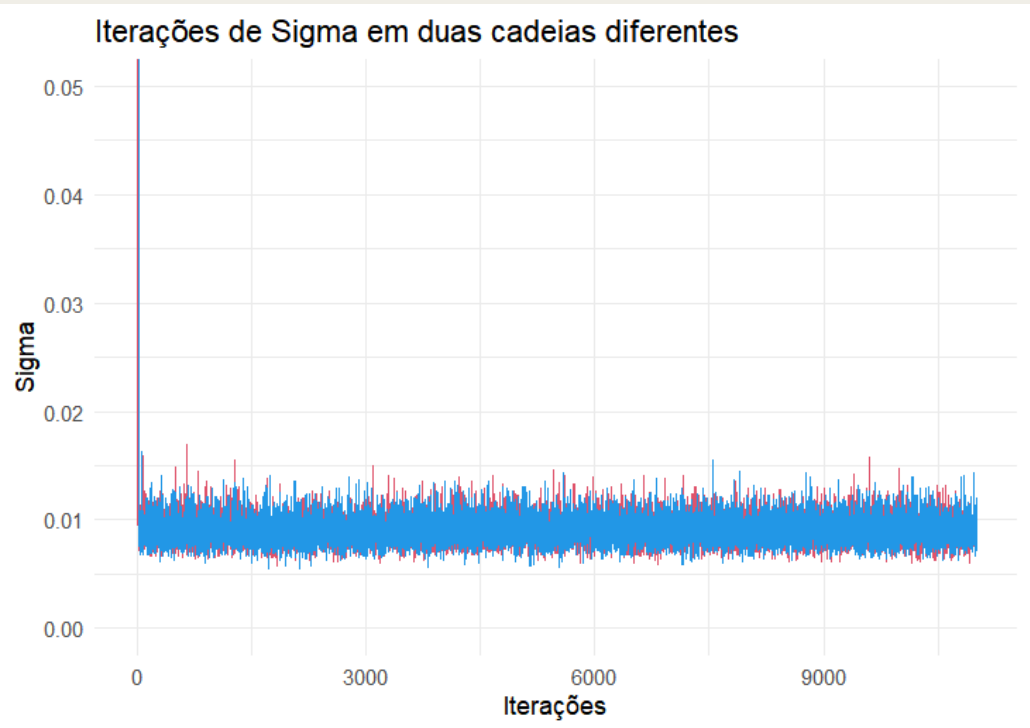
(a) Figura 3



(b) Figura 4



(c) Figura 5



(d) Figura 6

Após realizarmos um burn-in de 100 observações, calculamos a estatística R, que é dada pela Tabela 1 para cada um dos parâmetros. Nota-se que todos os valores estão muito próximos de 1, indicando que ambas as cadeias de Markov convergiram para os mesmos valores. As Figuras 7, 8 e 9 representam a distribuição a posteriori dos parâmetros. Podemos verificar na prática aquilo que já se esperava: todas as distribuições conjugam.

Parâmetro	R
B0	0.9999561
B01	1.0000145
B02	0.9999584
B03	1.0000197
B04	1.0000084
B05	1.0011130
B06	0.9999904
B07	1.0001532
B08	1.0000380
B09	0.9999520
B10	0.9999738
B11	1.0001008
B12	0.9999510
B13	1.0000076
Sigma	0.9999829

Tabela 1

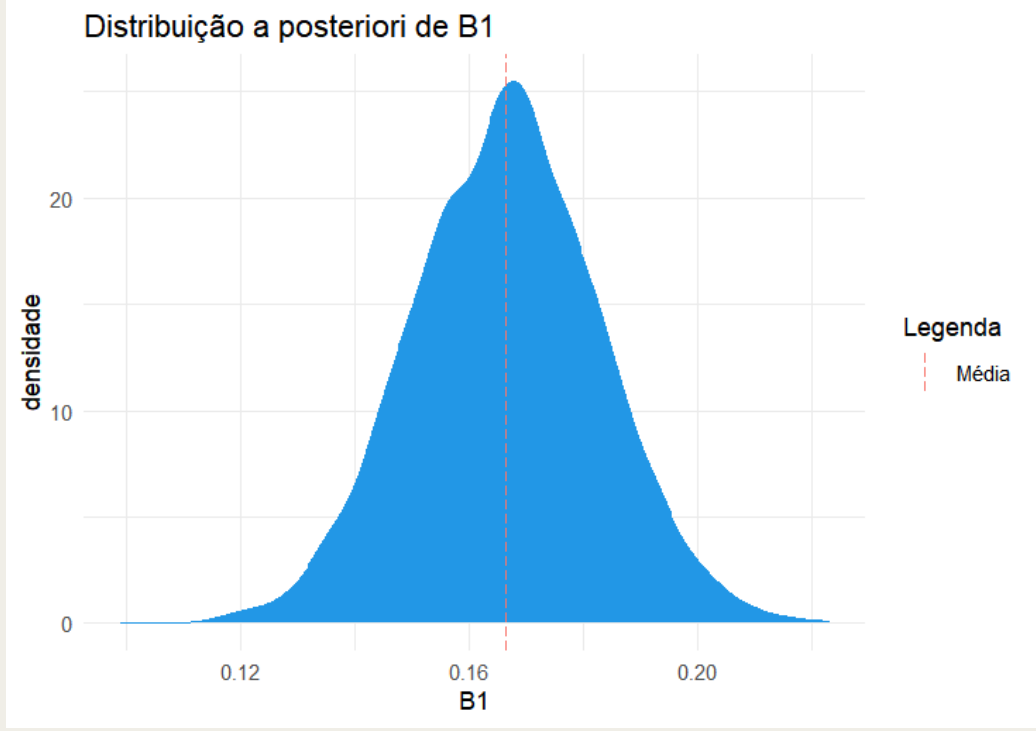


Figura 8

Parâmetro	2.5%	97.5%
B0	-1.4696	-0.5907
B01	0.1338	0.1976
B02	0.3794	0.5901
B03	0.5217	0.9383
B04	-0.1566	0.1073
B05	0.0129	0.4730
B06	-0.2417	0.5112
B07	0.0024	0.4286
B08	-0.1633	0.4376
B09	0.0909	0.4941
B10	-0.0159	-0.0077
B11	-0.0222	0.0129
B12	-0.0735	-0.0417
B13	0.0012	0.0020
Sigma	0.0070	0.0118

Tabela 2

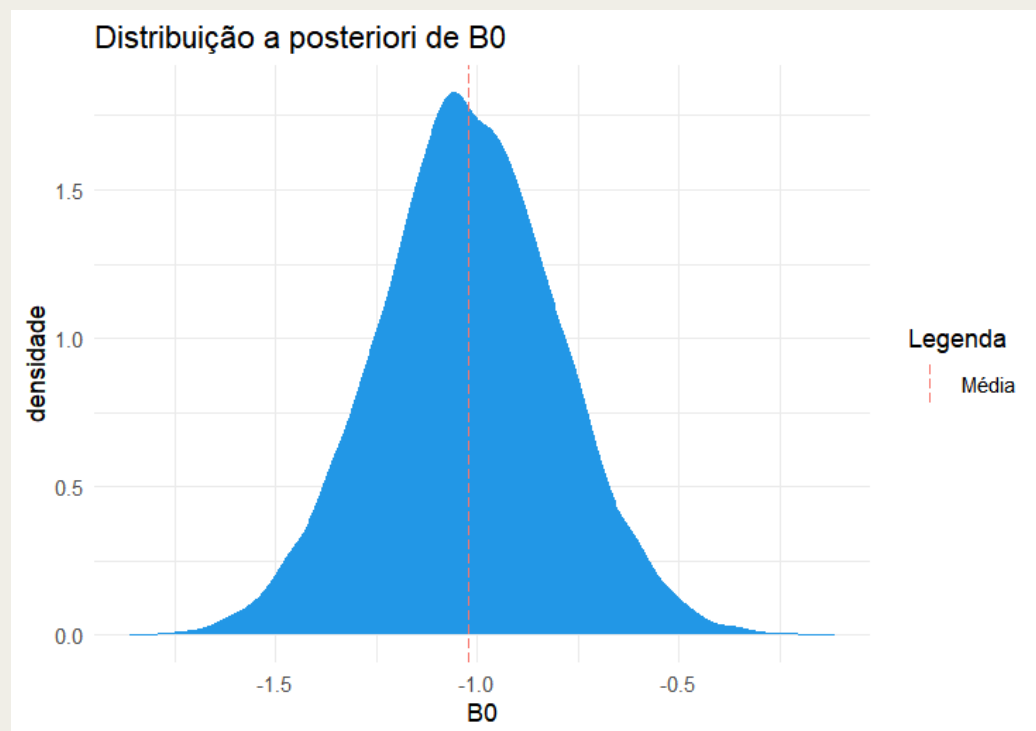


Figura 7

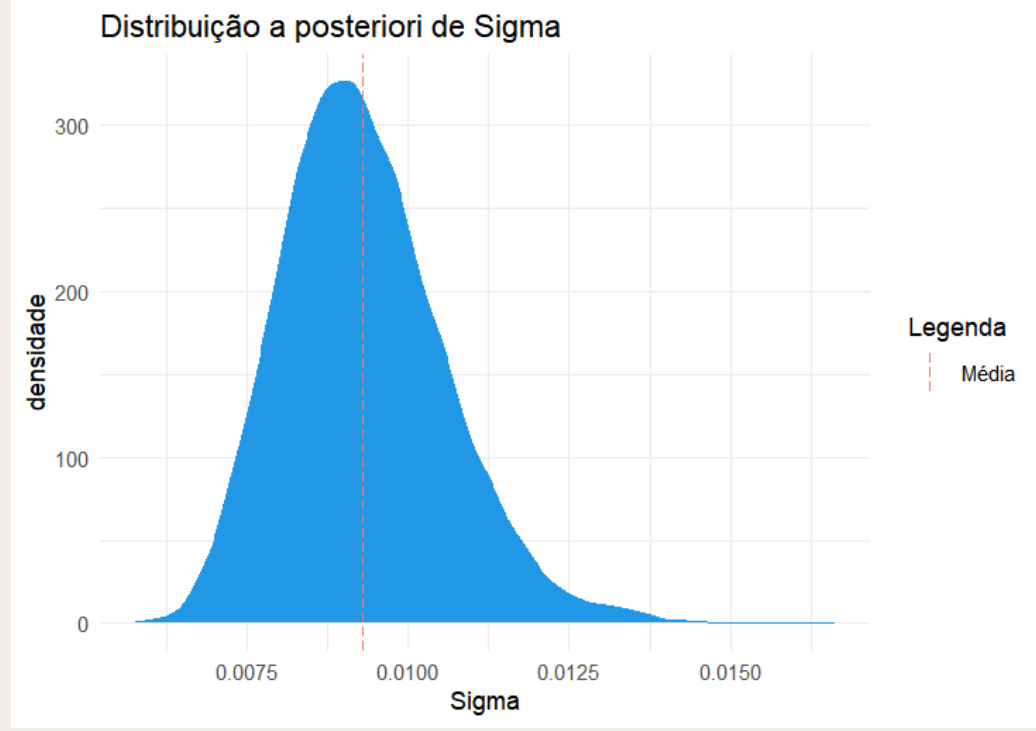


Figura 9

A Figura 10 e a Tabela 2 trazem os intervalos de credibilidade para cada um dos parâmetros. É possível observar que, com a exceção de algumas espécies de peixe, todas as variáveis são significativas. Devemos notar que peixes dessas espécies em questão — Parkki, Pike, Roach e Smelt — podem ser usados como observações para se ajustar um modelo preditivo para indivíduo da espécie Bream (que foi usada como referência).

A distribuição a posteriori da estatística R^2 é representada pela Figura 11. Por ela, podemos constatar que o modelo linear criado representa muito bem os dados.

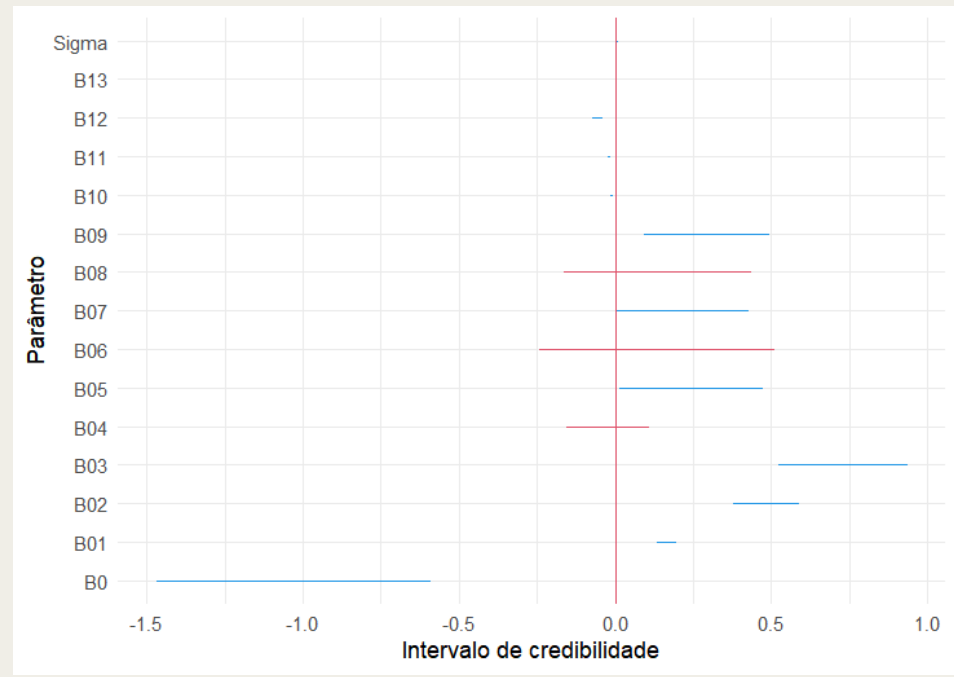


Figura 10

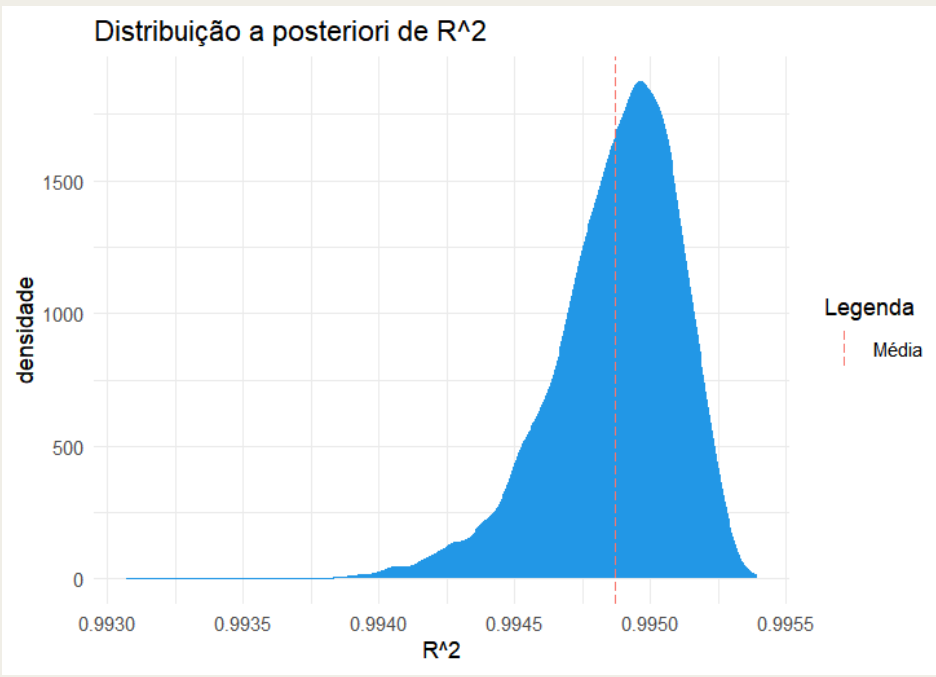


Figura 11

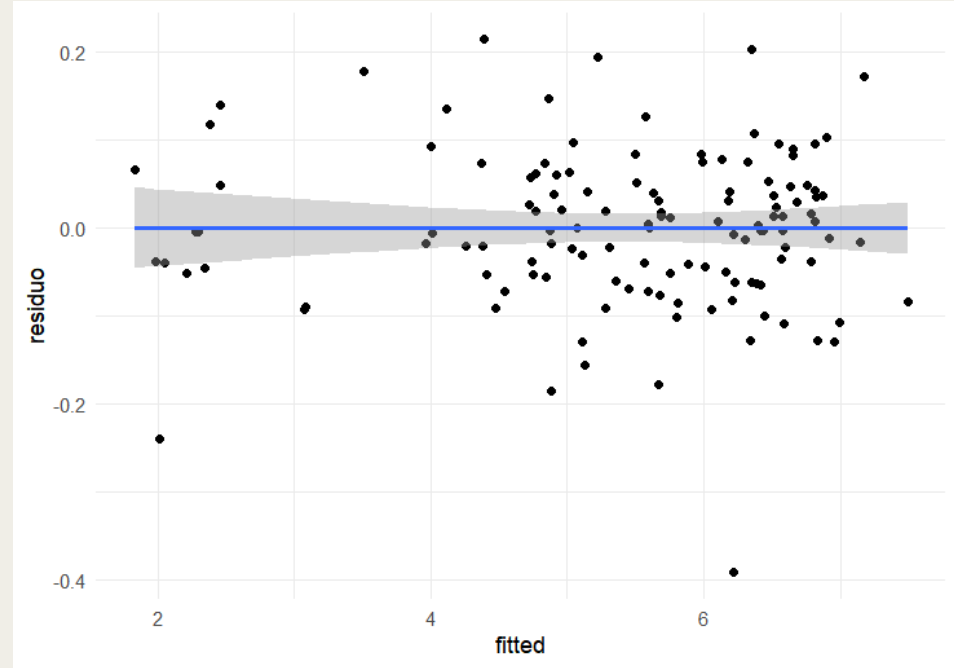


Figura 12

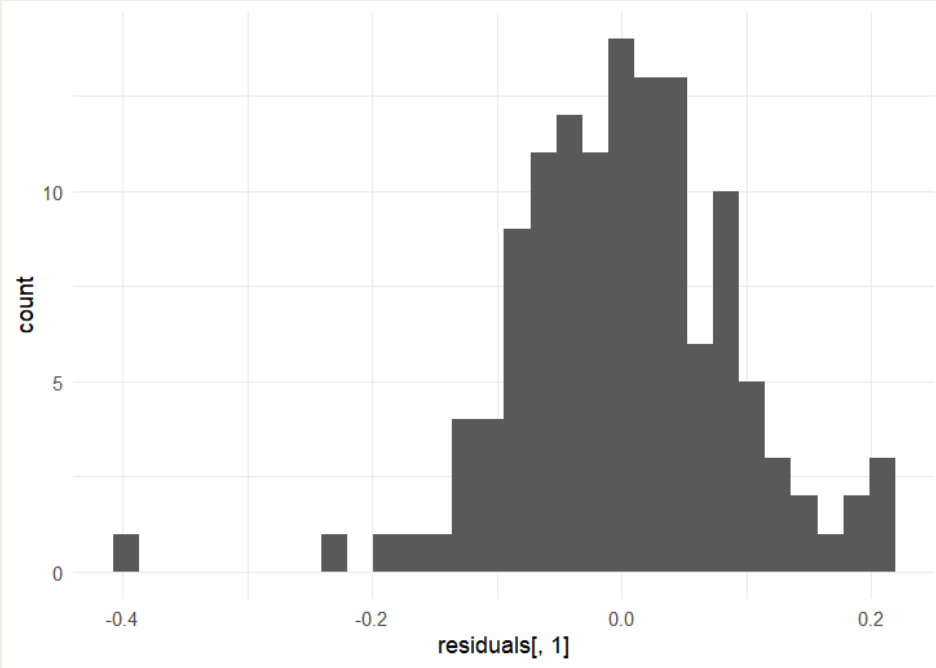


Figura 13

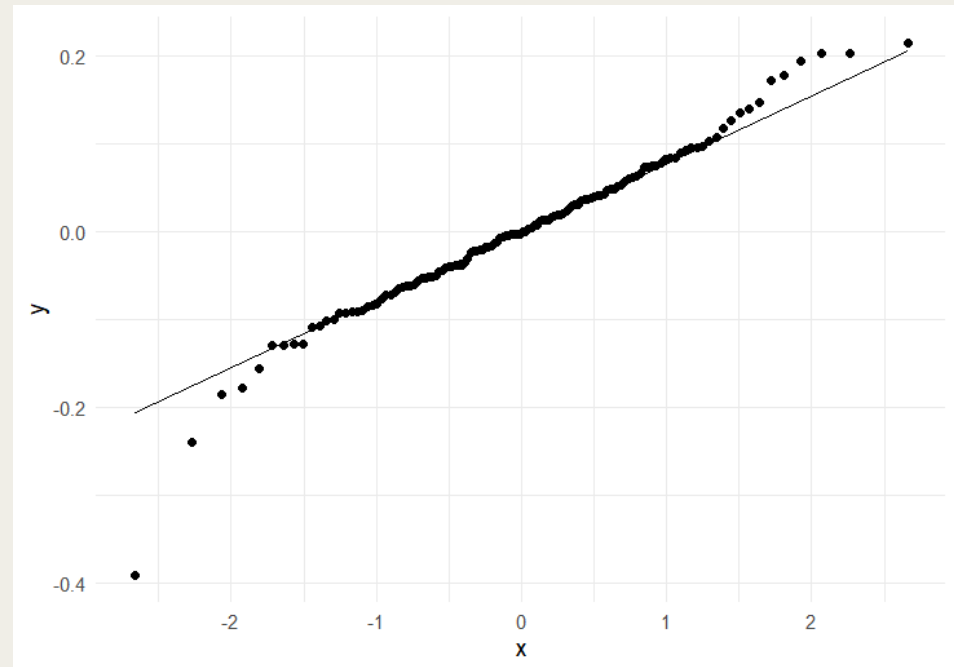


Figura 14

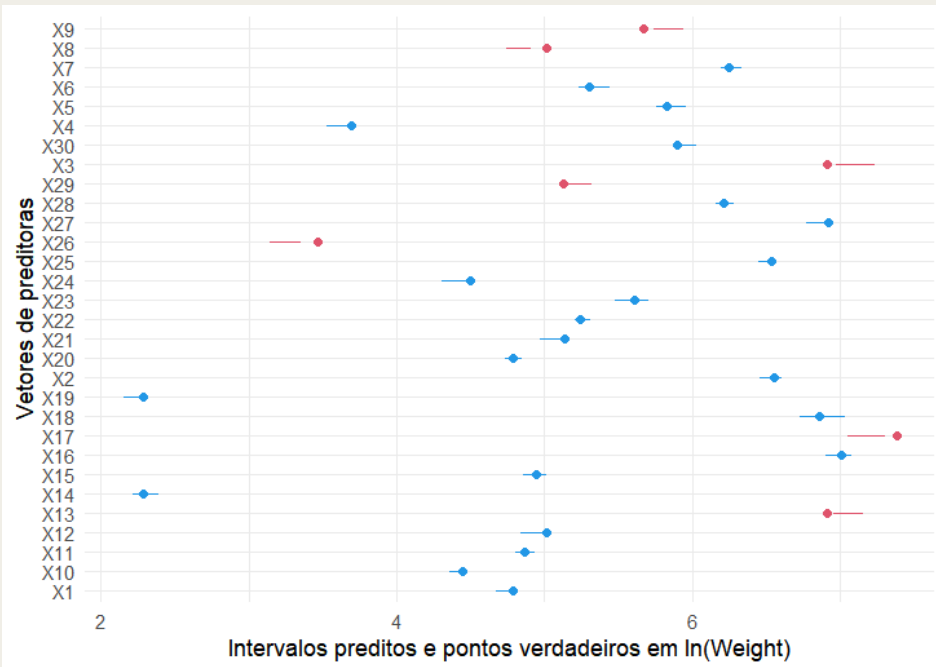


Figura 15

No mais, as Figuras 12, 13 e 14 nos permitem fazer uma análise gráfica do ajuste do modelo. Por elas, podemos verificar que não há nenhuma tendência visível entre valores do resíduo e o ajuste do modelo, que os erros são aproximadamente normal padrão, e que a variável resposta, dada as variáveis preditoras, também segue uma normal. Por fim, as distribuições preditivas dos parâmetros são dadas pelas Figura 15.

Conclusão

Dado os resultados obtidos, podemos afirmar que o método proposto pode ser útil para pesquisadores da área de ictiologia, pois, não apenas o modelo obteve um ajuste excelente em relação aos dados — propiciando uma boa previsão da massa de um peixe observado — como, caso esses pesquisadores tenham certeza da semelhança física entre duas espécies de peixes, uma rara e outra comum, por exemplo, um novo modelo pode ser ajustado, utilizando os dados dos indivíduos da espécie comum.