

**AVALIAÇÕES ANTIFRÁGEIS UMA COMPARAÇÃO ENTRE AS
INFERÊNCIAS BAYESIANA E CLÁSSICA**

AUTOR(A)

ianyqui falcao costa

REALIZAÇÃO

RESUMO

O processo embrionário da quinta revolução industrial, aliando o potencial criativo humano com a potencial analítico das máquinas modernas, nos proporcionou diversos desafios e perspectivas trazendo uma nova realidade comercial para as decisões gerenciais. Este trabalho apresenta uma perspectiva do método comparativo de dados de mercado aliado à riscos e incertezas sob a perspectiva da inferência bayesiana. No desenvolvimento, foram utilizados conceitos e informações da estatística básica, inferência estatística e simulações de Monte Carlo. Foi analisado a qualidade da definição do valor da avaliação do bem com a utilização da inferência clássica e a bayesiana.

Palavras-chave: Análise de risco, Simulação de Monte Carlo, Método comparativo.

ABSTRACT

The embryonic process of the fifth industrial revolution, combining human creative potential with the analytical potential of modern machines, provided us with several challenges and perspectives, bringing a new commercial reality to management's decisions. This work presents a perspective of the comparative's method combined with risks and uncertainties from the perspective of Bayesian inference. In the development, concepts and information from basics statistics, statistical inference and Monte Carlo simulations was used. The quality of the definition of the value of the appraisal of the asset was analyzed using classical and Bayesian inference.

Keywords: Risk analysis, Monte Carlo simulation, Comparative method.

1. INTRODUÇÃO

A engenharia vem cada vez mais sendo impactada por novas tecnologias e ferramentas de trabalho. Com a ideia de um modelo no qual engenheiros e arquitetos trabalham em total sinergia com máquinas inteligentes através da inteligência computacional. Robôs, computação nas nuvens, *webscrapper*, criatividade computacional, inteligência artificial, internet das coisas (IoT), inteligência de negócios, etc. são apenas alguns temas que rondam a prática da mesma nos últimos anos, exigindo dos profissionais uma adaptação à novos modelos que visam a tomada de decisões mais rápidas e precisas através de análises preditivas e inteligência operacional.

Nesse contexto, a inteligência artificial tornou-se um termo genérico para aplicativos que executam tarefas complexas que antes exigiam informações do usuário. O termo é frequentemente usado de forma intercambiável com seus subcampo, que incluem *Machine Learning* e *Deep Learning*. Esses, por sua vez, estão amplamente embasados na estatística bayesiana, em que ao contrário da abordagem frequentista, trata tudo como variável aleatória possibilitando uma modelagem bem mais flexível e poderosa, principalmente, em problemas mais complexo e realistas.

O objetivo central deste trabalho é comparar a estimativa de valor de mercado de um bem utilizando a inferência clássica e a inferência bayesiana. Espera-se que, como a solução da inferência bayesiana faz a utilização de simulações de Monte Carlo, a mesma apresente uma solução com maiores informações que a inferência clássica.

2. REVISÃO BIBLIOGRÁFICA E METODOLOGIA

Problemas de identificação de modelos são, em termos gerais, definidos como problemas inversos. Eles visam inferir algum desconhecido parâmetro θ de um certo modelo $\mathcal{M}$ dado algumas medições de y relacionadas com as respostas do modelo $g(\theta)$, dependente dos parâmetro do modelo. Sob a hipótese da adição de um ruído de observação ε , com parâmetros de ruído ϕ , resulta em:

$$y = \underbrace{g(\theta)}_{\text{modelo } \mathcal{M}} + \underbrace{e(\Phi)}_{\text{ruído}}$$

Em geral, o termo de ruído é considerado gaussiano, $e = N(0, \phi^{-1}I)$, com o grau de ruído desconhecido, Φ , a ser inferido.

Em métodos determinísticos para identificação de parâmetros, a solução é obtida minimizando a diferença entre os dados de medição y e as respostas do modelo $g(\theta)$:

$$k(\theta) = y - g(\theta)$$

Que podem ser solucionados através de mínimos quadrados, mínimos quadrados ponderados, etc.

Essas abordagens usam métodos de otimização para retornar a melhor estimativa para os parâmetros de interesse. Note que, as soluções são estimativas pontuais dos parâmetros, ou seja, os resultados não trazem nenhuma informação sobre quão confiável ou prováveis são.

Métodos de inferência probabilística, de outra forma, fornecem uma descrição probabilística das informações e crenças anteriores, permitindo incertezas. Isso é feito através do teorema de Bayes considerando todos os parâmetros desconhecidos como $\theta = [\theta, \phi]$

$$\underbrace{P(\theta|y, \mathcal{M})}_{\text{Posterior}} = \frac{\overbrace{P(y|\theta, \mathcal{M})}^{\text{verossimilhança}} \overbrace{P(\theta|\mathcal{M})}^{\text{priori}}}{\underbrace{\int P(y|\theta, \mathcal{M})P(\theta|\mathcal{M})d\theta}_{\text{termo de evidência } P(y|\mathcal{M})}}$$

Onde a função de densidade de probabilidade da priori reflete o conhecimento prévio dos parâmetros θ , a função de verossimilhança define a probabilidade das medições de y dado os parâmetros θ . O termo de normalização $P(y|\mathcal{M})$ descreve a evidência dos dados de y considerando o modelo $\mathcal{M}$. Note que, os parâmetros desconhecidos θ , incorpora tanto os parâmetros do modelo θ assim como os do erro ϕ . A posterior é uma função de densidade de probabilidade que reflete o conhecimento atualizado sobre os parâmetros após a observação dos dados. Sua solução no entanto, é extremamente complicada e custosa computacionalmente, principalmente quando $\dim(\theta) \gg 1$, no qual, necessita de aproximações.

Existem duas principais categorias para aproximações em métodos de inferência bayesiana: Métodos baseados em amostragem (por exemplo: Markov Chain Monte Carlo) e métodos baseados em otimização (por exemplo: Inferência Variacional).

Em métodos baseados em amostragem, os algoritmos mais conhecidos são: Metropolis-Hastings, Gibbs e Hamiltoniano Monte Carlo. Métodos que utilizam MCMC são ferramentas poderosas na solução de uma inferência bayesiana, eles fornecem informações assintoticamente não enviesadas. Uma outra vantagem é a natureza não paramétrica do método, na qual, não necessita da previa definição da forma do funcional da posteriori. Entretanto, existem algumas limitações com MCMC, geralmente exigem um alto custo computacional na amostragem, caso necessite de uma grande quantidade.

Métodos baseados em otimização, como a inferência variacional, de outra forma, são, geralmente, mais eficientes computacionalmente que os métodos baseados em amostragem, sendo mais apropriados para problemas de alta dimensão com modelos fisicamente bem postos. É fornecida uma aproximação para funções de densidade de probabilidade que são extremamente complicadas de serem resolvidas (integrais

intratáveis) através de otimização computacional. Em suas limitações, pode-se destacar que assim como a MCMC, a inferência variacional pode fornecer resultados enviesados e sua acurácia depende fortemente da família das funções de densidade de probabilidade adotada.

Com o objetivo de comparar o tratamento inferencial entre a estatística frequentista e a bayesiana, foi escolhido o tradicional método dos mínimos quadrados e o MCMC por Hamiltoniano Monte Carlo, respectivamente.

2.1. Procedimento Bayesiano

Como apresentado no item 2, a essência da abordagem bayesiana é tratar o parâmetro desconhecido θ como uma variável aleatória, especificar uma distribuição a priori para θ que represente as convicções sobre θ antes de ver os dados, usar o Teorema de Bayes para atualizar as convicções a priori na forma de probabilidades a posteriori e fazer inferências apropriadas. Portanto há quatro passos característicos da abordagem bayesiana:

1. Especificação da Verossimilhança do modelo $L(\theta|y) \equiv f(y|\theta)$;
2. Determinação da priori $f(\theta)$;
3. Cálculo da densidade à posteriori $f(\theta|y)$, através do Teorema de Bayes;
4. Extrair inferências da densidade à posteriori.

Portanto, a seguir explorar-se-á cada tópico.

2.1.1. Especificação da Verossimilhança

A escolha do modelo de Verossimilhança depende do que se assume por ser razoável sobre o mecanismo gerador de dados do problema em análise. Em geral, o conhecimento da estrutura pela qual os dados foram gerados pode sugerir um modelo apropriado, no entanto, o modelo será definido hipoteticamente para quê, em seguida, seja possível avaliar sua adequação no contexto dos dados.

Com o objetivo de reproduzir as condições da regressão linear por MMQ, foi adotado uma verossimilhança gaussiana, uma vez que a mesma é amplamente aceita como uma densidade representativa no contexto de avaliações de bens imóveis com o uso da inferência frequentista.

2.1.2. Definição das Prioris

Uma vez que no cálculo da posteriori é necessário a especificação de densidades a priori, a definição da mesma torna-se fundamental na inferência bayesiana. Portanto, sua escolha não deve ser leviana, necessitando de um cuidado considerável. A seguir será discutida alguns tipos de prioris.

2.1.2.1. Prioris Conjugadas

A principal dificuldade da estimativa da posteriori pelo Teorema de Bayes se situa no cálculo da constante normalizadora do denominador,

$$f(y) = \int f(y|\theta)f(\theta)d\theta,$$

O que na grande maioria dos casos não pode ser calculada analiticamente, restando apenas a análise numérica na sua estimativa.

No entanto, a depender do critério de escolha da priori, é possível calcular-se a posteriori sem a exigência de integrações numéricas. De outra forma, caso a densidade de probabilidade da priori pertencer à mesma família da posteriori, os cálculos podem ser simplificados. A esse conjunto de prioris dar-se o nome de prioris conjugadas.

Tabela 1 – Prioris Conjugadas Usuais

Verossimilhança	Priori	Posteriori
$Y \sim B(n, \mu)$	$\theta \sim Beta(p, q)$	$\theta y \sim Beta(p + y, q + n - y)$
$Y \sim P(\theta)$	$\theta \sim Ga(p, q)$	$\theta y \sim Ga\left(p + \sum_{i=1}^n y_i, q + n\right)$
$Y \sim \mathcal{N}(\theta, \tau^{-1}), (\tau \text{ conhecido})$	$\theta \sim \mathcal{N}(b, c^{-1})$	$\theta y \sim \mathcal{N}\left(\frac{cb + n\tau\bar{y}}{c + n\tau}, \frac{1}{c + n\tau}\right)$
$Y \sim Ga(k, \theta), (k \text{ conhecido})$	$\theta \sim Ga(p, q)$	$\theta y \sim Ga\left(p + nk, q + \sum_{i=1}^n y_i\right)$
$Y \sim Geo(\theta)$	$\theta \sim Beta(p, q)$	$\theta y \sim Beta\left(p + n, q + \sum_{i=1}^n y_i - n\right)$
$Y \sim BN(r, \theta)$	$\theta \sim Beta(p, q)$	$\theta y \sim Beta(p + r, q + y - r)$

Fonte: o Autor.

No entanto, precisa-se destacar que: prioris conjugadas são apenas um instrumento matematicamente conveniente de estimativa para as posteriores.

De modo geral, a utilização de prioris conjugadas é suficiente para representar convicções prévias para um nível de aproximação aceitável. Portanto, se não for este o caso, as prioris conjugadas não devem ser utilizadas apenas para facilitar o cálculo matemático.

2.1.2.2. Prioris não informativas

Quando não existe ou é muito vaga a informação sobre a densidade a priori, diversos estatísticos propuseram teorias que lidam com essa situação incorporando o que chamamos de prioris não informativas (*flat priors*). Tais densidades dão a mínima informação possível sobre o parâmetro. Definindo-as como uma densidade que não favorece nenhum valor específico de θ em relação a outros, ou seja, uma densidade uniforme com probabilidades equiprováveis para os possíveis valores, ou uma densidade gaussiana com média zero e desvio padrão extremamente alto, se assemelhando a algo próximo da densidade uniforme.

É comum, ajustar-se uma priori não informativa considerando a medida de informação de Fisher (Priori de Jeffreys). Analiticamente têm-se:

$$I(\theta) = E \left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta} \right)^2 | \theta \right] = -E \left[\frac{\partial^2 \ln f(X|\theta)}{\partial \theta^2} \right]$$

No caso de um vetor paramétrico $\theta = (\theta_0, \theta_1, \dots, \theta_k)$,

$$I(\theta)_{ij} = E \left[\left(\frac{\partial \ln f(X|\theta)}{\partial \theta_i} \right) \left(\frac{\partial \ln f(X|\theta)}{\partial \theta_j} \right) | \theta \right]$$

$$I(\theta)_{ij} = E \left[\left(\frac{\partial^2 \ln f(X|\theta)}{\partial \theta_i \partial \theta_j} \right) | \theta \right]$$

A priori de Jeffreys é definida como:

$$\pi(\theta) \propto |I(\theta)|^{1/2}$$

2.1.2.3. Mistura de Prioris

Uma pequena discursão sobre a utilização de prioris conjugadas foi tratada no item 3.1.2.1.. Entretanto, destaca-se que, por muitas vezes a utilização de famílias de prioris conjugadas naturais, como na tabela 1, podem ser bastante restritiva na modelagem da priori. Uma possibilidade de flexibilidade na especificação da priori é a utilização de uma priori que resulta da mistura de densidades. Tratando analiticamente:

Supondo que $f_1(\theta), \dots, f_n(\theta)$ são todas distribuições conjugadas para θ , levando a posterioris $f_1(\theta|y), \dots, f_n(\theta|y)$. Considere a família de mistura das distribuições:

$$f(\theta) = \sum_{i=1}^k p_i f_i(\theta)$$

Então, pelo Teorema de Bayes (sem a constante de normalização):

$$f(\theta|y) \propto f(\theta) f(y|\theta)$$

$$= \sum_{i=1}^k p_i f_i(\theta) f(y|\theta)$$

$$\propto \sum_{i=1}^k p_i^* f_i(\theta|y)$$

Sendo assim, a posteriori pertence à mesma família de misturas que a priori. No entanto, as proporções p_i^* de mistura na posteriori não necessariamente serão os mesmos da priori, p_i .

Os pesos da posteriori, p_i^* , são obtidos a partir dos pesos das prioris, p_i , e irá depender dos seus respectivos termos constantes para cada termo da mistura. A seguir apresentamos analiticamente:

Sejam as densidades da priori, $f_i(\theta) = C_i K_i(\theta)$, e da posteriori, $f_i(\theta|y) = C_i^* K_i(\theta|y)$, isolando os termos independentes têm-se:

$$f(\theta|y) \propto \sum_{i=1}^k p_i f_i(\theta) f(y|\theta)$$

$$\propto \sum_{i=1}^k p_i C_i K_i(\theta) L(\theta|y)$$

$$\propto \sum_{i=1}^k p_i C_i K_i(\theta|y)$$

$$\propto \sum_{i=1}^k p_i \frac{C_i}{C_i^*} f_i(\theta|y)$$

$$\propto \sum_{i=1}^k w_i f_i(\theta|y)$$

Portanto, os pesos na posteriori são:

$$p_i^* = \frac{w_i}{\sum_i w_i},$$

Em que:

REALIZAÇÃO

$$w_i = p_i \frac{C_i}{C_i^*}$$

2.1.2.4. Prioris Empíricas

Em um processo genuinamente bayesiano, a definição da densidade a priori deve ser formulada independentemente dos dados coletados. No entanto, métodos bayesiano empíricos são amplamente utilizados em análises bayesianas. Sendo assim, a especificação de prioris empíricas se sobressaem pela utilização de informações das amostras coletadas, com destaque na utilização da própria estatística descritivas na definição das mesmas.

2.1.3. Estimativa da posteriori por MCMC - HMC

O método de Monte Carlo Hamiltoniano (HMC - Hamiltonian Monte Carlo) é um método de Monte Carlo via Cadeia de Markov cuja implementação tem base em conceitos da área de física, para gerar a distribuição posterior de uma forma mais eficiente. Nele é utilizada a Dinâmica Hamiltoniana e o processo de aceitação do método de Metropolis.

A principal vantagem do HMC em relação aos métodos que realizam as simulações por passeio aleatório é que o processo de convergência, empiricamente, ocorre de maneira mais rápida, ou seja, com um número geralmente bem menor de iterações.

No algoritmo de Metropolis, define-se uma distribuição auxiliar, $q(\blacksquare|y)$, cuja função deve ser simétrica em y e x . Essa distribuição gera amostragens para um dado y^* , a partir de $q(\blacksquare|y_t)$, que pode ser rejeitado, analisando-se a razão:

$$r = \frac{p(y^*)}{p(y_t)},$$

sendo $p(\blacksquare)$ a distribuição de probabilidade alvo.

Se $r \geq 1$, define-se $y_{t+1} = y^*$, caso contrário define-se:

$y_{t+1} = y^*$, com probabilidade r ,

$y_{t+1} = y_t$, com probabilidade $1 - r$,

e assim realiza-se esse procedimento para um certo número de iterações.

O método de Monte Carlo Hamiltoniano utiliza variáveis de momento, ρ , incluídas artificialmente no modelo, e define-se a distribuição de probabilidade conjunta:

$$p(\rho, \theta) = p(\rho|\theta)p(\theta)$$

A distribuição de ρ é independente de θ , sendo geralmente atribuído uma distribuição normal multivariada $\rho \sim N(0, \Sigma)$, em que Σ é uma matriz de variância-covariância, simétrica, positiva-definida, e geralmente diagonal. Essa matriz usualmente é chamada de “matriz de massa”.

A função de Hamilton é definida a partir da distribuição conjunta $p(\rho, \theta)$:

$$H(\rho, \theta) = -\log p(\rho, \theta),$$

$$H(\rho, \theta) = -\log p(\rho|\theta) - \log p(\theta),$$

$$H(\rho, \theta) = T(\rho|\theta) + V(\theta),$$

em que $T(\rho|\theta) = -\log p(\rho|\theta)$ é chamado de “energia cinética” e $V(\theta) = -\log p(\theta)$ é a “energia potencial”. Aqui cabe o conceito de conservação de energia, de forma que a função de Hamilton permanece invariante.

Iniciando a partir de um estado da variável θ , define-se um valor para ρ , independente θ , e que não apresenta persistência ao longo das iterações. A dinâmica de ρ e θ ao longo do tempo t é definida pelas derivadas parciais da equação de Hamilton:

$$\frac{\partial \theta}{\partial t} = \frac{\partial H}{\partial \rho} = \frac{\partial T}{\partial \rho}$$

$$\frac{\partial \rho}{\partial t} = -\frac{\partial H}{\partial \theta} = -\frac{\partial T}{\partial \theta} - \frac{\partial V}{\partial \theta}$$

Como o momento ρ é independente de θ , $p(\rho|\theta) = p(\rho)$, e assim $-\partial T / \partial \theta = 0$. Dessa forma, as derivadas da equação de Hamilton, ficam:

$$\frac{\partial \theta}{\partial t} = \frac{\partial T}{\partial \rho}$$

$$\frac{\partial \rho}{\partial t} = -\frac{\partial V}{\partial \theta}$$

Para resolver as equações diferenciais para os dois estados, utiliza-se aqui o integrador Leapfrog, que é um algoritmo de integração numérica adaptado para fornecer resultados estáveis para o HMC. Esse algoritmo opera através de etapas de tempo discretas, de um pequeno intervalo ϵ , após definir o momento ρ inicial. Em seguida, ele alterna entre uma atualização em $t + \epsilon/2$ para ρ , e uma atualização em $t + \epsilon$ para θ e ρ , da seguinte forma:

$$\rho^{(t+\epsilon/2)} \leftarrow \rho^{(t)} - \frac{\epsilon}{2} \frac{\partial V^{(t)}}{\partial \theta}$$

$$\theta^{(t+\epsilon)} \leftarrow \theta^{(t)} + \epsilon \Sigma \rho^{(t+\epsilon/2)}$$

$$\rho^{(t+\epsilon)} \leftarrow \rho^{(t+\epsilon/2)} - \frac{\epsilon}{2} \frac{\partial V^{(t+\epsilon)}}{\partial \theta}$$

Após L repetições das etapas acima, é realizada uma simulação para $L\epsilon$ períodos discretos de tempo e o estado final resultante desse processo é denotado como (ρ^*, θ^*) . Como a atualização de uma variável depende somente da outra variável, as atualizações feitas pelo método de Leapfrog mantém o “volume” da distribuição constante após mapear seus pontos com o integrador.

A distribuição que se pretende amostrar é obtida a partir de outro conceito de física (função de energia potencial, pelo conceito de distribuição canônica da estatística mecânica). A função de densidade da distribuição conjunta, $P(\rho, \theta)$, fica definida como:

$$P(\rho, \theta) = \frac{1}{Z} \exp(-H(\rho, \theta)),$$

sendo Z uma constante para garantir que $\int_{-\infty}^{+\infty} P(\rho, \theta) = 1$.

No processo de integração de Leapfrog é aplicada a condição de aceitação do método de Metropolis. A probabilidade de se aceitar a transição do estado (ρ, θ) para o estado (ρ^*, θ^*) , após as L repetições é:

$$\min(1, \exp(H(\rho, \theta) - H(\rho^*, \theta^*)))$$

Se o estado proposto não for aceito, o parâmetro retorna para seu valor anterior, e a próxima iteração é iniciada a partir dele.

2.1.3.1. Convergência

Na prática, não é possível afirmar com certeza que, após um dado número de iterações, uma cadeia convergiu para uma distribuição estacionária. Porém, existem alguns testes que indicam que essa convergência pode ter ocorrido. Possivelmente o teste mais utilizado para fazer essa verificação, é o diagnóstico de Gelman-Rubin.

Sua aplicação ocorre após a simulação de n realizações para m cadeias paralelas, iniciando de pontos diferentes de um espaço π . Escolhendo uma função escalar ω , pertencente ao espaço π , ω_{ij} é a i -ésima realização para a j -ésima cadeia. A “variância” entre as cadeias é definida por:

$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\omega}_j - \bar{\omega})^2 = 1,$$

em que $\bar{\omega}$ é a média total para todas as mn realizações e $\bar{\omega}_j$ é a média das n realizações da cadeia j .

A variância dentro das cadeias é definida como uma média da variância dentro de cada cadeia:

$$W = \frac{1}{m(n-1)} \sum_{i=1}^n \sum_{j=1}^m (\omega_{ij} - \bar{\omega}_j)^2,$$

Os valores de B e W , se a cadeia for ergódica, com $n \rightarrow \infty$, serão estimações consistentes da variância de ω , σ^2 , assim como a média ponderada deles:

$$\hat{\sigma}_{\omega}^2 = \left(1 - \frac{1}{n}\right) W + \left(\frac{1}{n}\right) B,$$

A verificação de convergência é feita a partir da estatística de Gelman-Rubin:

$$R = \sqrt{\frac{\hat{\sigma}_{\omega}^2}{W}},$$

Se os pontos iniciais da cadeia estão mais dispersos relativamente à variância esperada para ω , a variância entre as cadeias B também ficará acima da variância de ω . E a variância dentro das cadeias W , no início, tende a ficar abaixo da variância de ω , já que ainda não explorou completamente o espaço de estado. Dessa forma, a estatística de Gelman-Rubin, R , tende a ser maior do que 1 no início da cadeia, mas se aproxima desse valor conforme aumenta o comprimento da cadeia.

2.1.4. Inferência

Uma análise bayesiana entrega uma inferência bem mais completa em termos de informação, no sentido de que todo o conhecimento acumulado a respeito de θ , incorporado na priori e nos dados, é representado pela densidade posteriori.

No entanto, em avaliações de bens imóveis, deseja-se inferir o valor do bem através de uma estimativa pontual e/ou intervalar. O que se torna possível através da estimativa dos coeficientes da posteriori.

3. ESTUDO DE CASO

Deseja-se avaliar um apartamento de 150m² de área privativa, com 2 quartos, sendo 2 suítes, 1 WC social, 3 vagas de garagem e 1 quarto com WC de empregada. Para avaliação foram coletados vinte dados de mercado relativos às Áreas Privativas em m² (AP), Quartos, Vagas de Garagem, número de WC com peso 0,5 para WC de empregada e 1 WC para suíte e os respectivos preços unitários em R\$/m² (PU), conforme tabela 1. Com base nos dados coletados, analisou-se o valor de mercado de compra e venda do apartamento avaliando utilizando a inferência frequentista e a bayesiana.

Tabela 1 – Dados coletados

Dados	Área	Quartos	Garagem	WC	Pu(R\$/m ²)
1	122.00	3.00	1.00	3.50	5081.97
2	157.00	4.00	3.00	3.50	6050.96
3	196.00	4.00	3.00	3.00	4591.84
4	175.00	4.00	2.00	5.00	5628.57
5	131.00	3.00	2.00	2.50	4961.83
6	200.00	4.00	2.00	2.50	4500.00
7	108.00	3.00	2.00	2.50	5833.33
8	137.00	4.00	2.00	5.50	6569.34
9	117.00	3.00	2.00	2.50	5555.56
10	167.00	4.00	2.00	7.50	5988.02
11	136.00	4.00	2.00	5.50	6577.21
12	163.00	4.00	3.00	4.50	6134.97
13	130.00	4.00	2.00	3.50	7067.20
14	132.00	4.00	2.00	3.00	6818.18
15	149.00	4.00	2.00	4.50	6577.18
16	158.00	4.00	2.00	3.50	5632.91
17	160.00	4.00	2.00	2.50	5562.50
18	127.00	4.00	2.00	2.50	7086.61
19	162.00	4.00	2.00	3.00	5833.33
20	120.00	3.00	1.00	2.50	5000.00

Fonte: o Autor.

3.1. Inferência Frequentista – Regressão Linear por MMQ

Em seguida, foi realizado a análise exploratória e possíveis mudanças de escalas (direta – x ou logarítmica – Ln x) com o objetivo de identificar o melhor ajuste inferencial. Portanto, temos que:

Tabela 2 – Correlações e aderência da melhor escala

	Área	Ln(Área)	Quartos	Ln(Quartos)	Garagem	Ln(Garagem)	WC	Ln(WC)	Pu(R\$/m²)	Ln(Pu(R\$/m²))
Área	1									
Ln(Área)	0.996	1								
Quartos	0.643	0.682	1							
Ln(Quartos)	0.643	0.682	1	1						
Garagem	0.511	0.518	0.522	0.522	1					
Ln(Garagem)	0.495	0.505	0.576	0.576	0.974	1				
WC	0.208	0.247	0.411	0.411	0.102	0.128	1			
Ln(WC)	0.217	0.260	0.453	0.453	0.139	0.155	0.986	1		
Pu(R\$/m²)	-0.383	-0.333	0.432	0.432	0.112	0.198	0.333	0.368	1	
Ln(Pu(R\$/m²))	-0.394	-0.342	0.418	0.418	0.111	0.197	0.351	0.386	0.998	1

Fonte: o Autor.

Tabela 3 – Dados utilizados

Dados	Área	Quartos	Ln(Garagem)	Ln(WC)	Ln(Pu(R\$/m²))
1	122.00	3.00	0.00	1.25	8.53
2	157.00	4.00	1.10	1.25	8.71
3	196.00	4.00	1.10	1.10	8.43
4	175.00	4.00	0.69	1.61	8.64
5	131.00	3.00	0.69	0.92	8.51
6	200.00	4.00	0.69	0.92	8.41
7	108.00	3.00	0.69	0.92	8.67
8	137.00	4.00	0.69	1.70	8.79
9	117.00	3.00	0.69	0.92	8.62
10	167.00	4.00	0.69	2.01	8.70
11	136.00	4.00	0.69	1.70	8.79
12	163.00	4.00	1.10	1.50	8.72
13	130.00	4.00	0.69	1.25	8.86
14	132.00	4.00	0.69	1.10	8.83
15	149.00	4.00	0.69	1.50	8.79
16	158.00	4.00	0.69	1.25	8.64
17	160.00	4.00	0.69	0.92	8.62
18	127.00	4.00	0.69	0.92	8.87
19	162.00	4.00	0.69	1.10	8.67
20	120.00	3.00	0.00	0.92	8.52

Fonte: o Autor.

Partindo então para a regressão através do método dos mínimos quadrados, temos os resultados apresentados na tabela 4:

Tabela 4 – Resultados da Regressão Linear

RESUMO DOS RESULTADOS								
<i>Estatística de regressão</i>								
R múltiplo	97.83%							
R-Quadrado	95.70%							
R-quadrado ajustado	94.56%							
Erro padrão	0.032							
Observações	20							
ANOVA								
	<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>			
Regressão	4	3.318E-01	8.295E-02	8.350E+01	0.000000046%			
Resíduo	15	1.490E-02	9.934E-04					
Total	19	3.467E-01						
	<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>	<i>Inferior 95.0%</i>	<i>Superior 95.0%</i>
Interseção	8.300	0.065	126.763	0.00%	8.161	8.440	8.161	8.440
Área	-	0.006	0.000	15.996	0.00%	0.007	0.005	0.007
Quartos	0.297	0.026	11.640	0.00%	0.243	0.352	0.243	0.352
Ln(Garagem)	0.085	0.033	2.595	2.03%	0.015	0.156	0.015	0.156
Ln(WC)	0.066	0.025	2.688	1.69%	0.014	0.119	0.014	0.119

Fonte: o Autor.

Apresentando resultados para os Valores-p, F de significação dentro das recomendações da NBR 14653-2, atingindo, inclusive, o GRAU III.

Um coeficiente de correlação(R^2) de 95,7%, deixando de representar apenas 4,3% da variabilidade dos preços praticados pelo mercado.

Analisando os gráficos exigidos para os pressupostos básico da regressão linear:

Gráfico 1 – Normalidade dos Resíduos



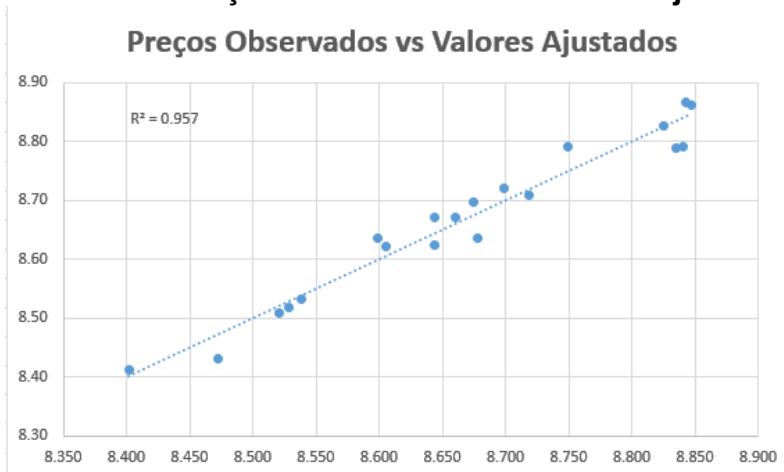
Fonte: o Autor.

Gráfico 2 – Resíduos Padronizados vs Valores Ajustados



Fonte: o Autor.

Gráfico 3 – Preços Observados vs Valores Ajustados



Fonte: o Autor.

Gráfico 4 – Histograma dos Resíduos



Fonte: o Autor.

REALIZAÇÃO

Uma vez que os pressupostos básicos estiverem de acordo com o exigido pela regressão linear é possível partir para a montagem da equação que representa a variabilidade dos dados, portanto a partir dos coeficientes da tabela 4:

$$Pu_{frequentista} = 4025,038 * 0,994^A * 1,346^{Quartos} * G^{0,085} * WC^{0,66}$$

Interpretação: Em condições de Ceteris Paribus:

A cada aumento de 1m² em A, o Pu diminui em 0,6%;

A cada aumento de 1 Quarto, o Pu aumenta em 34,6%;

A cada aumento de 1 Vaga de Garagem, o Pu aumento em 8,5%;

A cada aumento de 1 WC, o Pu aumenta em 66%

Realizando a avaliação do Apartamento em análise,

$$A = 150m^2, Quarto = 3, Vaga de Garagem = 3, WC = 3,5;$$

Substituindo os valores na equação do Pu:

$$Pu_{frequentista} = 4025,038 * 0,994^{(150)} * 1,346^{(3)} * (3)^{0,085} * (3,5)^{0,66}$$

$$Pu_{frequentista} = \frac{R\$9988,56}{m^2};$$

Portanto:

$$Pt_{frequentista} = Pu_{frequentista} * A;$$

$$Pt_{frequentista} = 9988,56 * 150;$$

$$Pt_{frequentista} = R\$1.498.284,54;$$

Com limites superior e inferior para o intervalo de confiança de 80% em torno do Pt :

$$LI = R\$1.441.652,45;$$

$$LS = R\$1.557.141,30;$$

Fornecendo um Grau de Precisão de:

$$GP = \frac{LS - LI}{Pt} = 7,71\%$$

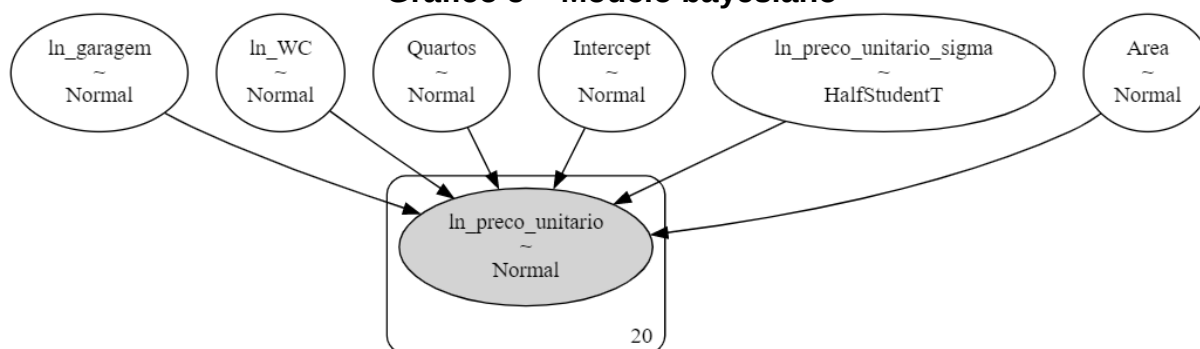
3.2. Inferência Bayesiana – Regressão Linear por MCMC - HMC

REALIZAÇÃO

Na estimativa do valor do apartamento avaliando através da inferência bayesiana, é necessário determinar os parâmetros do modelo da posteriori a partir da função de verossimilhança, das variáveis com suas respectivas funções de densidade de probabilidade a priori e os parâmetros de simulação.

Na modelagem, utilizando a biblioteca bambi em Python, com o objetivo de reproduzir uma a estimativa do valor de mercado do apartamento obtido no item 4.1., define-se o modelo a Posteriori, com seus parâmetros a serem inferidos, como sendo resultado uma p.d.f. gaussiana para a função de verossimilhança, na qual, tem em sua composição, todas as variáveis coletadas e o erro aleatório:

Gráfico 5 – Modelo bayesiano



Fonte: o Autor.

Matematicamente, tem-se:

$$y_i | x_i \sim f(y | x_i), \quad i = 1, \dots, n$$

$$f(y | x) = \mathcal{N}(y | x^T \beta, \sigma^2)$$

$$\beta_0 \propto 1$$

$$\beta_k | \sigma^2 \sim \mathcal{N}(0, \sigma^2 \nu_\beta), \quad k = 1, \dots, p$$

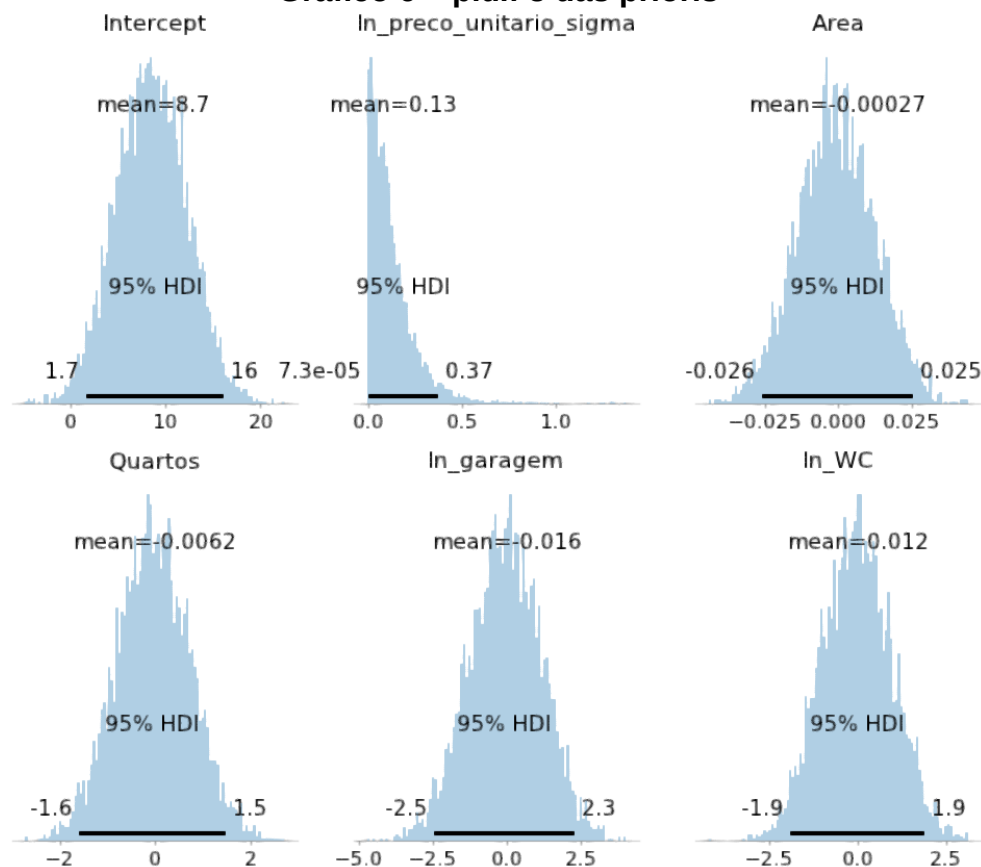
$$\sigma^2 \sim \mathcal{HT}(\alpha, \nu)$$

No qual, em seus parâmetros de simulação, definiu-se o algoritmo de amostragem como sendo Markov Chain Monte Carlo (MCMC), especificamente o NUTS, o mesmo pré definido pela biblioteca bambi em python, no qual adota um método de análise na definição de qual método entre o Hamiltoniano (HMC), Amostragem de Gibbs ou Metropolis-Hastings se adequa melhor aos dados. Neste artigo foi utilizado o HMC.

Definiu-se a utilização de 4 cadeias de Markov onde, seriam amostrados 10.000 realizações dos parâmetros para o ajuste/treinamento (*tune*), e em seguida, a amostragem de 20.000 realizações para o cálculo e definição da posteriori. Fazendo um total de 120.000 realizações total.

Em seguida define-se as funções de densidade de probabilidade das prioris dos parâmetros utilizados no modelo.

Gráfico 6 – p.d.f's das prioris



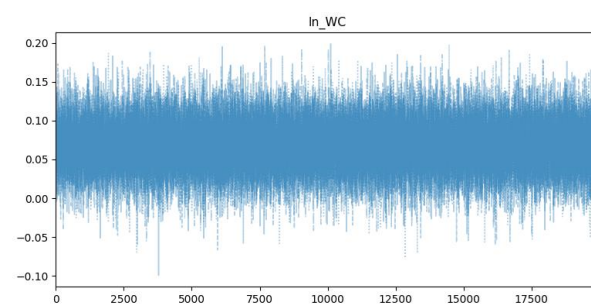
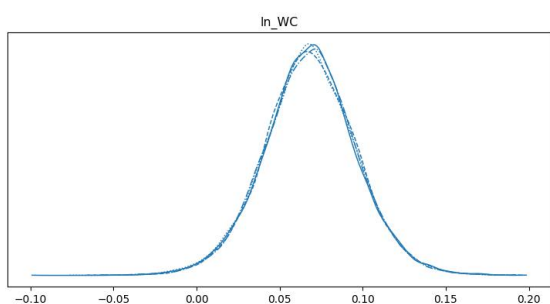
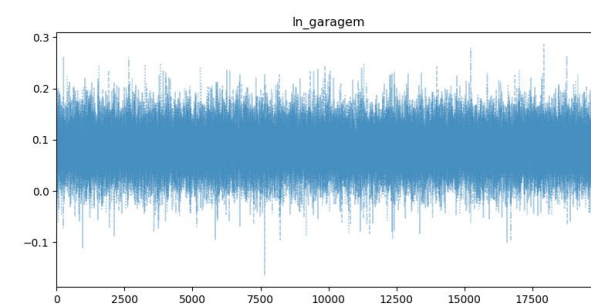
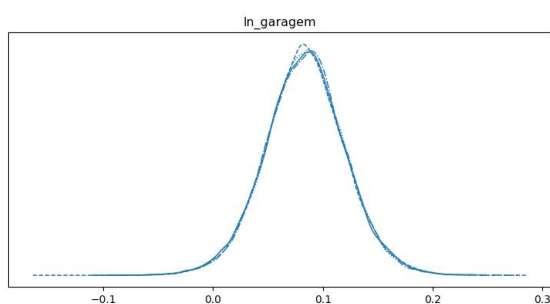
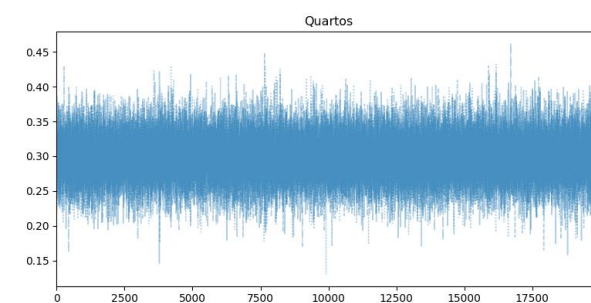
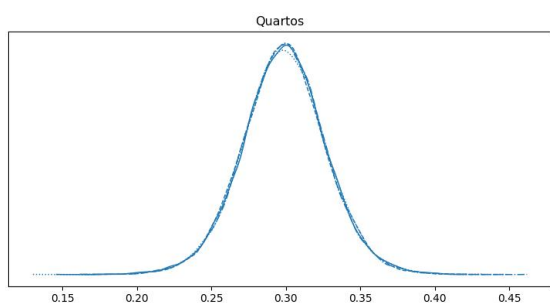
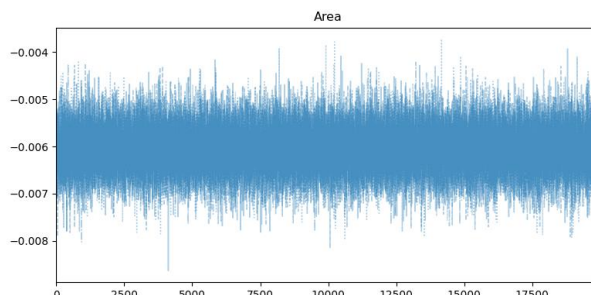
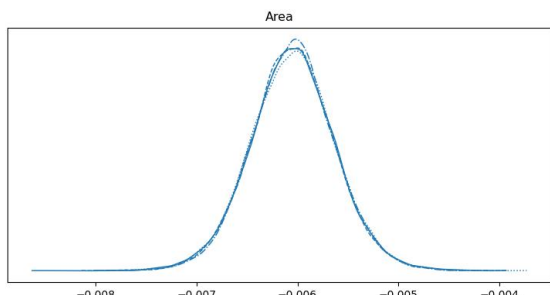
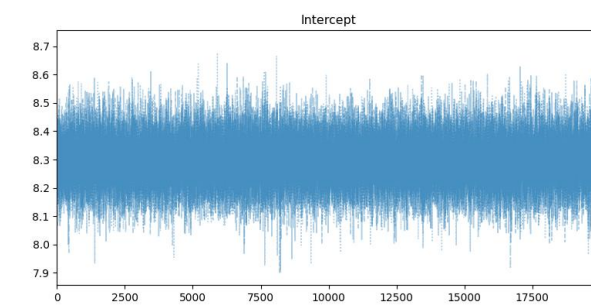
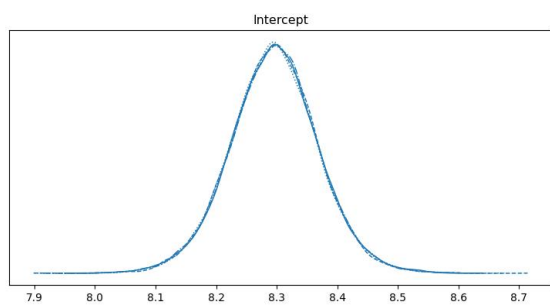
Fonte: o Autor.

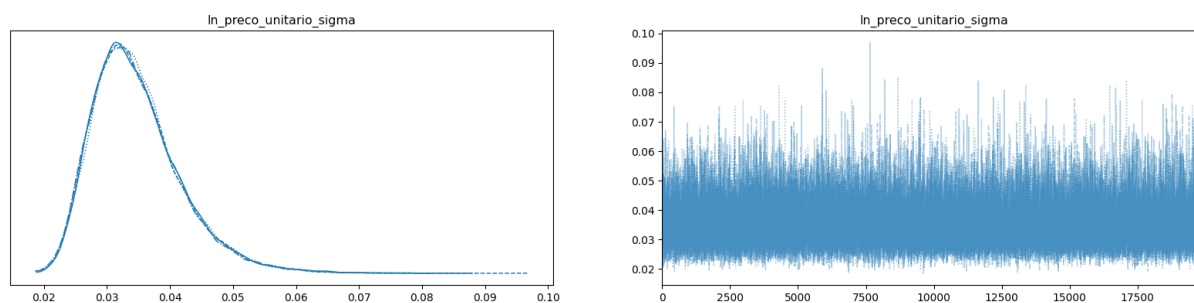
Na definição das p.d.f.'s a priori utilizamos o conceito das prioris pouco informativas (*Weakly Informative Priors*), implementada nas próprias bibliotecas bambi (Python) e na rstanarm (R), na qual seu uso se torna bem atraente na modelagem inferencial, uma vez que, define-se uma escala apropriada a partir dos próprios dados e às utiliza como princípio de regularização na análise.

Sendo assim, após a definição dos dados iniciais, é dado seguimento nas simulações para a obtenção da posteriori, portanto analisando os resultados da amostragem pelos gráficos 7 e 8, tem-se:

Gráfico 7 – Resultados da amostragem das prioris

REALIZAÇÃO

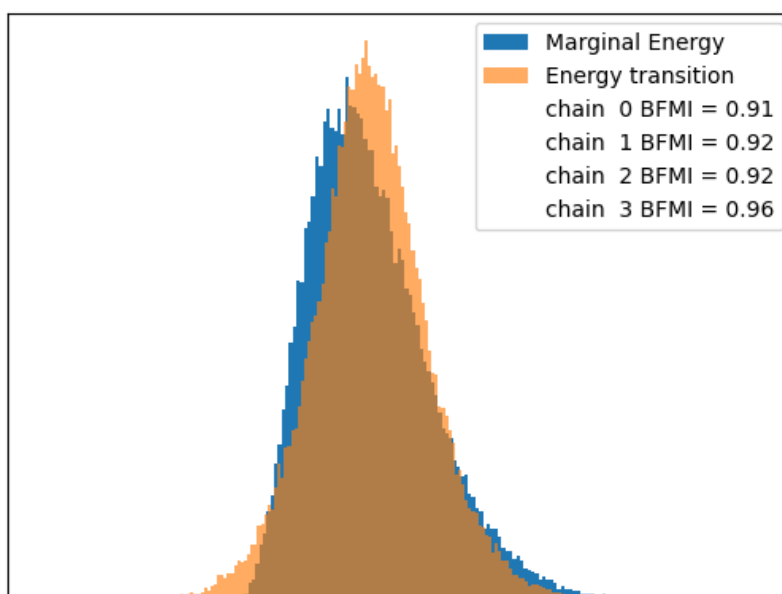




Fonte: o Autor.

Na amostragem, verifica-se que, nas quatro cadeias de Markov utilizadas, foi obtido a mesma convergência das p.d.f.s para as respectivas variáveis, indicando parâmetros semelhantes (lado esquerdo). Adicionalmente, não é verificado tendências ou padrões na série temporal da amostragem (lado direito). Indicando uma boa amostragem para as priors.

Gráfico 8 – Energia Marginal da amostragem vs Energia de Transição entre as simulações - NUTS



Fonte: o Autor.

Além da verificação de tendências e padrões na amostragem é necessário a análise do algoritmo utilizado, o NUTS, na qual a partir do gráfico 8 é possível verificar, com a sobreposição da Energia Marginal e Energia de Transição e o fator BFMI (*Bayesian fraction of missing information*), se o algoritmo explorou o máximo possível da densidade a posteriori.

Sendo assim, a sobreposição das p.d.f.s obtidas no gráfico 8, parecem se ajustar bem, o que pode ser confirmado através do fator BFMI para cada cadeia de Markov, conseguindo representar entre 91% a 96% da posteriori.

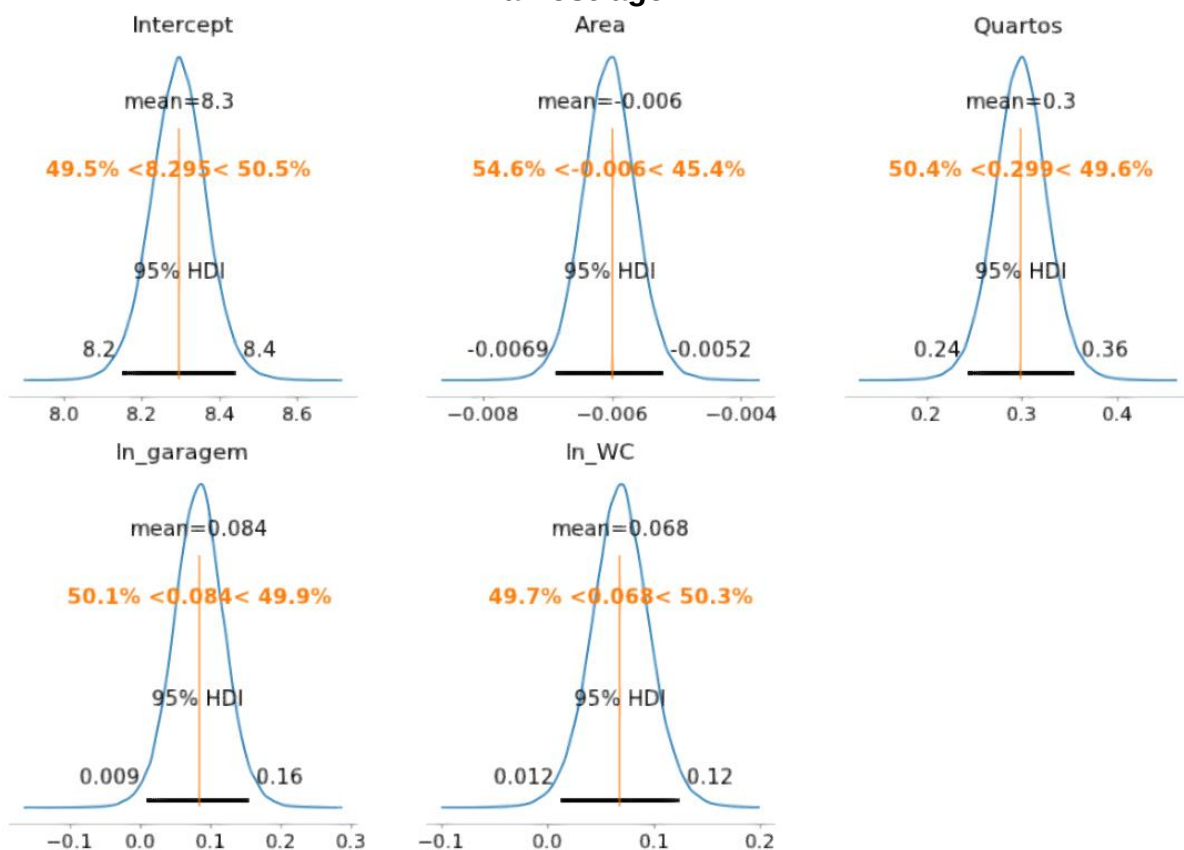
Uma vez que a amostragem esteja de acordo com o esperado, é dada sequência na obtenção dos coeficientes, seus intervalos de credibilidades e alguns testes de diagnóstico necessários para a verificação do ajuste da modelagem:

Tabela 5 – Resultados da Inferência Bayesiana

	mean	sd	Hdi_3 %	Hdi_97 %	Mcse_mean	Mcse_sd	Ess_bulk	Ess_tail	r-hat
Intercept	8,295	0,073	8,158	8,436	0,0	0,0	83740	67627	1
Area	-0,006	0,000	-0,007	-0,005	0,0	0,0	74646	55748	1
Quartos	0,299	0,028	0,244	0,352	0,0	0,0	60382	52870	1
Ln_garagem	0,084	0,037	0,015	0,154	0,0	0,0	77106	56080	1
Ln_WC	0,068	0,028	0,016	0,122	0,0	0,0	74005	53428	1
Sigma	0,035	0,007	0,023	0,048	0,0	0,0	48384	40519	1

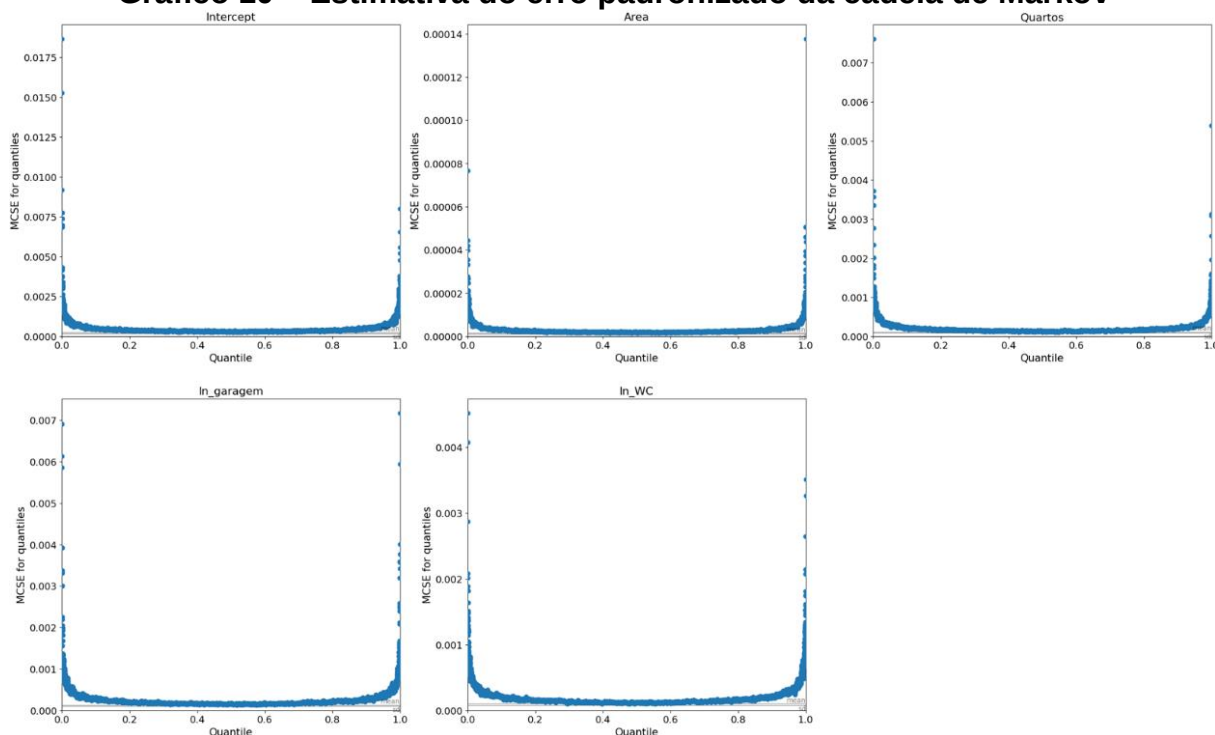
Fonte: o Autor.

Gráfico 9 – Estimativa das priors com intervalo de credibilidade após amostragem



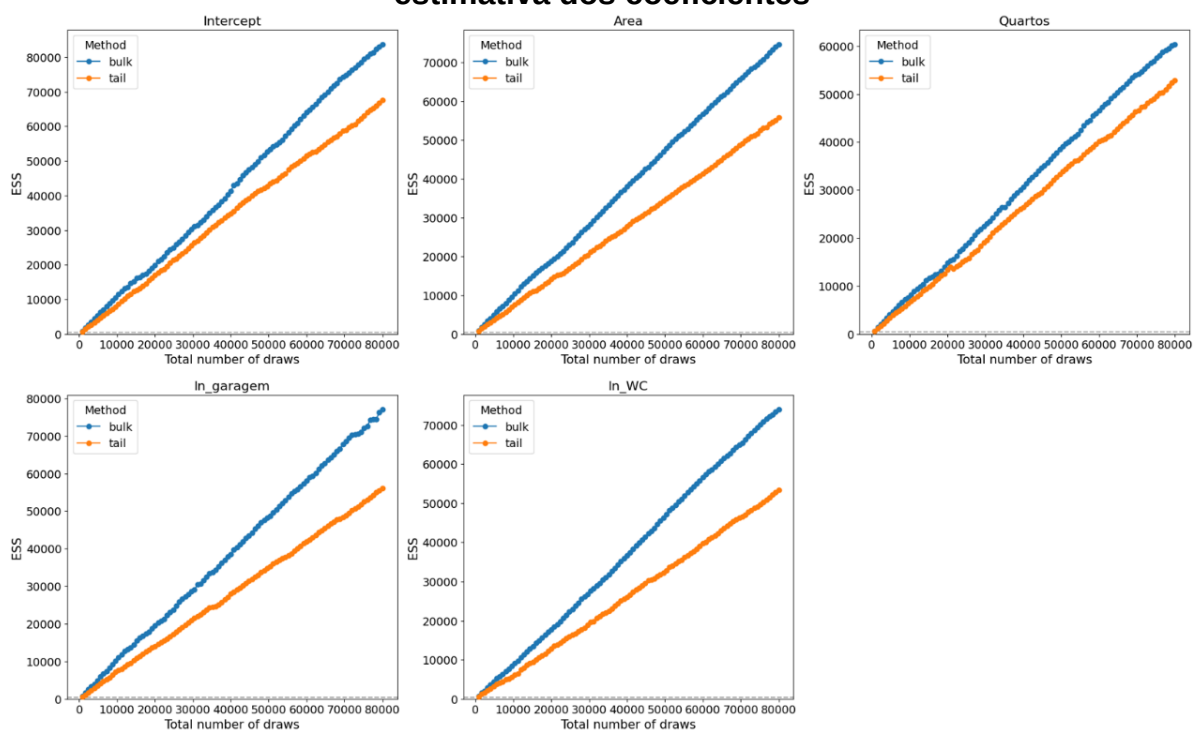
Fonte: o Autor.

Gráfico 10 – Estimativa do erro padronizado da cadeia de Markov



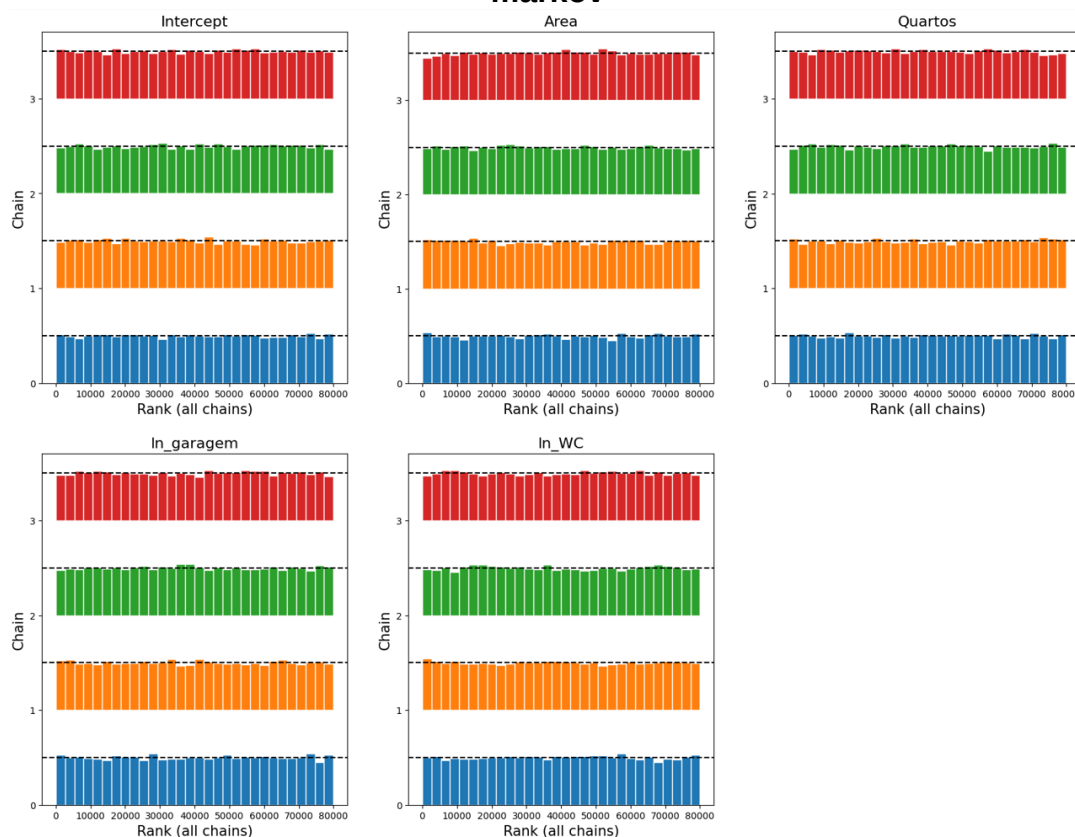
Fonte: o Autor.

Gráfico 11 – Estimativa da quantidade de realizações efetivamente utilizada na estimativa dos coeficientes



Fonte: o Autor.

Gráfico 12 – Estimativa do teste de diagnóstico de convergência da cadeia de markov



Fonte: o Autor.

Os resultados das médias (*mean*) obtidos para cada coeficiente na tabela 5 se assemelham aos coeficientes obtidos na tabela 4. No qual, nos indica uma grande possibilidade de semelhança na estimativa do valor do apartamento avaliando.

Da tabela 5, temos outras informações de grande importância para a inferência:

1. Assim como esperado, as variáveis na inferência bayesiana são tratadas como variáveis aleatórias representadas por p.d.f.s gaussianas com média (*mean*) e desvio padrão (*sd*);
2. O intervalo de credibilidade, gráfico 9, com 3% (*Hdi_3%*) e 97% (*Hdi_97%*) indica de forma probabilística à pertinência do parâmetro ao intervalo destacado;
3. O erro padronizado da cadeia de Markov, gráfico 10, medido em média (*Mcse_mean*) e desvio padrão (*Mcse_sd*), verifica-se que os mesmos apresentam valores próximos à 0;
4. A quantidade efetivamente utilizada de realizações da amostragem, gráfico 11, para cada parâmetro para representar o centro (*Ess_bulk*) e as caudas (*Ess_tail*). Diz-se que o modelo está convergindo adequadamente quando as duas linhas apresentam um comportamento aproximadamente linear;

5. O teste de diagnóstico de convergência normalizado das Cadeias de Markov ($\hat{r}$), gráfico 12, no qual a unidade representa a convergência entre as variâncias das cadeias.

Adotando os mesmos procedimentos da inferência frequentista, podemos realizar a estimativa do valor do apartamento avaliando através da equação de inferência utilizando os valores médios dos coeficientes obtidos da amostragem, portanto:

$$Pu_{bayes} = 4003,803 * 0,994^A * 1,349^{Quartos} * G^{0,084} * WC^{0,68}$$

Realizando a avaliação do Apartamento em análise,

$$A = 150m^2, Quarto = 3, Vaga de Garagem = 3, WC = 3,5;$$

Substituindo os valores na equação do Pu:

$$Pu_{bayes} = 4003,803 * 0,994^{(150)} * 1,349^{(3)} * (3)^{0,084} * (3,5)^{0,68}$$

$$Pu_{bayes} = \frac{R\$10.244,97}{m^2};$$

Portanto:

$$Pt_{bayes} = Pu * A;$$

$$Pt_{bayes} = 10.244,97 * 150;$$

$$Pt_{bayes} = R\$1.536.745,41;$$

Analisando o desvio dos valores do apartamento avaliando obtidos nas inferências bayesiana e frequentista, tem-se:

$$R = \frac{Pt_{bayes} - Pt_{frequentista}}{Pt_{bayes}};$$

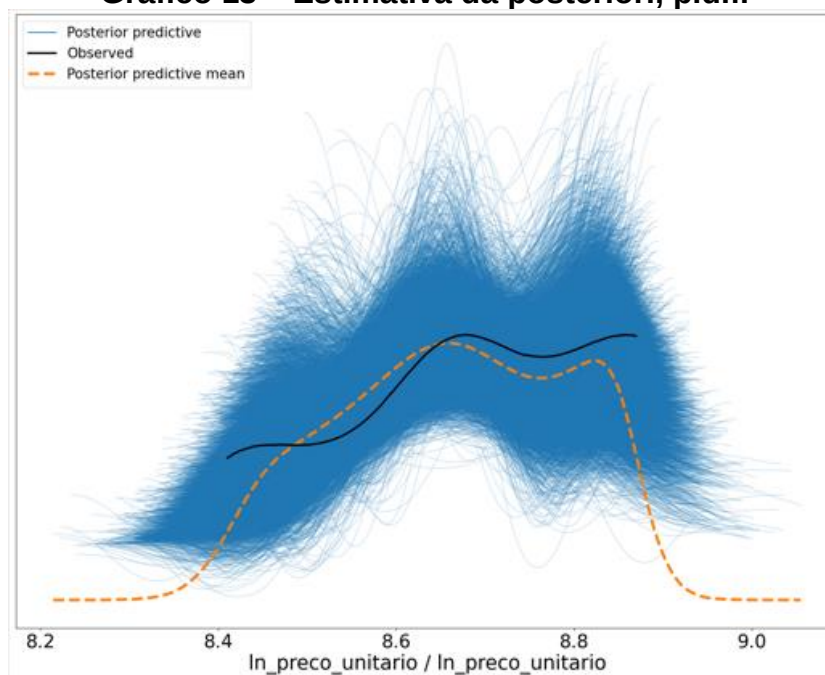
$$R = \frac{R\$1.536.745,41 - R\$1.498.284,54}{R\$1.536.745,41};$$

$$R = 2,503\%$$

Obtendo um desvio de apenas 2,5% na estimativa do valor de compra e venda do avaliando, pode-se validar a estimativa através da inferência bayesiana.

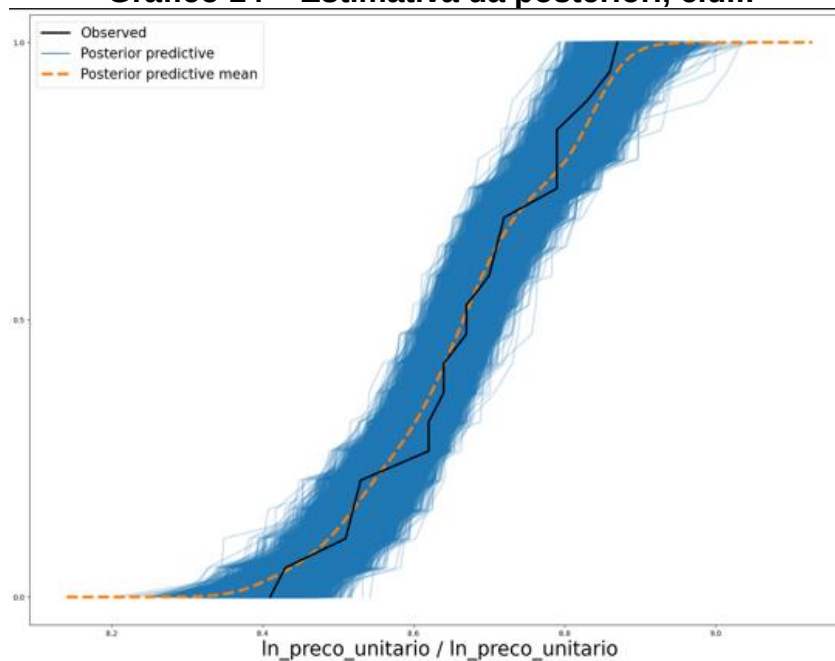
No entanto, através de uma abordagem bayesiana, na estimativa da p.d.f. da posteriori que represente a variabilidade dos preços pesquisados, verifica-se alguns detalhes importantes a serem observados:

Gráfico 13 – Estimativa da posteriori, p.d.f.



Fonte: o Autor.

Gráfico 14 – Estimativa da posteriori, c.d.f.



Fonte: o Autor.

Nos gráficos 13 e 14, verifica-se que:

1. Os dados observados (linha preta), carecem de informações sobre as caudas, o que pode ocasionar uma má interpretação na análise da presença de *outliers*, podendo levar a estimativas do apartamento avaliando enviesadas;
2. A amostragem a partir da MCMC (nuvem azul), assim como já analisado, percorre toda a extensão dos dados coletados, apresentando resultados bem precisos e acurados na inferência;
3. A curva obtida pela média da amostragem do MCMC (linha tracejada laranja) representa bem melhor a variabilidade dos preços praticados pelo mercado, inclusive as caudas, dando uma maior confiabilidade na estimativa dos preços.

Conclui-se então que, mesmo com os excelentes resultados dos testes de hipóteses obtidos pela inferência frequentista, os dados coletados carecem de informações, o que podem levar o analista a resultados tendenciosos, principalmente na análise de *outliers*.

No entanto, a inferência bayesiana, a partir do poder da análise das simulações de Monte Carlo consegue reproduzir cenários bem mais completos da variabilidade dos preços praticados pelo mercado, produzindo resultados bem mais confiáveis.

4. CONCLUSÕES E TRABALHOS FUTUROS

Assim como esperado, os resultados obtidos pela inferência bayesiana fornecem uma maior quantidade de informações ao analista, com a garantia dos pressupostos fazendo com que a tomada de decisão seja mais precisa e acurada.

Os coeficientes obtidos pelas inferências frequentista e bayesiana foram equivalentes, indicando que o valor encontrado pela estimativa pontual do imóvel avaliando pode ser determinada por quaisquer dos métodos.

A função de densidade de probabilidade representando a variabilidade dos preços de oferta praticados pelo mercado obtida pela simulação de monte carlo, na inferência bayesiana, foi bem mais completa do que a encontrada pela pesquisa do mercado, indicando inclusive a ausência de caudas bem definidas, o que pode levar a um problema na análise da presença de *outliers*, na inferência frequentista.

Recomenda-se, em trabalhos futuros, a análise da inferência bayesiana com a incorporação de diferentes tipos de variáveis qualitativas e outras p.d.f.s na representação dos parâmetros e da verossimilhança, além da gaussiana.

REFERÊNCIAS BIBLIOGRÁFICAS

Biblioteca em R: <https://mc-stan.org/rstanarm/articles/priors.html>

Biblioteca em Python: <https://bambinos.github.io/bambi/index.html>

CATTY, J. P., IFRS: Guia de Aplicação do Valor Justo, Ed. Bookman.

DANTAS, Rubens Alves; Engenharia de Avaliações: Uma Introdução à Metodologia Científica, 3ª edição, Ed. Pini.

GUJARATI, Damodar N., PORTER, Dawn C., Econometria Básica, 5ª edição, Ed. Bookman.

HOFFMAN; M. D.; GELMAN, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. Journal of Machine Learning Research, v. 15, p. 1351-1381, 2014.

Michael Betancourt, How the Shape of a Weakly Informative Prior Affects Inferences (mc-stan.org)

NEAL, R. M. MCMC using Hamiltonian dynamics. In BROOKS, S.; GELMAN, A.; JONES, G. L.; MENG, X. L. (ed.) Handbook of Markov Chain Monte Carlo. Boca Raton: Chapman and Hall–CRC Press, 2011. cap. 5, p. 113-162.

NBR 14653 - Avaliação de Bens parte 1 e 2: Procedimentos Gerais.

NBR 14653-2- Avaliação de Bens parte 2: Imóveis Urbanos.

Paulino, C.D., Turkman, M.A.A. and Murteira, B. (2003). Estatística Bayesiana, Fundação Calouste Gulbenkian, Lisboa.

Rafael L. B., Márcio P. L., Estimação da estrutura a termo da taxa de juros por Inferência Bayesiana: uma aplicação do método de Monte Carlo Hamiltoniano

Tomás Capretto, Camen Piho, Ravin Kumar, Jacob Westfall, Tal Yarkoni, Osvaldo A. Martin, Bambi: A simple interface for fitting Bayesian linear models in Python.

Nassim Nicholas Taleb, STATISTICAL CONSEQUENCES OF FAT TAILS - Real World Preasymptotics, Epistemology, and Applications, Stem Academy Press.