

Aplicação de modelos não lineares na precificação de passagens aéreas

Juliana de Melo 

Roberto Molina de Souza 

Resumo

Devido à autonomia dos usuários de transporte aéreo no que diz respeito à compra de bilhetes, a necessidade de intermediários, como agências de turismo, tornou-se desnecessária em muitos casos, sendo do usuário a tomada de decisão quanto à companhia aérea e tempo de antecedência da compra de sua passagem. Adquirir um bilhete aéreo com antecedência pode tornar a compra mais barata. Nesse sentido, este trabalho estuda o comportamento dos preços praticados pelas três maiores empresas aéreas brasileiras nos trechos Londrina/PR - Curitiba/PR e Londrina/PR - São Paulo/SP quanto ao tempo de antecedência na aquisição do bilhete. São apresentados gráficos do preço médio ao longo de trinta e sessenta dias de estudo, nos quais se observa que a relação entre trinta dias de antecedência e o valor cobrado é não linear, motivando assim o estudo de modelos não lineares para a modelagem dos dados obtidos, visto que com sessenta dias de antecedência, no geral, os preços estabilizam. Logo, para os trechos em estudo recomenda-se a compra da passagem com trinta dias de antecedência.

Palavras-chave: Modelos não lineares; Preços de passagens aéreas; Dias de antecedência.

Abstract

Due to the autonomy of air transport users with regard to ticket purchase, the need for intermediaries, such as travel agencies, has become unnecessary in many cases, being the user's decision regarding the airline and time in advance of your ticket purchase. Acquiring an air ticket in advance can make the purchase cheaper. In this sense, this work studies the behavior of the prices practiced by the three largest Brazilian airlines in the Londrina / PR - Curitiba / PR and Londrina / PR - São Paulo / SP flight stretch regarding the time in advance in ticket acquisition. Average price graphs are presented during thirty and sixty days. This study shows that the relation between thirty days in advance and the amount charged is non-linear, thus motivating the study of non-linear statistical models for the modeling of the data obtained, since at sixty days in advance prices generally stabilize. Therefore, for the flight stretch under study it is recommended to purchase the ticket thirty days in advance.

Keywords: Nonlinear models; airline tickets; days in advance

1. Introdução

A partir do surgimento da rede mundial de computadores como relevante canal de distribuição e comercialização dos bilhetes aéreos, diminuiu-se a necessidade da intermediação de agentes de

viagens, reduzindo os custos de pesquisa (*search costs*) e também as assimetrias de informação no mercado. Observou-se, assim, um aumento na competitividade do setor aéreo brasileiro [15].

Segundo Evans e Kessides (1993) [10] existem duas categorias nos estudos de análise empírica, que são aproximações para o estudo dos elementos que influenciam o nível das tarifas aéreas. A primeira categoria é a análise *inter-rota*, em que se inferem os efeitos das variáveis de estrutura de mercado em preços *cross-sectional* nas rotas, ou seja, os dados estão desagregados no nível da rota e referem-se a variáveis como distância de voo, número de companhias presentes, entre outras. A segunda categoria é a de análise *intrarota*, em que se introduz a heterogeneidade das empresas que operam em uma rota como as características específicas das companhias, quantidades de assentos, entre outros.

A gestão de oferta também conhecida como *Yield Management (YM)* é uma técnica de gestão baseada em modelos matemáticos utilizada para calcular a menor política de preços, por meio da otimização e maximização dos lucros gerados pela venda de um determinado produto ou serviço. O *YM* surgiu na década de 80, nos Estados Unidos, com a desregulamentação da indústria aérea americana e a tecnologia disponível a um custo aceitável, ou seja, as companhias aéreas puderam precificar seus assentos como bem entendessem. Além disso, houve abertura para novas companhias aéreas de custo mais baixo entrarem no mercado, aumentando a concorrência [14].

As primeiras empresas que investiram fortemente em *YM* como sua principal arma para manter a lucratividade continuaram no mercado, sendo elas: *American Airlines* e a *Delta Airlines*. Essas empresas criaram grandes sistemas centrais nas quais tinham o potencial de reduzir a concorrência, baseando-se em três situações: a dificuldade de empresas aéreas que não possuem sistemas centrais teriam para visualizar a tela de um agente de viagem; as cláusulas restritivas de contrato entre as agências de viagem, e as empresas proprietárias do sistema [22].

A primeira empresa da América Latina com o modelo de negócios "*Low Cost, Low Fare*" foi a Gol Linhas Aéreas, com entrada no mercado doméstico em janeiro de 2001, dando início a uma trajetória de crescimento impressionante do transporte aéreo mundial. A forte expansão da empresa provocou um incremento da contestabilidade ao mercado aéreo e uma maior competitividade do setor. Fatores como preços menores, publicidade e acesso ao aeroporto de Congonhas desde o primeiro ano de operações (2001) e ao aeroporto de Santos Dumont e à Ponte Aérea Rio de Janeiro - São Paulo desde o segundo ano (2002) contribuíram para o crescimento acelerado da Gol [16].

Em geral, existem fatores que influenciam na variação dos preços cobrados pelos bilhetes, como dia da semana, horário do voo, a empresa que se escolhe para voar e a antecedência da compra da passagem em relação ao voo etc.

Quanto à antecedência da compra, algumas indefinições podem ser consideradas [5]. A expectativa mais comum é de que os preços cobrados cresçam conforme a data do embarque se aproxima. É, porém, pertinente observar que a aproximação da data do voo coloca a empresa aérea em uma situação de risco.

Nesse caminho, o objetivo deste trabalho é estudar o menor valor diário de passagens aéreas em um período de 30 e 60 dias de antecedência da compra do bilhete, dos trechos Londrina/PR - Curitiba/PR (LDB-CWB) e Londrina/PR - São Paulo/SP (LDB-CGH), considerando as três maiores empresas aéreas que fazem esses trechos. Os dados coletados para a rota Londrina - São Paulo consideraram o aeroporto de Congonhas.

Este trabalho está organizado da seguinte maneira: na Seção 2 são apresentados os conceitos de

modelos de regressão linear e não linear, e, nas subseções, um resumo dos métodos de estimativa dos parâmetros e critério de comparação dos modelos; na Seção 3 são apresentados gráficos que resumem o conjunto de dados obtido, os modelos escolhidos para a estimação dos parâmetros, as aplicações e os resultados, e, finalmente, na Seção 4 as considerações finais.

2. Modelos de Regressão

A análise de regressão tem o objetivo de verificar a existência de uma relação funcional entre uma variável dependente com uma ou mais variáveis independentes. Em outras palavras, a análise de regressão consiste na obtenção de uma equação que busca explicar a variável dependente pela variação dos níveis das variáveis independentes [7]. Modelos estatísticos são capazes de explicar essa relação por meio de uma equação que relaciona tais variáveis.

2.1. Regressão linear

Na análise de regressão linear simples, os dados são modelados utilizando uma função linear do tipo $f(x) = \alpha + \beta x + \varepsilon$, sendo os parâmetros α e β desconhecidos e estimados a partir dos dados ε representa o que não é explicado pela variável independente, conhecido como erro aleatório, seguindo alguma distribuição de probabilidade.

Na análise de regressão linear múltipla, mais de uma variável independente pode ser considerada. Essa relação pode ser analisada como um processo no qual os valores das variáveis X_1, X_2, \dots, X_n podem ser chamados de variáveis regressoras, e Y de variável resposta [7]. Logo, tais modelos são da forma $f(x) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$.

2.2. Regressão não linear

Os modelos lineares são bastante utilizados para um grande número de aplicações em regressão, cujo propósito é relacionar uma variável resposta a uma ou mais variáveis explicativas. Todavia, em algumas situações um modelo não linear pode ser mais adequado, uma vez que muitos fenômenos geralmente não são lineares.

O surgimento dos modelos não lineares deu-se por meio da regressão geral na década de 20 com Ronald Fisher Aylmer e Winifred Mackenzie. Nos anos 70, esses modelos ganharam mais popularidade e passaram a ser mais bem estudados devido aos avanços computacionais. Com exceção de alguns trabalhos mais antigos, a maior parte da literatura referente a modelos não lineares surgiu depois dos anos 80, e, desde então, tem sido uma área de pesquisa bastante explorada [21].

Com o surgimento de novos algoritmos para ajustes de modelos de regressão não lineares, houve o aumento da popularidade dos modelos não lineares e a valorização dos processos inferenciais por eles gerados. Os modelos não lineares são usados por pesquisadores para modelar variáveis nas mais diversas áreas do conhecimento, como econometria, agricultura, biologia, química etc [21].

Dentre as diferenças entre as classes de modelos lineares e não lineares, a principal delas está relacionada as suas formulações. Por meio de um conjunto de observações, o modelo linear busca o modelo que melhor explique a relação, se existir alguma, entre as variáveis inerentes a um dado fenômeno [13]. Já no caso não linear, quase sempre, as formulações de possíveis modelos são baseadas em considerações teóricas referentes ao fenômeno de interesse em modelar [3, 19].

Um modelo de regressão é dito não linear se ao menos um dos seus parâmetros aparece de forma não linear. Por exemplo, os modelos:

$$y = \exp(\alpha + \theta x) \tag{1}$$

e

$$y = \alpha + \beta \exp(\theta x) \tag{2}$$

são não lineares. No modelo (1) α e θ são parâmetros não lineares, e no modelo (2) α e β são parâmetros lineares, enquanto θ é não linear.

Os modelos não lineares podem ser escritos da seguinte forma:

$$y_i = f(x_i, \theta) + \varepsilon_i, \tag{3}$$

em que, y_i e x_i , para $i = 1, 2, \dots, n$, são os vetores de variável resposta e variável explicativa, respectivamente, para uma amostra de tamanho n ; $f(x_i, \theta)$ é uma função das variáveis regressoras e dos parâmetros, chamada função de regressão, θ é o vetor dos parâmetros desconhecidos e ε_i é o vetor de erros aleatórios. Em geral, atribui-se que os erros são variáveis aleatórias independentes e identicamente distribuídos, seguindo distribuição normal com média zero e variância constante denotada por σ^2 .

2.3. Método dos mínimos quadrados

O Método dos Mínimos Quadrados (MMQ) é uma técnica utilizada no método iterativo de Gauss-Newton, em que seu objetivo é encontrar o melhor ajuste para um conjunto de dados, cujo objetivo é minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados.

Considerando que a função de regressão é contínua e admite derivadas de primeira e segunda ordens com relação aos parâmetros, define-se a soma de quadrados dos erros de (3) por:

$$SQRes(\theta) = \sum_{i=1}^n [y_i - f(x_i, \theta)]^2, \tag{4}$$

em que $SQRes(\theta)$ depende exclusivamente de θ . Denota-se por $\hat{\theta}$ os estimadores de mínimos quadrados de θ , ou seja, os valores de $\hat{\theta}$ que minimizam $SQRes(\theta)$. Para encontrar esses estimadores, é necessário fazer a diferenciação da equação (4) com relação a cada θ_i com $i = 1, 2, \dots, p$ e igualar as equações a zero, isto é:

$$\frac{\partial SQRes(\theta)}{\partial \theta_i} = \sum_{i=1}^n -2(Y_i - f(X_i, \theta)) \left[\frac{\partial f(X_i, \theta)}{\partial \theta_i} \right]. \tag{5}$$

Igualando as derivadas parciais definidas em (5), a zero, tem-se:

$$\sum_{i=1}^n Y_i \left[\frac{\partial f(X_i, \theta)}{\partial \theta_i} \right] - \sum_{i=1}^n f(X_i, \theta) \left[\frac{\partial f(X_i, \theta)}{\partial \theta_i} \right] = 0. \tag{6}$$

Dessa forma, determina-se p equações denominadas de equações normais ou homogêneas, sendo definidas por:

$$\sum_{i=1}^n [y_i - f(x_i, \theta)] \left[\frac{\partial f(x_i, \theta)}{\partial \theta_i} \right]_{\theta=\hat{\theta}} = 0, \quad (7)$$

e quando $\partial f(x_i, \theta)$ não depender de θ , ou seja, $\left[\frac{\partial f^2(x_i, \theta)}{\partial \theta_i^2} \right] = 0$, tem-se as equações de um modelo de regressão linear. Para funções não lineares, θ estará presente em pelo menos uma das derivadas parciais de $f(x_i, \theta)$.

Para modelos lineares esse sistema normalmente possui um vetor único de soluções para os parâmetros θ , em que cada parâmetro depende unicamente dos dados observados. Já para modelos não lineares, os estimadores $\hat{\theta}$ resultantes da última equação, em geral, não apresentam forma analítica, e as soluções precisam ser obtidas por algum método de aproximação iterativo. Dentre os principais, pode-se citar o método de Gauss-Newton [21] descrito a seguir.

2.3.1 Método iterativo de Gauss-Newton

Para minimizar $SQRes(\theta)$ descrito em (4), pode ser necessário o uso de algum método numérico iterativo. O método iterativo de Gauss-Newton permite estabelecer uma relação linear entre a função não linear e os dados observados. Essa relação é feita com base na expansão em série de Taylor.

Suponha que $\theta^{(k)}$ seja uma aproximação da estimativa de mínimos quadrados θ de $f(X_i, \theta)$. Tomando $\theta \approx \theta^{(k)}$, tem-se que a expansão em série de Taylor de primeira ordem é dada por:

$$f(X_i, \theta) \approx f(X_i, \theta^{(k)}) + \frac{\partial f(X_i, \theta)}{\partial \theta_j} (\theta - \theta^{(k)}), \quad (8)$$

em que a matriz com as derivadas da função não linear em relação a cada um dos parâmetros, $\frac{\partial f(X_i, \theta)}{\partial \theta_j}$, será denotada por $F(\theta)$. Seja $r(\theta) = Y_i - f(X_i, \theta)$ sendo o vetor de resíduos, na qual podem ser aproximados por:

$$r(\theta) \approx r(\theta^{(k)}) - F(\theta^{(k)}) (\theta - \theta^{(k)}). \quad (9)$$

Como o objetivo é minimizar a $SQRes = r(\theta)' r(\theta)$ pelo método de mínimos quadrados, em que $r(\theta)'$ é o vetor transposto de $r(\theta)$, tem-se:

$$SQRes = [r(\theta)' r(\theta)] \approx [r(\theta^{(k)}) - F(\theta^{(k)}) (\theta - \theta^{(k)})]' [r(\theta^{(k)}) - F(\theta^{(k)}) (\theta - \theta^{(k)})], \quad (10)$$

$$\begin{aligned}
 SQRes &= r(\theta^{(k)})' r(\theta^{(k)}) - r(\theta^{(k)})' F(\theta^{(k)}) (\theta - \theta^{(k)}) - \\
 & r(\theta^{(k)}) [F(\theta^{(k)}) (\theta - \theta^{(k)})]' + (\theta - \theta^{(k)})' F(\theta^{(k)})' F(\theta^{(k)}) (\theta - \theta^{(k)}).
 \end{aligned}$$

Dessa forma, SQRes será minimizada quando:

$$(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}^{(k)}) = [\mathbf{F}'(\boldsymbol{\theta}^{(k)}) \mathbf{F}(\boldsymbol{\theta}^{(k)})]^{-1} \mathbf{F}'(\boldsymbol{\theta}^{(k)}) \mathbf{r}(\boldsymbol{\theta}^{(k)}). \quad (11)$$

A cada iteração é gerada uma nova aproximação $\boldsymbol{\theta}^{(k+1)}$ para o vetor de parâmetros $\boldsymbol{\theta}^{(k)}$, através de:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + [\mathbf{F}'(\boldsymbol{\theta}^{(k)}) \mathbf{F}(\boldsymbol{\theta}^{(k)})]^{-1} \mathbf{F}'(\boldsymbol{\theta}^{(k)}) \mathbf{r}(\boldsymbol{\theta}^{(k)}), \quad (12)$$

resultando no processo iterativo conhecido como Método de Gauss-Newton. O processo (12) é repetido até obter a convergência.

Para a estimação dos parâmetros e ajuste dos modelos, pode ser utilizada a função *nls* do *software* livre R, em que o algoritmo padrão é Gauss-Newton. Essa função tem como estrutura os principais argumentos:

`nls(formula; data; start; algorithm),`

em que *formula* é a fórmula matemática do modelo que será utilizado; *data* é o conjunto de dados a ser utilizado; *start* é uma lista com valores iniciais para cada um dos parâmetros, e *algorithm* define o tipo de algoritmo utilizado. O algoritmo padrão é o de Gauss-Newton, já o algoritmo de Golub-Pereyra é utilizado para modelos parcialmente lineares, e a opção *port* é para um algoritmo que considera restrições nos parâmetros do modelo.

2.4. Critérios de informação

Os modelos estatísticos são simplificações da realidade e, ao fazer sua seleção, é preciso ter em mente que não existem modelos perfeitos, há apenas modelos aproximados da realidade na qual causam perda de informações. Dessa forma, para fazer a seleção do melhor modelo, dentre aqueles que foram ajustados, são utilizados os critérios de informação [9].

2.4.1 Critério de Akaike - AIC

Para a comparação dos ajustes de modelos pode ser utilizado o critério de Informação de Akaike [2]. O critério de Informação de Akaike foi formulado pelo estatístico japonês Hirotugu Akaike (1927-2009). Esse critério é um estimador da qualidade relativa de modelos estatísticos para um dado conjunto de dados. A partir de uma coleção de modelos para os dados, o AIC estima a qualidade de cada modelo, em relação a cada um dos outros modelos fornecendo assim um meio para a seleção dos mesmos. Entre vários modelos candidatos, deve ser escolhido aquele que apresentar o menor valor de AIC.

Akaike (1974) [2] mostrou que o viés é dado assintoticamente por p , em que p é o número de parâmetros a serem estimados no modelo, e definiu seu critério de informação como

$$\text{AIC} = -2 \log L(\hat{\boldsymbol{\theta}}) + 2(p), \quad (13)$$

em que $\log L(\hat{\boldsymbol{\theta}})$ representa o logaritmo da função de verossimilhança e p o número de parâmetros desconhecidos e estimados. A obtenção completa do resultado acima pode ser encontrada, por

exemplo, em Konishi e Kitagawa (2007)[12] e Emiliano (2009)[8].

De acordo com Sugiura (1978) [20] e Sakamoto *et. al* (1986) [17], o critério de Akaike pode ter um comportamento ruim se existirem muitos parâmetros, em comparação com o tamanho da amostra. Burnham (2002) [6] recomendam usar AIC para selecionar modelos quando o número de observações, n , for maior ou igual do que pelo menos 40 vezes o número de parâmetros p [4].

2.4.2 Critério de informação de Akaike corrigido - AICc

Hurvich e Tsai (1989) [11] desenvolveram o critério Critério de Akaike Corrigido - (AICc) que é indicado para seleção de modelos nos casos de respostas com distribuição normal e pequenas amostras, ou seja, quando o número de observações, n , for menor do que pelo menos 40 vezes o número de parâmetros, denotado por p [4]:

$$AIC_c = -2 \log L(\hat{\theta}) + 2(p) + 2 \frac{p(p+1)}{n-p-1}. \quad (14)$$

2.4.3 Critério de informação Bayesiano - BIC

O Critério de informação Bayesiano, também chamado de Critério de Schwarz, proposto por Schwarz (1978)[18], é um critério de avaliação de modelos definido em termos de probabilidade [8].

$$BIC = -2 \log L(\hat{\theta}) + p \log n, \quad (15)$$

em que $\log L(\hat{\theta})$ representa o logaritmo da função de verossimilhança, e p o número de parâmetros desconhecidos e estimados.

3. Aplicação

O conjunto de dados foi obtido considerando um período ininterrupto de 30 e 60 dias de antecedência. No período 30 dias de antecedência, os dados foram coletados no mês de setembro e novembro de 2017, para os trechos Londrina/PR - Curitiba/PR (LDB-CWB) e Londrina/PR - São Paulo/SP (LDB-CGH) respectivamente, e setembro de 2018, para o período de 60 dias de antecedência, em que a coleta ocorreu simultaneamente para ambos os trechos.

Os dados foram coletados por meio dos *websites* das três maiores companhias aéreas nacionais que fazem os trechos em estudo, denominadas por A, B e C. Considerou-se sempre a passagem com menor valor. Para 30 dias de antecedência, a coleta ocorreu de maneira prospectiva, em que sempre considerava o próximo dia até 30 dias à frente, gerando um total de 30 observações. Para 60 dias de antecedência, a coleta foi feita de 5 em 5 dias, ou seja, coletava-se o menor valor referente ao dia seguinte, e, depois, de 5 dias à frente, até completar 60 dias de antecedência, totalizando 13 observações.

3.1. Modelos não lineares

No intuito de explicar a relação entre a quantidade de dias de antecedência do bilhete adquirido e o valor cobrado, foram considerados os cinco modelos apresentados a seguir [1]. Cada modelo é uma função com seus respectivos parâmetros, em que x é a variável explicativa, ou seja, dias de antecedência, e y representa a média do menor valor diário, conhecida também como variável resposta.

Modelo 1 - Modelo de regressão linear simples:

$$y = \alpha + \beta x,$$

em que: α é coeficiente linear; e β é coeficiente angular

Modelo 2 - Exponencial:

$$y = \exp(\alpha + \beta x),$$

em que: α é o deslocamento no eixo x ; e:

se $\beta > 0$

$$\begin{aligned} x &\rightarrow \infty \\ y &\rightarrow 0, \end{aligned}$$

se $\beta < 0$

$$\begin{aligned} x &\rightarrow -\infty \\ y &\rightarrow 0. \end{aligned}$$

Modelo 3 - Modelo de crescimento exponencial:

$$y = \alpha + \beta \exp(\gamma x),$$

em que: α é o deslocamento no eixo y ; β a concavidade; e γ a rotação.

Modelo 4 - Soma de exponenciais:

$$y = \alpha \exp(\beta x) + \gamma \exp(\delta x),$$

em que: α é o deslocamento no eixo y e a concavidade; β é a rotação; γ é o deslocamento no eixo y , e δ é o fechamento da concavidade.

Modelo 5 - Logístico:

$$y = \alpha + \frac{\beta - \alpha}{1 + \exp\left[\frac{x - \gamma}{\delta}\right]}$$

em que: α é o máximo da função; β é a assintota horizontal; γ é a assintota vertical; e

se $\delta > 0$

$$\begin{aligned} x &\rightarrow \infty \\ y &\rightarrow 0, \end{aligned}$$

se $\delta < 0$

$$\begin{aligned} x &\rightarrow -\infty \\ y &\rightarrow 0. \end{aligned}$$

3.2. Resultados

Para exemplificar a estimação dos parâmetros da Tabela 1, tomando como exemplo o Modelo 2, $y = \exp(\alpha + \beta x)$, os dois parâmetros (α e β) são estimados utilizando o método iterativo de Gauss-Newton descrito na seção 2.3.1. Logo, seja

$$r_i = y_i - \exp(\alpha + \beta x_i), \quad (16)$$

em que $i = 1, 2, \dots, 30$. Na equação (16), y_i representa a média no menor valor diário em reais para x_i dias de antecedência.

Denotando-se por $\theta = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ o vetor dos parâmetros a serem estimados, tem-se:

$$\theta^{(k+1)} = \theta^{(k)} + \left[\mathbf{J}(\theta^{(k)})^T \mathbf{J}(\theta^{(k)}) \right]^{-1} \mathbf{J}(\theta^{(k)}) \mathbf{r}(\theta^{(k)})$$

em que k denota cada iteração do método até a sua convergência. $\mathbf{J}(\theta^{(k)})$ é a matriz jacobiana, dada por:

$$\mathbf{J}(\theta^{(k)}) = \begin{bmatrix} \frac{\partial r_1}{\partial \alpha} & \frac{\partial r_1}{\partial \beta} \\ \frac{\partial r_2}{\partial \alpha} & \frac{\partial r_2}{\partial \beta} \\ \vdots & \vdots \\ \frac{\partial r_{30}}{\partial \alpha} & \frac{\partial r_{30}}{\partial \beta} \end{bmatrix}$$

em que, nesse caso,

$$\begin{aligned} \frac{\partial r_i}{\partial \alpha} &= \exp(\alpha + \beta x_i) \\ \frac{\partial r_i}{\partial \beta} &= x_i \exp(\alpha + \beta x_i). \end{aligned}$$

A implementação das expressões acima, de forma iterativa, pode ser feita no *software* livre R conforme listagem abaixo:

Listagem 1: Script - Método iterativo de Gauss-Newton

```

y<- c(689.77, 675.27, 675.62, 582.04, 586.42, 598.77, 486.61, 488.44, 481.47, 408.01,
      389.36, 394.89, 406.19, 336.89, 333.67, 320.17, 310.47, 311.26, 319.21, 325.06,
      249.57, 237.09, 229.02, 225.92, 230.91, 244.14, 246.54, 191.88, 190.44, 194.46)
x<- seq(1:30)

```

```
a<- 6.85 #valor inicial para alfa
b<- -0.1 #valor inicial para beta

est.a<- vector()
est.b<- vector()

for(i in 1:10)
{
  r<- y-exp(a+b*x)
  da<- exp(a+b*x)
  db<- x*exp(a+b*x)
  Jac<- matrix(c(da,db),30,2)

  theta<- matrix(c(a,b),2,1) + ginv(t(Jac)%*%Jac) %*% (t(Jac)%*%r)

  a<- theta[1]
  b<- theta[2]

  est.a[i]<-a
  est.b[i]<-b
}
```

Após 10 iterações, pode-se observar a convergência do método a partir da terceira iteração, e os valores das estimativas para α (a) e β (b) obtidos são apresentados abaixo:

| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| α | 6,550 | 6,577 | 6,590 | 6,590 | 6,590 | 6,590 | 6,590 | 6,590 | 6,590 | 6,590 |
| β | -0.033 | -0.045 | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 | -0.048 |

De forma mais simples, e como citado na subseção 2.3.1, o *software* R disponibiliza a função *nls* que dispensa a implementação do algoritmo, e pode ser utilizada de forma bastante simples conforme exemplificado na listagem abaixo.

Listagem 2: Script - Uso da função *nls*

```
#Modelo 1:
M1<- lm(dados$A ~ dados$Dias.consecutivos)

#Modelo 2:
M2 = nls(y ~ exp(a + b*x), data = dados, start = list(a =6.85, b = -0.1))

#Modelo 3:
M3 = nls(y ~ a + b*exp(gama*x), data = dados, start = list(a =689, b = 40, gama=-0.09))

#Modelo 4:
M4 = nls(y ~ a*exp(b*x) + gama*exp(delta*Dias.consecutivos), data = dados,
start = list(a=1000, b=-0.09, gama=100, delta=-0.01))

#Modelo 5:
M5 = nls(y ~ a+(b-a)/(1+exp((x-gama)/delta)), data = dados,
start = list(a = 500, b = 1350, gama=0.5, delta=10))
```

A Tabela 1 apresenta um exemplo das estimativas obtidas para cada modelo com seu respectivo erro padrão e valor de AIC, AICc e BIC. Para todos os dois trechos e as três companhias consideradas foram construídas tabelas como esta, porém as mesmas foram omitidas do texto.

| Modelos | Parâmetros | Estimativas | AIC | AICc | BIC |
|---------|------------|-----------------|-----------------|-----------------|-----------------|
| 1 | α | 641,97(5,01) | 21822,28 | 21822,72 | 21838,77 |
| | β | -16,99(0,28) | | | |
| 2 | α | 6,59(0,009) | 21584,37 | 21584,81 | 21600,86 |
| | β | -0,05(0,0007) | | | |
| 3 | α | 111,48(16,35) | 21561,76 | 21562,68 | 21583,74 |
| | β | 647,45(12,88) | | | |
| | γ | -0,07(0,004) | | | |
| 4 | α | 698,99(144,64) | 21563,64 | 21565,24 | 21591,12 |
| | β | -0,06(0,01) | | | |
| | γ | 58,13(152,17) | | | |
| | δ | 0,015(0,06) | | | |
| 5 | α | 149,88(26,02) | 21562,46 | 21564,06 | 21589,93 |
| | β | 1940,09(1180,6) | | | |
| | γ | -7,28(11,37) | | | |
| | δ | 10,55(2,39) | | | |

Tabela 1: Estimativa dos parâmetros (erro padrão) dos modelos para o trecho LDB - CWB, empresa A, 30 dias de antecedência.

Optou-se pelo menor valor de AICc para representar os modelos melhores ajustados devido aos tamanhos pequenos de amostra, sendo que esses são apresentados na Tabela 2.

| Comp. Aérea/Dias de antec. | Trecho | Modelo |
|----------------------------|-----------|--------|
| A/30 dias | LDB - CWB | 3 |
| A/60 dias | LDB - CWB | 4 |
| B/30 dias | LDB - CWB | 3 |
| B/60 dias | LDB - CWB | 5 |
| C/30 dias | LDB - CWB | 2 |
| C/60 dias | LDB - CWB | 5 |
| A/30 dias | LDB - CHG | 5 |
| A/60 dias | LDB - CHG | 5 |
| B/30 dias | LDB - CHG | 5 |
| B/60 dias | LDB - CHG | 5 |
| C/30 dias | LDB - CHG | 3 |
| C/60 dias | LDB - CHG | 5 |

Tabela 2: Modelos mais bem ajustados segundo AICc.

Os modelos mais bem ajustados foram sobrepostos aos valores médios dos trechos em estudo para cada companhia aérea, como mostram as Figuras 1 a 6, em que os gráficos à esquerda são referentes a 30 dias de antecedência, e os da direita, 60 dias de antecedência. O nome da empresa e os respectivos modelos estão na legenda de cada figura.

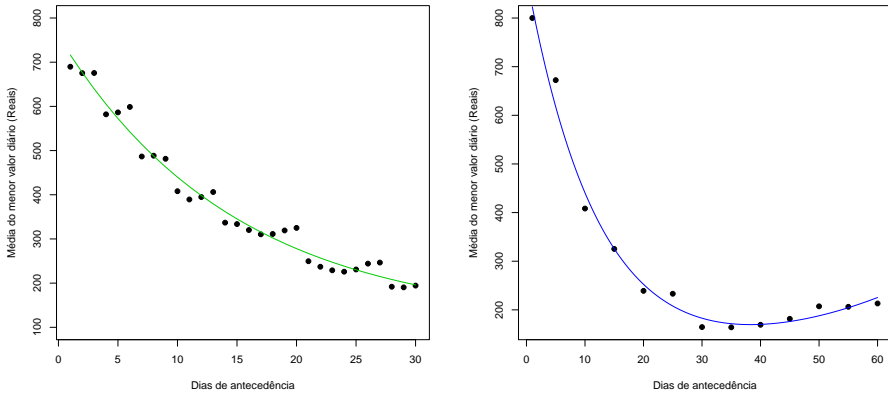


Figura 1: LDB-CWB, empresa A, modelos M3 e M4.

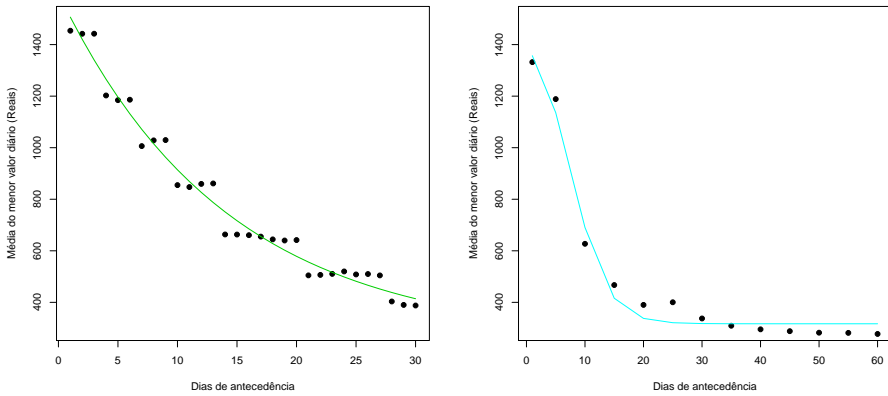


Figura 2: LDB-CWB, empresa B, modelos M3 e M5.

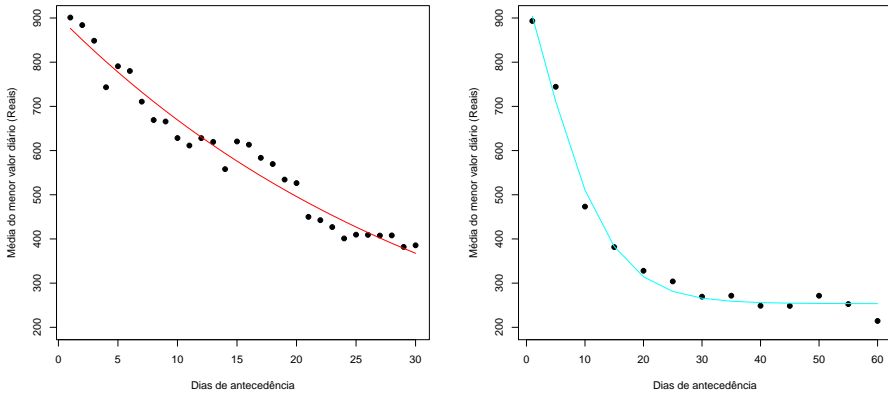


Figura 3: LDB-CWB, empresa C, modelos M2 e M5.

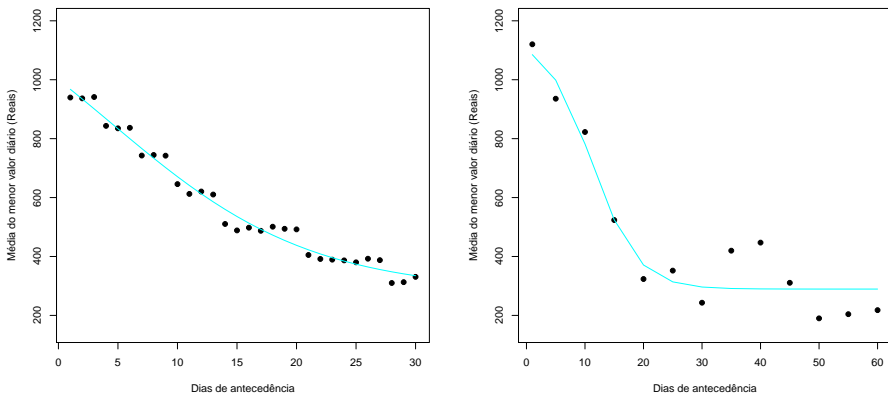


Figura 4: LDB-CGH, empresa A, modelos M5 e M5.

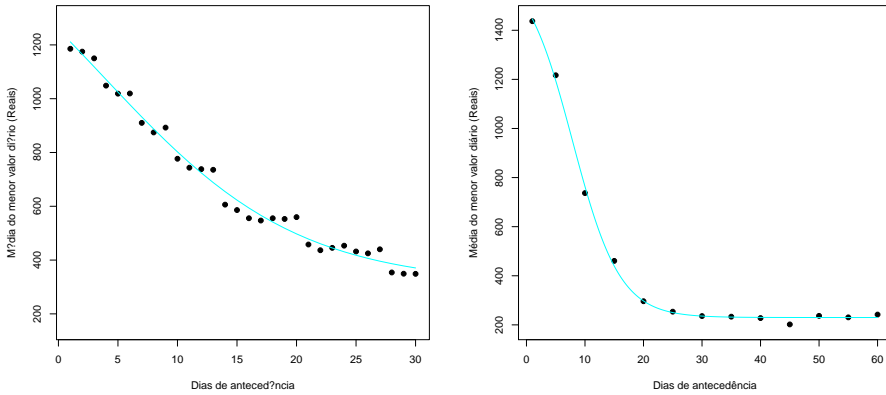


Figura 5: LDB-CGH, empresa B, modelos M5 e M5.

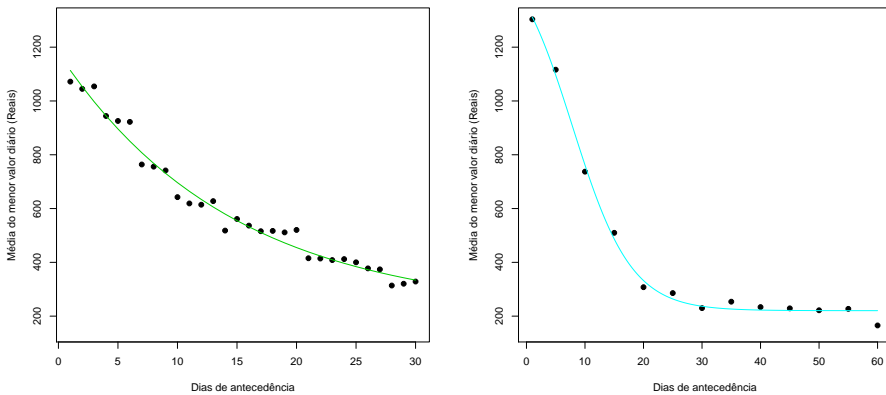


Figura 6: LDB-CGH, empresa C, modelos M3 e M5.

Dessa forma, considerando um período de 30 dias de antecedência, os valores das passagens comportam-se de forma não linear, ou seja, o valor pago por uma passagem adquirida com 15 dias de antecedência está mais próximo do valor pago na mesma com 30 dias de antecedência do que com 1 dia de antecedência da data da viagem; após esse intervalo, no geral, os preços das passagens não sofrem alterações significativas.

4. Considerações Finais

Embora existam poucos estudos disponíveis na literatura nacional sobre as formas de precificação dos trechos para companhias aéreas, entender como isso funciona pode trazer alguma economia na compra desses bilhetes.

Esse trabalho mostrou que, em um período de 30 dias de antecedência da compra do bilhete, os valores comportam-se de forma não linear, e, conseguinte a essa data, não há alterações significativas do mesmo, sendo assim, comprar passagem de forma antecipada pode trazer economia ao consumidor, além de pesquisar companhias que ofereçam os melhores preços.

Contudo, caso o consumidor não consiga adquirir a passagem com 30 dias de antecedência ainda é possível economizar comprando com até 15 dias antes da data da viagem. Do contrário, a compra realizada posterior a 30 dias de antecedência, não garante ao consumidor uma expressiva economia. Logo, recomenda-se a compra de passagens aéreas para o trecho Londrina - Curitiba ou Londrina - São Paulo com 30 dias de antecedência da viagem. Além disso, por meio dos ajustes, é possível prever o valor pago ao longo de 60 dias de antecedência, utilizando os modelos da Tabela 3 em que x é o número de dias de antecedência para a compra do bilhete. Obviamente, deve-se considerar uma possível inflação com o passar do tempo.

| Trecho | Aérea | Melhor Modelo Ajustado |
|---------|-------|---|
| | A | $\hat{y} = 832,76 \exp(-0,080x) + 53,15 \exp(0,024x)$ |
| LDB-CWB | B | $\hat{y} = 316,88 + \frac{1473,1-316,9}{1+\exp\left(\frac{x-7,73}{3,07}\right)}$ |
| | C | $\hat{y} = 253,51 + \frac{1454,4-253,5}{1+\exp\left(\frac{x-2,03}{6,13}\right)}$ |
| | A | $\hat{y} = 289,37 + \frac{1143,6-289,4}{1+\exp\left(\frac{x-11,21}{3,91}\right)}$ |
| LDB-CGH | B | $\hat{y} = 229,92 + \frac{1666,6-229,9}{1+\exp\left(\frac{x-7,88}{4,03}\right)}$ |
| | C | $\hat{y} = 220,34 + \frac{1589,1-220,3}{1+\exp\left(\frac{x-7,88}{5,02}\right)}$ |

Tabela 3: Tabela de predição dos preços para 60 dias de antecedência.

Referências

- [1] Achcar, J. A.; Oliveira, B. R. G. “Gompertz growth curves assuming stable distributions: An application to intrauterine growth for preterm infants”. *Revista Brasileira de Biometria*, v. 33, n°. 2, pp. 170-183, 2015.
- [2] Akaike, H. *A new look at the statistical model identification*. Automatic Control, IEEE Transactions on, v. 19, n°. 6, pp. 716-723, 1974.
- [3] Bates, D. M.; Watts, D. G. *Nonlinear Regression Analysis and Its Applications*. Wiley, (Wiley Series in Probability and Statistics), 2007
- [4] Bello, L. H. A. *Modelagem em Experimentos Mistura-Processo para Otimização de Processos Industriais*. Rio de Janeiro, 2010
- [5] Bilotkach, V. “Understanding Price Dispersion in the Airline Industry: Capacity Constraints and Consumer Heterogeneity”. Darin Lee, Org. *Advances in Airline Economics*, Vol. I. Elsevier. *Advances in Airline Economics*, 2006.
- [6] Burnham, K. P.; Anderson, D.R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer New York, 2002.
- [7] Draper, N. R.; Smith, H. *Applied regression analysis*. New York, U.S.A, 1998

- [8] Emiliano, P. C. *Fundamentos e aplicações dos critérios de informação: Akaike e Bayesiano*. Dissertação Mestrado em Estatística e Experimentação Agropecuária. Universidade Federal de Lavras, 2009
- [9] Emiliano, P. C. *Crítérios de informação de akaike versus bayesiano: análise comparativa*. 19º Simpósio Nacional de Probabilidade e Estatística, 2010
- [10] Evans, W.; Kessides, I. "Localized market power in the U.S. airline industry". *The Review of Economics and Statistics*, v. 75, n°. 1, pp. 66-75, 1993.
- [11] Hurvich, C. M.; Tsai, C.L. "Regression and time series model selection in small samples". *Biometrika*, v. 76, n°. 2, pp. 297-307, 1989.
- [12] Konishi, S.; Kitagawa, G. *Information Criteria and Statistical Modeling*. Springer Publishing Company, Incorporated, 2007.
- [13] Mazucheli, J.; Achcar, J. A. *Algumas considerações em regressão não linear*. *Acta Scientiarum : Technology*, v. 24, n°. 6, pp. 1761-1770, 2002.
- [14] Morais, L. G. A.; Huse, C. *Yield Management nos hotéis do Rio de Janeiro: levantamento e análise*. Dissertação Mestrado em Engenharia de Produção. Pontifícia Universidade Católica, 2002.
- [15] Oliveira, A. V. M.; Huse, C. *Localized Competitive Advantage and Price Reactions to Entry: Full-Service vs. Low-Cost Airlines in the Brazilian Market*. Acervo científico do núcleo de estudos em competição e regulação do transporte aéreo (Nectar), 2005.
- [16] Oliveira, A. V. M. *Estudo dos Determinantes dos Preços das Companhias Aéreas no Mercado Doméstico*. Anac - Agência Nacional de Aviação Civil, 2009.
- [17] Sakamoto, Y.; Ishiguro, M.; Kitagawa, G. *Akaike Information Criterion Statistics*. Reidel Publishing Company, 1986
- [18] Schwarz, G.; "Estimating the dimension of a model". *The Annals of Statistics*, v. 6, pp. 461-464, 1978.
- [19] Seber, G. A. F.; Wild, C. J. *Nonlinear Regression*. Wiley, (Wiley Series in Probability and Statistics), 2003
- [20] Sugiura, N. "Further analysts of the data by Akaike's information criterion and the infinite corrections". *Communications in Statistics-Theory an Methods.*, Ontario, v. 7, n°. 1, pp. 13-26, 1978.
- [21] Thomas, G. *Regressão não Linear*. Trabalho de conclusão do curso de pós-graduação em estatística e experimentação agropecuária. Escola superior de Agricultura Luiz de Queiroz - USP, 2016.
- [22] Weatherford, L. R. *Perishable Asset Revenue Management in General Business Situations*. Ph.D. thesis, Darden Graduate School of Business Administration, University of Virginia, Charlottesville, VA, 1991.

Juliana de Melo
Universidade Tecnológica Federal do Paraná
<jumello@hotmail.com>

Roberto Molina de Souza

Universidade Tecnológica Federal do Paraná
<rmolinasouza@utfpr.edu.br>

Recebido: 11/05/2020
Publicado: 06/08/2020