

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



Roberto Chaves Cavalcante Filho

**MÉTODOS QUANTITATIVOS APLICADOS À
ANÁLISE FINANCEIRA DE EMPRESAS DE
TRANSPORTE AÉREO**

Trabalho de Graduação
2019

Curso de Engenharia Civil-Aeronáutica

Roberto Chaves Cavalcante Filho

**MÉTODOS QUANTITATIVOS APLICADOS À
ANÁLISE FINANCEIRA DE EMPRESAS DE
TRANSPORTE AÉREO**

Prof. Dr. Alessandro Vinícius Marques de Oliveira – ITA

Orientador

ENGENHARIA CIVIL-AERONÁUTICA

SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA

2019

Dados Internacionais de Catalogação-na-Publicação (CIP)**Divisão de Informação e Documentação**

Cavalcante Filho, Roberto Chaves

Métodos quantitativos aplicados à análise financeira de empresas de transporte aéreo / Roberto Chaves Cavalcante Filho

São José dos Campos, 2019.

Número de folhas no formato 48f.

Trabalho de Graduação – Engenharia Civil-Aeronáutica – Métodos quantitativos aplicados à análise financeira de empresas de transporte aéreo – Instituto Tecnológico de Aeronáutica, 2019. Orientador: Prof. Dr. Alessandro Vinícius Marques de Oliveira

1. Regressão penalizada. 2. Tarifa de Bagagem. 3. Transporte Aéreo Instituto Tecnológico de Aeronáutica. III. Métodos quantitativos aplicados à análise financeira de empresas de transporte aéreo

REFERÊNCIA BIBLIOGRÁFICA

CAVALCANTE FILHO, Roberto Chaves; **Métodos quantitativos aplicados à análise financeira de empresas de transporte aéreo**. 2019. 51f. Trabalho de Conclusão de Curso. Graduação em Engenharia Civil-Aeronáutica – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: Roberto Chaves Cavalcante Filho

TÍTULO DO TRABALHO: Métodos quantitativos aplicados à análise financeira de empresas do setor aéreo.

TIPO DO TRABALHO/ANO: Graduação / 2019

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.



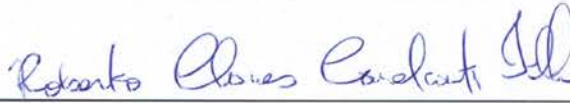
Roberto Chaves Cavalcante Filho

Rua H8A, apartamento 119

12228-460, São José dos Campos-SP

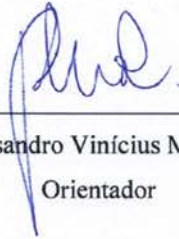
MÉTODOS QUANTITATIVOS APLICADOS À ANÁLISE FINANCEIRA DE EMPRESAS DE TRANSPORTE AÉREO

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação



Roberto Chaves Cavalcante Filho

Autor



Prof. Dr. Alessandro Vinícius Marques Oliveira

Orientador



Prof. Dr. Eliseu Lucena Neto

Coordenador do Curso de Engenharia Civil-Aeronáutica

São José dos Campos, 29 de novembro de 2019

Dedico este trabalho primeiramente a Deus e à Sua Santíssima Mãe, que sempre proveram mais do que eu merecia em todos os momentos. Em segundo lugar, aos meus pais, irmãos de sangue e irmãos do H8A 119, do grupo de Católicos do ITA e, finalmente, a todos os outros que me auxiliaram de alguma forma a trilhar este caminho.

Agradecimentos

Para fins deste trabalho, gostaria de agradecer ao meu orientador, Alessandro, que forneceu todo o apoio necessário para a execução deste.

Queria agradecer aos meus pais e irmãos, que nunca deixaram faltar nada durante a minha graduação e me deram o privilégio de poder focar nos estudos.

Também gostaria de agradecer pelos esforços dos bons professores do ITA, que forneceram uma motivação a mais para frequentar as aulas, em especial o Professor Eliseu, cujas aulas me faziam ir feliz ao instituto.

Não poderia esquecer também de meus amigos de Fortaleza, dos quais a distância não conseguiu diminuir a amizade, e que me ensinaram o significado de *idem velle, idem nolle*.

*"O ato mais específico da fortaleza, mais do que atacar, é
aguentar, isto é, manter-se imóvel em face do perigo."*

-Santo Tomás de Aquino

Resumo

Em 2017, após a liberação da tarifa de bagagem, começou a ganhar relevância no mercado de transporte aéreo a discussão sobre o impacto das outras formas de arrecadação das companhias aéreas no custo do bilhete e no custo efetivo para o consumidor. Além disso, em 2019 houve a interrupção das operações e pedido de recuperação judicial da quarta maior companhia em um mercado caracterizado por seu oligopólio e Este trabalho tem como objetivo identificar quais são as variáveis que efetivamente possuem impacto no preço final, contribuir para a literatura acerca do uso de regressões regularizadas em seleção de variáveis para modelos preditivos, além de oferecer subsídio para pesquisas futuras acerca da precificação de passagens e competição.

Abstract

In 2017, after the allowance of bag fare charging, the discussion about the overall impact of alternative revenue forms of airline companies on the customer effective cost and in the ticket pricing began to show relevance. At the same time, in 2019, there was the interruption of the operations of the 4th largest company in a market already characterized by its oligopoly and the objective of this paper is to identify important variables for the final price, contribute to the literature of regularized regressions in variable selection for predictive models, as well as offering subsidies for future research on competition and air travel pricing.

Lista de Figuras

Figura 2.1-1 Número de Passageiros Pagos (ANAC,2018).	16
Figura 2.2-1 Participação, em assentos oferecidos, das companhias aéreas (ANAC, 2018). ...	17
Figura 2.2-2 Participação das Companhias Aéreas por Receita (ANAC, 2018).	18
Figura 2.4-1 <i>Boxplot</i> da série temporal de preços do trecho FOR-GRU pela Azul versus a antecedência de compra (Das Chagas, 2018).	22
Figura 2.4-2 <i>Boxplot</i> da série temporal de preços do trecho FOR-GRU pela Avianca versus a antecedência de compra (Das Chagas, 2018).	23
Figura 2.4-3 <i>Boxplot</i> da série temporal de preços do trecho FOR-GRU pela LATAM versus a antecedência de compra (Das Chagas, 2018).	23
Figura 2.4-4 <i>Boxplot</i> da série temporal de preços do trecho FOR-GRU pela Gol versus a antecedência de compra (Das Chagas, 2018).	24
Figura 2.5-1 Modelo <i>underfit</i> , apropriado e <i>overfit</i> . (Scikit, 2019).....	25
Figura 3.2-1 Ambiente de Desenvolvimento do Rstudio, versão 1.1.456	31

Lista de Tabelas

Tabela 2.3-1 Participação das Fontes Alternativas de Receita para a Companhia Azul.	19
Tabela 2.3-2 Participação das Fontes Alternativas de Receita para a Companhia Gol.....	19
Tabela 2.3-3 Participação das Fontes Alternativas de Receita para a Companhia LATAM....	20
Tabela 2.3-4 Participação das Fontes Alternativas de Receita para a Companhia Avianca. ...	20
Tabela 2.3-5 Faturamento da Azul, segmentado por tipo (em milhões de Reais).....	20
Tabela 4.1-1 Coeficientes Obtidos (variáveis normalizadas, $\lambda=0.02$).....	36
Tabela 4.1-2 Dez rotas com maior movimentação de passageiros em 2018 (ANAC, 2018)...	37
Tabela 4.1-3 Retenção das variáveis selecionadas com a variação do parâmetro α	40
Tabela 4.2-1 Parâmetro λ para inclusão das variáveis de presença no modelo regularizado...	43
Tabela 4.2-2 Coeficientes associados às variáveis de presença no modelo de referência.....	44
Tabela 4.2-3 Níveis de significância Associados às Variáveis de presença no modelo de referência (* =99%, **=99.9%, ***=~100%)	44

Sumário

1	INTRODUÇÃO	14
1.1	Delimitação do Tema.....	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Setor Aéreo Brasileiro	16
2.1.1	História Recente do Setor	16
2.1.2	Retomada do Crescimento Econômico.....	16
2.2	Companhias Aéreas.....	17
2.3	Fontes de Receita das Companhias Aéreas	18
2.4	Dinâmica dos Preços.....	20
2.4.1	<i>Revenue Management, Yield Management</i> ou Gerenciamento de Receita	21
2.5	Regressões Penalizadas	24
2.5.1	Regressão Linear	25
2.5.2	Regressão <i>LASSO</i>	27
2.5.3	Regressão <i>Ridge</i>	26
2.5.4	Regressão <i>Elastic-Net</i>	28
2.5.5	Significância das Variáveis em Regressões Penalizadas.....	29
3	METODOLOGIA.....	30
3.1	Base de Dados Stataer	30
3.2	Linguagem R.....	31
3.2.1	Biblioteca <i>glmnet</i>	31
4	RESULTADOS E ANÁLISE.....	32
4.1	Tratamento dos dados e Análise Preliminar	35
4.1.1	Normalização dos dados	35
4.1.2	Seleção das cidades	36
4.1.3	<i>Tuning</i> dos Parâmetros	37
4.2	Análise Competitiva das Companhias Aéreas	41
4.2.1	Efeito da Concentração de Mercado.....	41
4.2.2	Posicionamento Estratégico das Companhias Aéreas	42
4.2.3	Efeito da cobrança de bagagem sobre os preços	45
4.2.4	Efeito da privatização de aeroportos no preço das passagens	47
5	CONCLUSÃO.....	49
5.1	Performance das regressões penalizadas.....	49

5.2	Fatores de influência	49
6	REFERÊNCIAS	51

1 Introdução

O mercado de transporte aéreo brasileiro sofreu mudanças estruturais profundas nos anos de 2010 a 2019. O mercado se preparava para um aumento do número de passageiros que vinha ocorrendo de maneira consistente de 2003 a 2015, até que, em 2016, houve a primeira redução do número de passageiros transportados ao ano, o que trouxe o congelamento de investimentos em infraestrutura que estavam programados e que viriam ser retomados até recentemente.

Caracterizado por um forte oligopólio, o mercado também perdeu uma das quatro grandes companhias aéreas operantes no país após a falência da Avianca, conhecida por oferecer serviços e aeronaves frequentemente mais bem avaliados que os das duas maiores (LATAM e Gol) e por oferecer preços competitivos.

O debate acerca das “receitas auxiliares”, descritas mais detalhadamente na seção 2.3, no mercado de transporte aéreo ganhou relevância no cenário brasileiro devido à liberação por parte dos órgãos competentes da cobrança de tarifa sobre despacho de bagagem que ocorreu em 2017, o que, na época, gerou insatisfação dos consumidores, mas as companhias, com exceção da Avianca, mantiveram a cobrança sob o pretexto de que as viagens manteriam os preços, mas seriam mais baratas para aqueles passageiros que viajassem apenas com bagagem de mão.

Entretanto, a precificação das passagens depende de fatores internos, relativos à operação de cada voo, e fatores externos, relativos ao mercado e indicadores macroeconômicos. Em paralelo, tem se tornado cada vez mais comum a utilização de métodos de aprendizagem de máquina para estudos quantitativos, tanto para fins de previsão, quanto classificação. Desse modo, a identificar a influência de variáveis em um sistema tão complexo quanto o mercado de transporte aéreo é uma questão por si só e demanda métodos mais sofisticados.

1.1 Delimitação do Tema

Devido à base de dados utilizada possuir milhares de variáveis relacionadas às linhas aéreas, é um problema *per se* identificar quais são as variáveis independentes relevantes ao problema. Para isso, foram testadas regressões penalizadas, a fim de identificar variáveis que impactam na precificação das passagens e como impactam. Desse modo, o trabalho visou ao esclarecimento de algumas questões:

- Existe correlação entre concessão ou privatização de aeroportos e o preço das passagens?

- Qual a estratégia de precificação das companhias?
- Qual o impacto da cobrança de bagagem na precificação?

Para análise do cenário brasileiro foram utilizados dados financeiros e operacionais que devem ser publicados por regulamentação da ANAC (Agência Nacional de Aviação Civil). Com esses dados, é possível obter informações qualitativas e fazer análise das tendências de mercado para embasar as conclusões derivadas dos resultados quantitativos.

Devido ao número de variáveis que influenciam na precificação das passagens ser distinto entre o mercado doméstico brasileiro e o mercado internacional, este trabalho limitou-se ao estudo do mercado interno brasileiro.

1.2 Objetivos Gerais

O objetivo final deste relatório serão (1) testar a capacidade de extração de informações importantes com o uso de regressões regularizadas, avaliando qual das regressões apresentadas mais se adequa ao problema e (2) utilizar as regressões para responder a perguntas específicas do setor aéreo, descritas na seção 1.1.

2 Revisão Bibliográfica

2.1 Setor Aéreo Brasileiro

2.1.1 História Recente do Setor

Sendo um país de dimensões continentais, o transporte aéreo figurou no Brasil como uma das alternativas principais de transporte durante a história recente. Atualmente, as principais empresas atuando no transporte doméstico de passageiros são a LATAM, Gol e Azul, havendo também a presença de outras empresas que representam, cada uma, menos do que 1% do volume financeiro movimentado pelo setor como Passaredo e Copa.

A aviação comercial no Brasil teve início em 1927 com a criação da VARIG (Viação Aérea Rio-Grandense) e da Condor Syndikat, logo após, na década de 30 com a criação da Viação Aérea São Paulo (VASP) e o começo da oferta da rota São Paulo-Rio de Janeiro (FERREIRA, 2017). Em sequência, houve o surgimento da TAM transportes aéreos na década de 70 e, em sequência, a criação da Infraero, a fim de regulamentar o setor. Durante a década de 80, a crise do setor aérea motivada por ações intervencionistas do governo de congelamento de tarifas durante a época da Hiperinflação. A Gol passou a operar em 2001, seguida pela Avianca em 2002 (na época, OceanAir) e pela Azul em 2008.

2.1.2 Retomada do Crescimento Econômico

Durante o período de 2000-2014, havia grande otimismo por parte de investidores institucionais no setor aéreo brasileiro. Devido a crescimentos consistentes ano após ano conforme pode ser visualizado nos dados da ANAC da Figura 2.1-1 Número de Passageiros Pagos (ANAC,2018).

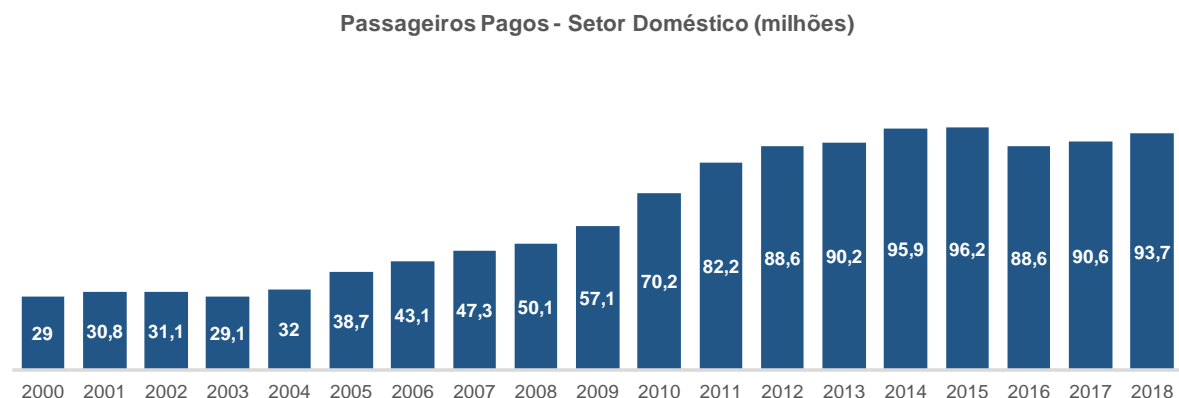


Figura 2.1-1 Número de Passageiros Pagos (ANAC,2018).

Entretanto, devido à crise econômica que o país atravessou durante os anos de 2015-2018, motivada, entre muitas coisas, pela instabilidade política, as taxas de crescimento se inverteram e vários investimentos em infraestrutura no setor foram cancelados ou adiados.

Todavia, a economia brasileira mostrou sinais de recuperação em 2019, o que significou, novamente, otimismo dos investidores, com projeções de crescimento de 2% a 3% do ministério de infraestrutura e no aumento expressivo do número de concessões aeroportuárias, com a concessão de 12 aeroportos em 3 blocos.

2.2 Companhias Aéreas

Atualmente, no país figuram 26 empresas aéreas registradas pela ANAC. Dessas, 3 são focadas em aviação comercial doméstica (Gol, LATAM, Azul), 3 são as subsidiárias das grandes companhias para transporte de carga (LATAM Cargo Brasil, Azul Cargo, Gollog), implicando que as outras 20 dividem uma fatia relativamente pequena do mercado.

Segundo dados da ANAC que podem ser visualizados na Figura 2.2-1 Participação, em assentos oferecidos, das companhias aéreas as quatro grandes companhias aéreas significam mais de 99% do transporte de passageiros em número de assentos oferecidos em 2018.

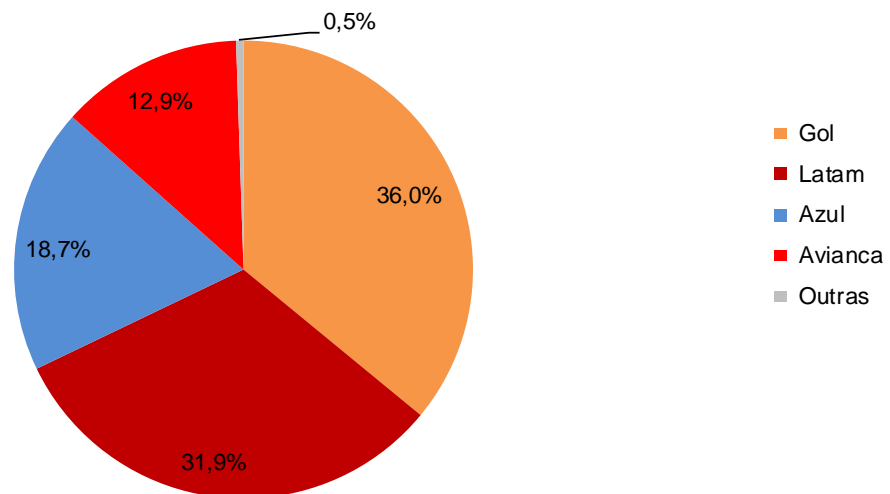


Figura 2.2-1 Participação, em assentos oferecidos, das companhias aéreas (ANAC, 2018).

Dada a relevância dessas quatro grandes empresas, foi posto em um gráfico a participação dessas, em termo da receita, para verificação do posicionamento estratégico antes e depois da falência da Avianca, que não tem seus dados contábeis do segundo trimestre de 2019 divulgados ainda. Esses dados podem ser visualizados na Figura 2.2-2.

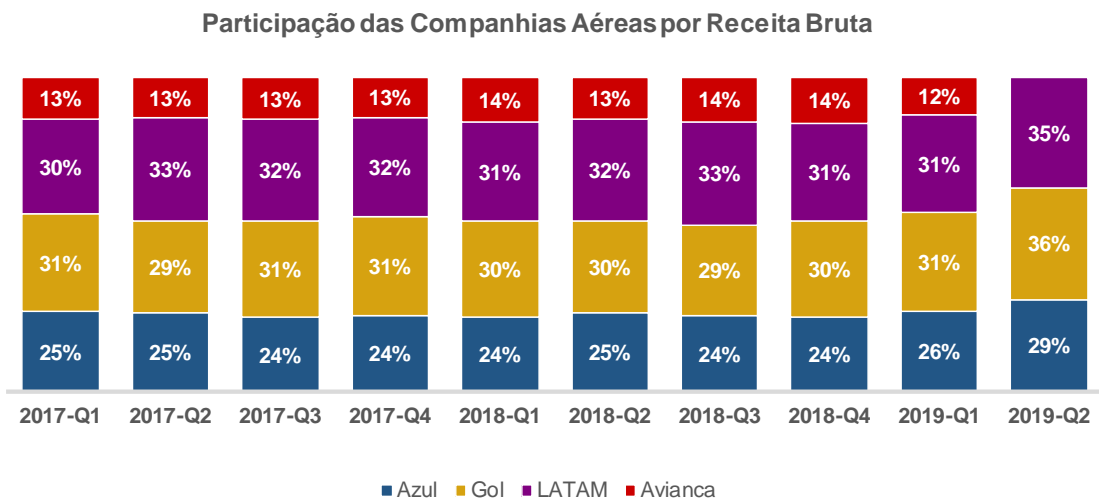


Figura 2.2-2 Participação das Companhias Aéreas por Receita (ANAC, 2018).

De acordo com Malveira (2018), devido às grandes pressões que as empresas sofrem por parte dos órgãos reguladores e sua sensibilidade a fatores macroeconômicos, torna-se inviável a que estas adiram a estratégias de competitividade por custos. Desse modo, conforme pode ser visualizado na Figura 2.2-2, percebe-se que:

- A participação das companhias apresenta pouca variação;
- Não há uma agressividade por parte delas na obtenção de novos passageiros;
- A Azul, Gol e LATAM mantiveram suas posições com relação uma às outras após a interrupção das operações da Avianca, indicando que ou não houve tempo suficiente para a realocação dessa demanda, ou a demanda foi redistribuída de maneira proporcional.

2.3 Fontes de Receita das Companhias Aéreas

No mercado de transporte aéreo, tornou-se prática comum para as companhias aéreas a de oferecer serviços extras durante o voo, tais como serviços multimídia ou alimentação, que outrora já foi inclusa no preço da passagem.

A ANAC exige das companhias com mais de 1% de participação em RPK (Refere-se ao volume de Passageiros Quilômetros Transportados, ou seja, a soma do produto entre o número de passageiros pagos e a distâncias das etapas) que sejam publicadas suas demonstrações contábeis trimestrais. Ao término deste trabalho, estavam publicadas, para as três maiores companhias (Gol, Azul e LATAM) até o segundo trimestre de 2019, já para a Avianca, devido ao seu pedido de recuperação judicial, tem-se as demonstrações até o primeiro trimestre de 2019.

Por normas da agência, as empresas devem segmentar as suas receitas entre vôos domésticos e internacionais, regulares e não regulares e, dentro dessas categorias, especificar as receitas provenientes de:

- Passageiros;
- Bagagem;
- Carga;
- Mala postal;
- Receitas Auxiliares (alimentação, internet, telefone, multimídia, acompanhamento, assistência médica, outras) e
- Penalidades (cobrança por cancelamento, cobrança por reembolso, cobrança por remarcação, cobrança por não comparecimento, outras).

Com os dados contábeis disponibilizados pela ANAC, pode ser verificado a participação das fontes de receita no faturamento de vôos domésticos na Tabela 2.3-1, Tabela 2.3-2, Tabela 2.3-3 e Tabela 2.3-4.

Tabela 2.3-1 Participação das Fontes Alternativas de Receita para a Companhia Azul.

	Azul									
	2017-Q1	2017-Q2	2017-Q3	2017-Q4	2018-Q1	2018-Q2	2018-Q3	2018-Q4	2019-Q1	2019-Q2
Transporte Aéreo Regular Doméstico	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	87,6%	86,2%	85,9%	85,0%	89,1%	85,9%	86,8%	86,5%	87,3%	86,3%
Bagagem	0,7%	0,7%	0,6%	1,9%	2,1%	2,0%	2,3%	2,3%	2,4%	2,4%
Carga	2,2%	2,9%	2,9%	2,9%	3,1%	3,8%	3,5%	3,9%	3,9%	4,7%
Mala Postal	0,0%	0,0%	0,0%	5,8%	1,4%	2,6%	3,0%	2,9%	2,1%	2,0%
Receitas Auxiliares	5,1%	5,4%	6,0%	4,4%	4,2%	5,7%	4,5%	4,4%	4,2%	4,5%
Penalidades	4,3%	4,9%	4,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
Transporte Aéreo Regular Internacional	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	91,2%	90,9%	91,4%	93,2%	90,6%	88,2%	88,6%	88,4%	89,3%	87,3%
Bagagem	0,2%	0,1%	0,2%	0,4%	0,4%	0,6%	0,6%	0,9%	1,0%	0,8%
Carga	3,9%	4,1%	3,9%	5,1%	4,0%	5,8%	6,0%	5,7%	4,7%	5,4%
Mala Postal	0,0%	0,0%	0,0%	0,6%	3,7%	3,1%	2,5%	3,0%	2,8%	3,1%
Receitas Auxiliares	4,0%	4,3%	3,9%	0,8%	1,3%	2,2%	2,3%	2,0%	2,1%	3,4%
Penalidades	0,8%	0,6%	0,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Tabela 2.3-2 Participação das Fontes Alternativas de Receita para a Companhia Gol.

	Gol									
	2017-Q1	2017-Q2	2017-Q3	2017-Q4	2018-Q1	2018-Q2	2018-Q3	2018-Q4	2019-Q1	2019-Q2
Transporte Aéreo Regular Doméstico	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	87,0%	86,1%	87,0%	87,9%	87,5%	84,7%	86,9%	86,6%	86,9%	86,9%
Bagagem	1,1%	1,0%	1,9%	1,9%	2,3%	2,3%	2,9%	2,7%	3,0%	2,5%
Carga	3,8%	4,6%	3,8%	3,9%	4,0%	4,8%	4,0%	3,9%	3,5%	3,7%
Mala Postal	0,0%	0,0%	0,0%	3,8%	1,1%	1,7%	1,5%	1,6%	1,6%	1,4%
Receitas Auxiliares	1,2%	1,4%	1,3%	4,6%	4,7%	6,0%	4,4%	4,8%	4,6%	5,2%
Penalidades	6,9%	6,8%	6,0%	0,5%	0,4%	0,4%	0,4%	0,4%	0,4%	0,3%
Transporte Aéreo Regular Internacional	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	96,9%	92,3%	95,5%	87,0%	96,0%	92,8%	94,9%	92,6%	94,4%	92,9%
Bagagem	0,1%	0,2%	0,6%	0,7%	0,4%	0,8%	1,0%	1,2%	1,0%	1,5%
Carga	0,5%	1,0%	1,0%	1,3%	0,5%	1,1%	0,9%	0,9%	0,5%	0,8%
Mala Postal	0,0%	0,0%	0,0%	11,6%	1,7%	3,3%	1,2%	3,7%	2,4%	3,2%
Receitas Auxiliares	0,3%	0,4%	0,4%	1,5%	0,6%	1,2%	1,1%	1,1%	0,9%	1,0%
Penalidades	2,2%	6,0%	2,5%	0,7%	0,7%	0,7%	0,9%	0,6%	0,9%	0,6%

Tabela 2.3-3 Participação das Fontes Alternativas de Receita para a Companhia LATAM.

	TAM									
	2017-Q1	2017-Q2	2017-Q3	2017-Q4	2018-Q1	2018-Q2	2018-Q3	2018-Q4	2019-Q1	2019-Q2
Transporte Aéreo Regular Doméstico	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	85,2%	82,5%	84,6%	84,3%	86,1%	83,2%	82,9%	84,1%	84,5%	82,1%
Bagagem	0,6%	0,5%	1,7%	2,3%	2,1%	1,8%	2,2%	2,3%	2,3%	0,5%
Carga	6,6%	7,4%	6,2%	6,3%	5,7%	7,0%	5,6%	6,1%	5,7%	7,1%
Mala Postal	0,0%	0,0%	0,0%	1,0%	1,0%	1,5%	0,3%	1,0%	0,2%	0,3%
Receitas Auxiliares	2,0%	3,2%	2,3%	3,6%	4,4%	2,2%	4,0%	2,4%	3,0%	2,4%
Penalidades	5,6%	6,4%	5,1%	2,6%	0,7%	4,3%	5,0%	4,0%	4,4%	7,5%
Transporte Aéreo Regular Internacional	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
Passageiro	91,4%	90,0%	89,8%	77,7%	86,9%	87,2%	88,6%	92,0%	86,4%	88,7%
Bagagem	0,2%	0,1%	0,5%	0,5%	0,5%	0,2%	0,0%	0,5%	0,7%	0,3%
Carga	6,5%	7,9%	6,8%	17,5%	9,4%	4,9%	4,4%	2,4%	3,9%	3,1%
Mala Postal	0,0%	0,0%	0,0%	-0,4%	-0,1%	0,1%	0,2%	2,2%	3,3%	3,6%
Receitas Auxiliares	0,4%	0,5%	0,2%	0,0%	-0,1%	1,8%	0,0%	0,5%	0,8%	0,7%
Penalidades	1,4%	1,4%	2,7%	4,9%	3,4%	5,7%	6,8%	2,4%	4,9%	3,5%

Tabela 2.3-4 Participação das Fontes Alternativas de Receita para a Companhia Avianca.

	Avianca									
	2017-Q1	2017-Q2	2017-Q3	2017-Q4	2018-Q1	2018-Q2	2018-Q3	2018-Q4	2019-Q1	2019-Q2
Transporte Aéreo Regular Doméstico	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	-
Passageiro	84,7%	81,7%	82,8%	83,8%	83,5%	80,5%	80,4%	69,9%	80,6%	-
Bagagem	0,9%	0,9%	0,9%	1,1%	1,4%	2,3%	2,0%	2,1%	2,9%	-
Carga	4,2%	4,9%	5,6%	2,5%	4,2%	5,8%	4,7%	4,5%	3,7%	-
Mala Postal	0,1%	0,1%	0,1%	1,6%	2,0%	-0,5%	-5,0%	5,6%	0,6%	-
Receitas Auxiliares	0,6%	1,2%	1,7%	3,6%	3,7%	5,5%	4,9%	4,6%	4,8%	-
Penalidades	9,5%	11,2%	8,9%	7,4%	5,1%	6,3%	13,0%	13,3%	7,5%	-
Transporte Aéreo Regular Internacional	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	-
Passageiro	50,6%	64,6%	92,0%	97,7%	95,5%	95,0%	88,2%	87,5%	86,8%	-
Bagagem	1,7%	1,6%	0,3%	0,8%	0,7%	0,9%	1,8%	2,3%	1,9%	-
Carga	1,9%	1,7%	0,3%	0,7%	1,2%	1,8%	0,5%	0,0%	0,0%	-
Mala Postal	0,0%	0,0%	0,0%	0,1%	0,0%	0,1%	0,8%	1,1%	1,1%	-
Receitas Auxiliares	1,1%	1,0%	0,2%	0,7%	2,6%	2,3%	2,2%	2,6%	10,2%	-
Penalidades	44,6%	31,1%	7,2%	0,0%	0,0%	0,0%	6,5%	6,5%	0,0%	-

Desse modo, percebe-se que, desde a liberação da cobrança de tarifa de bagagem, a participação dessa cobrança no faturamento das empresas subiu de 0,5% a 2-3%, à primeira vista, pode não parecer um crescimento relevante, mas deve ser levado em conta o próprio crescimento da receita como um todo que as empresas tiveram nesse período. A exemplo da Azul, que, conforme pode ser visto na Tabela 2.3-5, em 2017 faturava apenas R\$ 12 milhões trimestralmente com taxa de bagagem, passou a faturar R\$ 53 milhões no segundo trimestre de 2019.

Tabela 2.3-5 Faturamento da Azul, segmentado por tipo (em milhões de Reais)

	Azul									
	2017-Q1	2017-Q2	2017-Q3	2017-Q4	2018-Q1	2018-Q2	2018-Q3	2018-Q4	2019-Q1	2019-Q2
Transporte Aéreo Regular Doméstico	1.670,12	1.514,21	1.713,08	1.910,15	1.766,59	1.689,78	2.014,43	2.124,89	2.082,13	2.205,12
Passageiro	1.463,26	1.304,67	1.472,31	1.623,08	1.573,86	1.450,82	1.749,14	1.837,91	1.818,65	1.903,31
Bagagem	12,09	10,70	10,83	36,45	37,82	33,84	45,76	49,02	50,89	53,94
Carga	37,54	43,42	49,79	55,74	54,80	64,44	70,06	81,94	80,40	104,39
Mala Postal	0,05	0,00	0,00	110,48	25,32	43,76	59,60	62,17	43,98	44,23
Receitas Auxiliares	85,34	81,24	103,10	84,41	74,80	96,92	89,86	93,85	88,20	99,25

2.4 Dinâmica dos Preços

Percebe-se que, por mais que haja um investimento por parte das companhias em rentabilizar suas operações de outras maneiras, a venda dos bilhetes continua sendo a principal fonte de renda.

É sabido, ainda, que atualmente a precificação das passagens se dá de maneira dinâmica por algoritmos proprietários. Das Chagas (2018) mapeou o comportamento das séries temporais da precificação do trecho Fortaleza-Guarulhos durante o ano de 2018 a fim de verificar a viabilidade de se economizar na compra de bilhetes pela identificação de padrões.

Segundo Das Chagas (2018), há pouca divergência entre a média de preços para o trecho estudado, havendo um aumento na semana próxima à decolagem.

2.4.1 *Revenue Management, Yield Management* ou Gerenciamento de Receita

As companhias aéreas enfrentam uma delicada situação, onde oferecem um produto altamente perecível, pois tem prazo estritamente definido para ser comercializado, tornando sua receita volátil. Ainda, expostas à sazonalidade de demanda, são extremamente sensíveis a fatores macroeconômicos.

Segundo a ANAC (2018), em 2018, os custos com combustíveis e lubrificantes foram os mais expressivos, representando 31,4% dos gastos totais das companhias, seguidos dos custos com arrendamento (ou *leasing*), manutenção e seguro de aeronaves, 18,2% do total. Desse modo, a estrutura de custos das companhias é extremamente dependente do câmbio, devido ao preço do petróleo ser referenciado em dólar e dependente também das taxas de juros dos bancos centrais, que afetam o custo do *leasing* de aeronaves.

Desse modo, a fim de gerenciar os riscos e extrair o máximo de receita possível dos diferentes períodos do ano, as companhias aéreas recorrem à técnica do Gerenciamento de Receita, que, segundo Pak e Piersma (2002) é a arte de maximizar o lucro gerado a partir de uma capacidade limitada de um produto sob um horizonte finito, levando em conta o tempo, preço e consumidor.

A prática de *revenue management*, de acordo com McGill e Ryzin (2002), começou nos Estados Unidos nos anos 70. A fim de preencher assentos que iam vazios nos voos, as companhias começaram a oferecer descontos para a compra da passagem de um acompanhante, oferecendo uma segunda passagem por um preço menor e, com isso, diluindo os custos da viagem por passageiro.

Com o passar do tempo, surgiram diferentes técnicas de gerenciamento de receita, que incluem o oferecimento de tarifas distintas, penalidade de remarcação, taxa por cancelamento, cobrança de penalidade por não-comparecimento (*no show*), *overbooking* e outros. Atualmente, as companhias estabeleceram métodos muito mais elaborados de aplicar a técnica, adaptando, em tempo real, o preço dos bilhetes a variáveis como antecedência e lotação do voo. Segundo Das Chagas (2018), há uma grande correlação entre o aumento no

preço das passagens que acontece nos 15 dias antecedentes à decolagem e o esgotamento da oferta de bilhetes econômicos, restando ao passageiro que compra de última hora apenas tarifas mais caras. Esse aumento do preço pode ser verificado em todas as quatro companhias que ofereciam o trecho na época. A variabilidade dos preços pode ser verificada em Figura 2.4-1, Figura 2.4-2, Figura 2.4-3 e Figura 2.4-4.

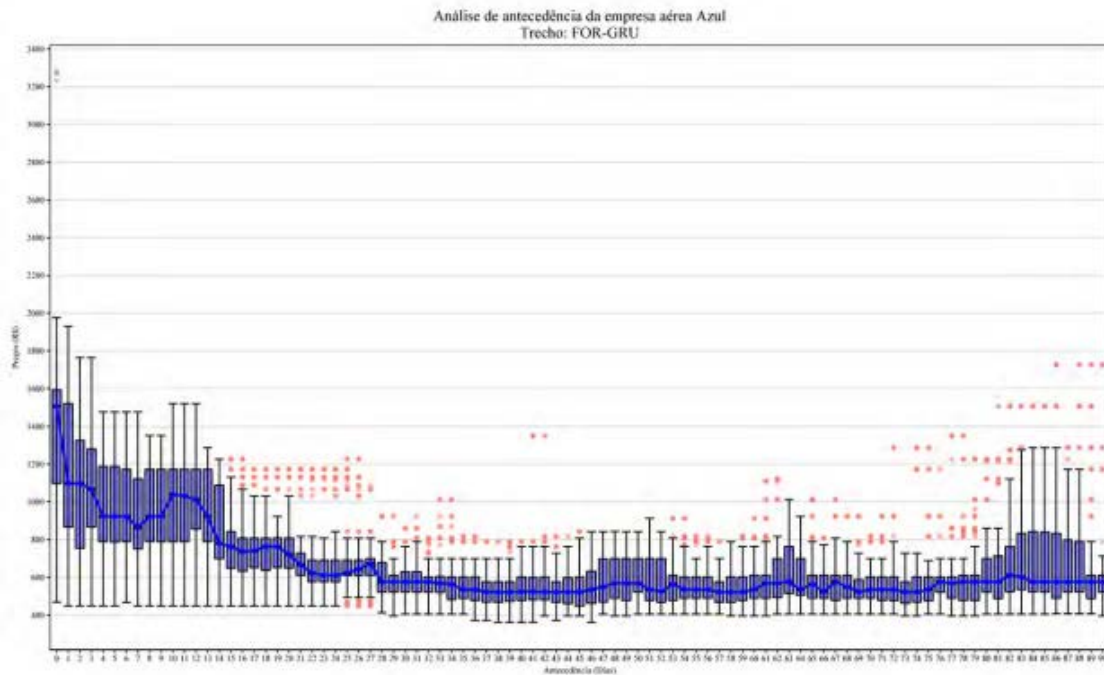


Figura 2.4-1 *Boxplot* da série temporal de preços do trecho FOR-GRU pela Azul versus a antecedência de compra (Das Chagas, 2018).

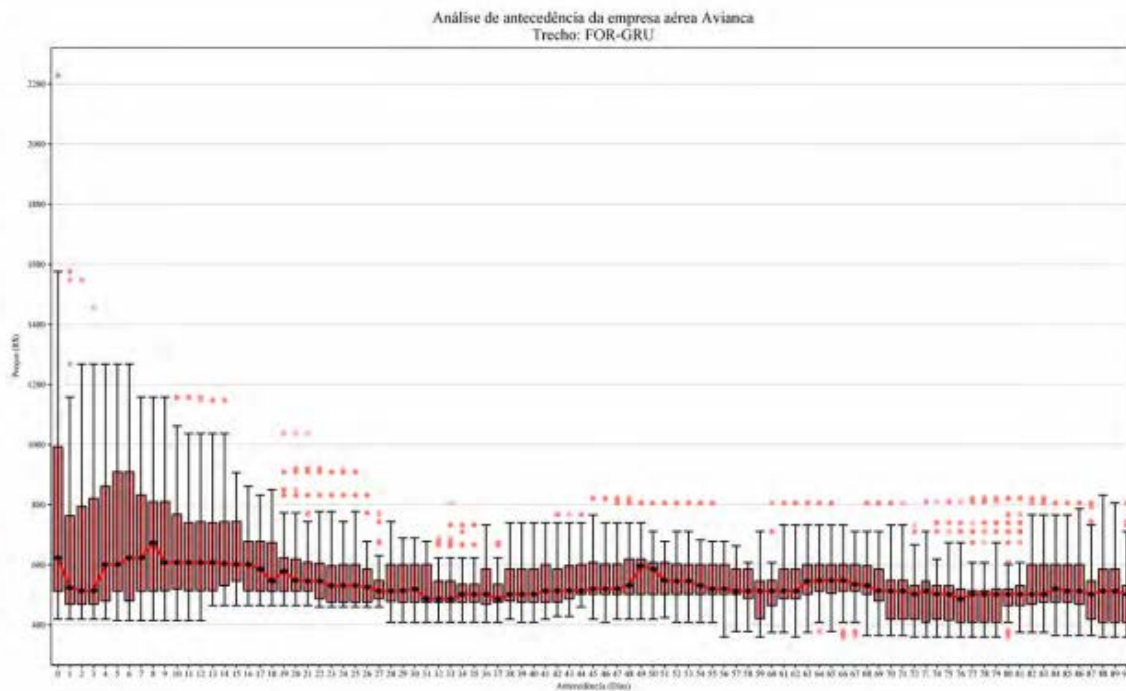


Figura 2.4-2 *Boxplot* da série temporal de preços do trecho FOR-GRU pela Avianca versus a antecedência de compra (Das Chagas, 2018).

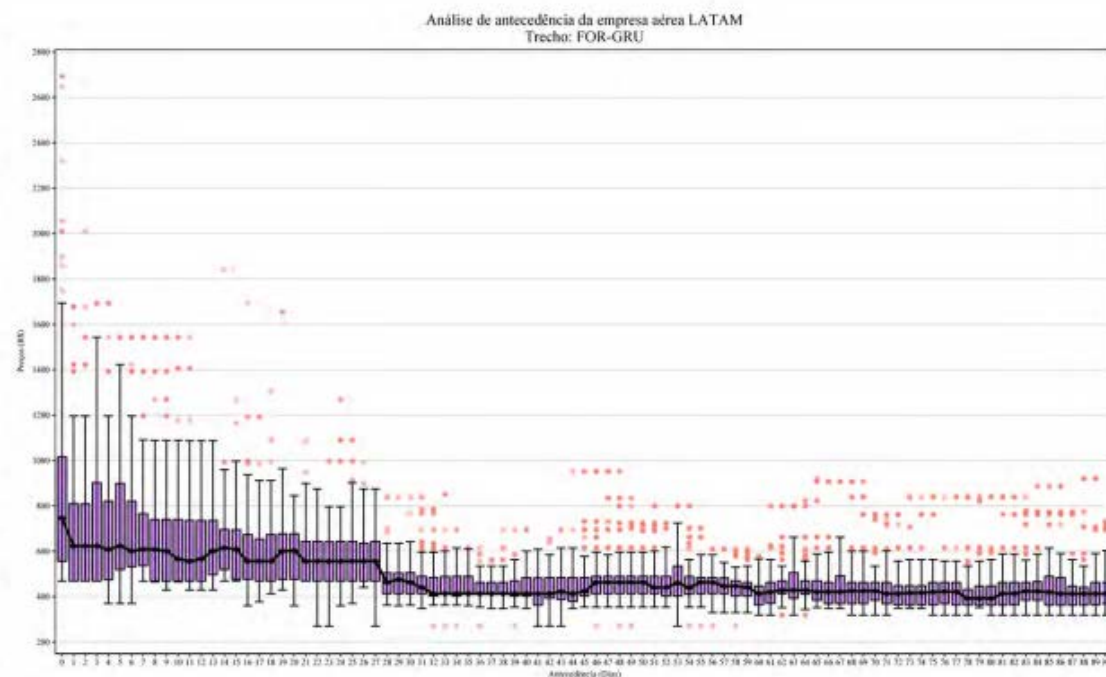


Figura 2.4-3 *Boxplot* da série temporal de preços do trecho FOR-GRU pela LATAM versus a antecedência de compra (Das Chagas, 2018).

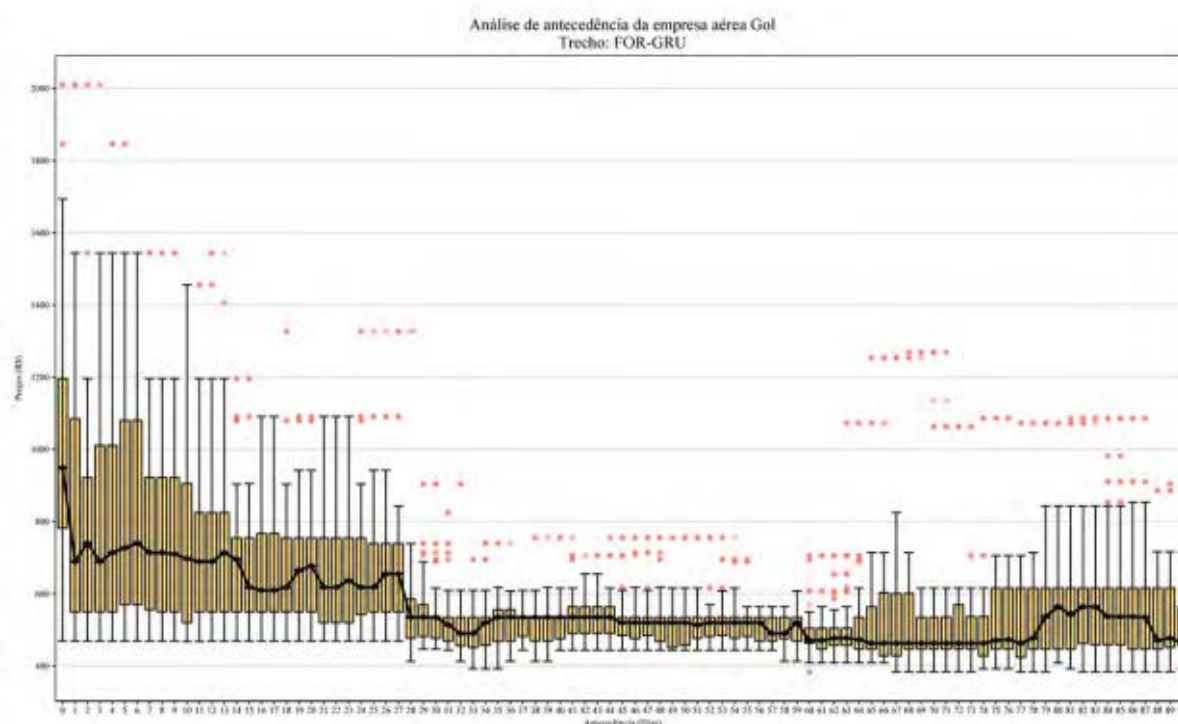


Figura 2.4-4 *Boxplot* da série temporal de preços do trecho FOR-GRU pela Gol versus a antecedência de compra (Das Chagas, 2018).

2.5 Regressões Penalizadas

Também chamadas, na literatura, de regressões regularizadas, elas são regressões que visam a modelagem através da diminuição do efeito do *overfitting* isso é, quando o modelo se adequa de maneira excessiva aos dados de treinamento (BURNHAM, 2002), apresentando grande viés e impossibilitando o modelo de prever corretamente situações alheias ao conjunto com o qual foi alimentado.

Desse modo, ao se adicionar variáveis de maneira excessiva em um modelo que contenha uma razão pequena de observações para variáveis, corre-se o risco de *overfitting*. Na Figura 2.5-1, as linhas azuis representam modelos de predição e as linhas laranja representam a função geradora dos pontos. Tem-se exemplos de três modelos, um com *underfit* (situação oposta ao *overfit*) ou seja, grande viés, mas com variância pequena, um modelo intermediário, que apresenta um comportamento mais próximo à função original e um modelo que padece do *overfit*, que, embora o modelo se adequa bem aos pontos de treinamento, ele se afasta bastante da função real no resto do domínio.

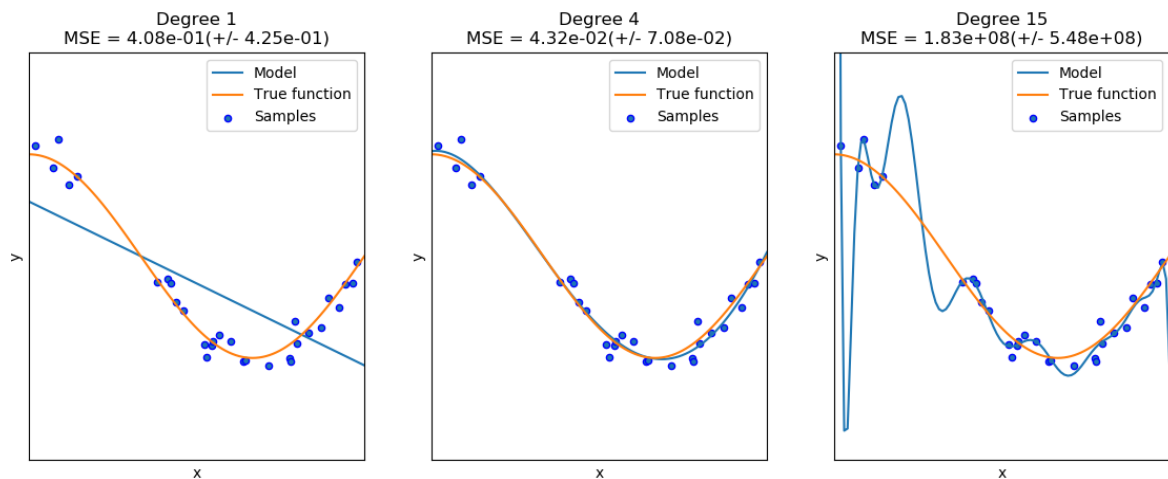


Figura 2.5-1 Modelo *underfit*, apropriado e *overfit*. (Scikit, 2019)

A fim de evitar os problemas citados, possibilidades são a redução de variáveis através da identificação de correlação entre elas, aumentar a quantidade de dados para o treinamento do modelo ou fazer uso da regularização.

2.5.1 Regressão Linear

Os modelos lineares são chamados assim estabelecem estimadores para Y , considerando linearidade nos preditores $X_1, X_2 \dots X_n$. Três aspectos são os principais ao se tratar delas:

- Apesar do nome, esses modelos permitem que sejam aplicadas transformações não-lineares nos preditores, preservando a linearidade da função;
- Os métodos lineares, tais como os MQO (mínimos quadrados ordinários) possuem desempenho muito competitivo quando comparado a modelos não-lineares mais sofisticados a um menor custo computacional e;
- De modo, mais importante para este trabalho, as regressões permitem que a visualização das relações entre as variáveis seja feita de maneira direta, dada a estrutura linear da equação.

Pelo método dos mínimos quadrados, a regressão linear visa estimar a variável dependente através de um conjunto de coeficientes.

Equação 1: Estimador da Regressão Linear comum

$$E(Y_i|X) = \beta_0 + \sum_{i=1}^j \beta_j X_{i,j}$$

O método MQO, visa à redução do erro quadrático médio entre os valores previstos e amostrais, segundo indica a Equação 2, onde y_i são as observações, β_0 o coeficiente angular, x_i o vetor de variáveis independentes da observação i , e β o vetor dos coeficientes.

Equação 2: Função Objetivo da Regressão de MQO

$$\min \left\{ \frac{1}{N} \sum_{i=1}^j (y_i - \beta_0 - x_i^T \beta)^2 \right\}$$

2.5.2 Regressão Ridge

A regressão Ridge, foi proposta por Hoerl e Kennard (1970), como uma alternativa ao método comum de mínimos quadrados, pois, segundo os autores, a regressão ordinária seria extremamente sensível e propensa a erros quando as variáveis de estudo padecessem do problema da multicolinearidade, a fim de estabilizar os sistemas de estudo, a regressão Ridge penaliza os coeficientes através da sua função objetivo conforme a Equação 3. Pelo fato de ser penalizada pela norma L2 do vetor de coeficientes, se diz que a regressão ridge é uma regressão regularizada em L2.

Equação 3: Função Objetivo da regressão Ridge

$$\min \left\{ \frac{1}{N} \sum_{i=1}^j (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{i=1}^m \beta_i^2 \right\}$$

Dado que λ é um número real e positivo, a regressão Ridge impõe sobre o vetor de coeficientes uma condição de contorno esférica, conforme pode ser visualizado na Figura 2.5-1.

Desse modo, a regressão Ridge consegue diminuir de maneira consistente a variância de modelos lineares com os coeficientes calculados com ela e possui também a propriedade de corrigir os sinais dos coeficientes calculados, de maneira que os estimadores possuem um comportamento convergente com o aumento de λ .

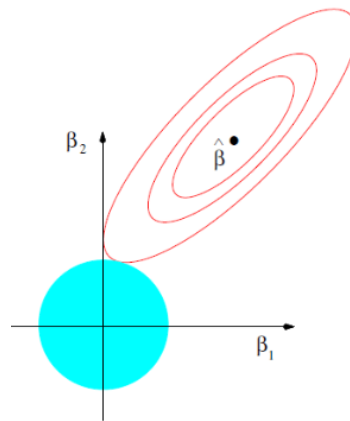


Figura 2.5-1 Condição de Contorno para Regressão Ridge (TIBSHIRANI, 1996)

2.5.3 Regressão LASSO

A Regressão LASSO, Lasso ou lasso deriva de *Least Absolute Shrinkage and Selection Operator*, que significa operador de seleção e encolhimento baseado em mínimos valores absolutos (módulos) foi proposta por Tibshirani (1996) como uma alternativa aos métodos vigentes de encolhimento de coeficientes, incluindo a Ridge. Desse modo, assim como a regressão linear comum, a variável dependente é estimada através de somas lineares, ela possui, entretanto, uma regularização de norma L1 do vetor dos coeficientes, isso a LASSO possui a sua função objetivo conforme a Equação 4. Para esse problema quando λ tende a zero, o problema se torna um problema ordinário de regressão e, quando λ tende a valores maiores, resta somente o coeficiente linear.

Equação 4: Função Objetivo da Regressão LASSO

$$\min \left\{ \frac{1}{N} \sum_{i=1}^j (y_i - \beta_0 - x_i^T \beta)^2 + \lambda \sum_{i=1}^m |\beta_i| \right\}$$

Desse modo, a Regressão LASSO delimita ao vetor de coeficientes uma condição de contorno que tem formato de diamante, conforme pode ser visualizada para o caso de duas variáveis no vetor de coeficientes na Figura 2.5-2.

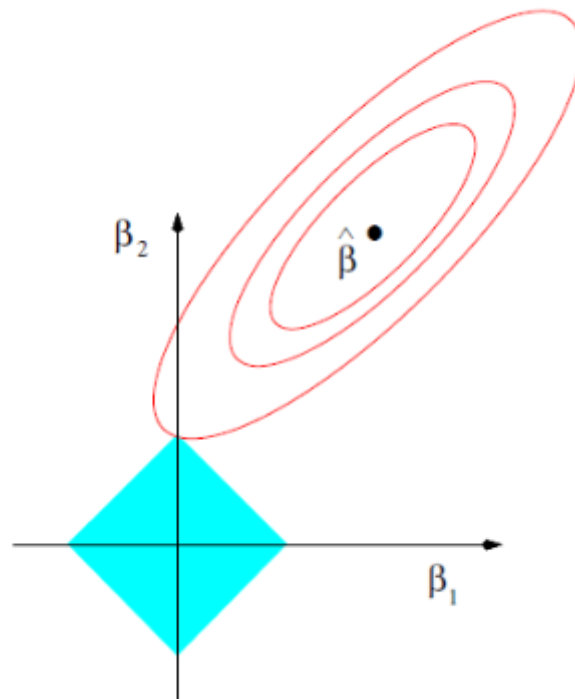


Figura 2.5-2 Condição de Contorno para Regressão LASSO (TIBSHIRANI, 1996)

Devido às características da regressão, ela se torna um bom algoritmo de seleção de variáveis, devido à sua capacidade de zerar os coeficientes. Além dessa propriedade, que é relevante para o problema de seleção de variáveis, verifica-se empiricamente que:

- Na LASSO, há a tendência de se aumentar coeficientes de variáveis correlacionadas;
- A LASSO é uma boa aplicação quando se tem uma quantidade pequena de parâmetros significantes (TIBSHIRANI, 1996);
- A Ridge se torna uma alternativa mais interessante quando há muitas variáveis que impactam de maneira menos significativa (TIBSHIRANI, 1996);
- Para variáveis correlacionadas, a LASSO irá descartar uma delas aleatoriamente, o que pode ser danoso no caso de populações onde essas variáveis não o são.

2.5.4 Regressão *Elastic-Net*

Dado que a regressão LASSO possui um problema com variáveis correlacionadas e que a regressão Ridge é menos eficiente em problemas que contenham poucas variáveis significativas, é utilizada como alternativa a regressão *Elastic-Net*, que, em poucos termos, é uma combinação convexa das outras.

Equação 5 Função Objetivo da Regressão *Elastic-Net*

$$L_{enet}(\hat{\beta}) = \frac{\sum_{i=1}^n (y_i - x_i' \hat{\beta})^2}{2n} + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^m \hat{\beta}_j^2 + \alpha \sum_{j=1}^m |\hat{\beta}_j| \right),$$

Verifica-se, portanto, que a regressão *Elastic-Net* possui um coeficiente a mais, além do λ , comum às outras duas, o fator α se torna um fator a mais de regularização da regressão e, quando α tende a 0, a regressão tende à regularização L2 e, quando α tende a 1, tem-se regularização de norma L1.

Em vista disso, a regressão *Elastic-Net* se mostra como uma alternativa para criação de modelos lineares, visto que, se calibrada corretamente, ela tenderá a excluir variáveis sem importância de maneira mais observável do que a LASSO, mas possuirá tendência maior de excluí-las do que a regressão *Ridge*.

2.5.5 Significância das Variáveis em Regressões Penalizadas

Um passo seguinte à seleção de variáveis utilizando modelos é a verificação da significância delas. O método mais utilizado para isso em modelos lineares de MQO é a verificação através do p-valor. Todavia, esse método possui algumas hipóteses anteriores:

1. O modelo linear está correto;
2. A variável dependente é homocedástica, isso é, possui variância constante;
3. Os resíduos possuem distribuição normal ou conhecida;
4. O vetor de parâmetros é esparso;
5. As variáveis independentes possuem fraca colinearidade entre si.

Para as regressões penalizadas, é considerado que algumas dessas hipóteses não são aplicáveis segundo Lockhart et al (2014), no artigo *A Significance Test for the LASSO* onde é proposto um teste de significância próprio para para a Regressão LASSO, entretanto, o modelo proposto carece ainda de validação e implementação numérica.

Alternativas comuns incluem uma pré-seleção de variáveis utilizando as regressões penalizadas e, com as variáveis selecionadas, executar regressão por MQO e aplicar, a partir disso verificar significância dos coeficientes obtidos.

3 Metodologia

3.1 Base de Dados Stataer

Como ponto de partida deste trabalho, foi utilizada a base de dados Stataer, de autoria do laboratório de transporte aéreo do ITA. A base possui informações e séries históricas a respeito de vôos ocorridos, retirando essas informações dos canais:

- AISWEB (fonte de dados da aeronáutica);
- ANAC (agência nacional de aviação civil);
- ANP (agência nacional de petróleo, gás natural e biocombustíveis);
- FIPE (fundação instituto de pesquisas econômicas);
- IPEA (instituto de pesquisa econômica aplicada);
- SINTAC (sistema integrado de aviação civil).

Desse modo, a base conta com dados acerca dos vôos que incluem fatores demográficos das cidades de origem e destino (IDH, População, Índice de Gini), fatores operacionais (idade das aeronaves, *load factor*, quantidade de vôos no ano, índice de concentração de demanda no aeroporto e na rota), fatores macroeconômicos (câmbio, preço do petróleo, taxa de juros).

Portanto, tem-se, pela base de dados Stataer, uma quantidade razoável de dados de diferentes tipos. A base seria um modo de descobrir, portanto, se algumas concepções razoáveis acerca dos preços das passagens estariam corretas, como, por exemplo, seria comum se esperar que cidades com maiores IDHs tenham bilhetes aéreos mais caros devido à estratégia de precificação das companhias.

Ainda, a base de dados possui variáveis numéricas e informações em forma de *string*, para execução das regressões neste trabalho, algumas variáveis de texto foram convertidas em *dummies* para viabilizar a execução do algoritmo.

Ao todo, a base possui mais de 6000 variáveis, podendo ser baixada no link <www.nectar.ita.br/stataer_19ed> e seu manual, descrevendo os dados que a compõe, pode ser visualizado em <www.nectar.ita.br/stataer_19ed_manual>, pertencente ao laboratório de transportes do Instituto Tecnológico de Aeronáutica (ITA, São José dos Campos – SP).

3.2 Linguagem R

Para confecção deste trabalho, foi utilizada a linguagem R, em sua versão 3.6.1, disponibilizada em 05 de julho de 2019, e, como IDE (Ambiente de Desenvolvimento) foi utilizado o Rstudio, versão 1.1.456, versão estudantil.

A linguagem R, foi desenvolvida em 1993 como uma linguagem *open source*, isso é, de contribuição pública e, atualmente é uma das preferidas para análises estatísticas, aplicações de aprendizado de máquina e usos acadêmicos de maneira geral.

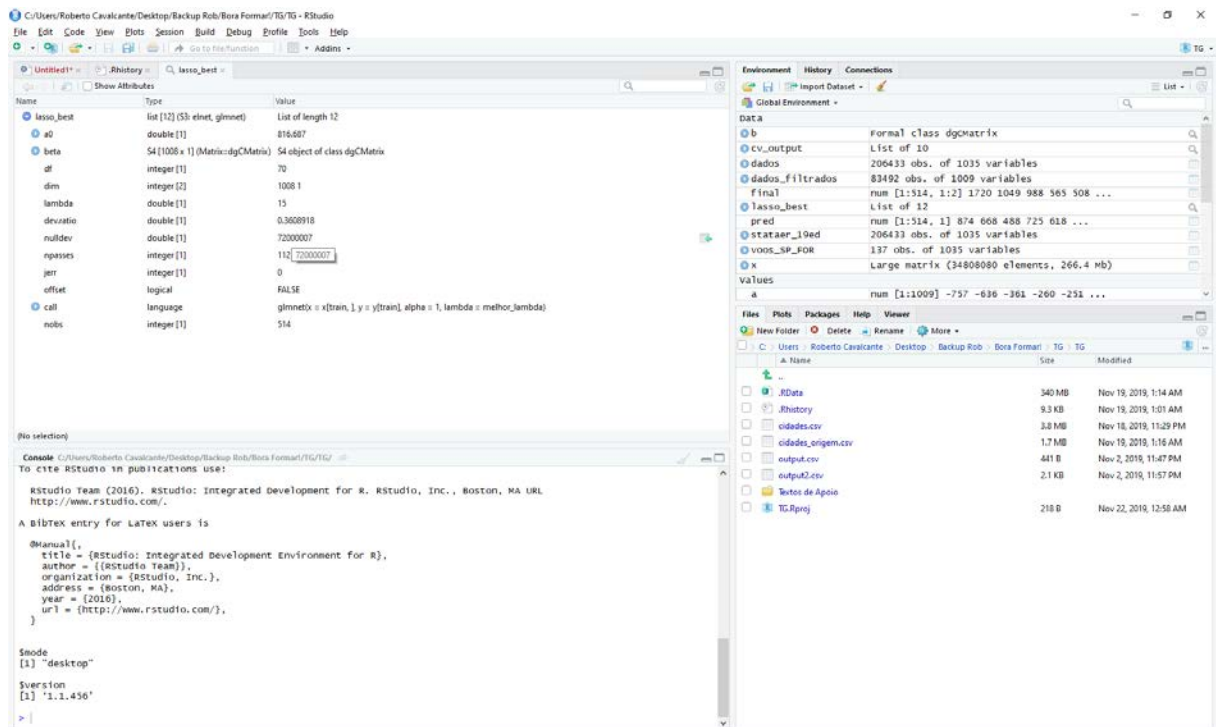


Figura 3.2-1 Ambiente de Desenvolvimento do Rstudio, versão 1.1.456

O atual time de desenvolvimento da linguagem é composto por pessoas de todo planeta, entretanto, o seu foco em análise estatística faz com que sejam desenvolvidas, de maneira independente, várias bibliotecas que tornam seu uso no meio acadêmico mais proeminente.

3.2.1 Biblioteca *glmnet*

Para processamento dos dados, utilizou-se a biblioteca *glmnet* (*generalized linear models Elastic Net*). A *glmnet* foi desenvolvida e divulgada pelos precursores da regressão LASSO, Jerome Friedman, Trevor Hastie e Rob Tibshirani (2010), que desenvolveram os algoritmos

baseados em descida coordenada (*coordinate descent*) para otimização da execução, sob a conclusão de que este método se mostra mais eficiente do que outros tradicionais, como, por exemplo, Netwon-Raphson, ao se aproveitar de esparsidade da matriz de variáveis independentes.

Desse modo, a função executa, dentre outras regressões regularizadas, a *Elastic-Net*, que pode ser calibrada para assumir a forma de LASSO ou Ridge, através do parâmetro “alpha”, que é o mesmo α da Equação 5.

3.3 Tratamento da base de Dados

A variável dependente foi a variável “p” da base de dados Stataer_19ed, que significa a média de preços mensal para a rota, já ajustada pelo IPCA (Índice de Preços do Consumidor Amplo). Os passos seguintes foram tomados para se ter modelos interpretáveis e estudar os efeitos desejados (cobrança de bagagem, privatização, concentração):

- 1) Remoção de observações que não possuíam a variável “p”
- 2) Remoção de variáveis não numéricas, pois o pacote *glmnet* as transformaria em variáveis descritivas, gerando variáveis redundantes e intensificando o problema de excesso de variáveis;
- 3) Delimitação da base de dados aos vôos das 10 rotas mais movimentadas, devido à poluição causada por *dummies* indicativas de rota;
- 4) Remoção das variáveis indicativas de aeronave, devido à poluição causada por *dummies* indicativas de aeronave.

A necessidade de remoção dessas *dummies* foi identificada ao longo do estudo do problema e sua justificativa se encontra na seção de resultados, ao final de tudo, o número de variáveis total se reduziu a 956, número ainda significativo.

3.4 Obtenção do modelo de regressões regularizadas

Foi utilizado o procedimento descrito no fluxograma da Figura 3.4-1. Com este procedimento, é possível avaliar a “importância” da variável de interesse de acordo com modelo, verificar quais variáveis foram escolhidas antes e depois dela e, com o coeficiente associado a esse parâmetro, interpretar a relação entre a variável dependente e a variável de interesse.

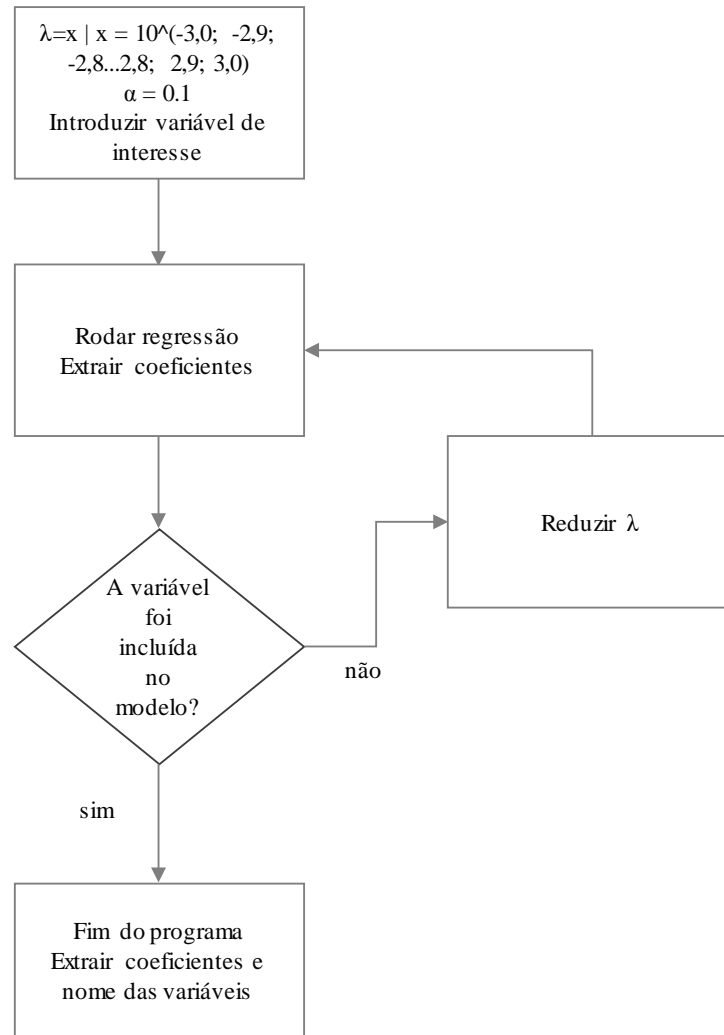


Figura 3.4-1 Fluxograma para obtenção dos coeficientes das variáveis de interesse.

3.5 Verificação de performance das regressões regularizadas

Para avaliação da significância das variáveis relacionadas à esta seção, definiu-se uma regressão de referência, com base nas recomendações de Oliveira, Vassallo e Bettini (2011), que realizaram o estudo com base nas variáveis de câmbio, percentual de passageiros em conexão, número de companhias aéreas, dentre outros fatores. Entretanto, por conta da quantidade diferenciada de aeroportos considerados, levou-se em conta também a distância, ICMS, proporção de passageiros viajando por programas de fidelidade e preço das passagens de ônibus da rota. Verifica-se, na Figura 3.5-1 que todas as variáveis da regressão padrão possuem alta significância para o problema.

Equação 6: Modelo de referência tal que X_i são as variáveis de interesse para cada estudo.

$$Y_{preço} = \beta_0 + \beta_1 \cdot \text{Câmbio} + \beta_2 \cdot \text{Distância} + \beta_3 \cdot \text{n}^\circ \text{ de cias. na rota} + \beta_4 \cdot \text{ICMS}(\%) + \beta_5 \text{Pax fidelidade}(\%) + \beta_6 \text{Preço}_{ônibus} + \sum \beta_i \cdot X_i$$

```

Residuals:
    Min       1Q   Median       3Q      Max
-401.50  -80.75  -12.75   63.25  928.45

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  225.073448  11.965130  18.811 < 2e-16 ***
nUSD_d       32.839191   3.217298  10.207 < 2e-16 ***
avkm         0.125090   0.001458  85.820 < 2e-16 ***
icms_d      214.999437  18.778867  11.449 < 2e-16 ***
ffp         -62.357085  13.263596  -4.701 2.61e-06 ***
nairlines   -45.360420   0.991840 -45.734 < 2e-16 ***
p_bus       0.090466    0.009381   9.643 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.4 on 12591 degrees of freedom
Multiple R-squared:  0.4514,    Adjusted R-squared:  0.4512
F-statistic: 1727 on 6 and 12591 DF,  p-value: < 2.2e-16

```

Figura 3.5-1 Resultados da regressão de referência (*output* não reflete todas as variáveis)

O processo de validar os resultados das regressões regularizadas, em suma, é inserir as variáveis selecionadas no modelo de referência e verificar se há uma melhora do ajuste e se as variáveis inseridas são significantes para o problema.

4 Resultados e Análise

4.1 Tratamento dos dados e Análise Preliminar

Primeiramente, dado que o foco deste trabalho é o estudo da precificação das passagens, foram retiradas as observações em que não havia preço. Desse modo, observações onde o preço da passagem era nulo ou aproximado, foram descartadas.

4.1.1 Normalização dos dados

Devido à formulação da regressão, variáveis que são de maiores ordens de grandeza superior, como, por exemplo a distância da rota, seria penalizada de maneira desigual com relação a variáveis de dimensão inferior, como *dummies*.

Desse modo, a fim de se captar a real influência das variáveis sob o preço e não descartar fatores potencialmente explicativos apenas por sua escala, as variáveis foram normalizadas com a função *scale*, nativa da linguagem R, que possui a propriedade de escalar os valores para que tenham média 0 e desvio padrão igual a 1.

Isso é de suma importância para verificação da validade da análise, conforme pode ser visto em Figura 4.1-1 e Figura 4.1-2, nas quais as anotações representam as 10 variáveis que requeriram um aumento maior do lambda para serem zeradas. Percebe-se, também, na Figura 4.1-2, que todas as variáveis escolhidas foram do tipo *dummy* ou ou câmbio, ou seja, sendo variáveis com menores escalas.

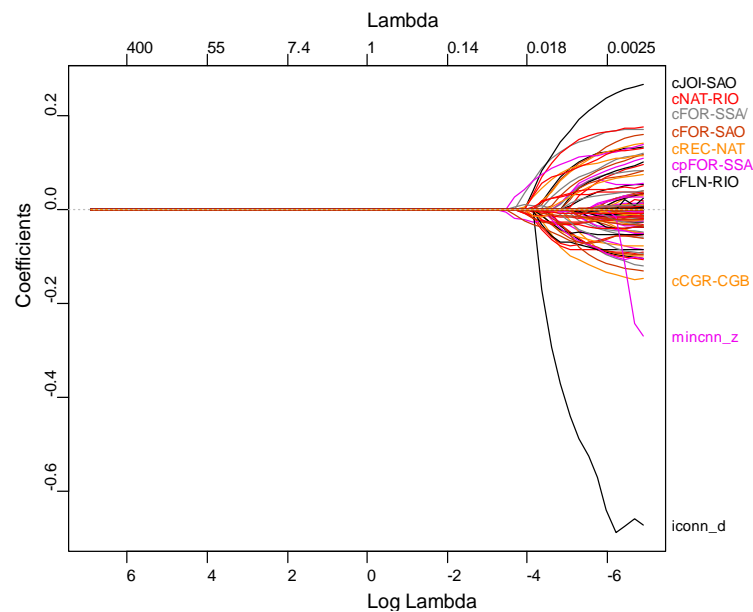


Figura 4.1-1 Encolhimento dos coeficientes para variáveis normalizadas.

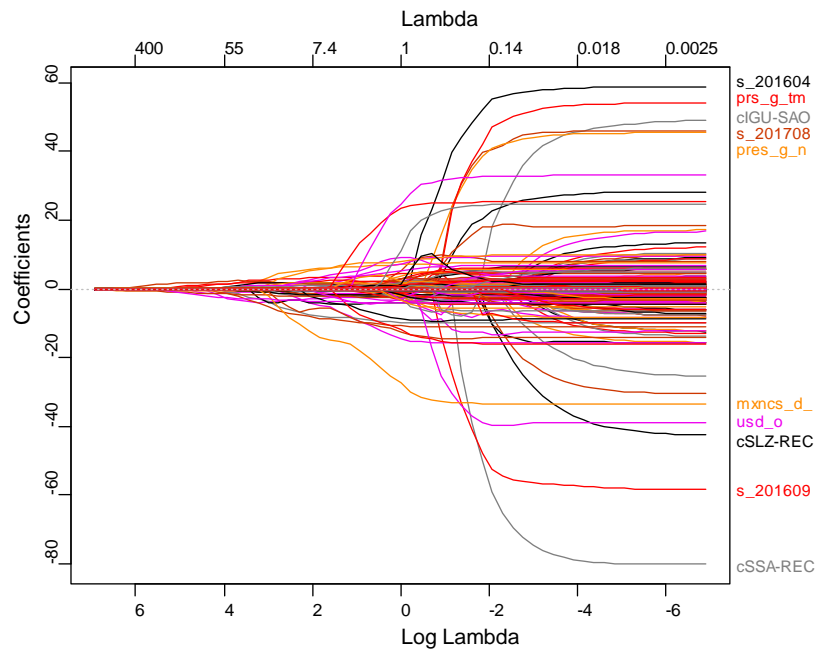


Figura 4.1-2 Encolhimento dos coeficientes para variáveis não-normalizadas.

4.1.2 Seleção das cidades

Uma regressão preliminar foi executada, com a variável dependente sendo o ticket médio das passagens vendidas por rota num determinado mês. Os coeficientes obtidos, com o parâmetro que minimiza a variância do modelo, podem ser visualizados na Tabela 4.1-1.

Tabela 4.1-1 Coeficientes Obtidos (variáveis normalizadas, $\lambda=0.02$)

Coeficiente	Valor	Descrição
(Intercept)	2,55E-01	coeficiente linear
cpCWB-SAO	2,03E-03	Rota Curitiba-São Paulo
cSAO-REC	2,02E-02	Rota São Paulo-Recife
coFOR	1,99E-02	Cidade de Origem Fortaleza
cdFOR	2,26E-02	Cidade de Destino Fortaleza
ffp	3,08E-02	Proporção de Passageiros por Programas de milhas
pres_g_one_o	-2,28E-03	Presença da Avianca na cidade de origem
pres_g_pam_o	-4,64E-03	Presença da MAP na cidade de origem
pres_g_ptb_o	-4,67E-03	Presença da Passaredo na cidade de origem
agemod	-6,55E-04	Idade da Aeronave
BSB_pos	-7,90E-03	Tempo de Turnaround Cidade de Brasília

Desse modo, a regressão trouxe resultados que podem ser interpretados como:

- A relação positiva entre o preço e o aeroporto de Fortaleza ou São Paulo;
- A presença de mais passageiros viajando por programa de milhas pode encarecer as passagens como um todo;
- Indícios da presença da Avianca na rota provocar uma redução de preços;
- É razoável pensar que vôos em aeronaves mais velhas podem significar preços menores.

Entretanto, pode ser visto nos resultados das Tabela 4.1-1, Figura 4.1-1, Figura 4.1-2 o excesso de variáveis relacionados apenas às cidades e rotas (coFOR, CWB-SAO, por exemplo). Esses coeficientes podem sim ter alguma relação com a precificação da passagem, todavia, por si só, eles não adicionam à análise, servindo apenas de indicativos para uma análise mais profunda (turismo, feriados e afins).

Desse modo, foram selecionadas rotas específicas para nelas serem apuradas as variáveis de interesse. As rotas se encontram na Tabela 4.1-2 e, segundo os dados da ANAC, representam as rotas mais movimentadas.

Tabela 4.1-2 Dez rotas com maior movimentação de passageiros em 2018 (ANAC, 2018).

Rota	Passageiros Pagos Transportados 2017*	Passageiros Pagos Transportados 2018*
Rio De Janeiro - Santos Dumont / São Paulo - Congonhas	4.093.857	4.225.692
Porto Alegre / São Paulo - Guarulhos	1.888.322	2.212.481
Brasília / São Paulo - Congonhas	2.042.799	2.085.926
Recife / São Paulo - Guarulhos	1.806.149	2.014.023
Salvador / São Paulo - Guarulhos	1.794.490	1.998.738
Porto Alegre / São Paulo - Congonhas	1.737.474	1.813.608
Belo Horizonte - Confins / São Paulo - Congonhas	1.810.854	1.774.711
Fortaleza / São Paulo - Guarulhos	1.496.762	1.750.976
Belo Horizonte - Confins / São Paulo - Guarulhos	1.311.234	1.651.004
Curitiba / São Paulo - Guarulhos	1.544.731	1.633.026
Total	19.526.672	21.160.185

4.1.3 Tuning dos Parâmetros

A fim de conciliar as propriedades das penalizações LASSO e Ridge, tem-se que regular os parâmetros λ e α , de modo a reduzir a variância do modelo linear gerado. Os parâmetros foram regularizados com a função *cv.glmnet*, que faz o teste de *cross-validation* para os dados de amostra e retorna o λ que minimiza a variância calculada.

A função *cv.glmnet* necessita de um subconjunto aleatório da amostra, executa as regressões lineares para o um vetor que contém os valores de λ a serem testados e retorna, dentre outros valores, o valor do parâmetro que reduz a variância do modelo.

Para determinar o valor de α , foram executadas algumas regressões para verificar o comportamento do modelo frente aos dados fornecidos. Conforme pode ser verificado na Figura 4.1-5 , as variáveis independentes que foram selecionadas foram mantidas em sua maioria.

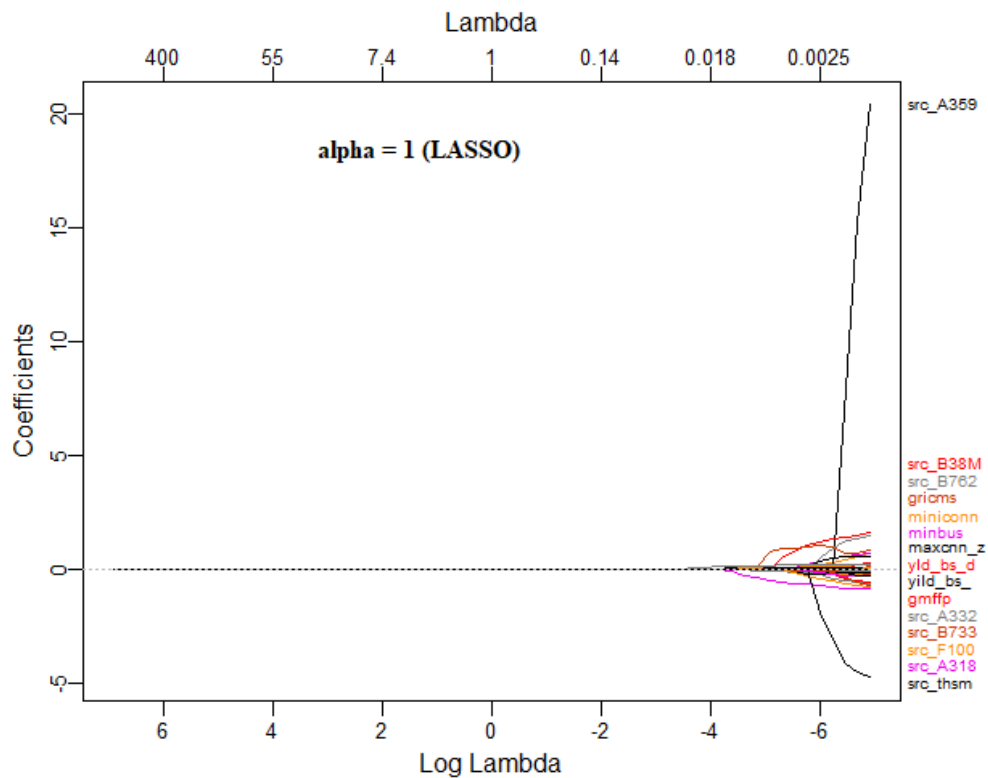


Figura 4.1-3 Encolhimento dos coeficientes para $\alpha = 1$ (variáveis descritas são as 10 últimas a serem mantidas no modelo com o aumento de λ).

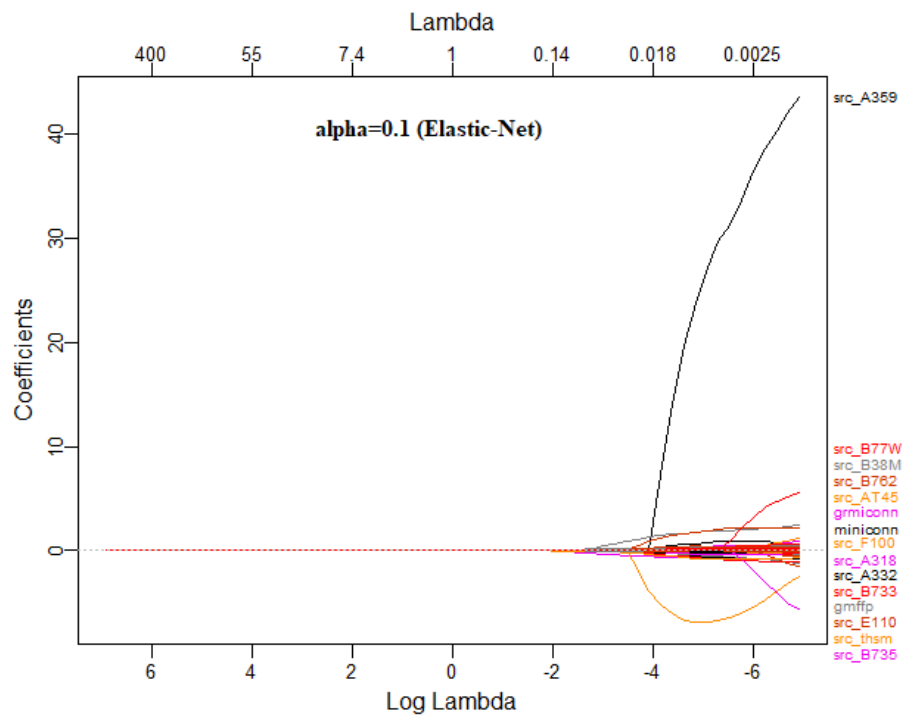


Figura 4.1-4 Encolhimento dos coeficientes para $\alpha = 0.1$ (variáveis descritas são as 10 últimas a serem mantidas no modelo com o aumento de λ).

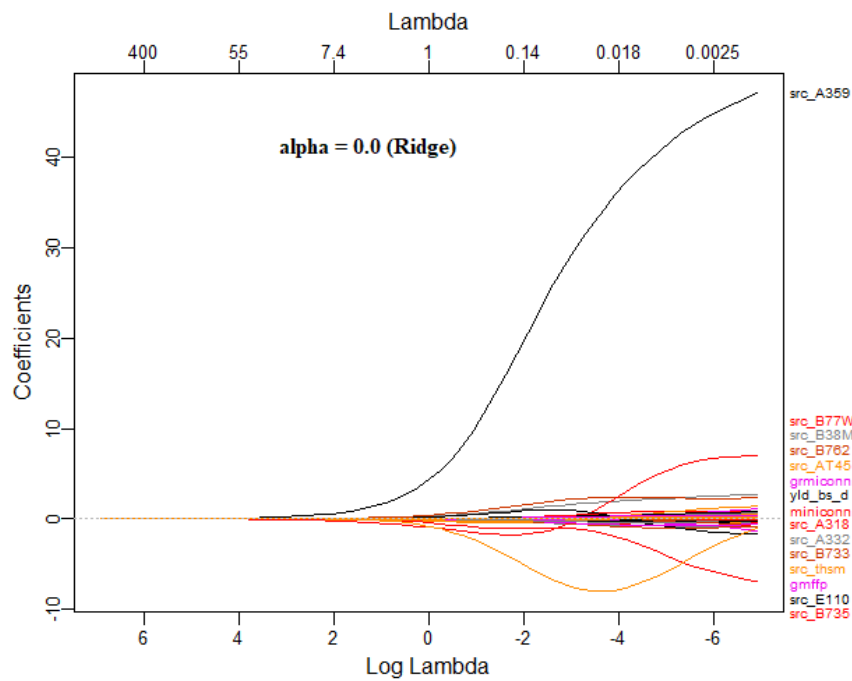


Figura 4.1-5 Encolhimento dos coeficientes para $\alpha = 0$ (variáveis descritas são as 10 últimas a serem mantidas no modelo com o aumento de λ , de maneira assintótica).

Tabela 4.1-3 Retenção das variáveis selecionadas com a variação do parâmetro α .

Variável	Descrição	alpha = 1	alpha = 0.1	alpha = 0
src_A359	Aeronave Utilizada é A359	presente	presente	presente
src_B38M	Aeronave Utilizada é B_38M	presente	presente	presente
src_B762M	Aeronave Utilizada é B762M	presente	presente	-
gricms	Proxy para icms cobrado sobre gasolina de aviação nos estados da rota	presente	-	-
miniconn	Proporção de passageiros em conexão internacional	presente	-	presente
minbus	Proporção de passageiros que já viajaram naquela rota de ônibus	presente	-	-
maxcnm_z	Proporção de passageiros em conexão doméstica e internacional	presente	-	-
yld_bs_d	yield da operação de ônibus	presente	-	presente
gmffp	Média geométrica de passageiros utilizando programas de fidelidade	presente	presente	presente
src_A332	Aeronave Utilizada é A332	presente	presente	presente
src_B733	Aeronave Utilizada é B733	presente	presente	-
src_F100	Aeronave Utilizada é F100	presente	presente	-
src_A318	Aeronave Utilizada é A318	presente	presente	presente
src_thsm	Proporção de vôos com aeronaves pequenas na rota	presente	presente	presente
src_B77W	Aeronave Utilizada é B77W	-	presente	presente
src_AT45	Aeronave Utilizada é ATR-45	-	presente	presente
grminconn	Proporção de passageiros em conexão (modelo gravitacional)	-	presente	-
src_E110	Aeronave Utilizada é E110	-	presente	presente
src_B735	Aeronave Utilizada é B735	-	presente	presente
src_B762	Aeronave Utilizada é B762M	-	-	presente

Diante do exposto, podem ser feitas algumas inferências sobre os dados e o comportamento do modelo com a variação do α :

- A redução do parâmetro faz com que os coeficientes sejam encolhidos de maneira mais suavizada;
- A alteração do parâmetro faz com que sejam escolhidas variáveis diferentes, mas bastante correlacionadas, como *grminconn* e *minconn*. Conforme previsto na fundamentação teórica. A regressão LASSO escolheu variáveis correlacionadas, a regressão Elastic-Net escolheu uma delas e a Ridge também não mostrou variáveis correlacionadas.
- A aeronave utilizada mostrou ser bastante relevante na precificação da passagem;
- O fato da rota poder ter o ônibus como uma alternativa se mostrou relevante e impacta de maneira a reduzir o preço da passagem;
- O a alíquota do ICMS sobre a querosene de aviação nas cidades revela uma correlação positiva com a precificação. O que está em consonância com o dado que 30% dos custos operacionais das companhias aéreas eram oriundos de combustível e lubrificantes.

Portanto, a regressão utilizada será a Elastic-Net, dado que ela supre as deficiências da regressão Ridge e da LASSO. Ao contrário da penalização L1, ela tende a excluir variáveis correlacionadas e, ao contrário da penalização L2, ela tem a capacidade de zerar os coeficientes.

4.2 Análise Competitiva das Companhias Aéreas

Para análise da estratégia competitiva das companhias, foram retiradas da base de dados as *dummies* relacionadas às aeronaves pois, conforme o output da Tabela 4.1-3, as *dummies* de aeronaves acabam por poluir a análise e não nos permitem entender qual fator das aeronaves interfere realmente na precificação (tamanho, assentos ofertados, consumo).

4.2.1 Efeito da Concentração de Mercado

É hipótese comum da economia ortodoxa, o fato de que monopólios ou oligopólios tendem a elevar preços de bens e serviços. E, de fato, é uma discussão bastante corrente no Brasil por conta do reduzido número de companhias operando que operam e suas grandes fatias no mercado doméstico, conforme explicitado na Figura 2.2-1.

Em sequência, foi executada uma regressão Elastic-Net, 106 das 956 variáveis foram selecionadas e três dessas estão relacionadas à concentração de oferta. A variável *maxhhi*, que

significa o máximo valor do índice de Herfindal-Hirschman entre as cidades da rota, com um coeficiente associado de $8,19e+02$, a variável *tri* que é uma dummy que indica tripólio na rota desejada, com um coeficiente associado $7,82e+00$ e a variável *qua*, variável dummy que indica quadripólio na rota, com coeficiente associado de $-3,10e+01$. Desse modo, há a indicação de que maiores concentrações de oferta estão positivamente correlacionadas com o preço, com exceção da oferta por quatro companhias, que tem um efeito redutor no preço.

```

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  38.816150  14.345037   2.706  0.00682 **
nUSD_d       30.697237   3.163329   9.704 < 2e-16 ***
avkm         0.134565   0.001533  87.783 < 2e-16 ***
icms_d      161.889066  18.798879   8.612 < 2e-16 ***
ffp         -38.720542  13.092714  -2.957  0.00311 **
p_bus        0.075832   0.009269   8.181 3.08e-16 ***
maxhhi      101.901089   6.232390  16.350 < 2e-16 ***
mon          89.030390   8.677644  10.260 < 2e-16 ***
duo          39.342582   8.534827   4.610 4.07e-06 ***
tri           9.242816   8.458128   1.093  0.27451
qua        -12.200632   8.655187  -1.410  0.15867
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 121.3 on 12587 degrees of freedom
Multiple R-squared:  0.4704,    Adjusted R-squared:  0.47
F-statistic: 1118 on 10 and 12587 DF,  p-value: < 2.2e-16

```

Figura 4.2-1 Resultado da regressão após inserção de variáveis de concentração.

Conforme os resultados explícitos na Figura 4.2-1, verifica-se a significância do HHI no preço da passagem e sua relação positiva, o mesmo pode ser dito para as variáveis de monopólio e duopólio. Contudo as variáveis de tripólio e quadripólio não representam grande significância, contudo, deve-se atentar para o fato de que o quadripólio implica uma redução do preço da passagem.

Percebe-se também, que a inserção destas variáveis no modelo linear acarretou um maior valor de R-quadrado, indicando uma melhor adequação do modelo aos dados.

Nesse caso, as regressões regularizadas se mostraram uma ferramenta útil ao tentar definir variáveis de interesse, pois conseguiram revelar coeficientes e seus comportamentos, filtrando-as dentre o décimo superior das variáveis.

4.2.2 Posicionamento Estratégico das Companhias Aéreas

Primeiro, para testar a eficácia da regularização da regressão em determinar fatores relevantes para o modelo, foi testado alterar a variável λ , até que aparecessem as variáveis de presença das companhias aéreas no modelo.

Desse modo, para $\alpha = 0.1$, conforme definido anteriormente neste trabalho, os parâmetros λ para os quais as variáveis de presença de linhas aéreas estão na Tabela 4.2-1.

Tabela 4.2-1 Parâmetro λ para inclusão das variáveis de presença no modelo regularizado

λ	Variável	Coefficiente
7,66	pres_j_azul	21,2
9,58	pres_j_glo	-56,2
2,59	pres_j_one	-11,2
9,51	pres_j_tam	-46,3
3,71	pres_j_pam	-125,2

Por consequência, poderíamos supor que:

- A variável de presença de companhia mais significativa deveria ser, em tese, a gol, pois foi incluída no modelo pela regularização antes das demais;
- A Azul foi a única companhia a ter coeficiente positivo, indicando um encarecimento das passagens da companhia;

Para testar os resultados obtidos do modelo regularizado, as variáveis foram postas, duas a duas e individualmente no modelo de referência. Desse modo, puderam ser obtidos os coeficientes relativos às companhias e os níveis de significância a exemplo da Figura 4.2-2. Também foram testadas as variáveis individualmente nos modelos. Os resultados estão contidos em Tabela 4.2-2 e Tabela 4.2-3.

```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 201.501198  11.836224  17.024 < 2e-16 ***
nUSD_d      29.805789   3.163105   9.423 < 2e-16 ***
avkm        0.135057   0.001467  92.066 < 2e-16 ***
icms_d     200.133481  18.502211  10.817 < 2e-16 ***
nairlines  -20.158068   1.623640 -12.415 < 2e-16 ***
pres_j_azul 11.223320   3.012108   3.726 0.000195 ***
pres_j_tam  -74.778615  3.563284 -20.986 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 121.2 on 12591 degrees of freedom
Multiple R-squared:  0.4707,    Adjusted R-squared:  0.4704
F-statistic: 1866 on 6 and 12591 DF,  p-value: < 2.2e-16
    
```

Figura 4.2-2 Modelo de referência com a adição das dummies de presença da Azul e LATAM.

Tabela 4.2-2 Coeficientes associados às variáveis de presença no modelo de referência.

Coeficientes	Gol	LATAM	Avianca	Azul
Gol	-85,0 -42,2/-90,1	-70,4/-50,4	-85,1/-7,1	
LATAM	-90,1/-42,2	-109,1	-98,2/-35,2	-74/11
Avianca	-50,4/-70,4 -35,2/-98,2		-77	-77,5/-1,1
Azul	-7,1/-85,1	11/-74	-1,1/-77,5	33

Tabela 4.2-3 Níveis de significância Associados às Variáveis de presença no modelo de referência (* =99%, **=99.9%, ***=~100%) .

Significâncias	Gol	LATAM	Avianca	Azul
Gol	***	******	******	****
LATAM	******	***	******	******
Avianca	******	******	***	***\
Azul	****	******	***\	***

Desse modo, podem ser tiradas as seguintes conclusões:

- Parece que a Azul é, dentre todas, a companhia mais cara, pois possui os maiores coeficientes;
- Comparando-se os resultados das regressões regularizadas e não regularizadas, verifica-se que as duas divergem quanto à significância das variáveis, dado que, pelo modelo Elastic-Net, a variável de presença da Azul foi incluída antes da Avianca, entretanto, nos testes de significância, a Avianca se mostrou mais significativa.
- Neste caso, a regressão penalizada mostrou um comportamento relativamente pior quando se utiliza dela como uma ferramenta para seleção de variáveis.

Contudo, segundo Das Chagas (2018), a Azul possui o modelo de precificação mais complexo e a maior variância nos preços dentre as quatro grandes companhias aéreas, o que deve ser levado em conta para justificar que, na média, seus preços são mais elevados, mas é possível que sejam observados trechos onde a Azul não é a mais cara, devido ao

4.2.3 Efeito da cobrança de bagagem sobre os preços

Do mesmo modo, buscou-se o λ ($\alpha = 0,1$), para o qual a variável *dbagcharge* fosse inclusa no modelo e, dentre uma lista de 956 variáveis, a *dummy* que indica o começo da cobrança da tarifa de bagagem foi a 744ª variável inclusa, com $\lambda=1,2e-03$ e com um coeficiente associado de -1,63. O que já seria um indício de uma pouca relevância dessa variável para o modelo. Assim como nos outros casos, ela foi inclusa no modelo linear de referência para verificação da significância.

```
Residuals:
    Min       1Q   Median       3Q      Max
-410.36  -81.50  -13.08   63.11  912.43

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  222.885207   11.996494   18.58  <2e-16 ***
nUSD_d       33.031285    3.284892   10.06  <2e-16 ***
avkm         0.127468    0.001427   89.33  <2e-16 ***
icms_d      215.906843   18.839290   11.46  <2e-16 ***
nairlines   -43.936274    0.977283  -44.96  <2e-16 ***
dbagcharge   -1.578838    2.289673   -0.69    0.49
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 124 on 12592 degrees of freedom
Multiple R-squared:  0.4464,    Adjusted R-squared:  0.4462
F-statistic: 2031 on 5 and 12592 DF,  p-value: < 2.2e-16
```

Figura 4.2-3 Modelo de referência com *dummy* de cobrança de bagagem inclusa

E, conforme o resultado posto na Figura 4.2-3, verificou-se que *dbagcharge* não possui uma significância elevada, e, ao contrário do que se esperava, haveria indícios de que a cobrança de bagagem teria diminuído o preço. Contudo, devido a estratégias de gerenciamento de receita, haveria a possibilidade de que o passageiro prefira aderir a uma tarifa mais cara do que comprar uma tarifa econômica e pagar o adicional de bagagem. Desse modo, foi testada a variável *bagppax*, que significa quantos quilogramas de bagagem foram levados, em média por passageiro. Resultando, desse modo, no *output* da Figura 4.2-4.

```

Residuals:
    Min       1Q   Median       3Q      Max
-536.99  -73.48  -11.55   58.52  898.98

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.611035  13.574417   1.076   0.282
nUSD_d       41.340178   3.189123  12.963 <2e-16 ***
avkm         0.092811   0.001811  51.250 <2e-16 ***
icms_d      195.659578  18.231701  10.732 <2e-16 ***
nairlines   -40.594923   0.951836 -42.649 <2e-16 ***
bagppax      18.512937   0.626467  29.551 <2e-16 ***
dbagcharge   38.803482   2.601983  14.913 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 119.9 on 12591 degrees of freedom
Multiple R-squared:  0.4823,    Adjusted R-squared:  0.4821
F-statistic: 1955 on 6 and 12591 DF,  p-value: < 2.2e-16

```

Figura 4.2-4 Modelo de Referência com *dummy* de cobrança e bagagem média por passageiro.

Desse modo, o modelo indicaria que:

- Passageiros que levam maior quantidade de bagagem optam por tarifas mais caras, com bagagem inclusa;
- A cobrança da tarifa estimula os passageiros a levarem menos bagagem;
- O modelo regularizado não foi capaz de captar a significância da *dummy dbagcharge* associada com a variável *bagppax*.

Contudo, a fim de se testar os efeitos imediatos da liberação da bagagem, os dados foram filtrados por ano entre 2013 e 2017 e a regressão foi executada novamente.

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	567.682387	5.382766	105.463	< 2e-16	***
nusd_d	-62.164206	1.463651	-42.472	< 2e-16	***
avkm	0.169493	0.001151	147.297	< 2e-16	***
icms_d	209.570803	14.403924	14.550	< 2e-16	***
nairlines	-66.513543	0.829823	-80.154	< 2e-16	***
ffp	-93.200694	6.635822	-14.045	< 2e-16	***
p_bus	0.160535	0.007825	20.516	< 2e-16	***
dbagcharge	-7.026219	2.172329	-3.234	0.00122	**

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figura 4.2-5 Modelo com *dummy* de cobrança de bagagem para os anos de 2013 a 2017

O que é um indicativo de que, realmente, após um primeiro momento, a cobrança de bagagem teve efeito de “desacoplar” o preço do ticket e da bagagem, contudo, por outros motivos, as empresas aéreas tiveram de elevar os preços (alta do dólar, alta do preço do petróleo).

4.2.4 Efeito da privatização de aeroportos no preço das passagens

Segundo McKinsey & Company (2010), o efeito da privatização sobre o preço das passagens foi contrário a o que acontece nos mercados regulares. Devido à característica monopolística do setor de transporte aéreo aeroportos privatizados tenderam a ter um aumento no custo final para o consumidor, pois, segundo a companhia, as empresas envolvidas na privatização colocariam o retorno para o acionista, a geração de caixa e retorno de longo prazo em detrimento do custo para o consumidor final. Todavia, no Brasil, pôde ser percebido uma melhora na avaliação da experiência do passageiro dentro do aeroporto, o que poderia ser um reflexo dos níveis de serviço impostos pelo governo aos contratos de privatização.

Contudo, deve-se atentar que, em 2019, não houve privatizações de aeroportos, mas concessões. Essas diferem da privatização no aspecto de que o tempo de controle da companhia aérea sobre o aeroporto tem um prazo estabelecido e há margem para que o impacto delas seja nos preços seja diferente das privatizações.

Desse modo, já havendo uma expectativa sobre a significância da privatização para os preços, foi executada a regressão Elastic-Net, variando-se λ , até que uma variável de privatização fosse incluída no modelo. E para λ igual a 3.94, a variável *dpriv* (*dummy* de aeroporto privatizado na cidade de rota) foi incluída ao modelo, dentre 78 de 956 variáveis, com um coeficiente associado de 20,4.

Seguindo o mesmo procedimento, ela foi adicionada ao modelo linear de referência segundo e os resultados podem ser vistos na Figura 4.2-6. O que indica que a variável é bastante significativa para o problema e tende a aumentar o preço da passagem.

```

                Estimate Std. Error t value Pr(>|t|)
(Intercept) 196.94110    12.09809   16.279  <2e-16 ***
nUSD_d       32.31619     3.21361   10.056  <2e-16 ***
avkm         0.12803     0.00142   90.191  <2e-16 ***
icms_d      181.25966    18.95639    9.562  <2e-16 ***
nairlines   -45.91594     0.98580  -46.578  <2e-16 ***
dpriv       42.30141     3.54012   11.949  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 123.3 on 12592 degrees of freedom
Multiple R-squared:  0.4526,    Adjusted R-squared:  0.4524
F-statistic: 2082 on 5 and 12592 DF,  p-value: < 2.2e-16

```

Figura 4.2-6 Dummy de aeroporto privatizado na rota.

5 Conclusão

5.1 Performance das regressões penalizadas

Diante do exposto, percebe-se que as regressões penalizadas são uma ferramenta valiosa na seleção de variáveis para composição de um modelo. Todavia, é necessário, ainda, um julgamento crítico sobre os resultados.

Os modos de penalização diferentes apresentaram comportamentos semelhantes, podendo ser recomendada a utilização para o problema de seleção. Contudo, pelo fato da regressão Ridge não zerar os coeficientes, ela se torna menos eficiente na identificação em si.

Como era previsto por Tibshirani (1996), a regressão Ridge se mostrou mais eficiente ao tratar do problema de multicolinearidade dentro da amostra, enquanto a regressão LASSO se mostrou mais eficiente

Todavia, no meio acadêmico, há várias críticas ao modelo de regressão LASSO e são apresentadas várias alternativas ao método, como, a regressão LARS (*Least Angle Regression*), baseada em reduzir o produto vetorial dos vetores de coeficiente e de variáveis, a regressão SLOPE, que possui um parâmetro λ para cada variável, penalizando, portanto as variáveis de maneiras diferentes. E é sugerido que este trabalho seja continuado testando a adequação desses métodos à base de dados Stataer, que é riquíssima.

5.2 Fatores de influência

O estudo conseguiu comprovar, novamente, a influência de vários fatores acerca da precificação e verificou uma variável recente, que é a cobrança de bagagem.

Uma limitação das análises realizadas foi o fato de se trabalhar com a média dos preços na rota, não podendo ser captada com precisão a variabilidade proveniente das políticas de *revenue management*.

Contudo, corroboram-se as conclusões de Das Chagas (2018), com relação à estratégia das companhias, que afirma que a Azul, apesar de se denominar uma *low cost*, possui um modelo sofisticado e consegue elevar sua receita por passageiro. Outro fator que foi largamente atestado pelas simulações foi a idade da frota, contudo, é necessário um estudo mais aprofundado sobre essa variável, pois as empresas mais caras (Azul e Avianca) são as que possuem uma frota mais nova, o que não indicaria uma relação de causalidade.

Dados os eventos de 2019 e os resultados deste trabalho, é de se esperar um aumento progressivo nos preços dos vôos domésticos, dada a exclusão de um grande *player*, que era a Avianca, aumentando, desse modo, a concentração de oferta em uma parte significativa das

rotas aéreas. Também é razoável inferir que as concessões irão provocar um aumento nos preços devido ao aumento na taxa aeroportuária.

Desse modo, é sugerido que este estudo seja aprofundado levando-se em conta também dados microscópicos e variáveis relacionadas à compra (timing) a fim de ter um modelo adequado de previsão da precificação.

6 Referências

- ANAC, Agência Nacional de Aviação Comercial. Dados Estatísticos. Disponível em <<https://www.anac.gov.br/assuntos/dados-e-estatisticas/dados-estatisticos/dados-estatisticos>> Acesso em 31 Agosto. 2019.
- BURNHAM, K.P; ANDERSON, D.R. 2002. **Model selection and multimodel inference**. A practical information - theoretic approach. Springer, New York.
- DAS CHAGAS, Luis Victor Camilo: **Modelagem computacional e econômico financeiro da evolução temporal dos preços de passagens aéreas: um estudo sob a perspectiva das companhias aéreas nacionais**. 2018. 116f. Trabalho de Conclusão de Curso. (Graduação em Engenharia Civil-Aeronáutica) – Instituto Tecnológico de Aeronáutica, São José dos Campos.
- FERREIRA, J. C. 2017. Um breve histórico da aviação comercial brasileira. In: XII Congresso Brasileiro de História Econômica. 13ª Conferência Internacional de História de Empresas. Niterói: Associação Brasileira de Pesquisadores em História Econômica.
- FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. **Regularization Paths for Generalized Linear Models via Coordinate Descent**, 2008.
- HOERL, A. E.; KENNARD, R. W. Ridge regression: biased estimation for nonorthogonal problems. **Technometrics**. Washington, DC., v. 12 n.1, 55-67, 1970.
- LOCKHART, R.; TAYLOR, J.; TIBSHIRANI, Ryan J.; TIBSHIRANI, Robert; A significance test for the LASSO. **The Annals of Statistics**, v. 42 n. 2, 413 – 468, 2014.
- MALVEIRA, Davi de Anchieta Sousa da Costa. **Análise das estratégias competitivas das principais empresas aéreas que atuam no Brasil**. 2018. 51f. Trabalho de Conclusão de Curso. (Graduação em Engenharia Civil-Aeronáutica) – Instituto Tecnológico de Aeronáutica, São José dos Campos.
- MCKINSEY & COMPANY. **Estudo do Setor de Transporte Aéreo do Brasil**: Relatório Consolidado. Rio de Janeiro, 2010. 381p.
- MCGILL, J.I.; RYZIN, G. J. V.; Revenue Management: Research overview and prospects. *Transportation science*, **Inform**s, v. 33, n. 2, p. 233-256, 1999.
- PAK, K; PIERSMA, N. **airline revenue management**, 2002.
- TIBSHIRANI, R.; Regression Shrinkage and Selection via the LASSO/**Journal of the Royal Statistical Society – Series B**, v. 58, n. 1, p. 267-288, 1996.

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TC	2. DATA 21 de novembro de 2019	3. REGISTRO N° DCTA/ITA/TC-122/2019	4. N° DE PÁGINAS 51
5. TÍTULO E SUBTÍTULO: Métodos quantitativos aplicados à análise financeira de empresas de transporte aéreo.			
6. AUTOR(ES): Roberto Chaves Cavalcante Filho			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica – ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: 1. Regressão Penalizada 2. Tarifa de Bagagem. 3. Transporte Aéreo			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Tarifas; Bagagem; Operações de linhas aéreas; Transporte aéreo; Transporte.			
10. APRESENTAÇÃO: ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Prof. Dr. Alessandro Vinícius Marques de Oliveira. Publicado em 2019.			
11. RESUMO: Em 2017, após a liberação da tarifa de bagagem, começou a ganhar relevância no mercado de transporte aéreo a discussão sobre o impacto das outras formas de arrecadação das companhias aéreas no custo do bilhete e no custo efetivo para o consumidor. Além disso, em 2019 houve a interrupção das operações e pedido de recuperação judicial da quarta maior companhia em um mercado caracterizado por seu oligopólio e Este trabalho tem como objetivo identificar quais são as variáveis que efetivamente possuem impacto no preço final, contribuir para a literatura acerca do uso de regressões regularizadas em seleção de variáveis para modelos preditivos, além de oferecer subsídio para pesquisas futuras acerca da precificação de passagens e competição.			
12. GRAU DE SIGILO: (X) OSTENSIVO () RESERVADO () SECRETO			