

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas (ICEX)

Geovane Rainoni Silva

**Modelo de estimação do valor residual de até cinco anos para
veículos com potencial de locação.**

Belo Horizonte

2018

Geovane Rainoni Silva

**Modelo de estimação do valor residual de até cinco anos para
veículos com potencial de locação.**

Trabalho de conclusão de curso, apresentado como pré-requisito de avaliação do curso de especialização em Estatística Aplicada da Universidade Federal do Estado de Minas Gerais, para a obtenção definitiva do título de especialista.

**Professora Orientadora: Ilka Afonso
Reis**

Belo Horizonte

2018

RESUMO

O setor de aluguel de veículos vem apresentando crescimento ao longo dos últimos anos, mesmo em um cenário macroeconômico adverso. Visando manter esse ritmo, é necessário adotar medidas que melhorem a competitividade e minimizem as incertezas futuras. Neste escopo, conseguir mensurar com maior acurácia o valor residual de um veículo (valor de venda após certo tempo de uso) aumenta a possibilidade de rentabilidade de uma companhia de aluguel de veículos de ponta a ponta, ou seja, garantindo a escolha das melhores opções de modelos de veículo para compra, criando competitividade ao aluguel e assegurando rentabilidade no momento da venda. Nesse sentido, este trabalho tem por objetivo estimar o valor residual de um modelo de veículo em até cinco anos, dado seu conjunto de características. Para compreendermos qual a relação existente entre as características de um veículo e sua desvalorização ao longo do tempo, a técnica estatística escolhida para se trabalhar foi a análise de regressão. Foram selecionados 49 modelos de veículos e, para a construção do banco de dados, foram utilizadas seis características dos veículos: Montadora, Motorização, Potência, Segmento, Nº Itens de Segurança e Nº Itens de Conforto. Foram construídos cinco modelos de regressão, um para cada período de tempo analisado. Os modelos de regressão foram capazes de explicar a influência das características de um modelo de veículo em sua desvalorização e também de estimar qual a desvalorização média esperada de um modelo de veículo após determinado período de utilização. A partir disto, espera-se que a tomada de decisão de uma companhia de aluguel de veículos em relação a compra de um determinado modelo de veículo e a definição do preço de aluguel possam ser baseadas em informações mais detalhadas sobre o valor residual do modelo de veículo analisado.

Palavras-chaves: Modelos de Veículos; Valor Residual; Desvalorização; Análise de Regressão; Modelos de Regressão; Aluguel de Veículos.

SUMÁRIO

1. INTRODUÇÃO	7
2. OBJETIVOS.....	9
2.1 Objetivo Geral	9
2.2 Objetivos Específicos	9
3. MATERIAL E MÉTODO.....	10
3.1 Banco de dados	10
3.1.1 Descrição das Variáveis	15
3.2 Análise de Regressão com variáveis explicativas categóricas.....	21
4. RESULTADOS.....	23
4.1 Análise exploratória dos dados	23
4.2 Análise dos modelos de regressão	50
4.2.1 Modelo de regressão para o banco de dados de um ano	50
4.2.2 Modelo de regressão para o banco de dados de dois anos	58
4.2.3 Modelo de regressão para o banco de dados de três anos.....	66
4.2.4 Modelo de regressão para o banco de dados de quatro anos	72
4.2.5 Modelo de regressão para o banco de dados de cinco anos	78
5. CONCLUSÃO	83
6. REFERÊNCIAS BIBLIOGRÁFICAS	86
7. APÊNDICE.....	87

LISTA DE TABELAS

1. Tabela 1 - Distribuição dos veículos conforme período de dados	10
2. Tabela 2 – Modelos de veículos presentes no trabalho, Período de coleta dos dados, Número de versões e Número de dados por modelo	12
3. Tabela 3 - Distribuição das observações conforme banco de dados	13
4. Tabela 4 - Estatísticas descritivas das variáveis resposta do banco um ao cinco.	26
5. Tabela 5 – Tabela de frequência da variável Montadora nos bancos de dados de um a cinco anos de depreciação.....	30
6. Tabela 6 – Tabela de frequência da variável Segmento nos bancos de dados de um a cinco anos de depreciação.....	33
7. Tabela 7 – Tabela de frequência da variável Motorização nos bancos de dados de um a cinco anos de depreciação.....	39
8. Tabela 8 – Tabela de frequência da variável Potência nos bancos de dados de um a cinco anos de depreciação.....	43
9. Tabela 9 – Tabela de frequência da variável N ^o Itens Segurança nos bancos de dados de um a cinco anos de depreciação.....	48
10. Tabela 10 – Tabelas de frequência da variável N ^o Itens Conforto nos bancos de dados de um a cinco anos de depreciação.....	50
11. Tabela 11- de correlação das variáveis N ^o Itens de Segurança, N ^o Itens de conforto nos bancos de dados de um a cinco anos de depreciação.....	52

LISTA DE GRÁFICOS

1. Gráfico 1 – Distribuição de frequências do Valor Residual do Veículo (em proporção) nos bancos de dados de um a cinco anos de depreciação.....	27
2. Gráfico 2 – Valor residual médio dos veículos (em %) ao longo dos cinco anos de uso.....	28
3. Gráfico 3 – Boxplot do Valor Residual dos Veículos (em proporção) segmentados pela variável Montadora nos bancos de dados de um a cinco anos de depreciação.....	31
4. Gráfico 4 – Boxplot do Valor Residual dos Veículos (em proporção) seccionado pela variável Segmento nos bancos de dados de um a cinco anos de depreciação.....	35
5. Gráfico 5 – Boxplot do Valor Residual dos Veículos (em proporção) seccionado pela variável Segmento, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação.....	37
6. Gráfico 6 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Motorização nos bancos de dados de um a cinco anos de depreciação.....	39
7. Gráfico 7 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Motorização, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação.....	42
8. Gráfico 8 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Potência nos bancos de dados de um a cinco anos de depreciação.....	44
9. Gráfico 9 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Potência, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação.....	46

10. Gráfico 10 – Dispersão do Valor Residual dos Veículos (em proporção) segundo a variável N ^o Itens de Segurança nos bancos de dados de um a cinco anos de depreciação.....	49
11. Gráfico 11 – Dispersão do Valor Residual dos Veículos (em proporção) segundo a variável N ^o Itens de Conforto nos bancos de dados de um a cinco anos de depreciação.....	51
12. Gráfico 12 – Análise de Resíduos do modelo 1.2 (banco de dados de um ano).	58
13. Gráfico 13 – Análise de Resíduos do modelo 1.3 (banco de dados de um ano).	60
14. Gráfico 14 – Análise de Resíduos do modelo 2.2 (banco de dados de dois anos).....	67
15. Gráfico 15 – Análise de Resíduos do modelo 2.3 (banco de dados de dois anos).....	69
16. Gráfico 16 – Análise de Resíduos do modelo 3.2 (banco de dados de três anos).....	76
17. Gráfico 17 – Análise de Resíduos do modelo 4.2 (banco de dados de quatro anos).....	82
18. Gráfico 18 – Análise de Resíduos do modelo 5.2 (banco de dados de cinco anos).....	87

LISTA DE QUADROS

1. Quadro 1 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Montadora do Veículo	16
2. Quadro 2 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Segmento.....	17
3. Quadro 3 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Motorização do Veículo.....	18
4. Quadro 4 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Potência.	19
5. Quadro 5 – Modelo de regressão (1.1) da variável resposta VRV.01 (banco de dados de um ano)	54
6. Quadro 6 – Modelo de regressão (1.2) da variável resposta VRV.01 (banco de dados de um ano).	56
7. Quadro 7 – Modelo de regressão (1.3) da variável resposta VRV.01 (banco de dados de um ano).	59
8. Quadro 8 – Modelo de regressão (2.1) da variável resposta VRV.02 (banco de dados de dois anos).....	63
9. Quadro 9 – Modelo de regressão (2.2) da variável resposta VRV.02 (banco de dados de dois anos)	65
10. Quadro 10 – Modelo de regressão (2.3) da variável resposta VRV.02 (banco de dados de dois anos).....	68
11. Quadro 11 – Modelo de regressão (3.1) da variável resposta VRV.03 (banco de dados de três anos).....	73
12. Quadro 12 – Modelo de regressão (3.2) da variável resposta VRV.03 (banco de dados de três anos).....	75
14. Quadro 13 – Modelo de regressão (4.1) da variável resposta VRV.04 (banco de dados de quatro anos).....	79
15. Quadro 14 – Modelo de regressão (4.2) da variável resposta VRV.04 (banco de dados de quatro anos).....	81
16. Quadro 16 – Modelo de regressão (5.1) da variável resposta VRV.05 (banco de dados de cinco anos)	85

17. Quadro 17 – Modelo de regressão (5.2) da variável resposta VRV.05 (banco de dados de cinco anos) 86

1. INTRODUÇÃO

O setor de aluguel de veículos vem se destacando no cenário nacional pelo seu crescimento e relevância. Segundo os dados do anuário brasileiro do setor de locação de veículos de 2016, que apresenta os resultados do setor no ano de 2015 e foi elaborado pela ABLA (Associação Brasileira de Locadoras de Automóveis), no ano de 2015 este setor apresentou um faturamento total de R\$ 16,2 bilhões, gerou um montante de R\$ 5,3 bilhões de impostos aos cofres públicos e encerrou o ano com 413 mil trabalhadores diretos e indiretos. Outro ponto que corrobora com a importância do setor no âmbito nacional foi a aquisição de veículos nos segmentos de automóveis e comerciais leves, sendo o setor de locação de veículos responsável por adquirir 13,66% de todos os novos veículos licenciados nessas duas categorias (338.840 unidades).

Neste contexto, uma das estratégias cruciais para o sucesso de uma empresa no setor de locação de veículos é como a mesma lida com seu principal ativo, os veículos. É necessário renovar a frota a fim de reduzir custos operacionais e criar a vantagem competitiva de dispor com perenidade de veículos novos para o aluguel.

Para que se alcance esse objetivo, é vital saber o momento correto de se vender os veículos usados, com intenção de gerar caixa para possibilitar a aquisição dos novos veículos. Nesta etapa é importante definir qual modelo será comprado e saber mensurar qual será o volume de compra. Uma das variáveis chaves desta definição é o valor residual do veículo após um ano, ou seja, qual será o preço de venda do veículo após um ano de uso.

A estimação do valor residual é o objetivo deste trabalho, que não se limita a estimar essa quantidade apenas pelo período de um ano, mas tem como proposta estimar este preço de venda futuro pelo prazo de cinco anos, haja vista que os valores de venda para dois, três, quatro e cinco anos são insumos na definição de preços dos contratos de aluguel, principalmente nas operações de locações de frotas, cujo prazo dos contratos tende a se encaixar nos períodos citados.

Como exposto, a variável resposta deste trabalho é o valor residual de um veículo. Para que se tenha maior assertividade no cálculo deste preço levando em consideração o prazo de cinco anos, iremos elaborar cinco modelos de regressão distintos, um para cada ano de depreciação, e assim estudar a relação da variável resposta de cada modelo de regressão com as variáveis candidatas a preditoras, como montadora, categoria do modelo do automóvel, conforto, entre outras que serão apresentadas mais detalhadamente ao longo do texto.

Este estudo se justifica na medida em que a estimação do valor residual de um veículo é importante para mitigar riscos em relação à aquisição dos veículos de uma companhia do setor de alugueis, bem como para criar competitividade nos contratos, uma vez que a estimativa de preço futuro do veículo seja mais coerente com o valor que irá se concretizar.

2. OBJETIVOS

2.1 Objetivo Geral

Construir modelos de regressão que sejam capazes de estimar o valor residual de um veículo no período de um, dois, três, quatro ou cinco anos em função de variáveis preditoras relacionadas às características dos veículos.

2.2 Objetivos Específicos

- Entender o comportamento do valor residual de um veículo em função das características desse veículo por meio da análise exploratória dos dados;
- Identificar o valor que cada característica do veículo agrega na formação do preço futuro desse veículo;

3 Material e Método

3.1 - Banco de dados

Como o objetivo deste trabalho é estudar o valor residual dos veículos no período de cinco anos dados suas características, foi necessária a construção de cinco bancos de dados diferentes. Os dados foram coletados compreendendo o ano modelo dos veículos de 2012 a 2016 nos websites CarrosnaWeb, Icarros e Webmotors. Contudo, como o mercado automotivo é muito dinâmico e ocorrem várias reestilizações nos modelos de veículos, não foi possível utilizar o período de 2012 a 2016 para todos os modelos de veículos presentes no banco de dados.

Foram selecionados os principais modelos de veículos de 10 categorias da Federação Nacional da Distribuição de Veículos Automotores (Fenabreve), sendo elas, veículos de Entrada, Hatch Pequeno, Hatch Médio, Sedan Pequeno, Sedan Compacto, Sedan Médio, Sedan Grande, SW (Space Wagon), Monocab e SUV (Sport Utility Vehicle). Dois critérios foram levados em consideração na escolha dos modelos de veículos: suas vendas no mercado de zero quilômetro ao longo do ano de 2017 e ter sido lançado antes de 2017, para que se conseguisse analisar sua depreciação no período mínimo de um ano. Desta forma, foram selecionados 49 modelos de veículos e a Tabela 1 apresenta como ficaram distribuídos em relação ao período de coleta dos dados.

Tabela 1 - Distribuição dos modelos de veículos conforme período de dados

Período	Nº Modelos de Veículos
Dados no período de 2015 a 2016	11
Dados no período de 2014 a 2016	8
Dados no período de 2013 a 2016	14
Dados no período de 2012 a 2016	9
Dados no período de 2012 a 2015	5
Dados no período de 2013 a 2015	1
Dados no período de 2012 a 2014	1
	49

Elaboração Própria

Como mencionado anteriormente, os modelos de veículos sofrem reestilizações, que podem ou não alterar o veículo de forma significativa. Mudanças de menor impacto na indústria automobilística são definidas como *facelift* e podem incluir mudanças em pacotes de acessórios, alterações em elementos de lataria, lanternas ou modificações mecânicas. Contudo, esteticamente o veículo permanece bem próximo à versão anterior e normalmente não ocorrem reposicionamentos significativos em preço. Já em uma reestilização por completo, os modelos de veículos sofrem alterações estéticas relevantes, mudanças na família de motores e, ao contrário do *facelift*, normalmente ocorre um reposicionamento de preço, distanciando de forma substancial a nova versão da anterior.

Os 49 modelos de veículos foram avaliados individualmente em relação às suas reestilizações e a definição do melhor período de coleta dos dados se deu de forma empírica. Em alguns casos, optou-se por trabalhar após uma mudança significativa, como é o caso dos 11 modelos em que o período escolhido foi de 2015 a 2016. Em outros modelos de veículos, o julgamento foi por utilizar os dados antes de uma reestilização completa, situação em que se encontram os modelos selecionados nos períodos de 2013 a 2015 e de 2012 a 2014. Nenhum modelo de veículo foi utilizado antes e após uma reestilização completa: eles foram selecionados ou antes da reestilização ou após.

Foram selecionadas duas versões por modelo de veículo, uma mais básica e a outra intermediária ou completa, para que se conseguisse entender melhor como as características de um veículo, seja em relação à sua motorização, conforto ou segurança impactam na depreciação ao longo dos anos. Contudo, para sete modelos (Civic, C180, Focus Sedan, A200, IX35, Tucson e Tiguan), não foi possível selecionar mais de uma versão, pois, no período selecionado para coleta de dados, existia apenas uma versão ou porque, devido às várias mudanças de versões ao longo do tempo, só foi possível utilizar uma no período selecionado.

A Tabela 2 apresenta todos os modelos de veículos elencados no banco de dados utilizado neste trabalho, o período selecionado para cada um e o número de versões. Como são vários anos para cada veículo, o número de “indivíduos” presentes no banco de dados varia de 2 até 10 por modelo de veículo. Por exemplo,

para o modelo de veículo C180, os dados foram coletados no período de 2015 a 2016 e esse modelo só contou com uma versão. Desta forma, para este modelo de veículo, há apenas dois indivíduos, e como o objetivo deste trabalho é estudar o valor residual dos veículos no período de cinco anos dadas as suas características, foi necessária a construção de cinco bancos de dados diferentes.

Tabela 2 - Modelos de veículos presentes no trabalho, Período de coleta dos dados, Número de versões e Número de dados por modelo

Modelo	Período	Nº Versões	Nº Ano Modelo	Total de "indivíduos"
320i	2015 a 2016	2	2	4
A3 Sedan	2015 a 2016	2	2	4
A4	2015 a 2016	2	2	4
C180	2015 a 2016	1	2	2
City	2015 a 2016	2	2	4
Corolla	2015 a 2016	2	2	4
Fit	2015 a 2016	2	2	4
Golf	2015 a 2016	2	2	4
Ka	2015 a 2016	2	2	4
KA Sedan	2015 a 2016	2	2	4
Space fox	2015 a 2016	2	2	4
208	2014 a 2016	2	3	6
A3	2014 a 2016	2	3	6
Fiesta	2014 a 2016	2	3	6
Focus	2014 a 2016	2	3	6
Focus Sedan	2014 a 2016	1	3	3
HB20S	2014 a 2016	2	3	6
Logan	2014 a 2016	2	3	6
Sentra	2014 a 2016	2	3	6
C3	2013 a 2016	2	4	8
Doblo	2013 a 2016	2	4	8
Ecosport	2013 a 2016	2	4	8
Etios HB	2013 a 2016	2	4	7
Etios Sedan	2013 a 2016	2	4	8
Fusion	2013 a 2016	2	4	8
Gol	2013 a 2016	2	4	8
HB20	2013 a 2016	2	4	8
Onix	2013 a 2016	2	4	8
Prisma	2013 a 2016	2	4	8
Siena	2013 a 2016	2	4	8
Tiguan	2013 a 2016	1	4	4
Voyage	2013 a 2016	2	4	7
Weekend	2013 a 2016	2	4	8
308	2012 a 2016	2	5	8
Civic	2012 a 2016	1	5	5
Cruze HB	2012 a 2016	2	5	10
Cruze Sedan	2012 a 2016	2	5	10
Fox	2012 a 2016	2	5	10
March	2012 a 2016	2	5	10
Palio	2012 a 2016	2	5	10
Sandero	2012 a 2016	2	5	10
Uno	2012 a 2016	2	5	10
Cobalt	2012 a 2015	2	4	8
Duster	2012 a 2015	2	4	8
IX35	2012 a 2015	1	4	4
Jetta	2012 a 2015	1	4	4
Passat	2012 a 2015	1	4	3
A200	2013 a 2015	1	3	3
Versa	2012 a 2014	2	3	6
				312

Elaboração Própria

Os dados foram coletados compreendendo o ano modelo dos veículos de 2012 a 2016 nos websites CarrosnaWeb, Icarros e Webmotors. Contudo, como o mercado automotivo é muito dinâmico e ocorrem várias reestilizações nos modelos de veículos, não foi possível utilizar o período de 2012 a 2016 para todos os veículos presentes no banco de dados.

As 312 observações foram alocadas, cada qual, em seu respectivo banco de dados considerando que foram elaborados cinco bancos, um para cada período de tempo de depreciação a ser analisado, de um até cinco anos. A tabela 3 apresenta essa distribuição:

Tabela 3 - Distribuição das observações conforme banco de dados	
Período	Nº Observações
Banco de dados 1 Ano	80
Banco de dados 2 Anos	88
Banco de dados 3 Anos	69
Banco de dados 4 Anos	51
Banco de dados 5 Anos	24
	312

Elaboração Própria

A variável resposta, que será o foco de estudo deste trabalho, é o valor residual do veículo e outras seis características de cada modelo de veículo no ano de referência, foram escolhidas como candidatas a variáveis explicativas. São elas: Montadora, Motorização, Potência, Segmento, Segurança e Conforto.

Na obtenção dessas seis variáveis categóricas, foi utilizado o CarrosnaWeb, sítio que é referência para pesquisas de veículos em relação as suas versões, preços, itens opcionais e características, dado o seu extenso banco de dados. O sítio Icarros também foi utilizado, pois, detém um excelente banco de dados contendo informações detalhadas sobre cada veículo. O acesso aos sítios é gratuito e os dados foram retirados através de suas ferramentas de consulta. Para a obtenção do preço atual dos modelos, foi utilizado o sítio Webmotors, que é líder em anúncios para venda de veículos na internet. Todos os anúncios dos modelos presentes neste trabalho foram extraídos no mês de novembro de 2017 e,

selecionadas as mesmas configurações de modelo e versão, foi utilizada a média de preços encontrada, sendo esta considerada o preço de venda atual do modelo.

Como exposto no início desta seção, para escolha dos modelos de veículos para a composição do banco de dados, um dos critérios utilizados foi o volume de vendas no ano de 2017. Para obter esses dados, foi utilizado o relatório de emplacamento do mercado de zero quilometro emitido mensalmente pela Fenabreve. Deste relatório também foram retirados os segmentos em que cada veículo se enquadra.

3.1.1 Descrição das variáveis

É importante destacar que todas as variáveis selecionadas para comporem os modelos de regressão foram utilizadas em todos os cinco intervalos de depreciação estudados. Desta forma, todas tiveram um código atribuído em sua nomenclatura para identificação no banco de dados. Para as variáveis presentes no banco de dados de um determinado ano, o código que as identifica é .01, no banco de dados de dois anos .02 e assim sucessivamente até o banco de dados 5, cujo código é .05. Para fins genéricos, na descrição que segue os códigos das variáveis serão apresentados com .0X. A seguir, apresentamos a descrição de cada uma das variáveis.

A) Variável Resposta: % Valor residual do veículo – “VRV.0X”

Essa é a variável dependente do modelo de regressão, é do tipo continua e sua função é quantificar a desvalorização de um veículo ao longo do tempo. Pode ser descrita também como o valor residual de um veículo, isto é, o valor monetário que um determinado modelo apresenta após um período de uso. Essa variável é definida pela razão entre o preço de venda atual do veículo e o preço público de compra do veículo zero quilômetro atual (referência novembro 2017), como exposto na expressão abaixo:

$$\%ValorResidual = \frac{(Preçodevendaatual)}{(Preçopúblicodomoodelo0KMatual)} - 1$$

Ela será expressa em percentual para facilitar a interpretação e, caso seja de interesse identificar o valor absoluto do veículo após o período de depreciação, basta efetuar uma multiplicação simples do percentual encontrado pelo preço público atual do veículo do veículo.

Optou-se por atuar com o preço público do zero quilômetro, pois, caso fosse considerado o preço público do veículo na época de sua compra, teríamos a necessidade de reajustar o preço público passado do veículo a algum índice de inflação. A indústria automobilística não possui um índice em específico, então foram feitos alguns testes com o IPCA e notou-se que vários modelos com o valor corrigido, ultrapassavam seu preço público atual. Diante deste cenário adotamos então o preço público do veículo zero quilômetro, conforme exposto, como referência para o cálculo da desvalorização do modelo de veículo. Um exemplo de como se deu este cálculo é:

Considerando um modelo de veículo qualquer X 1.0 e que seu preço público atual (novembro de 2017) seja R\$ 40.000,00 e os seus preços de venda atuais são:

- X 1.0 – 2016 – PV (Preço de Venda): R\$ 34.200
- X 1.0 – 2015 – PV: R\$ 32.100
- X 1.0 – 2014 – PV: R\$ 30.200
- X 1.0 – 2013 – PV: R\$ 28.300
- X 1.0 – 2012 – PV: R\$ 27.000

Diante destes valores, o cálculo do valor residual deste veículo ficou desta forma:

- X 1.0 2016 – com 1 ano de uso: $Y = (34.200 / 40.000) - 1 = -14,5\%$
- X 1.0 2015 – com 2 anos de uso: $Y = (32.100 / 40.000) - 1 = -19,8\%$
- X 1.0 2014 – com 3 anos de uso: $Y = (30.200 / 40.000) - 1 = -24,5\%$
- X 1.0 2013 – com 4 anos de uso: $Y = (28.300 / 40.000) - 1 = -29,3\%$
- X 1.0 2012 – com 5 anos de uso: $Y = (27.000 / 40.000) - 1 = -32,5\%$

B) Montadora

Variável categórica que apresenta a montadora fabricante do modelo. No banco de dados, estão presentes 14 montadoras e a definição das variáveis indicadoras utilizadas no modelo de regressão é apresentada no Quadro 1.

Quadro 1 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Montadora do Veículo.

Montadora	Nome da Variável Indicadora da Montadora*
Ford	<i>Categoria de referência**</i>
Volkswagen	M-Volkswagen.0X
Toyota	M-Toyota.0X
Fiat	M-Fiat.0X
Hyundai	M.Hyundai.0X
General Motors (Chevrolet)	M-GM.0X
Renault	M-Renault.0X
Nissan	M-Nissan.0X
Citroen	M-Citroen.0X
Peugeot	M-Peugeot.0X
Audi	M-Audi.0X
Honda	M-Honda.0X
Mercedes Benz	M-MercedesBenz.0X
BMW	M.BMW.0X

* Recebe valor 1 se o veículo pertence à montadora e 0, caso contrário.

** Com exceção ao banco de dados de cinco anos, cuja referência é a montadora Nissan

C) Segmento

Variável categórica que define qual é o tipo de carroceria de um modelo de veículo. Neste estudo foram utilizados os segmentos apresentados nos relatórios de venda mensais da Fenabrave, onde são apresentados 13 segmentos. No entanto, no banco de dados deste trabalho, foram utilizados apenas 10 segmentos, pois não tivemos nenhum modelo que se enquadrava nos segmentos SW Grande, Sport e Grandcab. O Quadro 2 apresenta os 10 segmentos presentes no trabalho e suas respectivas variáveis indicadoras.

Quadro 2 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Segmento.

Segmento	Nome da Variável Indicadora do Segmento*
Entrada	<i>Categoria de referência</i>
Hatch Pequeno	Seg-HPequeno.0X
Hatch Médio	Seg-HMedio.0X
Sedan Pequeno	Seg.Spequeno.0X
Sedan Compacto	Seg.Scompacto.0X
Sedan Médio	Seg.Smedio.0X
Sedan Grande	Seg.Sgrand.0X
SW (Space Wagon) Médio	Seg.SWmedio.0X
Monocab	Seg.Monocab.0X
SUV (Sport Utility Vehicle)	Seg.SUV.0X

* Recebe valor 1 se o segmento do veículo se encaixa na categoria e 0, caso contrário.

D) Motorização

Variável categórica que indica a motorização de um veículo, que é a capacidade que um motor tem de aspirar ar/combustível e que, de forma dedutível, nos induz a crer que, quando maior sua marcação (1.0, 1.5, 2.0), mais potência tende a ter o motor. Contudo, na atualidade, este senso comum nem sempre é correto, uma vez que, dependendo da tecnologia empregada no motor e em seus sistemas de controle, um motor com uma marcação menor pode vir a ser mais potente que um de marcação superior. No banco de dados utilizado neste trabalho, temos presentes oito possíveis motorizações, distribuídas em quatro categorias da seguinte maneira conforme apresentado no Quadro 3.

Quadro 3 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Motorização do Veículo.

Motorização	Nome da Variável Indicadora da Motorização*
1.0	<i>Categoria de referência</i>
> 1.0 e ≤ 1.4	Mot.1.0e1.4.0X
> 1.4 e ≤ 1.8	Mot.1.4e1.8.0X
> 1.8	Mot.>1.8.0X

* Recebe valor 1 se a motorização do veículo se encaixa na categoria e 0, caso contrário.

E) Potência

Variável categórica que expressa a “força” que o motor de um veículo possui, sendo expressa em CV (cavalo-vapor), que é a medida comumente usada na indústria automobilística brasileira para definir a potência, por mais que o CV não pertença ao sistema internacional de medidas. As medidas da potência variam de acordo com o combustível utilizado em veículos do tipo flex, e para estes modelos, foi adotada a potência mensurada com o combustível etanol. O estudo contou com 46 variações de potências entre os modelos, distribuídas em sete categorias, que são apresentadas no Quadro 4:

Quadro 4 – Definição das variáveis indicadoras utilizadas no modelo de regressão para representação da característica Potência.

Potência	Nome da Variável Indicadora da Potência*
< 100 cv	<i>Categoria de referência</i>
> 100 cv <= 130 cv	Pot>100<=130.0X
> 130 cv <= 160 cv	Pot>130<=160.0X
> 160 cv <= 190 cv	Pot>160<=190.0X
> 190 cv <= 220 cv	Pot>190<=220.0X
> 220 cv <= 250 cv	Pot>220<=250.0X
> 250 cv	Pot>250.0X

* Recebe valor 1 se a potência do veículo se encaixa na categoria e 0, caso contrário.

F) Segurança

Variável quantitativa discreta que representa a quantidade de itens de segurança que determinado modelo apresenta. Como são vários os itens de segurança presentes em um veículo, foram considerados neste estudo os mais relevantes, sendo eles: Freio ABS / Airbag Duplo, Cinto de três pontos para todos os passageiros, Encosto de cabeça para todos os passageiros, mais de dois airbags presentes no veículo, Isofix, Assistente de partida em rampa, Controle de estabilidade e controle de tração.

Esses itens foram escolhidos, pois, são considerados os principais quando se fala em segurança de um automóvel. O Freio ABS e Airbag duplo se tornaram obrigatórios para todos os veículos fabricados no país e importados desde 2014. Contudo, como este trabalho aborda veículos fabricados antes deste ano, este item entrou na contagem do número de itens de segurança de um modelo de veículo. O cinto de três pontos para todos os passageiros e encosto de cabeça são itens apontados por diversos estudos como necessários para evitar lesões mediante uma colisão do veículo. Diante disto, o Contran (Conselho Nacional de Trânsito) redigiu a resolução 518 de 2015, que determina a inclusão no CTB (Código de Trânsito

Brasileiro) da obrigatoriedade, para modelos inéditos fabricados a partir de 2018 e modelos já fabricados a partir de 2020, terem como itens de série cinto de três pontos e encosto de cabeça para todos os ocupantes, além do ISOFIX, sistema que permite acoplar cadeirinhas de crianças no veículo sem a necessidade de utilizar o cinto de segurança.

O Latin Ncap (Programa de Avaliação de Carros Novos para América Latina e o Caribe) considera que o maior avanço da indústria automobilista em segurança, após o cinto, foi o desenvolvimento do controle de estabilidade e controle de tração. Essa tecnologia evita a perda do controle do automóvel. Para se obter uma nota 4 ou 5 no Latin Ncap, cuja escala vai de 1 a 5, é necessário que o veículo possua a tecnologia de controle de estabilidade e de tração. A presença de mais de dois airbags em um veículo, seja de cortina, joelho ou outras opções, tende a reforçar mais ainda a segurança do automóvel.

O sistema de assistência de partida em rampa evita colisões e derrapagens em terrenos íngremes.

G) Conforto

Variável quantitativa discreta que representa a quantidade de itens de conforto que um determinado modelo dispõe. Tendo sido estruturada de forma similar à variável Segurança. Na atualidade existem inúmeros itens opcionais de conforto e, para este estudo, foram selecionados os seguintes: Ar condicionado / Direção elétrica ou hidráulica / Vidro elétrico dianteiro e trava elétrica; Vidro elétrico traseiro e retrovisor elétrico; Câmbio automático; Central Multimídia e Volante Multifuncional; Roda de liga leve; Banco de couro; Teto solar ou panorâmico.

A grande maioria dos carros fabricados nos dias de hoje tem por configuração básica ar condicionado, direção hidráulica ou elétrica, trava elétrica e vidros dianteiros elétricos. Sendo assim, esses opcionais foram considerados em apenas um item. Alguns modelos de veículo deste trabalho não apresentam esses quatro itens e, para estes modelos, foi considerado como se não houvesse nenhum item de conforto.

3.2 Análise de Regressão

A técnica estatística escolhida neste trabalho para entendermos qual é a relação existente entre as características de um veículo e seu valor residual é a análise de regressão.

Com a análise de regressão, é possível estudar a dependência de uma variável Y (resposta) em relação a uma ou várias variáveis X (explicativas ou preditoras). Segundo Pagano (2004, p.367.), o objetivo principal da análise de regressão é prever ou estimar o valor da resposta associada a um valor fixo e conhecido da variável explicativa. Ainda conforme Pagano (2004, p.398), caso um modelo apresente mais de uma variável explicativa, o objetivo se mantém: estimar o valor da variável Y associada quando todas as demais variáveis X_n permanecem constantes.

3.2.1 Considerando variáveis explicativas do tipo categóricas

As variáveis explicativas em um modelo de regressão podem ser quantitativas ou qualitativas. Caso sejam variáveis qualitativas, as mesmas devem ser representadas no modelo por meio de variáveis indicadoras, também conhecidas como *dummies*. Gujarati (2008) reforça essa ideia ao citar que é possível quantificar atributos formulando variáveis artificiais que assumem valores de 1 ou 0, em que 1 indica a presença do atributo e 0 a ausência dele.

Modelos que assumem todas as suas variáveis explicativas como categóricas são chamados de modelos de análise de variância (ANOVA) e, conforme Gujarati (2008), “os modelos ANOVA devem ser usados para avaliar o significado estatístico da relação entre um regressando quantitativo e regressores binários ou qualitativos”. Supondo que esteja sendo avaliada a relação entre uma variável quantitativa Y qualquer com uma variável explicativa categórica X, tendo ela dois atributos (A e B), em termos gerais, um modelo simples de regressão contendo esta variável qualitativa X seria escrito como:

$$y_i = B_0 + B_1X_1 + \varepsilon_i$$

Onde X_1 assume o valor 1 se o indivíduo possuir o atributo B, caso contrário, assume o valor 0; B_0 é o intercepto, valor médio que a variável resposta assume caso a variável explicativa categórica X_1 possua o atributo A, também chamado de grupo de referência, comparação, controle ou base. Caso a variável X_1 assumira o valor 1, então o coeficiente angular B_1 será acrescido ao B_0 , deslocando o valor médio da resposta para uma nova posição. Dessa forma, o coeficiente angular apresenta o quanto o valor médio da resposta irá diferir do valor médio no grupo de referência caso o indivíduo possua o atributo B. Já ε_i é a parte da variável resposta não explicada pelo modelo, ou seja, o erro.

Se a variável categórica possuir mais de duas categorias, será necessário representar a mesma com mais de uma variável binária. Considerando que a variável explicativa categórica X tenha agora 3 atributos, sendo eles A, B e C, a representação de um modelo de regressão é:

$$y_i = B_0 + B_1X_1 + B_2X_2 + \varepsilon_i$$

onde:

B_0 – Representa o valor médio na categoria de referência, ou seja, se ela possuir o atributo A, neste caso X_1 e X_2 tem valor 0.

B_1 – Representa o acréscimo na categoria de referência, caso a variável possua o atributo B, neste caso $X_1 = 1$ e $X_2 = 0$.

B_2 – Representa o acréscimo na categoria de referência, caso a variável possua o atributo C, neste caso $X_1 = 0$ e $X_2 = 1$.

Desta forma notamos que, se uma variável possuir N categorias, é necessário a introdução de N-1 variáveis categóricas para representá-la. Gujarati (2008) descreve essa regra como: “para cada regressor qualitativo, o número de variáveis binárias introduzidas deve ser um a menos que as categorias daquela variável”. É essencial observar esta regra a fim de evitar a colinearidade perfeita, situação em que ocorre uma relação linear exata entre as variáveis explicativas.

Entretanto, é possível representar uma variável que tenha N categorias com N variáveis binárias. Para isso, é necessário suprimir o intercepto do modelo, de forma que os coeficientes B_n irão representar diretamente os valores médios das diversas categorias. Formular o modelo de regressão desta forma é pouco usual, pois se perde a capacidade de comparabilidade das categorias em relação a uma referência. Em nosso trabalho, optou-se por utilizar uma categoria referência, valendo-se da regra de N-1 variáveis binárias para representar uma variável explicativa categórica.

3.2.2 Considerando variáveis quantitativas

Caso as variáveis explicativas sejam quantitativas, não se faz necessário adotar uma variável indicadora (*dummie*) como no caso das variáveis qualitativas. A abordagem nesta situação, em termos gerais, pode ser exemplificada com um modelo de regressão linear simples, descrito de forma similar ao contextualizado no caso das variáveis qualitativas, mas a interpretação se dará de maneira distinta. A equação abaixo descreve a relação entre uma variável quantitativa Y qualquer com um variável explicativa quantitativa X:

$$y_i = B_0 + B_1X_1 + \varepsilon_i$$

onde B_0 , o intercepto, representa o valor da variável resposta quando X_1 for 0, já B_1 ou coeficiente angular representa o aumento na variável resposta à medida em que se aumenta uma unidade de X_1 . De forma análoga ao modelo com variáveis qualitativas, temos a presença do termo ε_i , que representa o erro, ou seja, a parte da variabilidade da variável resposta não explicada pelo modelo. Caso tenhamos um modelo de regressão linear múltipla, com variáveis quantitativas, a interpretação exposta acima se mantém. Se tivéssemos um termo positivo B_2X_2 , o valor de B_2 iria representar o acréscimo da variável resposta à medida em que se aumenta uma unidade de X_2 . E, nessa situação, B_0 passaria a representar o valor da variável resposta quando X_1 e X_2 fossem iguais a zero.

3.2.3 Pressupostos para o ajuste de um modelo de regressão

Para que consigamos obter modelos que sejam satisfatórios são necessários, que sejam observados alguns pressupostos básicos em análise de regressão linear, como normalidade dos erros, linearidade da relação entre a variável resposta e as explicativas, homocedasticidade dos erros, e ausência de multicolinearidade. Esses pressupostos serão investigados em todos os modelos a fim de garantir que não estejam sendo violados e, desta forma, esperamos aumentar a possibilidade de identificarmos modelos com um bom ajuste.

3.2.3.1 Normalidade dos erros

O pressuposto de normalidade dos erros é uma suposição essencial para a validação de um modelo de regressão linear. Supor que um modelo atenda a esse pressuposto significa assumirmos que os erros do modelo (os desvios dos valores preditos em relação aos valores observados pela variável resposta) seguem uma distribuição normal. Existem vários testes que auxiliam na validação deste pressuposto, como teste de Anderson-Darling ou teste de Kolmogorov-Smirnov, entre outros.

3.2.3.2 Linearidade

Para a correta especificação do modelo, devemos assumir que a relação entre a variável resposta e as variáveis explicativas é linear, isto é, a reta é o melhor ajuste aos dados no caso de regressão simples. Para a validação deste pressuposto, podemos utilizar o gráfico de dispersão dos resíduos do modelo com os valores preditos pelo modelo.

3.2.3.3 Homocedasticidade dos erros

O termo homocedasticidade é utilizado para representar a variância constante dos erros de um modelo de regressão. Esse pressuposto é de grande relevância para a validação de um modelo de regressão linear, pois significa, que caso seja atendido, a variabilidade da variável resposta é a mesma independentemente do

nível da variável explicativa. Quando um modelo de regressão não apresenta variância constante nos erros, ou seja, ele é heterocedástico, os valores preditos pelo modelo não terão uma variabilidade idêntica para cada nível da variável resposta. Desta forma o modelo de regressão não irá conseguir prever a resposta para valores mais altos da variável explicativa da mesma forma como faz para valores mais baixos, ou vice e versa. Para avaliarmos esse pressuposto, podemos nos valer do mesmo gráfico de dispersão citado na análise de linearidade.

3.2.3.4 Ausência de multicolinearidade

Quando duas ou mais variáveis explicativas do modelo estão fortemente correlacionadas, podemos supor que estaremos incorrendo em um modelo com a presença de multicolinearidade e se faz necessária a correta averiguação deste pressuposto. Quando incorremos em um modelo com a presença de multicolinearidade, podemos ter erros nas estimativas dos parâmetros, elas podem ser tornar instáveis e variáveis explicativas que seriam estatisticamente significantes passam a não ser ou então que são positivamente associadas à variável resposta e apresentam um coeficiente negativo. Para a detecção da multicolinearidade, utilizamos os fatores de inflação de variância.

4. Resultados

4.1 Análise exploratória dos dados

Nesta seção, iremos compreender melhor as variáveis abordadas neste estudo, analisando a relação entre a variável resposta (valor residual de um veículo) e as variáveis explicativas, além de analisar individualmente cada uma das variáveis. Estas análises irão compreender os cinco bancos de dados para que consigamos entender o comportamento das variáveis em cada um dos períodos selecionados, de um a cinco anos. Como estamos tratando do valor residual de um veículo, o sinal de negativo foi retirado de todas as observações, pois atuar com o valor absoluto facilita a interpretação das variáveis, uma vez que, quanto maior a depreciação de um veículo, menor será o seu valor residual.

A) Variável resposta: Valor Residual do Veículo (em proporção) nos bancos de dados de um a cinco anos de depreciação

A Tabela 4 apresenta as estatísticas descritivas do Valor Residual do veículo (em proporção) nos bancos de dados de um a cinco anos de depreciação. Podemos notar que, como esperado, na medida em que se passa o tempo de utilização de um veículo, aumenta a desvalorização do mesmo. O modelo de veículo que teve a menor desvalorização (o Honda City LX 1.5 Automático ano modelo 2016) está no banco de dados de um ano e perdeu 10,14% do seu valor em relação ao “zero quilômetro” atual. Já o modelo de veículo com maior desvalorização (Peugeot 308 2.0 Allure ano modelo 2012) pertence ao banco de dados de cinco anos, perdeu 65,90% do seu valor.

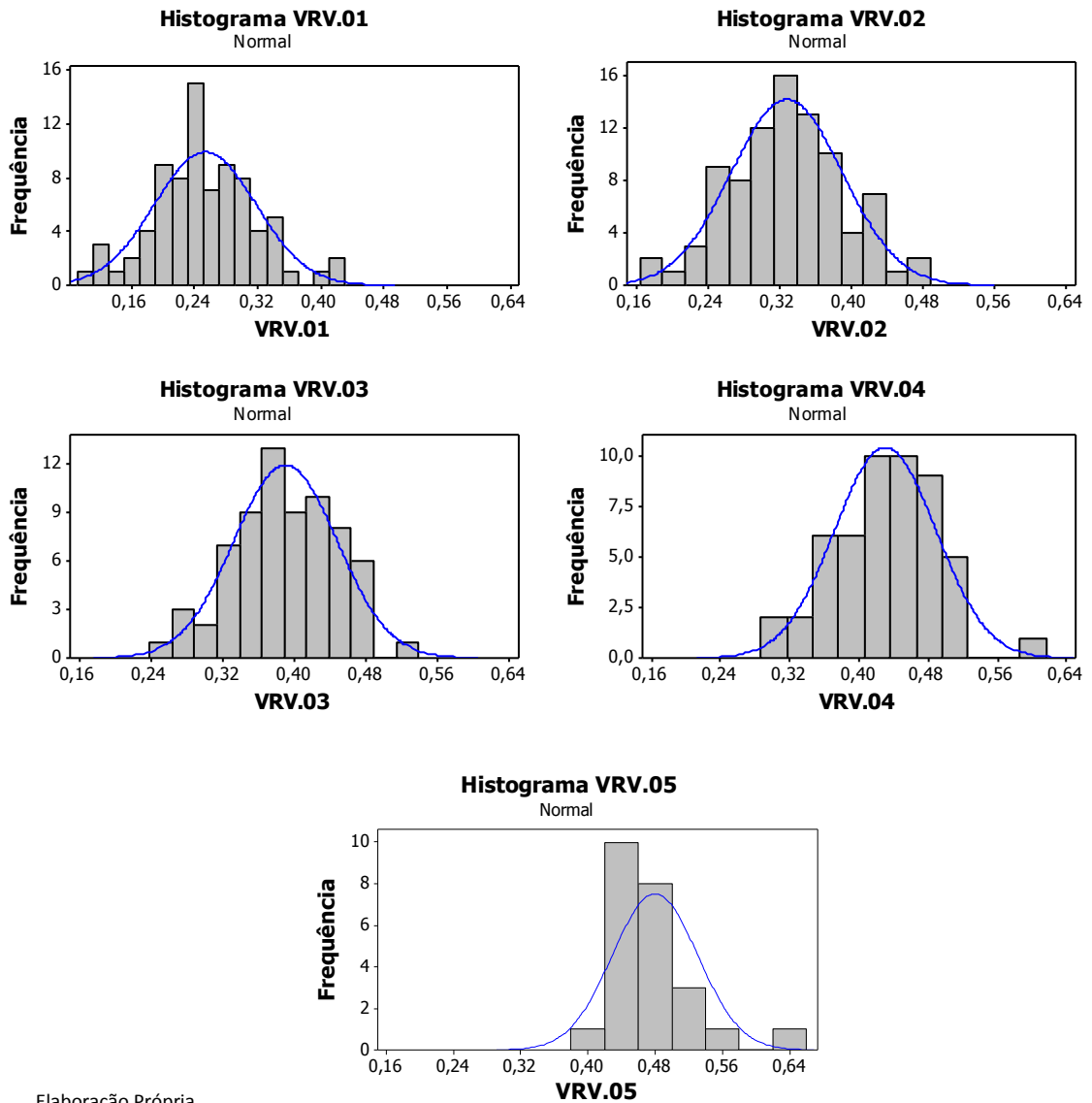
Notamos que, em média, os modelos de veículos utilizados por um ano perdem 25,23% do seu valor em relação ao valor do veículo “zero quilômetro” atual. Já a desvalorização dos modelos de veículos com cinco anos de utilização quase dobra em relação aos veículos com um ano de uso, tendo uma desvalorização média de 47,98%. Observando as medianas, nota-se que seus valores são muito próximos dos valores das médias, indicando simetria nas distribuições de frequências, o que pode ser visto no Gráfico 1. Além disto, ao observar a média e mediana percebemos que, na medida em que se utiliza mais um veículo, ele tende a depreciar mais. Contudo, com passar do tempo, também ocorre uma suavização na desvalorização de um veículo. O Gráfico 2 facilita a compreensão desta observação.

Tabela 4 – Estatísticas descritivas para o valor residual de um a cinco anos

Banco de Dados	Mínimo	1º Quartil	Mediana	3º Quartil	Máximo	Média	Variância	Desvio Padrão	Coefficiente de Variação	Assimetria	Nº Observações
1 Ano	0,1014	0,2088	0,2438	0,2954	0,4292	0,2523	0,0042	0,0645	25,5603	0,2720	80
2 Anos	0,1773	0,2856	0,3314	0,3703	0,4772	0,3268	0,0039	0,0622	19,0419	-0,0365	88
3 Anos	0,2560	0,3517	0,3847	0,4317	0,5351	0,3890	0,0033	0,0578	14,8689	-0,0841	69
4 Anos	0,2998	0,3956	0,4296	0,4736	0,5932	0,4292	0,0053	0,0588	13,6945	0,0228	51
5 Anos	0,3950	0,4472	0,4830	0,4970	0,6590	0,4798	0,0026	0,0508	10,5932	1,8344	24

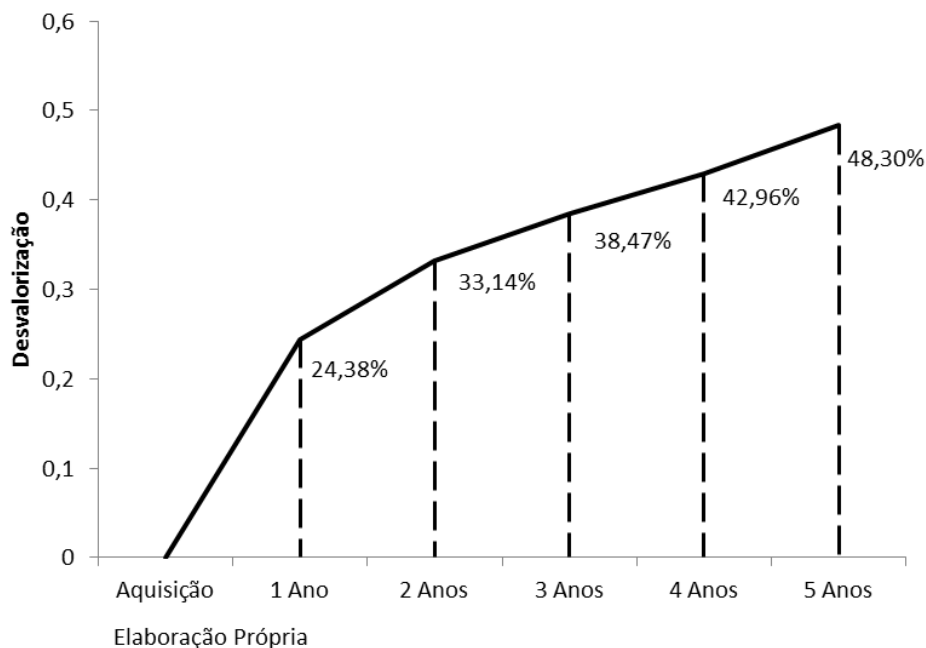
Elaboração Própria

Gráfico 1 - Distribuição de frequências do Valor Residual do Veículo (em proporção) nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

O Gráfico 2 - Valor residual médio dos veículos (em %) ao longo dos cinco anos de uso



Observando os valores do coeficiente de variação do valor residual médio dos veículos (em %) ao longo dos cinco anos de uso (tabela 4), notamos que a variabilidade da variável resposta diminui com o passar dos anos, o que é reforçado também pelos valores da amplitude. A amplitude da variável resposta no primeiro ano é de 32,77% e, no quinto ano, é de 26,4%. Os histogramas também mostram essa concentração dos dados com o passar dos anos.

Analisando os histogramas, notamos que a distribuição da variável resposta nos bancos um, dois, três e quatro são bem simétricas, enquanto a distribuição dos dados do banco de cinco anos tem uma leve assimetria à esquerda.

B) Variável Montadora: Montadoras dos modelos de veículo (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 5 apresenta a distribuição de frequências da variável montadora em todos os bancos de dados. Notamos que os modelos de veículos da montadora Ford, Volkswagen, Fiat e Chevrolet são os mais presentes em todos os bancos de dados, o que já era de se esperar, pois, são as montadoras que juntas têm maior

participação de mercado. Ainda assim, a Ford não contou com nenhuma observação no banco de dados de cinco anos.

Já Toyota, Hyundai, Nissan e Renault apresentaram um número próximo de observações em todos os bancos de dados. Essas quatro montadoras não ultrapassaram seis modelos de veículo em nenhum banco e, considerando o banco de um ano a quatro, não tiveram menos que três observações.

No banco de dados de cinco anos a Toyota, não teve nenhum dado e Hyundai apenas uma observação. Citroen teve duas observações nos bancos de um a quatro anos, no de cinco não apresentou nenhum dado, a Peugeot por sua vez teve quatro dados nos três primeiros bancos, e apenas um nos dois últimos. Audi e Honda são montadoras com razoável representatividade nos bancos de um e dois anos tendo a Audi sete observações em cada e a Honda cinco. Já nos demais bancos, elas perdem representatividade: a Audi tem duas observações no banco de dados de três anos e nenhuma nos demais; a Honda por sua vez tem uma observação apenas, nos bancos de três, quatro e cinco anos.

A Mercedes-Benz não contou com nenhum dado nos bancos de um e cinco anos, e, nos demais, teve apenas um modelo de veículo. A BMW por sua vez teve duas observações no banco de dados de um ano e o mesmo número no de dois anos, nos demais não apresentou nenhuma observação. Esse baixo volume de dados da Mercedes-Benz e BMW já era de se esperar, pois, são montadoras cujo posicionamento de preço médio dos seus veículos restringe o acesso aos mesmos, tendo um baixo volume de amostras no mercado.

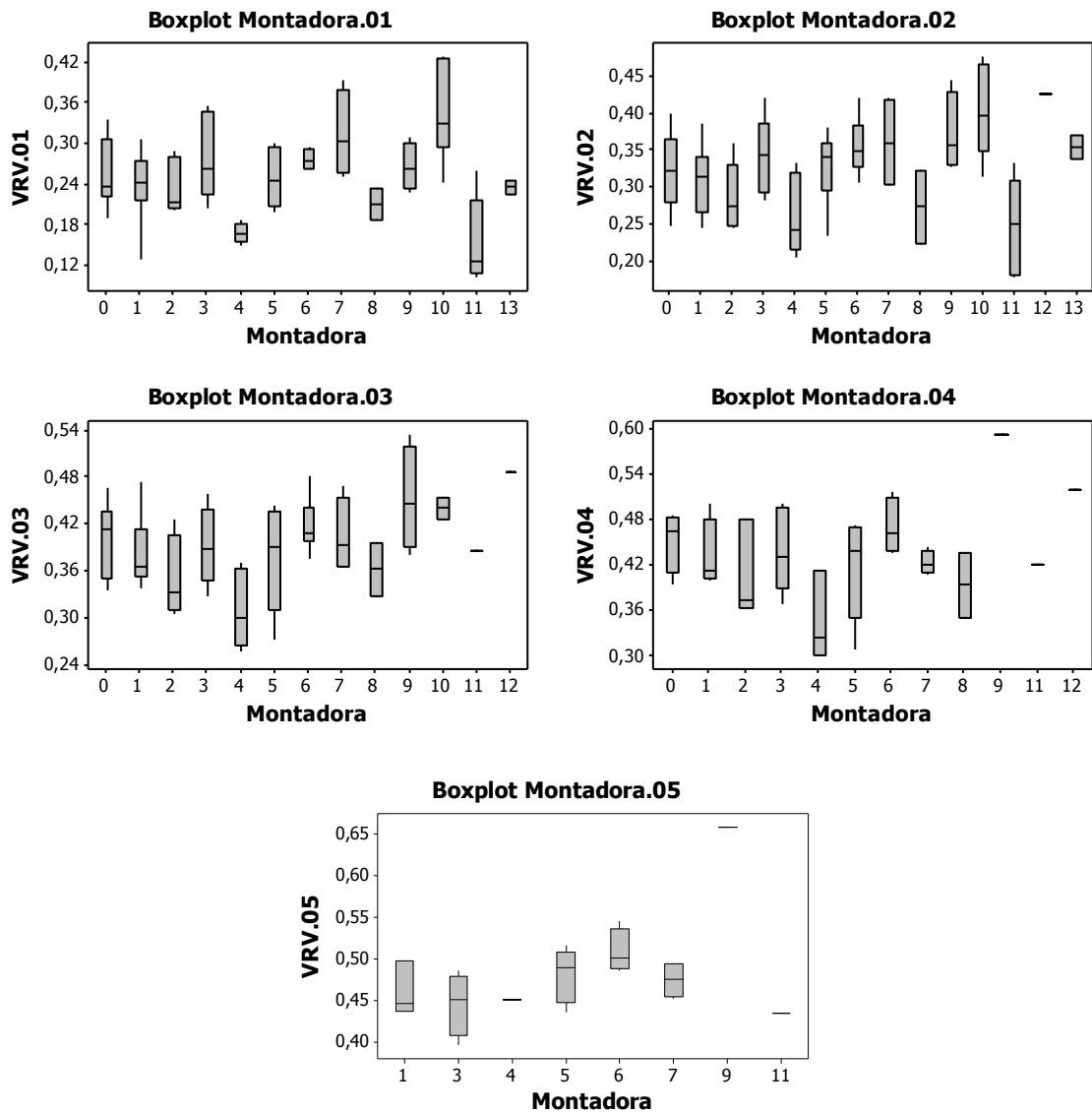
Tabela 5 – Tabela de frequência da variável Montadora nos bancos de dados de um a cinco anos de depreciação.

Frequência da variável Montadora do banco um ao cinco					
Montadora	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0 - Ford	24 (30%)	13 (14,8%)	9 (13%)	4 (7,8%)	0 (0%)
1 - Volkswagen	0 (0%)	13 (14,8%)	9 (13%)	8 (15,7%)	3 (12,5%)
2 - Toyota	6 (7,5%)	6 (6,8%)	4 (5,8%)	3 (5,9%)	0 (0%)
3 - Fiat	10 (12,5%)	10 (11,4%)	10 (14,5%)	10 (19,6%)	4 (16,7%)
4 - Hyundai	4 (5%)	5 (5,7%)	5 (7,2%)	3 (5,9%)	1 (4,2%)
5 - Chevrolet	8 (10%)	10 (11,4%)	10 (14,5%)	10 (19,6%)	6 (25%)
6 - Renault	4 (5%)	6 (6,8%)	6 (8,7%)	4 (7,8%)	4 (16,7%)
7 - Nissan	4 (5%)	4 (4,5%)	6 (8,7%)	4 (7,8%)	4 (16,7%)
8 - Citroen	2 (2,5%)	2 (2,3%)	2 (2,9%)	2 (3,9%)	0 (0%)
9 - Peugeot	4 (5%)	4 (4,5%)	4 (5,8%)	1 (2%)	1 (4,2%)
10 - Audi	7 (8,8%)	7 (8%)	2 (2,9%)	0 (0%)	0 (0%)
11 - Honda	5 (6,3%)	5 (5,7%)	1 (1,4%)	1 (2%)	1 (4,2%)
12 - Mercedes Benz	0 (0%)	1 (1,1%)	1 (1,4%)	1 (2%)	0 (0%)
13 - BMW	2 (2,5%)	2 (2,3%)	0 (0%)	0 (0%)	0 (0%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

O Gráfico 3 nos apresenta os boxplots da variável Montadora em relação a variável resposta, valor residual do veículo, para todos os cinco bancos de dados. Notamos que, no banco de dados de um ano, a montadora com os menores valores de desvalorização é a Honda, tendo sua mediana bem próxima do primeiro quartil, demonstrando uma assimetria dos dados com concentração à esquerda. A maior desvalorização em um ano ocorreu na montadora Audi (mediana de 32,94%), que também apresentou a maior amplitude interquartílica e a desvalorização máxima de 42,91%. A Hyundai apresentou uma mediana de 16,5%, próxima do 1º e 3º quartil, tendo assim uma amplitude interquartílica pequena e com baixa variabilidade dos dados. O mesmo é possível notar com a BMW e Renault, a essas três pode-se imputar essa concentração devido ao baixo volume de observações

Gráfico 3 – Boxplot do Valor Residual dos Veículos (em proporção) segmentados pela variável Montadora nos bancos de dados de um a cinco anos de depreciação



Referência de códigos da variável Montadora

Ford	Volkswagen	Toyota	Fiat	Hyundai	Chevrolet	Renault	Nissan	Citroen	Peugeot	Audi	Honda	Mercedes Benz	BMW
0	1	2	3	4	5	6	7	8	9	10	11	12	13

Elaboração Própria

Observando o boxplot do banco de dados de dois anos, percebemos que com exceção da Hyundai, Peugeot e Volkswagen, o comportamento da desvalorização dos veículos de cada montadora foi bem similar ao banco de dados um. A menor mediana foi da Hyundai (24,29%), contudo a Honda apresentou novamente o modelo de veículo com a menor desvalorização em dois anos, com perda de seu valor em relação ao modelo zero quilômetro de 17,73%. A Audi também se manteve como a montadora com a maior desvalorização de um modelo de veículo em dois anos (47,71%) e também apresentou a pior mediana dentre todas as montadoras analisadas. Vale destacar que o banco de dados de dois anos é o único que contou com todas as montadoras.

No banco de dados de três anos notamos que a menor mediana foi da Hyundai 29,92%. A Peugeot foi a montadora que apresentou uma alta desvalorização de seus modelos de veículos em três anos. Como evidenciado pela Tabela 5, a Honda contou apenas com uma observação no banco de dados de três anos e, devido a isso, teve um comportamento bem diferente do que foi presenciado nos bancos de dados anteriores.

Notamos que a Hyundai é a montadora que, em média, tem a menor desvalorização após quatro anos de uso, visto que possui a menor mediana e também o veículo que menos desvalorizou. A Hyundai, Chevrolet, Fiat e Toyota apresentam uma variabilidade próxima, com valores de amplitude interquartílica similares. A Peugeot, Honda e Mercedes Benz, apresentaram apenas um dado neste banco (o modelo de veículo 308 2.0 Allure), sendo que a observação da Peugeot é atípica em relação aos demais dados do banco de quatro anos.

Após cinco anos de utilização, a mediana das montadoras Volkswagen, Fiat, Honda e Hyundai ficaram em linha. Contudo, a mediana da Honda foi a menor entre as montadoras do banco de dados de cinco anos. Vale destacar que Honda e Hyundai contaram apenas com uma observação. Fiat foi a montadora que teve o modelo de veículo com menor desvalorização. Já a Peugeot teve apenas uma observação também, mas manteve o mesmo padrão de dado atípico já percebido no banco de dados de quatro anos, até por se tratar do mesmo modelo de veículo.

C) Variável Segmento: Segmentos dos modelos de veículo (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 6 apresenta a distribuição de frequências da variável segmento em todos os bancos de dados. Em todos os anos, o segmento que tem maior representatividade é o de Hatch Pequeno, seguido por Sedan Pequeno e os veículos de Entrada. Juntos eles somam 50,0% dos modelos de veículos avaliados no banco de um ano, 45,4% no banco de dois anos, 38,8% no banco de três anos, 54,9% no banco de quatro anos e 50,0% no banco de cinco anos. Os Hatchs Médios têm um número de dados similar nos bancos de um a três anos (dez observações nos dois primeiros e oito no terceiro). Já no quarto e quinto anos, ocorre uma queda e este segmento passa a contar com três observações no banco de quatro anos e três no de cinco anos.

Tabela 6 – Tabela de frequência da variável Segmento nos bancos de dados de um a cinco anos de depreciação.

Frequência da variável Segmento do banco um ao cinco					
Segmento	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0 - Entrada	40 (50%)	10 (11,4%)	38 (55,1%)	28 (54,9%)	10 (41,7%)
1 - Hatch Pequeno	0 (0%)	16 (18,2%)	0 (0%)	0 (0%)	0 (0%)
2 - Hatch Medio	10 (12,5%)	10 (11,4%)	8 (11,6%)	3 (5,9%)	3 (12,5%)
3 - Sedan Pequeno	0 (0%)	14 (15,9%)	0 (0%)	0 (0%)	4 (16,7%)
4 - Sedan Compacto	2 (2,5%)	4 (4,5%)	2 (2,9%)	2 (3,9%)	0 (0%)
5 - Sedan Medio	13 (16,3%)	24 (27,3%)	10 (14,5%)	6 (11,8%)	4 (16,7%)
6 - Sedan Grande	4 (5%)	5 (5,7%)	0 (0%)	0 (0%)	0 (0%)
7 - SW	4 (5%)	0 (0%)	5 (7,2%)	3 (5,9%)	0 (0%)
8 - Monocab	4 (5%)	5 (5,7%)	0 (0%)	3 (5,9%)	0 (0%)
9 - SUV	3 (3,8%)	0 (0%)	6 (8,7%)	6 (11,8%)	3 (12,5%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

O Sedan Compacto, SW, Monocab e o Sedan Grande têm pouca expressividade em todos os bancos. Esses segmentos realmente apresentam um número menor de veículos no mercado, o que dificulta identificar quais modelos de

veículos atendem aos critérios para estarem presentes no banco de dados. Os Sedans Compactos tiveram quatro observações no máximo e duas no mínimo. Já os Sedans Grandes e Monocab contaram com no máximo cinco observações e, no banco de cinco anos, ambos não tiveram nenhuma.

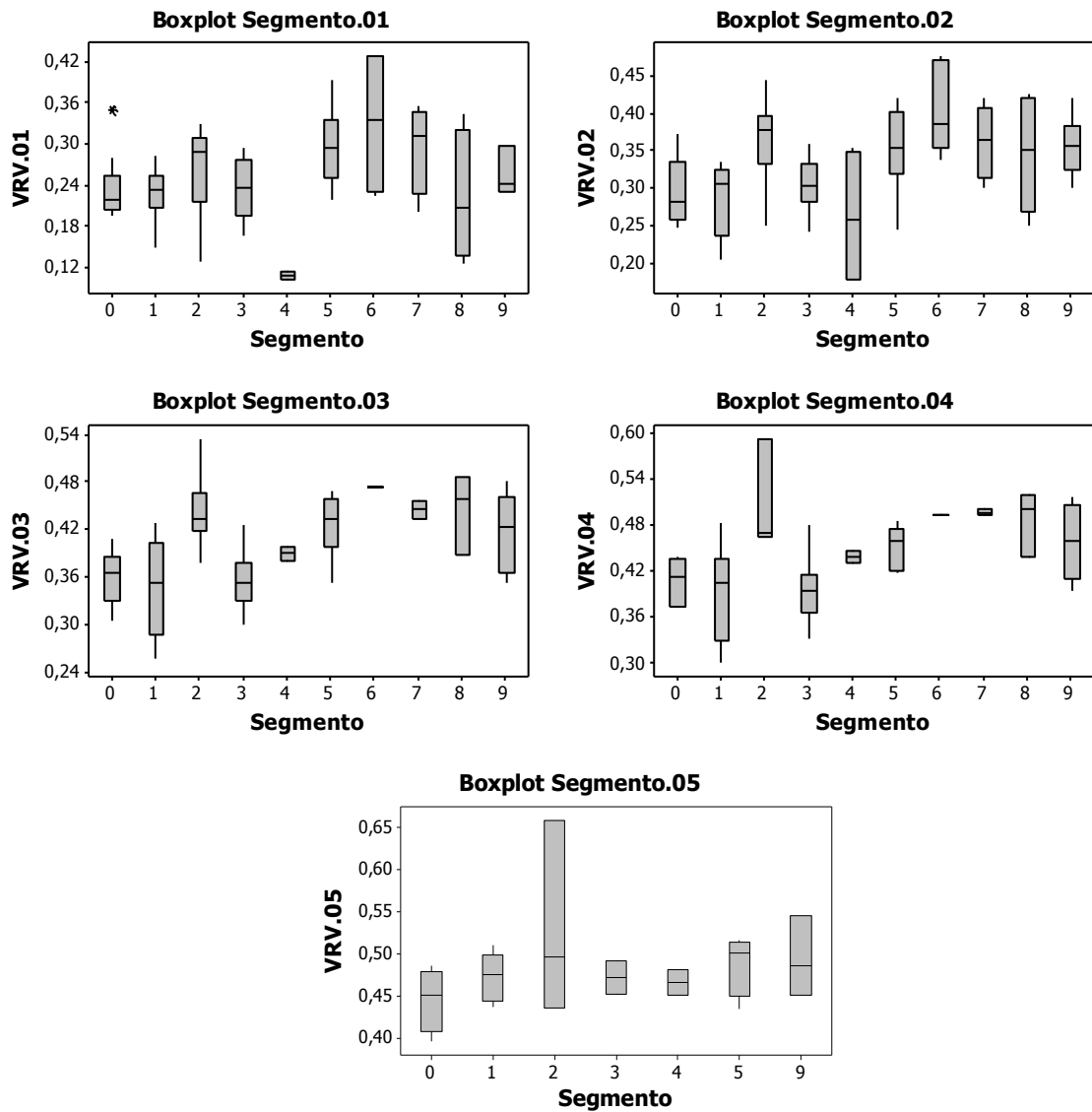
Nos bancos de dados de três e quatro anos, os Sedans Grandes apresentaram apenas uma única observação, o que pode vir a prejudicar a análise desta categoria. Dessa forma, nestes dois bancos de dados, iremos unir os Sedans Grandes, no de três anos a categoria dos Sedans Médios e no de quatro anos a categoria do SW, conseguindo assim viabilizar a utilização destas observações sem o prejuízo da análise das mesmas.

O SW também não teve nenhuma observação no banco de dados de cinco anos e contou no máximo com quatro dados nos bancos de um e dois anos. Os Sedans Médios, segmento com maior volume de amostras entre os sedans, teve uma participação expressiva nos bancos de um, dois e cinco anos, neste último teve a mesma participação do segmento de entrada, 16,66% dos dados, ficando atrás apenas do Hatch Pequeno. O segmento de SUV teve pouca expressividade nos bancos de dados, no de dois, três e quatro anos apresentou seis observações, nos outros dois apenas três. É importante destacar que segmentos como o de SUV na atualidade têm aumentando expressivamente sua participação no mercado, contudo, como este banco é composto por veículos com ano modelo de 2016, neste ano o segmento começava a aumentar sua representatividade e vários modelos sofreram reestilizações e outros foram lançados em 2017, por isso o baixo volume de dados desse segmento no banco.

O Gráfico 4 apresenta a distribuição dos valores residuais dos veículos segundo seus segmentos para os bancos de dados de um a cinco anos. Notamos que, no banco de dados de um ano, o segmento que apresentou o maior valor residual, ou seja, a menor desvalorização, foi o Sedan Compacto, com mediana de 10,81%, enquanto o Sedan Grande apresentou a maior desvalorização entre os segmentos selecionados. O segmento Monocab apresentou uma amplitude interquartilica alta, de forma similar ao Sedan Grande, mostrando uma alta variabilidade nos valores residuais deste tipo de veículo. Os segmentos de Entrada,

Hatch Pequeno e Sedan Pequeno têm um comportamento similar e, foram agrupados em um única categoria. O Gráfico 5 apresenta a distribuição dos valores residuais dos veículos segundo seus segmentos com as categorias agrupadas .

Gráfico 4 – Boxplot do Valor Residual dos Veículos (em proporção) seccionado pela variável Segmento nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

Referência de códigos da variável Segmento

Entrada	Hatch Pequeno	Hatch Médio	Sedan Pequeno	Sedan Compacto	Sedan Médio	Sedan Grande	SW	Monocab	SUV
0	1	2	3	4	5	6	7	8	9

Ainda no Gráfico 4, observamos que, no banco de dados de dois anos, os Sedans Compactos tendem a ter a menor desvalorização entre todos os segmentos, mas foi o segmento que apresentou a maior variabilidade dos dados visto a sua amplitude interquartilica. Já o segmento de Sedan Grande apresentou a mediana com maior desvalorização entre todas as categorias (35,47%). O boxplot evidência também que os segmentos cujos veículos tem um menor preço público, como Entrada, Hatch Pequeno, Sedan Pequeno e Sedan Compacto, apresentam medianas menores do que segmentos em que os veículos têm um preço público maior, como Hatch Médio, Sedan Médio, Sedan Grande, SW, Monocab e SUV. O comportamento da desvalorização dos segmentos Sedan Médio, SW e SUV neste banco de dados é bem próximo, motivo pelo qual esses segmentos foram agrupados (Gráfico 5).

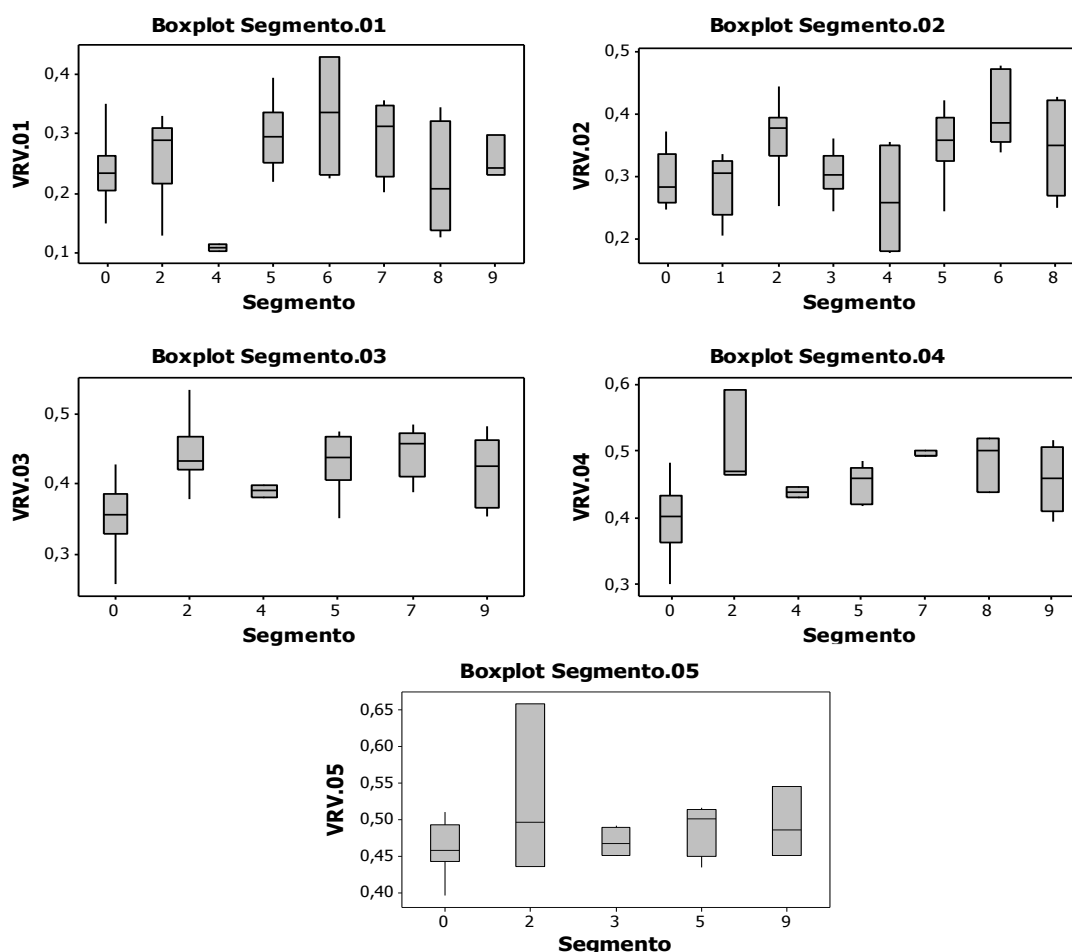
O comportamento da variável segmento no banco de dados de três anos é similar ao de dois anos, dado que os modelos de veículos pertencentes à categoria de Entrada, Hatch Pequeno e Sedan Pequeno, que são segmentos cujos veículos têm menor custo de aquisição, apresentaram uma mediana inferior a todos os demais segmentos. Isto reforça a ideia já exposta de que, quanto maior o preço de compra de um modelo de veículo, maior seria a depreciação desse modelo. Observando o comportamento dos segmentos de Entrada, Hatch Pequeno e Sedan Pequeno, no banco de dados de três anos, notamos uma semelhança entre eles e, por este motivo, estas três categorias foram agrupadas (Gráfico 5). A categoria Sedan Grande apresentou apenas um dado e, para que não se tenha prejuízo de sua análise, ela foi agregada a categoria dos Sedans Médios.

O Gráfico 4 evidencia que, após a utilização de um veículo pelo período de quatro anos, caso ele seja um veículo de Entrada, sua depreciação será similar a dos Hatchs Pequenos e Sedans Pequenos. Como essas três categorias apresentaram uma desvalorização similar para quatro anos, elas foram agrupadas em uma só categoria (Gráfico 5). A categoria de Hatch Pequeno apresentou o modelo de veículo com a menor desvalorização após quatro anos de utilização (33,02%). O modelo de veículo que teve a maior desvalorização (47,02%) pertence à categoria dos Hatch Médio. Novamente o Sedan Grande contou apenas com uma

observação, e a categoria SW, com duas observações bem similares em relação ao comportamento da desvalorização do Sedan Grande. Por motivo, as categorias Sedan Grande e SW foram agrupadas (Gráfico 5).

No banco de dados de cinco anos, notamos que a maior variabilidade de dados está na categoria dos Hatch Médio. Contudo, vale lembrar que esta categoria contou apenas com três observações. O segmento de Entrada apresenta a menor desvalorização mediana em cinco anos, enquanto os Sedans Médios apresentaram a maior desvalorização. O veículo com a maior desvalorização após cinco anos de uso está presente na categoria dos Hatchs Médios, já o veículo com a menor desvalorização em cinco anos pertence ao segmento de Entrada.

Gráfico 5 – Boxplot do Valor Residual dos Veículos (em proporção) seccionado pela variável Segmento, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

Referência de códigos da variável Segmento.01								
Entrada	Hatch Pequeno	Hatch Medio	Sedan Compacto	Sedan Medio	Sedan Grande	SW	Monocab	SUV
0	2	4	5	6	7	8	9	

Referência de códigos da variável Segmento.02							
Entrada	Hatch Pequeno	Hatch Medio	Sedan Pequeno	Sedan Compacto	Sedan Medio SW SUV	Sedan Grande	Monocab
0	1	2	3	4	5	6	8

Referência de códigos da variável Segmento.03						
Entrada	Hatch Pequeno	Hatch Medio	Sedan Compacto	Sedan Medio Sedan Grande	SW Monocab	SUV
0	2	4	5	7	9	

Referência de códigos da variável Segmento.04							
Entrada	Hatch Pequeno	Hatch Medio	Sedan Compacto	Sedan Medio	SW Sedan Grande	Monocab	SUV
0	2	4	5	7	8	9	

Referência de códigos da variável Segmento.05					
Entrada	Hatch Pequeno	Hatch Medio	Sedan Pequeno	Sedan Medio	SUV
0	2	3	5	9	

Elaboração Própria

Notamos que o segmento de Entrada e o Hatch Pequeno têm um comportamento muito próximo em relação à depreciação de seus modelos de veículos. Assim, essas duas categorias foram agrupadas (Gráfico 5). As categorias Sedan Pequeno e Sedan Compacto também foram agrupadas, visto o comportamento similar da depreciação entre as duas.

D) Variável Motorização: motorizações dos modelos de veículo (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 7 apresenta a distribuição de frequência das modelos de veículos segundo sua motorização. Os veículos cujo motor pertence à categoria “> 1.4 e ≤ 1.8” são os mais presentes em todos os bancos de dados e representam 60% das observações no banco de dados de um ano, e nos bancos de dois a cinco anos, representam, respectivamente, 56,8%, 55,1%, 47,1% e 50,0%. Apenas no banco de dados de quatro anos, essa categoria de motorização representou menos da metade dos dados.

As demais motorizações têm uma participação similar, oscilando pouco de um banco de dados para o outro. Os modelos de veículos com motorização “ ≥ 2.0 ” têm a segunda colocação nos bancos de dados um, dois e três anos. Já no banco de quatro anos, os modelos de veículo que assumem essa posição são aqueles cujo motor está na categoria “ > 1.0 e ≤ 1.4 ”. No banco de cinco anos, a segunda colocação ficou com os modelos de veículos da categoria “ ≤ 1.0 ”.

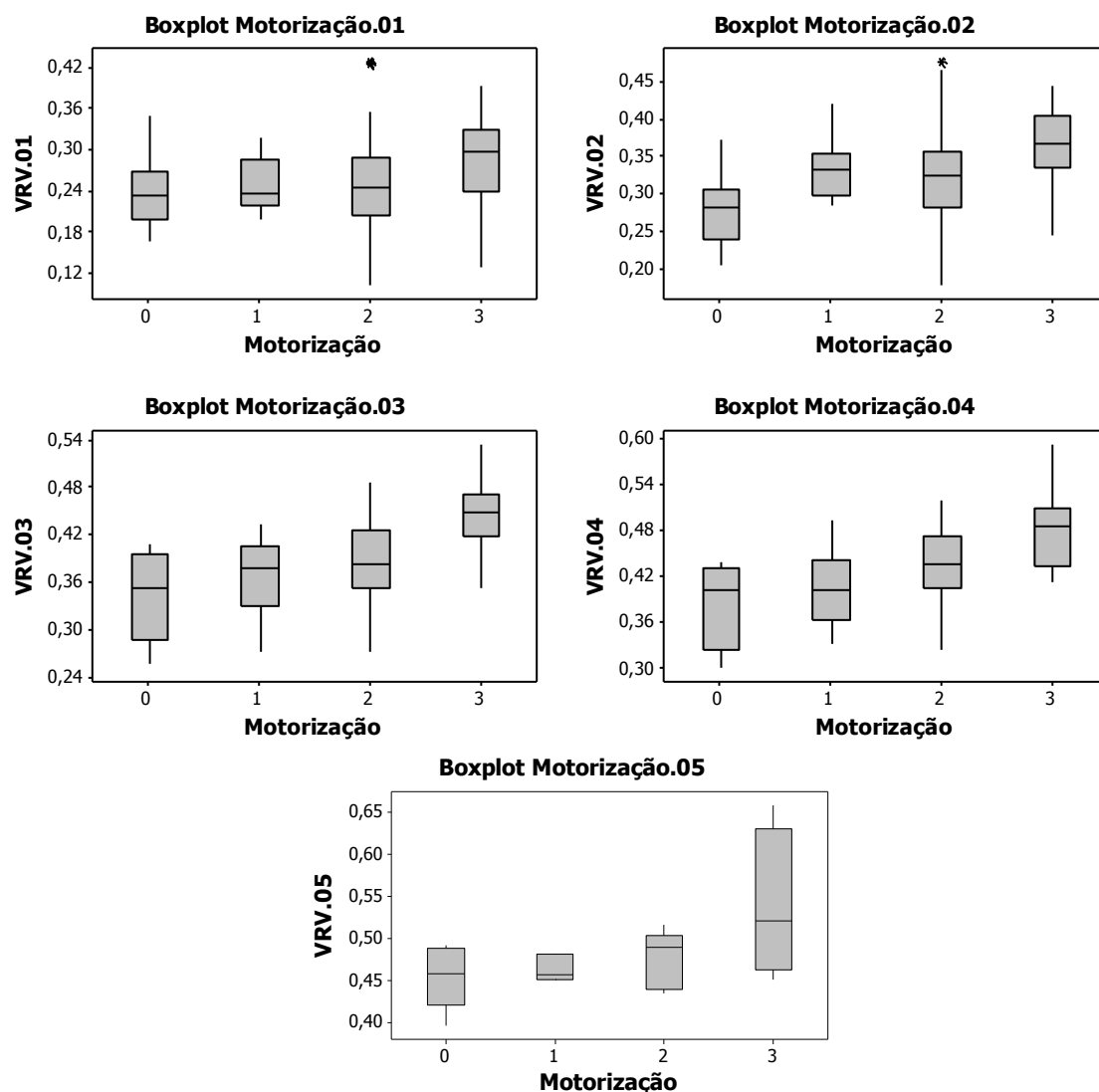
Tabela 7 – Tabela de frequência da variável Motorização nos bancos de dados de um a cinco anos de depreciação.

Frequência da variável Motorização do banco um ao cinco					
Motorização	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0 - " ≤ 1.0 "	67 (83,8%)	9 (10,2%)	56 (81,2%)	18 (35,3%)	8 (33,3%)
1 - " $> 1.0 \leq 1.4$ "	0 (0%)	62 (70,5%)	0 (0%)	0 (0%)	0 (0%)
2 - " $> 1.4 \leq 1.8$ "	0 (0%)	0 (0%)	0 (0%)	24 (47,1%)	12 (50%)
3 - " ≥ 2.0 "	13 (16,3%)	17 (19,3%)	13 (18,8%)	9 (17,6%)	4 (16,7%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

Na cultura popular, tem-se que veículos com maior motorização e, conseqüentemente, um maior preço de compra tendem a ter uma desvalorização maior. No Gráfico 6, notamos que existe embasamento nesta percepção, uma vez que há uma leve tendência de aumento da desvalorização à medida em que a motorização de um veículo aumenta. Esta tendência fica mais evidente com o passar dos anos de utilização do veículo. No banco de dados de um ano, existe uma semelhança no comportamento da desvalorização dos modelos de veículos que estão nas categorias de motorização “igual a 1.0”, “ > 1.0 e ≤ 1.4 ” e “ > 1.4 e ≤ 1.8 ”. Sendo assim, agrupamos estas categorias, ficando apenas com duas no banco de dados de um ano: motores até 1.8 e motores ≥ 2.0 (Gráfico 7).

Gráfico 6 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Motorização nos bancos de dados de um a cinco anos de depreciação



Referência de códigos da variável Motorização			
1.0	>1.0 <= 1.4	>1.4 <= 1.8	>=2.0
0	1	2	3

Elaboração Própria

Observando o banco de dados de dois anos, já se torna mais clara a tendência observada no banco de um ano, pois a mediana dos modelos de veículo cujo motor é igual a 1.0 é 28,21%, enquanto a de um veículo na categoria 2.0 é 36,77%. Nota-se também que as categorias de motorização “> 1.0 e ≤ 1.4” e “> 1.4 e

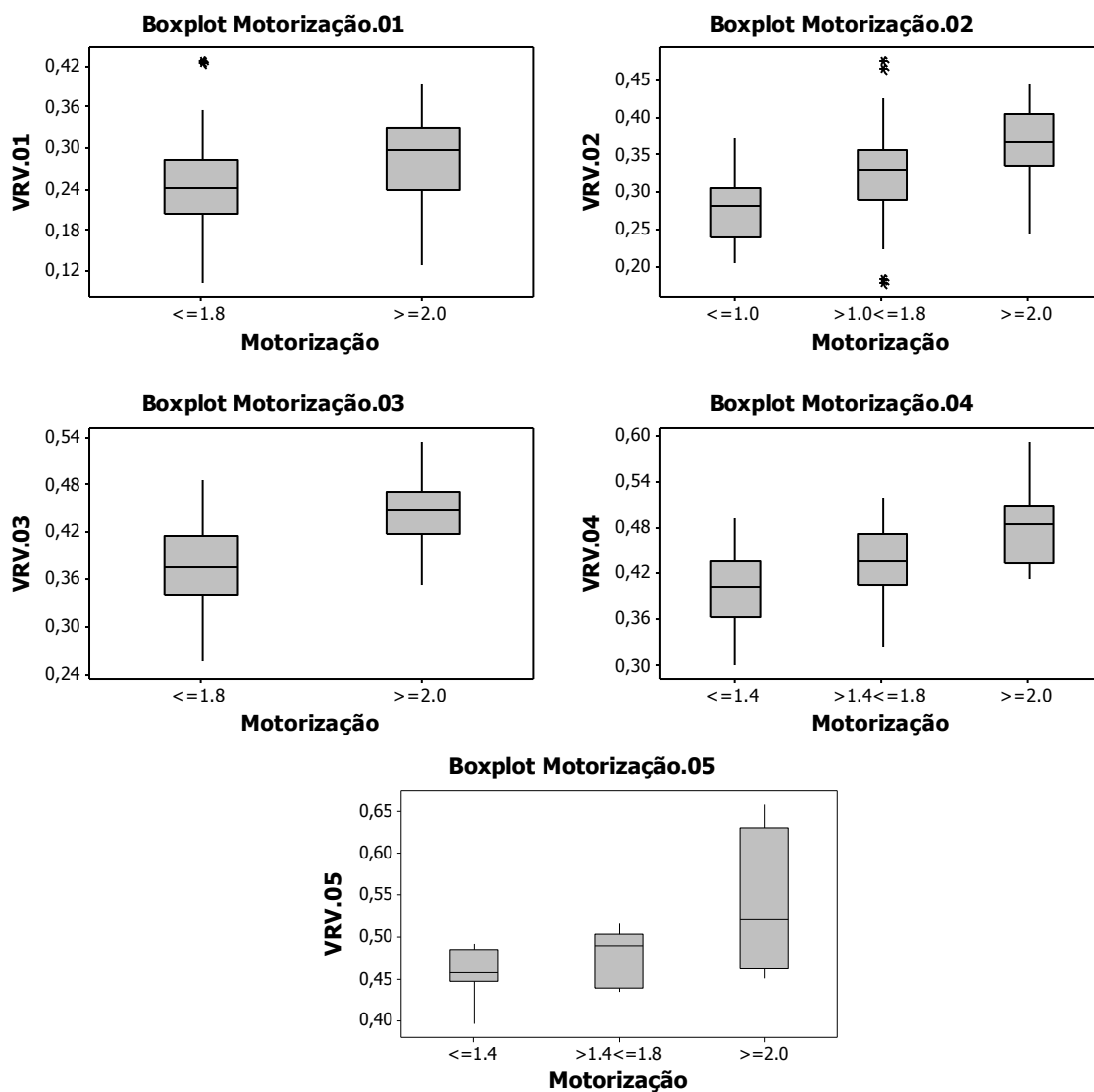
≤ 1.8 ” têm bastante similaridade e por isso foram agrupadas em um novo conjunto de motorização: “ > 1.0 e ≤ 1.8 ”.

No banco de dados de três anos, notamos que existe uma situação semelhante ao banco de dados de um ano, onde as três categorias de motorização, que são menores do que 2.0 têm o mesmo comportamento de desvalorização e, diante disto agrupamos, no banco de dados de três anos as categorias de motorização de forma similar ao banco de dados de um ano, considerando apenas duas categorias: “ ≤ 1.8 e ≥ 2.0 ”.

O modelo com a menor desvalorização após quatro anos (o HB20 Comfort Plus) pertence à categoria dos modelos de veículo até 1.0, cuja desvalorização foi de 29,97%. Já aquele com a maior desvalorização (o 308 2.0 Allure) está na categoria maior ou igual a 2.0, é com 59,32% de desvalorização. As categorias de motorização igual a 1.0 e “ >1.0 e ≤ 1.4 ” no banco de dados de quatro anos têm o mesmo comportamento. Por este motivo, foram agrupadas em “ ≤ 1.4 ”, conforme apresenta o Gráfico 7.

No banco de dados de cinco anos, seguimos com a mesma tendência já identificada nos bancos anteriores. O modelo de veículo com a menor desvalorização segue presente na categoria dos veículos com motor 1.0 e aquele que teve a maior desvalorização está na categoria dos modelos de veículos com motor ≥ 2.0 . Se observamos as categorias 1.0 e “ > 1.0 e ≤ 1.4 ”, notamos como a desvalorização das mesmas tem uma proximidade grande. Sendo assim, agrupamos estas duas em uma categoria apenas, sendo ela “ ≤ 1.4 ”.

Gráfico 7 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Motorização, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação



E) Variável Potência: potências dos modelos de veículo (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 8 apresenta a distribuição de frequência dos modelos de veículos segundo sua potência. Notamos que os modelos de veículo cuja potência é superior a 100 cv e menor ou igual a 130 cv são os mais presentes em todos os bancos de

dados, com a representatividade desses modelos de veículos chegando próximo a 40% dos dados em todos os bancos.

Outras duas categorias que também têm expressividade e se alternam entre segundo e terceiro lugar na representatividade dos bancos de dados é a dos modelos de veículos cuja potência é menor que 100 cv e aqueles com potência maior que 130 cv e menor ou igual a 160 cv, que, em média, oscilaram de 20 a 30% de representatividade nos bancos de dados. Notamos também que a categoria dos modelos de veículos com potência maior ou igual a 220 cv, apresentou apenas um dado nos quatro primeiros bancos, o veículo Fusion 2.0 Turbo, que não está presente no bando de dados de cinco anos. O fato de termos apenas uma observação nesta categoria dificulta a análise e interpretação, inclusive em uma abordagem via boxplot. Desta forma, vamos retirar essa categoria e migrar esta observação para a categoria anterior, que passa a ser dos veículos com potência acima de 190 cv.

Tabela 8 – Tabela de frequência da variável Potência nos bancos de dados de um a cinco anos de depreciação.

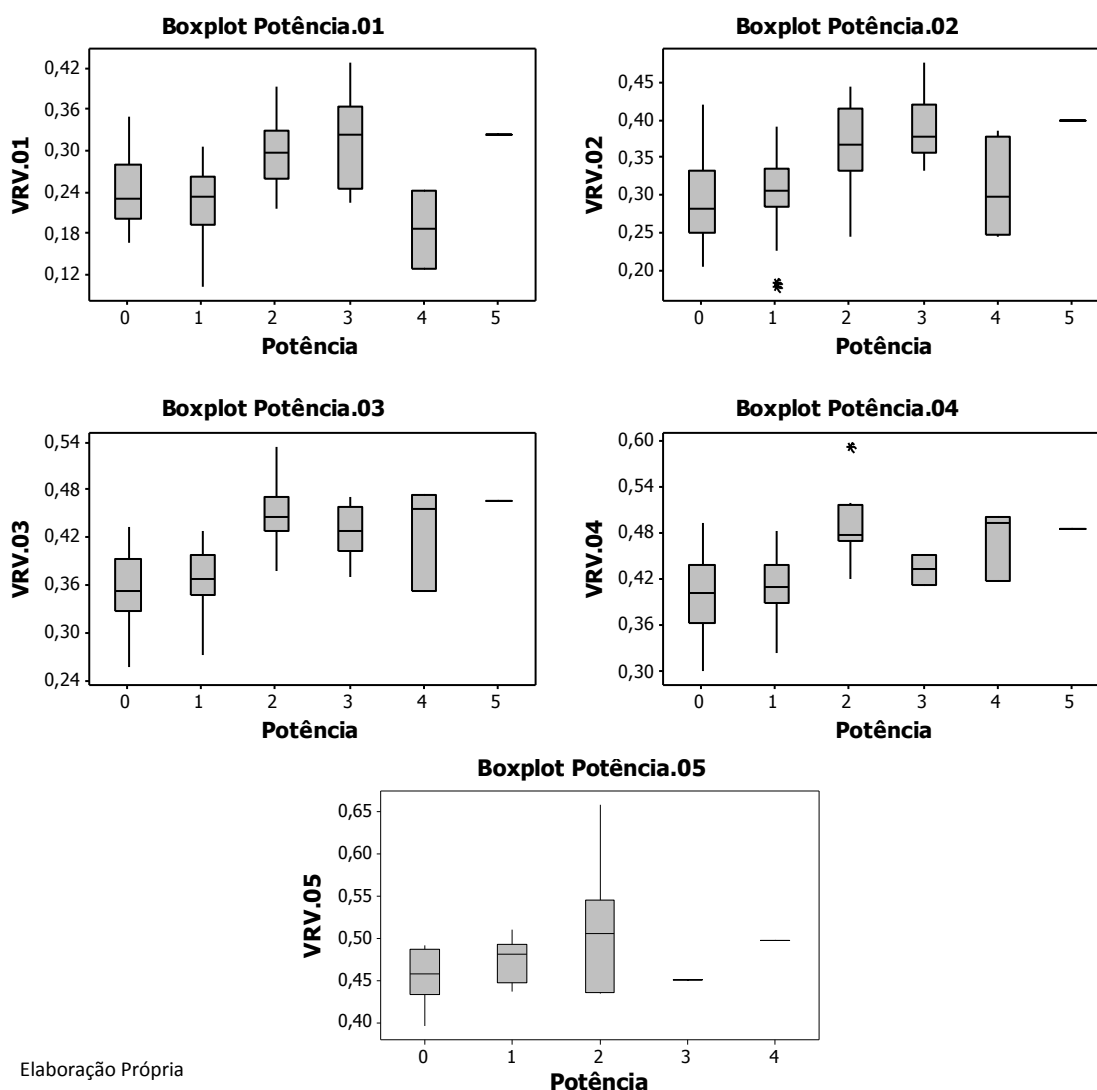
Frequência da variável Potência do banco um ao cinco					
Potência	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0 - " ≤ 100 "	51 (63,8%)	53 (60,2%)	45 (65,2%)	34 (66,7%)	15 (62,5%)
1 - " $> 100 \leq 130$ "	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
2 - " $> 130 \leq 160$ "	16 (20%)	30 (34,1%)	20 (29%)	11 (21,6%)	7 (29,2%)
3 - " $> 160 \leq 190$ "	10 (12,5%)	0 (0%)	0 (0%)	2 (3,9%)	2 (8,3%)
4 - " $> 190 \leq 220$ "	3 (3,8%)	5 (5,7%)	4 (5,8%)	4 (7,8%)	0 (0%)
5 - " ≥ 2.0 "	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

O Gráfico 8 apresenta a distribuição dos valores residuais segundo a potência do modelo de veículo para os bancos de dados de um a cinco anos. Analisando a distribuição dos dados do banco do ano um, é possível notar que o valor residual mediano da categoria " > 190 cv e ≤ 220 cv" é o menor dentre todas as categorias de potência, apresentando uma desvalorização de -18,52%, contudo, o banco de dados contém apenas duas observações para essa categoria. Já a categoria " > 100 cv e \leq

130 cv” apresentou uma mediana de -23,23% e é composta pelo veículo com a menor desvalorização em um ano (-10,14%), além de ter um comportamento de desvalorização similar aos modelos de veículo com potência menor do que 100 cv, levando a criarmos um novo agrupamento, a categoria “≤ 130 cv”. Já a categoria dos veículos com potência maior do que 220 cv apresentou apenas uma observação e, como já mencionado, vamos agrupar na categoria anterior e criar uma nova para os modelos de veículo com potência acima de 190 cv.

Gráfico 8 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Potência nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

Referência de códigos da variável Potência					
< 100	> 100 <= 130	> 130 <= 160	> 160 <= 190	> 190 <= 220	> 220
0	1	2	3	4	5

No banco de dados de dois anos, a categoria de potência que apresenta maior variabilidade dos dados é a “>190 cv e ≤220 cv” devido a ter a maior amplitude interquartílica. Já as categorias “>130 cv e ≤ 160 cv”, “> 160 cv ≤190 cv” e “> 220 cv” são as que apresentaram as maiores medianas, tendo uma maior desvalorização no período de dois anos. Neste banco de dados, agrupamos por questão de comportamento similar de desvalorização as categorias “< 100 cv” e “> 100 cv e ≤ 130”, que se tornou a categoria “≤ 130 cv”. Foram também agrupadas as categorias “> 130 cv e ≤ 160 cv” e “> 160 cv e ≤ 190 cv”, e a categoria “> 220 cv” transformou-se em “> 190 cv”, como ocorrido no banco de um ano. O Gráfico 9 apresenta a distribuição dos valores residuais usando a nova categorização para a variável potência.

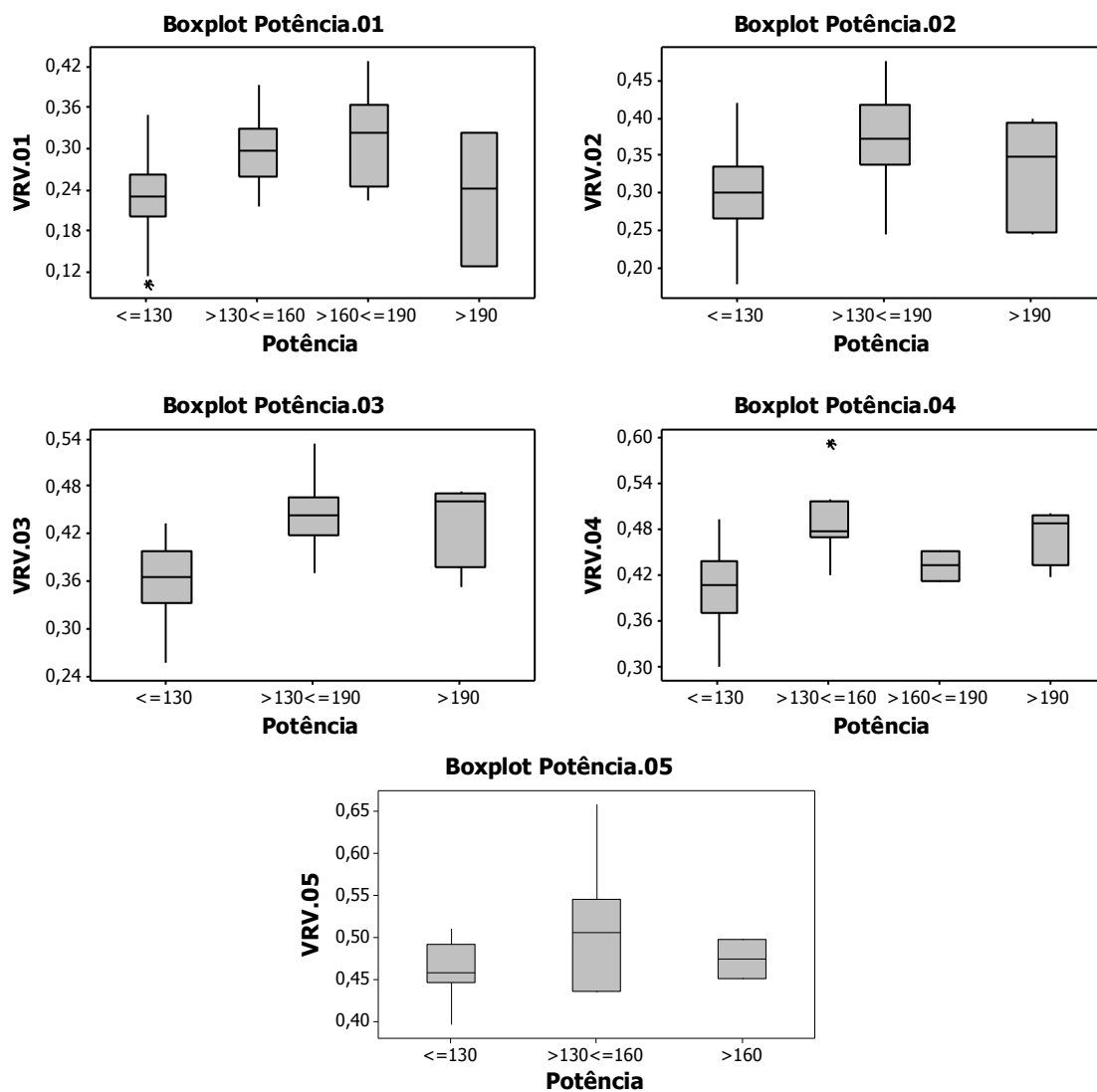
No banco de dados de três anos, o modelo de veículo com a menor desvalorização pertence a categoria “<100 cv”, cuja desvalorização foi de 25,6%. Já o modelo de veículo com a maior desvalorização (53,50%) foi na categoria “> 130 cv e ≤160 cv”. Notamos neste banco o mesmo comportamento apresentado no banco de dados de dois anos e fizemos as mesmas adequações em relação às categorias.

No banco de dados de quatro anos, identificamos que modelos de veículos cuja potência é menor que 130 cv também apresentam similaridade na desvalorização: a mediana da categoria “≤ 100 cv” ficou em 40,16%, enquanto a mediana da próxima categoria (“>100 cv e ≤130 cv” é 41,10%. Desta forma também agrupamos essas categorias neste banco. Foi observado um dado atípico na categoria “> 130 cv e ≤160 cv”, que também é o veículo com a maior desvalorização no período de quatro anos (308 2.0 Allure). O Gráfico 9 apresenta também o agrupamento da categoria “> 220 cv” com a anterior, “> 190 cv e < 220 cv”, como já exposto na análise desta variável.

Analisando a distribuições dos dados do banco de cinco anos (Gráfico 8), notamos que a menor mediana é do a modelo de veículo pertencente à categoria “>160 cv e ≤190 cv” e os modelos de veículo cuja potência está entre “>130 cv e ≤160 cv” apresentaram a maior variabilidade e a amplitude interquartílica reforça esta informação. Novamente foram agrupadas as duas primeiras categorias, sendo

criada a categoria " ≤ 130 cv". Agrupamos também as duas últimas categorias, lembrando que este banco não contou com nenhum modelo de veículo cuja potência fosse maior que 220 cv.

Gráfico 9 – Boxplot do Valor Residual dos Veículos (em proporção) segmentado pela variável Potência, após agrupamento das categorias, nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

F) Variável Nº Itens Segurança: número de Itens de Segurança dos modelos de veículos (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 9 apresenta a distribuição dos modelos de veículos segundo o número de itens de segurança. Os modelos de veículos que contam com um ou dois itens de segurança são aqueles com maior participação nos bancos de dados de um, dois, três e quatro anos. No bando de dados de cinco anos, permanece com maior representatividade os modelos de veículos com dois itens de segurança.

Observando o banco de dados de um ano, notamos que os modelos de veículo com oito itens de segurança são a terceira categoria com em número de observações, tendo uma representatividade de 17,5%. Esses modelos de veículo se mantêm em terceiro lugar no banco de dados de dois e três anos, contudo, a partir do banco de quatro anos, ocorre uma redução do número de observações, chegando a ficar sem nenhum dado no banco de dados de cinco anos.

Veículos que não possuem nenhum item de segurança não estão presentes no banco de dados de um e dois anos, pois, dentre os itens elencados como de segurança, apenas veículos mais antigos não possuem ao menos um deles. Os modelos de veículos que possuem de dois a sete itens de segurança têm menor presença nos bancos de dados. O banco de dados de três anos não contou com nenhum modelo com seis itens de segurança e o de quatro e cinco anos com nenhum de cinco ou seis itens.

Tabela 9 – Tabela de frequência da variável N° Itens Segurança nos bancos de dados de um a cinco anos de depreciação.

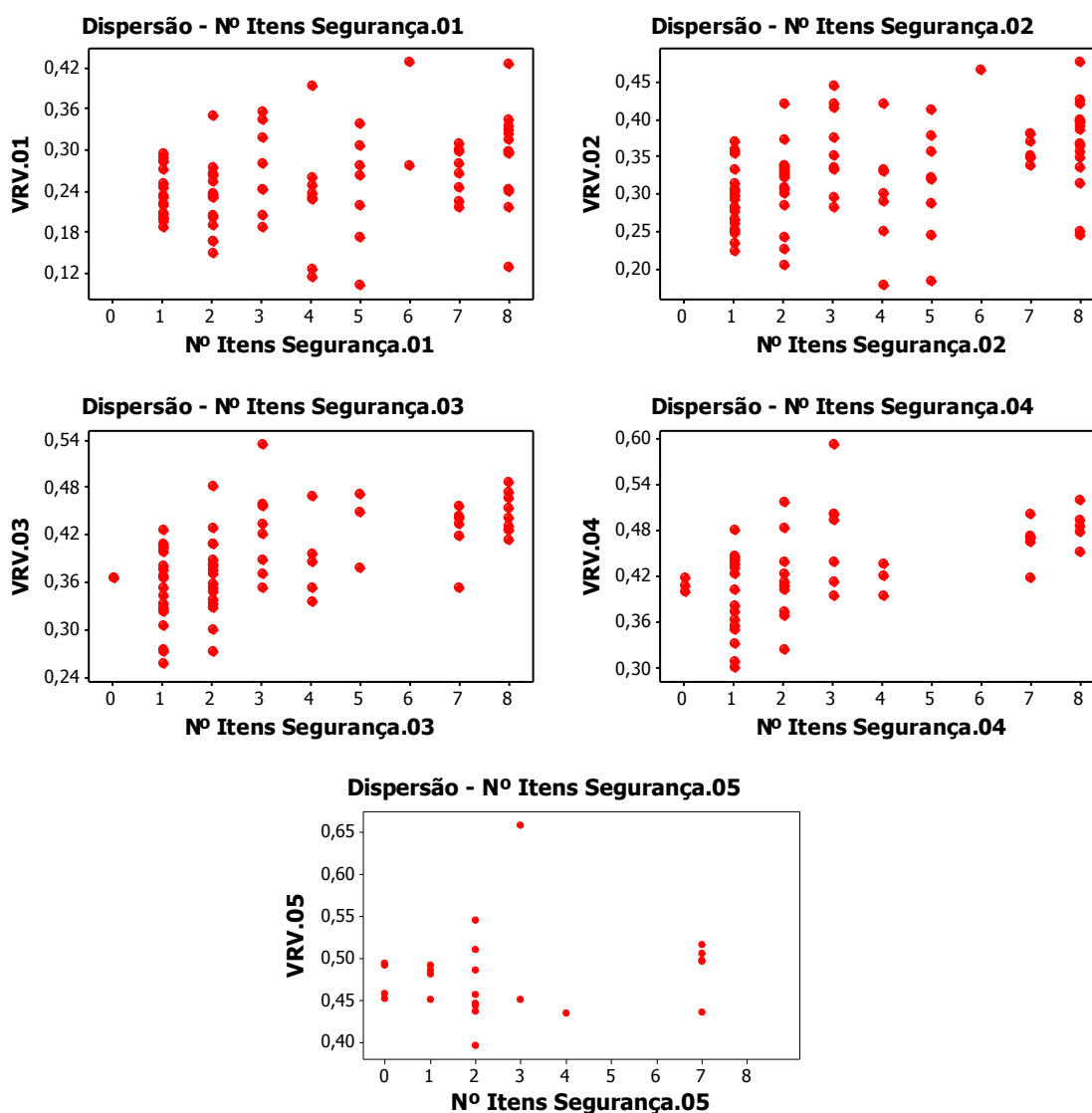
Frequência da variável N° Itens Segurança do banco um ao cinco					
N° Itens Segurança	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0	0 (0%)	0 (0%)	2 (2,9%)	3 (5,9%)	4 (16,7%)
1	18 (22,5%)	23 (26,1%)	19 (27,5%)	17 (33,3%)	4 (16,7%)
2	15 (18,8%)	17 (19,3%)	17 (24,6%)	10 (19,6%)	8 (33,3%)
3	8 (10%)	9 (10,2%)	8 (11,6%)	7 (13,7%)	2 (8,3%)
4	8 (10%)	7 (8%)	5 (7,2%)	3 (5,9%)	1 (4,2%)
5	7 (8,8%)	8 (9,1%)	3 (4,3%)	0 (0%)	0 (0%)
6	2 (2,5%)	1 (1,1%)	0 (0%)	0 (0%)	0 (0%)
7	8 (10%)	7 (8%)	6 (8,7%)	6 (11,8%)	5 (20,8%)
8	14 (17,5%)	16 (18,2%)	9 (13%)	5 (9,8%)	0 (0%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

O Gráfico 10 apresenta a distribuição dos valores residuais segundo o número de itens de segurança para os bancos de dados de um a cinco anos. Analisando o Gráfico 10, notamos que a maior parte dos dados está concentrada nos modelos de veículos que possuem até quatro itens de segurança, o que acontece em praticamente todos os bancos de dados. Nos bancos de dados de um e dois anos, é notada uma baixa variabilidade nos dados para os modelos de veículos que possuem apenas um item de segurança e essa variabilidade aumenta na medida em que se têm mais itens de segurança. Logo os modelos veículos com um número menor de itens de segurança tendem a ter uma desvalorização menor, já os modelos de veículos que possuem mais itens de segurança podem ter uma alta desvalorização como podem chegar a desvalorizar menos que os modelos de veículos com poucos itens. No banco de dados de três anos, essa variabilidade se reduz um pouco e é notada uma tendência ao aumento da desvalorização na medida em que se aumenta o número de itens de segurança. Essa ideia se reforça, pois veículos com um maior número de itens de segurança tem uma predisposição a ter um preço de aquisição maior e, como já vimos, esses modelos de veículos têm uma tendência a depreciar mais do que veículos com um menor custo de aquisição. Como já exposto, existe uma concentração de dados nos modelos de veículo que possuem até quatro itens de segurança e, nos bancos de dados de quatro e cinco anos, esta afirmação ganha mais força. Nestes dois bancos, também notamos a

presença de uma observação atípica (o Peugeot 308 2.0 Allure), com três itens de segurança um modelo de veículo destoa dos demais. A tendência à desvalorização na medida em que se aumenta o número de itens de segurança é notada no banco de dados de quatro anos, Já no bando de ados de cinco anos isto não acontece e os dados estão dispersos sem nenhum padrão visível.

Gráfico 10 – Dispersão do Valor Residual dos Veículos (em proporção) segundo a variável Nº Itens de Segurança nos bancos de dados de um a cinco anos de depreciação



Elaboração Própria

G) Variável Nº Itens Conforto: número de Itens de Conforto dos modelos de veículos (em %) nos bancos de dados de um a cinco anos de depreciação

A Tabela 10 apresenta a distribuição dos modelos de veículos segundo o número de itens de conforto. É possível perceber que os bancos de dados de um, dois e três anos não possuem nenhum modelo de veículo que não tenha ao menos um item de conforto.

Os modelos de veículos com seis itens de conforto têm a maior representatividade no banco de dados de um e dois anos (22,5% e 23,9%, respectivamente), seguidos pelos modelos de veículo com um e dois itens de conforto. Nos bancos de três, quatro e cinco anos, os modelos de veículos que possuem um item de conforto passam a ser os mais representativos. Os demais modelos de veículos com, três, quatro, cinco e sete itens de conforto apresentaram um volume menor de dados em praticamente todos os bancos.

Tabela 10 – Tabela de frequência da variável Nº Itens Conforto nos bancos de dados de um a cinco anos de depreciação.

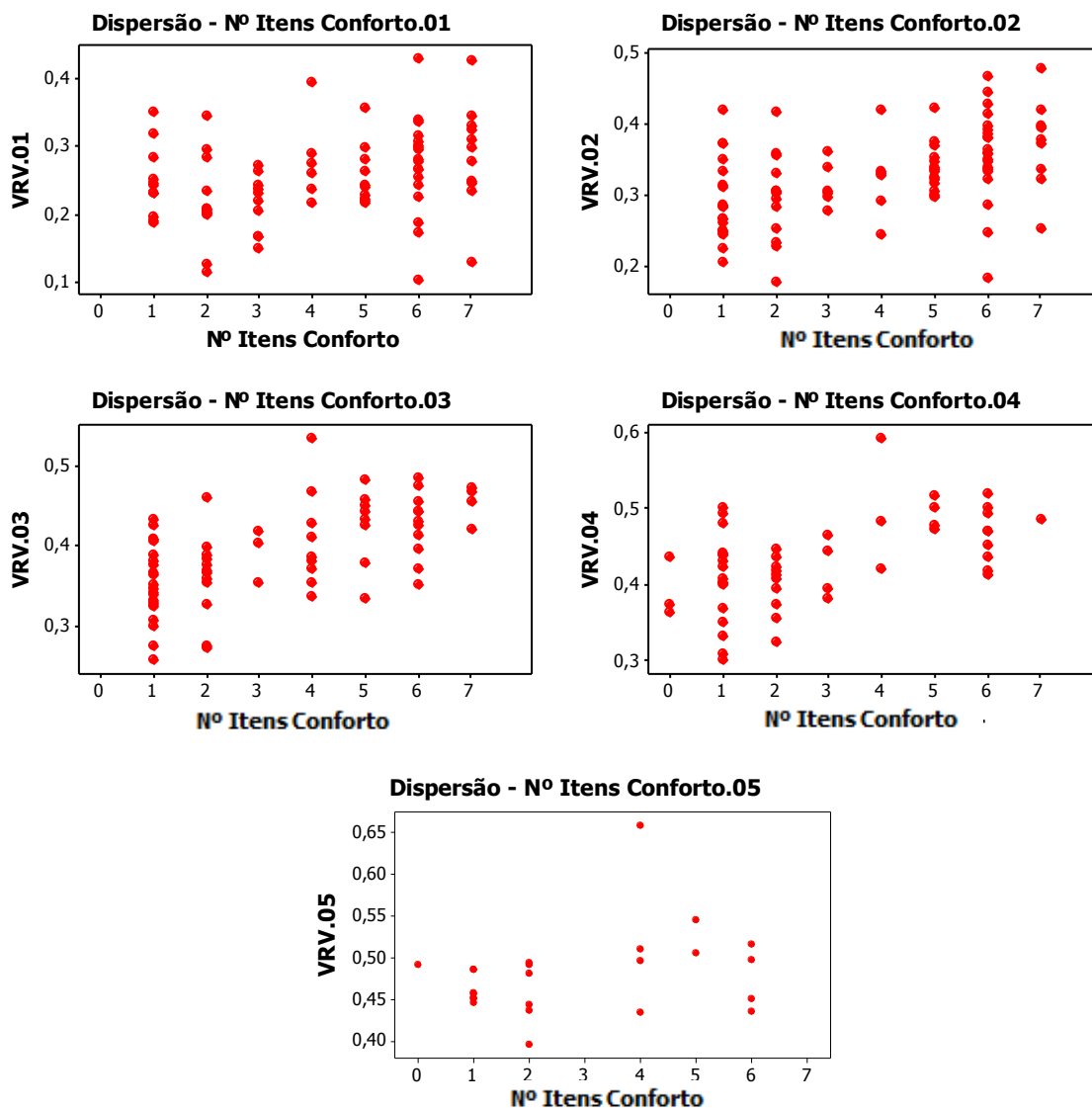
Frequência da variável Nº Itens Conforto do banco um ao cinco					
Nº Itens Conforto	1 Ano	2 Anos	3 Anos	4 Anos	5 Anos
0	0 (0%)	0 (0%)	0 (0%)	3 (5,9%)	1 (4,2%)
1	12 (15%)	19 (21,6%)	22 (31,9%)	17 (33,3%)	7 (29,2%)
2	13 (16,3%)	13 (14,8%)	12 (17,4%)	10 (19,6%)	6 (25%)
3	10 (12,5%)	6 (6,8%)	3 (4,3%)	4 (7,8%)	0 (0%)
4	6 (7,5%)	6 (6,8%)	9 (13%)	3 (5,9%)	4 (16,7%)
5	10 (12,5%)	14 (15,9%)	8 (11,6%)	4 (7,8%)	2 (8,3%)
6	18 (22,5%)	21 (23,9%)	11 (15,9%)	9 (17,6%)	4 (16,7%)
7	11 (13,8%)	9 (10,2%)	4 (5,8%)	1 (2%)	0 (0%)
Total	80 (100%)	88 (100%)	69 (100%)	51 (100%)	24 (100%)

Elaboração Própria

O observando o Gráfico 11 apresenta a distribuição dos valores residuais dos veículos segundo seu número de itens de conforto. Notamos que não existe uma concentração de dados nos bancos de um e dois anos, pois eles apresentam uma dispersão bem simétrica e, apenas no banco de dados de dois anos, é possível notar uma leve tendência que aponta para o aumento da desvalorização em função do aumento do número de itens de conforto. Esta tendência é reforçada e se torna

mais nítida no banco de dados de três anos, que já demonstra uma maior concentração de modelos de veículo com até quatro itens de conforto. O comportamento dos dados do banco de quatro anos é bem similar ao do banco de três anos. No banco de quatro anos, notamos a presença de um dado atípico contendo quatro itens de conforto, que se repete no banco de dados de cinco anos. O diagrama de dispersão dos dados do banco de cinco anos não apresentou nenhum padrão.

Gráfico 11 – Dispersão do Valor Residual dos Veículos (em proporção) segundo a variável Nº Itens de Conforto nos bancos de dados de um a cinco anos de depreciação



H) Correlação entre as variáveis Nº Itens de Segurança, Nº de Itens de conforto e as variáveis respostas nos bancos de dados de um a cinco anos.

Notamos na tabela 11 que existe uma correlação positiva e estatisticamente significativa nos bancos de dados de um a quatro anos entre as variáveis candidatas a explicativas Nº de Itens de Segurança e Nº de Itens de Conforto com as variáveis resposta em cada um dos anos de utilização. Ela parte de uma correlação fraca e, com o passar dos anos de utilização, chega um nível moderado. Isso reforça o conceito de que, com o passar dos anos, os veículos tendem a se depreciar mais. No banco de dados de cinco anos, a correlação se mantém positiva, mas perde força e pelo seu p-valor se mostra estatisticamente não significativa.

Tabela 11 – Tabela de correlação das variáveis Nº Itens de Segurança, Nº Itens de conforto nos bancos de dados de um a cinco anos de depreciação.

Correlações entre as variáveis Nº Itens de Segurança e Conforto e as variáveis resposta					
	VRV.01	VRV.02	VRV.03	VRV.04	VRV.05
Nº Itens de Segurança	0,318	0,416	0,564	0,499	0,128
p-valor	0,004	0,000	0,000	0,000	0,550
Nº Itens de Conforto	0,319	0,479	0,593	0,532	0,279
p-valor	0,004	0,000	0,000	0,000	0,186

4.2 Análise dos Modelos de regressão

Nesta seção iremos construir os cinco modelos, um para cada banco de dados, a fim de estimar qual é o preço de venda de um veículo, em um determinado prazo de utilização, dado suas características.

Iremos utilizar as seis variáveis já apresentadas para a elaboração dos modelos, contudo para as variáveis segmento, potência e motorização, iremos trabalhar com os agrupamentos após a análise das mesmas na seção de análise descritivas, ou seja, com seu novo agrupamento.

A abordagem adotada para a construção dos modelos foi a de saída uma-a-uma. Assim, se faz necessário avaliar a presença de multicolinearidade logo no início, pois ela pode afetar os testes que determinam a saída das variáveis. Após a construção do primeiro modelo, vamos utilizar um procedimento similar à seleção

backward. Iremos avaliar o modelo construído e retirar as variáveis que forem consideradas estatisticamente não significantes, mas uma a uma, tendo como critério de escolha aquelas com maior p-valor. Desta forma iremos retirar uma variável, construir o modelo novamente e avaliar, repetindo esse processo até que se atinja um modelo em que com todas as variáveis sejam consideradas estatisticamente não-nulas e que atenda aos pressupostos necessários a um bom ajuste.

4.2.1 Modelo de regressão para o banco de dados de um ano

O Quadro 5 apresenta o modelo inicial para o banco de dados de um ano, considerando todas as variáveis. Vale ressaltar que, para esse banco, as categorias de segmento, Entrada, Hatch Pequeno e Sedan Pequeno foram agrupadas; as de potência “ ≤ 100 ” e “ $>100 \leq 130$ ” se tornaram “ ≤ 130 ”, e as categorias “ $>190 \leq 220$ ” e “ >220 ”, foram agrupadas em “ >190 ”; e motorização, com as categorias “1.0”, “ $>1.0 \leq 1.4$ ” e “ $>1.4 \leq 1.8$ ” se tornando “ ≤ 1.8 ”. Neste primeiro modelo a categoria de referência difere do que foi apresentado na seção de descrição de variáveis devido aos agrupamentos realizados. Observando o Quadro 5, notamos que a montadora Ford, as categorias de segmento Entrada, Hatch Pequeno e Sedan Pequeno, a categoria de motorização “ ≤ 1.8 ” e a categoria de potência “ ≤ 130 ” não possuem coeficientes, isto porque estas formam as categorias de referência, representadas pelo β_0 ou intercepto do modelo.

Quadro 5 – Modelo de regressão (1.1) da variável resposta VRV.01 (banco de dados de um ano).

VRV.01 = 0,230 - 0,00687 N°_Itens_Segurança.01 + 0,00796 N°_Itens_Conforto.01
 - 0,0096 M-Volkswagen.01 - 0,0308 M-Toyota.01 + 0,0276 M-Fiat.01
 - 0,0774 M-Hyundai.01 - 0,0194 M-GM.01 + 0,0267 M-Renault.01
 + 0,0318 M-Nissan.01 - 0,0414 M-Citroen.01 - 0,0066 M-Peugeot.01
 + 0,0441 M-Audi.01 - 0,0698 M-Honda.01 - 0,168 M-BMW.01
 - 0,0162 Seg-Hatch Médio.01 - 0,0527 Seg-Sedan Compacto.01
 + 0,0081 Seg-Sedan Médio.01 + 0,0904 Seg-Sedan Grande.01
 + 0,0369 Seg-SpaceWagon.01 + 0,0020 Seg-Monocab.01 + 0,0094 Seg-SUV.01
 + 0,0202 Mot->= 2.0.01 + 0,0571 Pot-> 130 < = 160.01
 + 0,0597 Pot-> 160 < = 190.01 - 0,0108 Pot-> 190

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,22978	0,01798	12,78	0,000	
N°_Itens_Segurança.01	-0,006874	0,005790	-1,19	0,240	12,787
N°_Itens_Conforto.01	0,007961	0,003784	2,10	0,040	3,537
M-Volkswagen.01	-0,00957	0,01820	-0,53	0,601	2,215
M-Toyota.01	-0,03083	0,02245	-1,37	0,175	1,971
M-Fiat.01	0,02759	0,01927	1,43	0,158	2,290
M-Hyundai.01	-0,07744	0,02396	-3,23	0,002	1,537
M-GM.01	-0,01937	0,02162	-0,90	0,374	2,373
M-Renault.01	0,02667	0,02526	1,06	0,296	1,709
M-Nissan.01	0,03176	0,02649	1,20	0,236	1,880
M-Citroen.01	-0,04137	0,03181	-1,30	0,199	1,391
M-Peugeot.01	-0,00664	0,02484	-0,27	0,790	1,653
M-Audi.01	0,04413	0,02706	1,63	0,109	3,298
M-Honda.01	-0,06980	0,03260	-2,14	0,037	3,512
M-BMW.01	-0,16816	0,04545	-3,70	0,001	2,840
Seg-Hatch Médio.01	-0,01621	0,03494	-0,46	0,644	7,530
Seg-Sedan Compacto.01	-0,05270	0,04249	-1,24	0,220	2,482
Seg-Sedan Médio.01	0,00810	0,03493	0,23	0,817	9,365
Seg-Sedan Grande.01	0,09043	0,04912	1,84	0,071	6,463
Seg-SpaceWagon.01	0,03688	0,02269	1,63	0,110	1,380
Seg-Monocab.01	0,00200	0,02890	0,07	0,945	2,237
Seg-SUV.01	0,00941	0,03536	0,27	0,791	2,545
Mot->= 2.0.01	0,02023	0,02186	0,93	0,359	3,669
Pot-> 130 < = 160.01	0,05709	0,02246	2,54	0,014	4,553
Pot-> 160 < = 190.01	0,05970	0,02715	2,20	0,032	4,548
Pot-> 190	-0,01080	0,03541	-0,31	0,761	2,552

S = 0,0376629 R-Sq = 76,7% R-Sq(adj) = 65,9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	25	0,252036	0,010081	7,11	0,000
Residual Error	54	0,076599	0,001418		
Total	79	0,328635			

Elaboração Própria

Vamos investigar inicialmente a multicolinearidade do modelo de regressão 1.1 e para isso iremos nos valer do cálculo dos VIFs (*Variance Inflation Factor*) de

cada variável. Variáveis que apresentem um valor de VIF superior a cinco podem apresentar problemas de multicolinearidade, se fazendo necessário então retirar a variável com o maior valor e elaborar o modelo novamente para avaliar o impacto nas demais variáveis. A variável que se destaca nesta primeira avaliação é Nº de Itens de Seguraça.01 , que apresentou um VIF de 12,787. Logo, vamos retirá-la do modelo e, então, construir o modelo novamente, até que nenhuma outra variável possua um VIF superior a cinco. Outra forma de se avaliar a multicolinearidade é observando a análise de variância do modelo. Nela está presente o p-valor de estatística do teste F, que tem como hipótese nula a afirmação de que nenhuma variável do modelo é estatisticamente diferente de zero. Caso o resultado do teste F caia em contradição com os testes T das variáveis do modelo, ou seja, rejeitamos a hipótese nula na ANOVA, mas o teste T não consegue apontar nenhuma variável como sendo estatisticamente significativa para explicar a resposta, então teremos um indício de multicolinearidade.

Analisando o resultado do p-valor da estatística do teste F presente na análise de variância, temos evidências suficientes para rejeitar a hipótese nula em favor da alternativa. E como algumas categorias de variáveis se apresentaram estatisticamente significantes no teste T, não temos contradição entre o teste F da análise de variância e todos os testes T das variáveis presentes no modelo.

Observamos também que várias variáveis têm o seu p-valor da estatística do teste T, acima de um nível de significância de 5%. Após esgotarmos a possibilidade de multicolinearidade, o segundo passo foi retirar do modelo as variáveis que são estatisticamente nulas e manter apenas aquelas estatisticamente significativas. Este procedimento foi realizado de forma similar ao método *Backward*. Após esse procedimento, chegamos ao modelo de regressão presente no Quadro 6, o modelo 1.2, que não apresentou multicolinearidade e todas as variáveis preditoras são estatisticamente significantes.

Quadro 6 – Modelo de regressão (1.2) da variável resposta VRV.01 (banco de dados de um ano).

```

VRV.01 = 0,209 + 0,00626 N°_Itens_Conforto.01 - 0,0315 Seg-Hatch Médio.01
+ 0,104 Seg-Sedan Grande.01 + 0,0534 Pot-> 130 < = 160.01
+ 0,0739 Pot-> 160 < = 190.01 + 0,0423 M-Fiat.01 - 0,0658 M-Hyundai.01
+ 0,0435 M-Renault.01 + 0,0555 M-Nissan.01 - 0,0904 M-Honda.01
- 0,192 M-BMW.01

```

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,20907	0,01104	18,94	0,000	
N°_Itens_Conforto.01	0,006260	0,002561	2,44	0,017	1,630
Seg-Hatch Médio.01	-0,03155	0,01515	-2,08	0,041	1,424
Seg-Sedan Grande.01	0,10399	0,03161	3,29	0,002	2,692
Pot-> 130 < = 160.01	0,05342	0,01236	4,32	0,000	1,387
Pot-> 160 < = 190.01	0,07394	0,01854	3,99	0,000	2,133
M-Fiat.01	0,04234	0,01383	3,06	0,003	1,187
M-Hyundai.01	-0,06580	0,01989	-3,31	0,002	1,066
M-Renault.01	0,04346	0,01989	2,18	0,032	1,066
M-Nissan.01	0,05553	0,02019	2,75	0,008	1,098
M-Honda.01	-0,09045	0,01790	-5,05	0,000	1,065
M-BMW.01	-0,19206	0,03756	-5,11	0,000	1,950

S = 0,0375562 R-Sq = 70,8% R-Sq(adj) = 66,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	11	0,232723	0,021157	15,00	0,000
Residual Error	68	0,095912	0,001410		
Lack of Fit	31	0,063338	0,002043	2,32	0,008
Pure Error	37	0,032573	0,000880		
Total	79	0,328635			

Elaboração Própria

No modelo inicial 1.1 tínhamos vinte e cinco variáveis preditoras, mas, após efetuarmos o procedimento de retirada das variáveis, conseguimos construir um modelo com onze variáveis preditoras. Para a construção do modelo 1.2 as variáveis foram sendo retiradas na seguinte ordem:

1. N° Itens de Segurança.01
2. Seg-Hatch Médio.01 (Após a retirada desta segunda variável eliminamos a multicolinearidade do modelo, nenhuma variável apresentou VIF maior que 5)

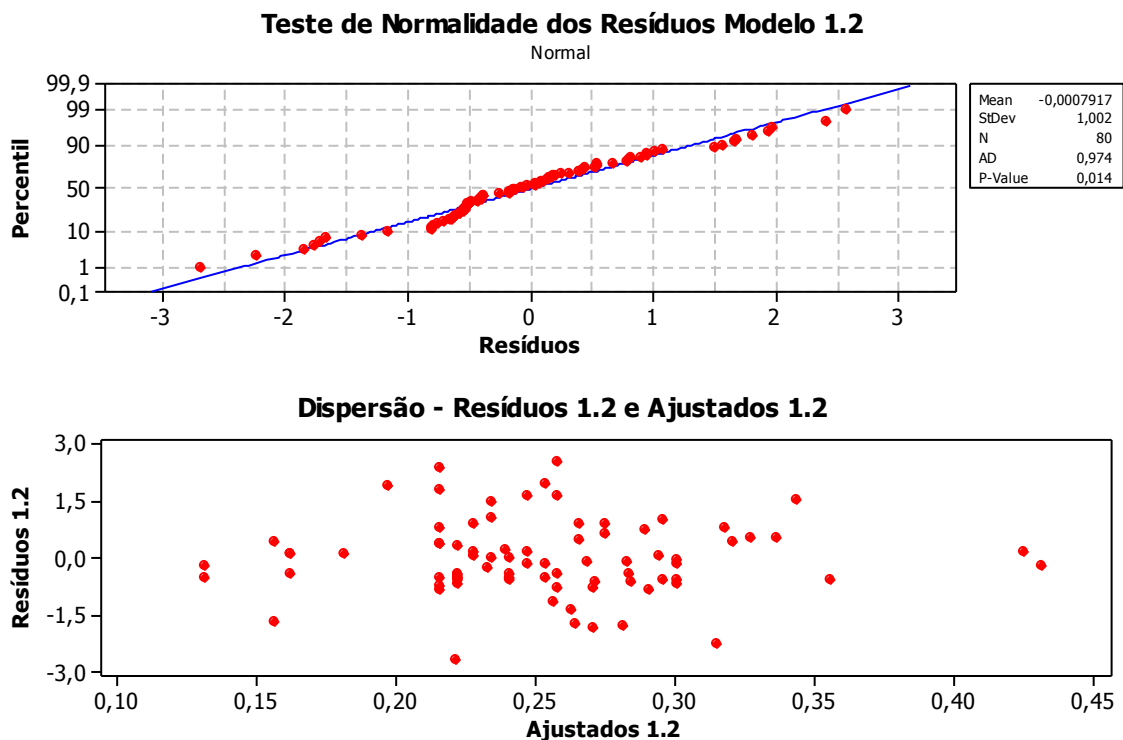
3. M-Peugeot.01 (A partir desta variável fomos eliminando pelo critério do maior p-valor)
4. Seg-SUV.01
5. M-Volkswagen.01
6. Seg-Monocab.01
7. M-GM.01
8. M-Citroen.01
9. M-Toyota.01
10. Pot-> 190
11. Mot- \geq 2.0.01
12. Seg-Sedan Compacto.01
13. Seg-SpaceWagon.01
14. M-Audi.01

Observando o Quadro 6, notamos que não há contradição entre os resultados dos testes F e T individuais para cada variável explicativa, o que afasta a hipótese de existência de multicolinearidade entre as variáveis deste modelo. O coeficiente de determinação ajustado nos evidencia que 66,1% da variância do valor residual dos veículos no primeiro ano de uso é explicada pelas variáveis preditoras, o que, para o caso em questão, pode ser considerado satisfatório. O Modelo 1.2 não teve a presença das variáveis Nº de Itens de Segurança e Motorização.

Uma vez atendido o pressuposto de ausência de multicolinearidade e tendo todas as variáveis explicativas estatisticamente significantes, vamos agora avaliar os resíduos, para concluir a investigação dos pressupostos necessários a um bom ajuste do modelo de regressão.

O gráfico 12 avalia a suposição de normalidade dos erros do modelo por meio do gráfico de probabilidade normal dos resíduos padronizados e também mostra relação entre os valores ajustados do modelo e os resíduos. Notamos que, pelo p-valor do teste de Anderson-Darling aplicado aos resíduos, temos evidências para rejeitar a hipótese de normalidade dos erros, não atendendo ao pressuposto de normalidade.

Gráfico 12 – Análise de Resíduos do modelo 1.2 (banco de dados de um ano)



Elaboração Própria

Como não conseguimos validar o pressuposto de normalidade dos erros, se faz necessário dar continuidade a possíveis ajustes no modelo, a fim de conseguirmos atender a esse pressuposto. Através do gráfico 12 do modelo 1.2, notamos que existem pontos que podem ser considerados discrepantes e podem estar influenciando nos resultados dos testes. Desta forma, partimos para uma análise dos valores de Hi e distância de Cook e identificamos algumas observações com estes valores elevados. Fizemos algumas simulações considerando a retirada de uma observação e o modelo foi sendo refeito para verificarmos o impacto da saída desta variável. Ao concluirmos essas simulações, retiramos apenas a observação “Honda Civic LSX 1.8 – 2015”, que detém um elevado valor de Cook e chegamos ao modelo 1.3, apresentado no Quadro 7.

Com a saída da observação “Honda Civic LSX 1.8 – 2015”, a variável “Seg-Hatch Médio.01” passou a apresentar um coeficiente estatisticamente nulo e foi retirada do modelo.

Quadro 7 – Modelo de regressão (1.3) da variável resposta VRV.01 (banco de dados de um ano).

```

VRV.01 = 0,210 + 0,00566 N°_Itens_Conforto.01 + 0,120 Seg-Sedan Grande.01
+ 0,0395 Pot-> 130 < = 160.01 + 0,0611 Pot-> 160 < = 190.01
+ 0,0457 M-Fiat.01 - 0,0644 M-Hyundai.01 + 0,0448 M-Renault.01
+ 0,0637 M-Nissan.01 - 0,104 M-Honda.01 - 0,192 M-BMW.01

```

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,20993	0,01102	19,06	0,000	
N°_Itens_Conforto.01	0,005665	0,002512	2,26	0,027	1,568
Seg-Sedan Grande.01	0,11986	0,03066	3,91	0,000	2,532
Pot-> 130 < = 160.01	0,03946	0,01247	3,16	0,002	1,341
Pot-> 160 < = 190.01	0,06108	0,01802	3,39	0,001	2,012
M-Fiat.01	0,04570	0,01376	3,32	0,001	1,174
M-Hyundai.01	-0,06443	0,01985	-3,25	0,002	1,061
M-Renault.01	0,04483	0,01985	2,26	0,027	1,061
M-Nissan.01	0,06374	0,02000	3,19	0,002	1,078
M-Honda.01	-0,10447	0,01987	-5,26	0,000	1,064
M-BMW.01	-0,19206	0,03754	-5,12	0,000	1,949

S = 0,0375441 R-Sq = 70,8% R-Sq(adj) = 66,5%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	0,232736	0,023274	16,51	0,000
Residual Error	68	0,095850	0,001410		
Lack of Fit	25	0,042946	0,001718	1,40	0,165
Pure Error	43	0,052904	0,001230		
Total	78	0,328586			

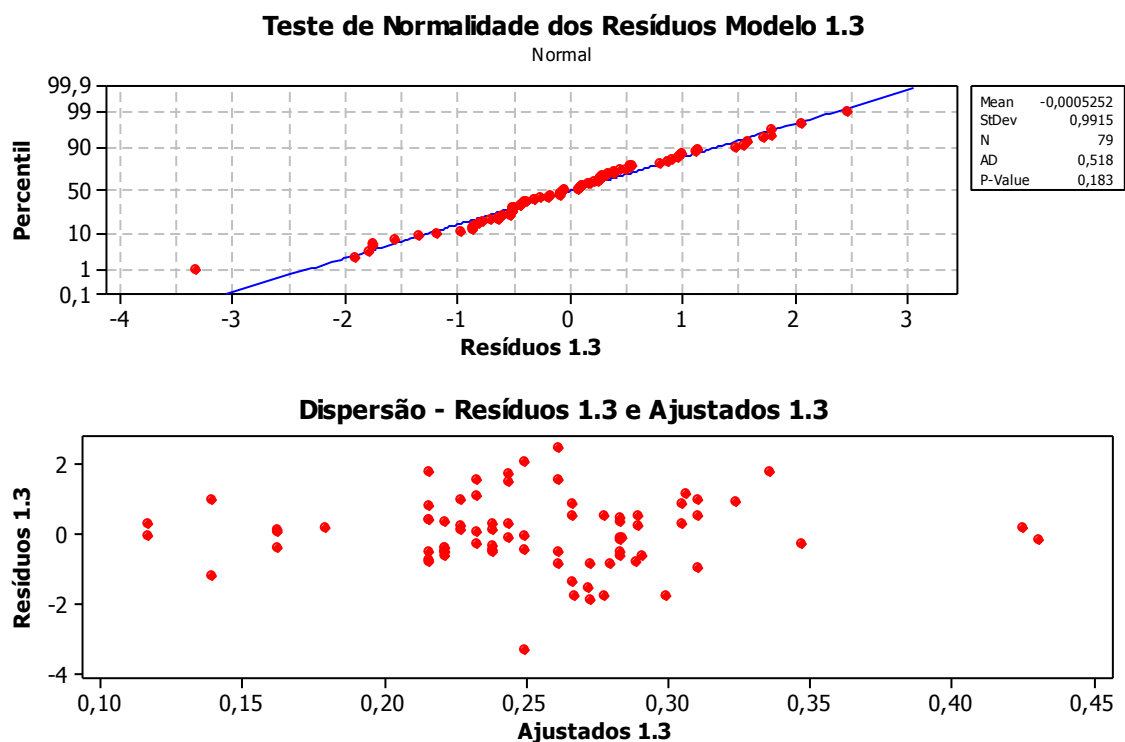
Elaboração Própria

Notamos que o modelo 1.3 apresenta todas as variáveis estatisticamente significantes. Não existe contradição entre os resultados do teste F e dos testes T individuais para cada variável preditora, além de todos os VIF's estarem abaixo de cinco, o que afasta a hipótese de multicolinearidade. O teste de ajuste do modelo da ANOVA, de forma distinta ao modelo 1.2, não rejeita a hipótese que o modelo linear está bem ajustado. O coeficiente de determinação ajustado atende às necessidades do trabalho. O modelo 1.3 não contou com as variáveis N° de Itens de Segurança e Motorização.

Partimos então para análise de resíduos deste modelo, com intuito de verificarmos se o modelo 1.3 atende as suposições de normalidade dos erros, linearidade da relação entre a variável resposta e as explicativas, além de homocedasticidade dos erros.

O gráfico 13 contém o gráfico de probabilidade normal dos resíduos acompanhado de um teste de Anderson Darling e um diagrama de dispersão dos resíduos contra os valores preditos pelo modelo 1.3. O fato de termos retirado uma observação discrepante e elaborado o modelo novamente fez com que chegássemos ao modelo 1.3 que conseguiu validar o pressuposto de normalidade dos erros, pois, analisando o p-valor do teste de Anderson Darling aplicado aos resíduos, não temos evidências para rejeitar a hipótese de normalidade dos erros.

Gráfico 13 – Análise de Resíduos do modelo 1.3 (banco de dados de um ano)



Elaboração Própria

Observando o diagrama de dispersão dos resíduos contra os valores ajustados do modelo, validamos o pressuposto de linearidade da relação entre a variável resposta e as explicativas, uma vez que não notamos uma correlação entre resíduos e ajustados do modelo. Em relação à homocedasticidade dos erros, os pontos do diagrama de dispersão estão bem distribuídos e não apresentam nenhuma tendência crescente ou decrescente, desta forma validamos também a suposição de homocedasticidade dos erros. Analisamos também os resíduos do modelo 1.3 contra as variáveis preditoras e não identificamos nenhum padrão de “funil”. (Apêndice A)

Diante das análises apresentadas, podemos concluir que o modelo 1.3 atende às suposições básicas para o ajuste de um modelo de regressão e, desta forma, ele é modelo final do banco de dados de um ano.

A equação do modelo final de regressão do banco de dados de um ano é:

$$\begin{aligned} \text{VRV.01} = & 0,210 + 0,00566 \text{ N}^{\circ}\text{Itens_Conforto.01} \\ & + 0,120 \text{ Seg-Sedan Grande.01} + 0,0395 \text{ Pot-} > 130 < = 160.01 \\ & + 0,0611 \text{ Pot-} > 160 < = 190.01 + 0,0457 \text{ M-Fiat.01} \\ & - 0,0644 \text{ M-Hyundai.01} + 0,0448 \text{ M-Renault.01} \\ & + 0,0637 \text{ M-Nissan.01} - 0,104 \text{ M-Honda.01} - 0,192 \text{ M-BMW.01} \end{aligned}$$

O intercepto representa o valor médio para a desvalorização de um modelo de veículo em um ano, caso todas as variáveis preditoras sejam iguais a zero, ou seja, se estivermos avaliando um modelo de veículo que pertença à categoria de referência de todas as variáveis explicativas. Assim, a desvalorização média após um ano de um veículo que seja das montadoras Ford, Peugeot, Volkswagen, General Motors, Citroën, Toyota ou Audi, segmentos de Entrada, Hatch Pequeno, Hatch Médio, Sedans pequenos, Compactos ou Médios, SUV, Monocab ou SW, e com potência igual ou inferior a 130 ou superior a 190 e nenhum item de conforto é de 21%.

Analisando o efeito das variáveis explicativas na desvalorização médio do veículo em um ano, vamos tomar como exemplo os coeficientes das variáveis preditoras “Seg-Sedan Grande.01” e “M-Hyundai.01”.

- Seg-Sedan Grande.01: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a categoria Sedan Grande, será acrescido, em média, de 12,0 pontos percentuais (p.p.) na sua desvalorização em um ano em comparação com os veículos da categoria de referência (segmentos de Entrada, Hatch ou Sedan pequenos). Desta forma, notamos que estar nesta categoria de seguimento aumenta a depreciação de um modelo de veículo em relação às outras categorias.
- M-Hyundai.01: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer à montadora Hyundai, espera-se que a

desvalorização após um ano diminua, em média, 6,44 p.p. em relação à desvalorização dos veículos das montadoras de referência (Ford, Peugeot, Volkswagen, General Motors, Citroën, Toyota e Audi). Sendo assim, ser da montadora Hyundai reduz a depreciação de um modelo de veículo em relação aos veículos das montadoras de referência.

- N°_Itens_Conforto.01: como essa variável é do tipo discreta (número de itens de conforto presentes em um modelo de veículo), temos que, para cada item de conforto que estiver presente no modelo de veículo, a desvalorização do veículo aumentará, em média, 0,566 p.p. na variável resposta, considerando todas as outras variáveis constantes.

É importante destacar que, de todas as variáveis preditoras estatisticamente significantes presentes no modelo 1.3, a variável “M-BMW.01”, referente à montadora BMW, foi a única que apresentou um coeficiente estranho ao conhecimento que temos do problema em questão, pois, não é razoável crer que, mantendo todas as demais variáveis constantes, o fato de um modelo de veículo ser da montadora BMW irá reduzir sua depreciação média após um ano em 19,20 pontos percentuais em relação às montadoras de referência. Optamos por manter esta variável no modelo final, pois, entendemos que seu comportamento pode estar sendo influenciado pelo baixo número de observações no banco de dados.

A seguir, daremos um exemplo de utilização do modelo final para o banco de dados de um ano considerando um modelo de veículo com as seguintes características: Montadora Honda; Segmento Sedan Médio; Motorização maior do que 1.4 e menor ou igual a 1.8; Potência maior do que 130 e menor ou igual a 160; 4 Itens de Segurança e 4 Itens de Conforto. Sendo assim, teríamos a seguinte equação para este modelo:

$$VRV.01 = 0,210 + 0,00566 * (4) + 0,120 * (0) + 0,0395 * (1) + 0,0611 (0) + 0,0457 * (0) - 0,0644 * (0) + 0,0448 * (0) + 0,0637 * (0) - 0,104 * (1) - 0,192 *(0)$$

$$VRV.01 = 0,21 + 0,02264 + 0,0395 - 0,104 = 0,1681$$

A desvalorização média para esse veículo em um ano de uso seria de 16,81%.

4.2.2 Modelo de regressão para o banco de dados de dois anos

O Quadro 8 apresenta o modelo inicial para o banco de dados de dois anos considerando todas as variáveis explicativas.

Quadro 8 – Modelo de regressão (2.1) da variável resposta VRV.02 (banco de dados de dois anos).

```

VRV.02 = 0,275 - 0,00540 N°_Itens_Segurança.02 + 0,00760 N°_Itens_Conforto.02
- 0,0135 M-Volkswagen.02 - 0,0361 M-Toyota.02 + 0,0302 M-Fiat.02
- 0,0508 M-Hyundai.02 + 0,0046 M-GM.02 + 0,0294 M-Renault.02
+ 0,0299 M-Nissan.02 - 0,0252 M-Citroen.02 + 0,0272 M-Peugeot.02
+ 0,0284 M-Audi.02 - 0,0894 M-Honda.02 + 0,0655 M-Mercedes Benz.02
- 0,111 M-BMW.02 - 0,0151 Seg-Hatch Pequeno.02
+ 0,0498 Seg-Hatch Médio.02 + 0,0100 Seg-Sedan Pequeno.02
+ 0,0070 Seg-Sedan Compacto.02 + 0,0475 Seg-Sedan Médio-SW-SUV.02
+ 0,136 Seg-Sedan Grande.02 + 0,0645 Seg-Monocab.02
+ 0,0160 Mot->1.0<= 1.8.02 + 0,0401 Mot->= 2.0.02
+ 0,0038 Pot-> 130 < = 190.02 - 0,0505 Pot-> 190

```

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,27486	0,02022	13,59	0,000	
N°_Itens_Segurança.02	-0,005404	0,003914	-1,38	0,172	7,132
N°_Itens_Conforto.02	0,007600	0,003397	2,24	0,029	3,580
M-Volkswagen.02	-0,01354	0,01643	-0,82	0,413	2,278
M-Toyota.02	-0,03606	0,01935	-1,86	0,067	1,594
M-Fiat.02	0,03020	0,01776	1,70	0,094	2,129
M-Hyundai.02	-0,05083	0,02076	-2,45	0,017	1,548
M-GM.02	0,00456	0,01838	0,25	0,805	2,279
M-Renault.02	0,02940	0,02097	1,40	0,166	1,872
M-Nissan.02	0,02992	0,02288	1,31	0,196	1,522
M-Citroen.02	-0,02521	0,03087	-0,82	0,417	1,418
M-Peugeot.02	0,02724	0,02498	1,09	0,280	1,813
M-Audi.02	0,02839	0,02302	1,23	0,222	2,599
M-Honda.02	-0,08938	0,02596	-3,44	0,001	2,419
M-Mercedes Benz.02	0,06551	0,04852	1,35	0,182	1,772
M-BMW.02	-0,11137	0,03852	-2,89	0,005	2,208
Seg-Hatch Pequeno.02	-0,01509	0,01804	-0,84	0,406	3,242
Seg-Hatch Médio.02	0,04980	0,02638	1,89	0,064	4,696
Seg-Sedan Pequeno.02	0,01003	0,01738	0,58	0,566	2,706
Seg-Sedan Compacto.02	0,00699	0,02760	0,25	0,801	2,214
Seg-Sedan Médio-SW-SUV.02	0,04746	0,02190	2,17	0,034	6,372
Seg-Sedan Grande.02	0,13576	0,03348	4,06	0,000	4,023
Seg-Monocab.02	0,06448	0,02706	2,38	0,020	2,628
Mot->1.0<= 1.8.02	0,01604	0,01681	0,95	0,344	3,939
Mot->= 2.0.02	0,04014	0,02648	1,52	0,135	7,320
Pot-> 130 < = 190.02	0,00376	0,01837	0,20	0,838	5,082
Pot-> 190	-0,05048	0,03149	-1,60	0,114	3,560

S = 0,0362445 R-Sq = 76,2% R-Sq(adj) = 66,1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	26	0,256792	0,009877	7,52	0,000
Residual Error	61	0,080133	0,001314		
Lack of Fit	59	0,079615	0,001349	5,21	0,174
Pure Error	2	0,000518	0,000259		
Total	87	0,336925			

Elaboração Própria

Destaca-se que, neste banco de dados, as categorias da variável segmento Sedan Médio, SW e SUV foram agrupadas em uma só, denominada “Seg-Sedan Médio-SW-SUV.02”; as categorias de potência foram reagrupadas e tornaram-se “ ≤ 130 ” (referência), “ >130 e ≤ 190 ” e “ >190 ”; para a variável motorização as categorias são “ >1.0 e ≤ 1.8 ” e “ ≥ 2.0 ”.

Analisando o modelo 2.1 em relação a multicolinearidade, notamos que quatro variáveis preditoras apresentaram VIF maiores que cinco. Ao retirarmos a com maior VIF do modelo (“Mot- $\geq 2.0.02$ ”) e construirmos o modelo novamente, ainda tivemos variáveis com VIF acima de cinco, somente após retirarmos mais duas variáveis, “Seg-Sedan Médio-SW-SUV.02” e “Nº Itens Segurança.02” é que conseguimos obter um modelo cujo VIF de todas as variáveis preditoras é menor que cinco. Observando o p-valor da estatística do teste F da tabela ANOVA e o p-valor do teste T, observamos que não existe contradição entre os resultados dos dois testes.

Após eliminarmos as possibilidades de multicolinearidade no modelo, eliminando três variáveis preditoras, adotamos a saída uma-a-uma das variáveis preditoras utilizando como critério aquelas com maior p-valor da estatística do teste T. A partir desta etapa o modelo foi refeito doze vezes até se atingir o modelo 2.2 presente no Quadro 9, que apresenta todas as variáveis estaticamente significantes.

Quadro 9 – Modelo de regressão (2.2) da variável resposta VRV.02 (banco de dados de dois anos).

$$\begin{aligned} \text{VRV.02} = & 0,293 + 0,00895 \text{ N}^\circ \text{ Itens Conforto.02} - 0,0396 \text{ M-Toyota.02} \\ & - 0,0478 \text{ M-Hyundai.02} + 0,0389 \text{ M-Renault.02} + 0,0398 \text{ M-Nissan.02} \\ & - 0,110 \text{ M-Honda.02} - 0,114 \text{ M-BMW.02} - 0,0404 \text{ Seg-Hatch Pequeno.02} \\ & + 0,0889 \text{ Seg-Sedan Grande.02} + 0,0567 \text{ Seg-Monocab.02} \\ & + 0,0277 \text{ Pot-} \rightarrow 130 < = 190.02 - 0,0428 \text{ Pot-} \rightarrow 190 \end{aligned}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,293442	0,009888	29,68	0,000	
N° Itens Conforto.02	0,008953	0,002628	3,41	0,001	2,035
M-Toyota.02	-0,03964	0,01642	-2,41	0,018	1,090
M-Hyundai.02	-0,04782	0,01765	-2,71	0,008	1,062
M-Renault.02	0,03894	0,01617	2,41	0,018	1,056
M-Nissan.02	0,03985	0,01991	2,00	0,049	1,094
M-Honda.02	-0,11022	0,01895	-5,82	0,000	1,224
M-BMW.02	-0,11356	0,03452	-3,29	0,002	1,684
Seg-Hatch Pequeno.02	-0,04035	0,01205	-3,35	0,001	1,373
Seg-Sedan Grande.02	0,08890	0,02294	3,88	0,000	1,794
Seg-Monocab.02	0,05668	0,01879	3,02	0,003	1,204
Pot- > 130 < = 190.02	0,02775	0,01256	2,21	0,030	2,254
Pot- > 190	-0,04276	0,02096	-2,04	0,045	1,498

S = 0,0371947 R-Sq = 69,2% R-Sq(adj) = 64,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	12	0,233167	0,019431	14,05	0,000
Residual Error	75	0,103758	0,001383		
Lack of Fit	35	0,039266	0,001122	0,70	0,861
Pure Error	40	0,064492	0,001612		
Total	87	0,336925			

Elaboração Própria

O modelo 2.1 contou com 26 variáveis preditoras, mas, após avaliarmos a multicolinearidade e depois retirarmos as variáveis que eram estatisticamente nulas, chegamos a um modelo 2,2, que possui 12 variáveis preditoras e a ordem em que as variáveis foram sendo retiradas e o modelo foi sendo refeito se deu da seguinte maneira:

1. Mot->2.0
2. N° Itens Segurança.02
3. Seg-Sedan Médio-SW-SUV.02
4. M-GM.02
5. Seg-Sedan Compacto.02
6. M-Mercedes Benz.02

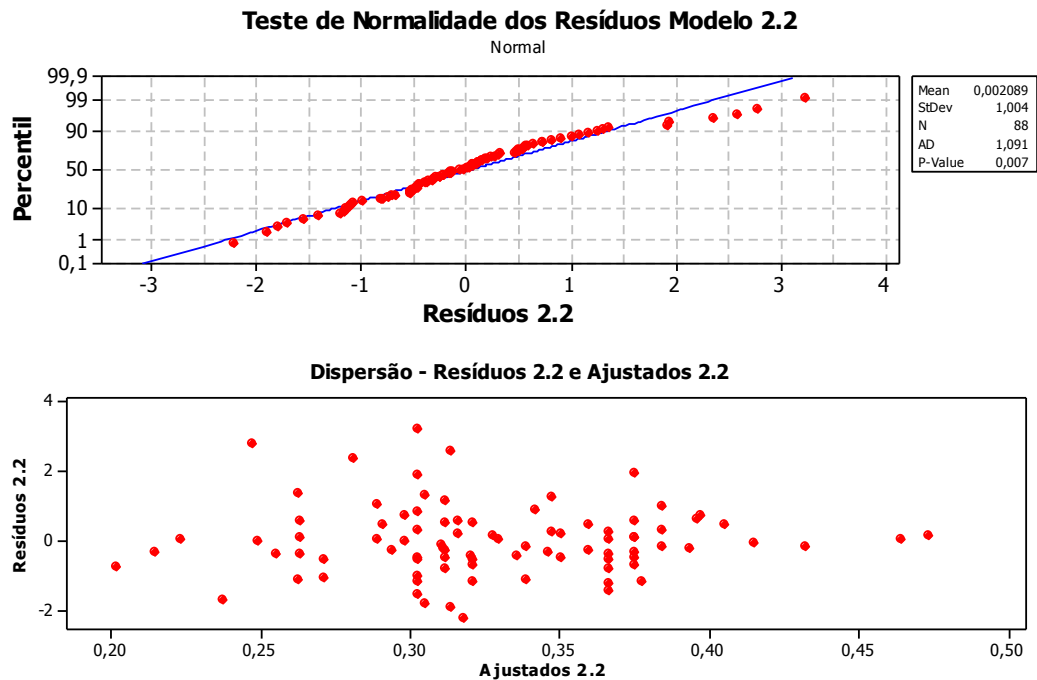
7. Mot->1.0 ≤ 1.8.02
8. Seg-Hatch Médio.02
9. Seg-Sedan Pequeno.02
- 10.M-Citroen.02
- 11.M-Audi.02
- 12.M-Volkswagen.02
- 13.M-Fiat.02
- 14.M-Peugeot.02

Analisando o Quadro 9, notamos que não há contradição entre os resultados dos testes F e T, o que é mais uma evidência da ausência de multicolinearidade no modelo. Além disto, todas as variáveis preditoras são estatisticamente significantes. O coeficiente de determinação ajustado deste modelo nos evidencia que 64,3% da variância da desvalorização do veículo após dois anos é explicada pelas variáveis preditoras e este coeficiente atende a necessidade do nosso trabalho.

Necessitamos agora validar os pressupostos de normalidade, linearidade e homocedasticidade para então eleger o modelo 2.2 como o modelo final do banco de dados de dois anos. Sendo assim, passamos a seguir para a análise dos resíduos do modelo.

No gráfico 14, estão presentes o gráfico de probabilidade normal dos resíduos, assim como o teste de Anderson Darling, e um diagrama de dispersão dos resíduos contra os valores ajustados pelo modelo 2.2. O teste de normalidade dos resíduos nos permite rejeitar a hipótese de normalidade dos erros e, desta forma, o ajuste deste modelo estaria violando o pressuposto de normalidade.

Gráfico 14 – Análise de Resíduos do modelo 2.2 (banco de dados de dois anos)



Elaboração Própria

Contudo utilizamos os valores de Hi e a distância de Cook para analisarmos possíveis valores discrepantes na amostra do modelo 2.2. Identificamos algumas observações com valores de Hi e Cook elevados e foram realizados alguns testes retirando uma observação e refazendo o modelo para verificar o impacto de sua retirada. Após esses testes retiramos quatro observações cujos valores de Hi ou Cook estavam elevados e chegamos ao modelo 2.3, presente no Quadro 10, cujo coeficiente de determinação ajustado se mostrou superior ao do modelo 2.2. Contudo a variável preditora “Pot-> $130 \leq 190.02$ ” passou a ser estatisticamente não significativa e foi retirada do modelo e a montadora BMW saiu do modelo, pois, retiramos suas duas observações.

As observações retiradas foram os seguintes modelos de veículos: “Honda Civic LSX 1.8 – 2015”, “320i Sport 2.0 Turbo – 2015”, “320i M Sport 2.0 Turbo – 2015” e “Weekend Attractive 1.4 – 2015”.

Quadro 10 – Modelo de regressão (2.3) da variável resposta VRV.02 (banco de dados de dois anos).

```

VRV.02 = 0,283 + 0,0138 N°_Itens_Conforto.02 - 0,0328 M-Toyota.02
        - 0,0432 M-Hyundai.02 + 0,0398 M-Renault.02 + 0,0500 M-Nissan.02
        - 0,150 M-Honda.02 - 0,0481 Seg-Hatch Pequeno.02
        + 0,0945 Seg-Sedan Grande.02 + 0,0788 Seg-Monocab.02
        - 0,0642 Pot-> 190
  
```

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,282995	0,009004	31,43	0,000	
N°_Itens_Conforto.02	0,013762	0,001838	7,49	0,000	1,183
M-Toyota.02	-0,03277	0,01468	-2,23	0,029	1,091
M-Hyundai.02	-0,04316	0,01576	-2,74	0,008	1,061
M-Renault.02	0,03976	0,01441	2,76	0,007	1,050
M-Nissan.02	0,05001	0,01748	2,86	0,006	1,058
M-Honda.02	-0,15023	0,01882	-7,98	0,000	1,225
Seg-Hatch Pequeno.02	-0,048093	0,009810	-4,90	0,000	1,132
Seg-Sedan Grande.02	0,09454	0,02037	4,64	0,000	1,090
Seg-Monocab.02	0,07884	0,01710	4,61	0,000	1,249
Pot-> 190	-0,06422	0,01639	-3,92	0,000	1,148

S = 0,0331777 R-Sq = 75,3% R-Sq(adj) = 72,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	0,245510	0,024551	22,30	0,000
Residual Error	73	0,080356	0,001101		
Lack of Fit	31	0,024954	0,000805	0,61	0,923
Pure Error	42	0,055401	0,001319		
Total	83	0,325866			

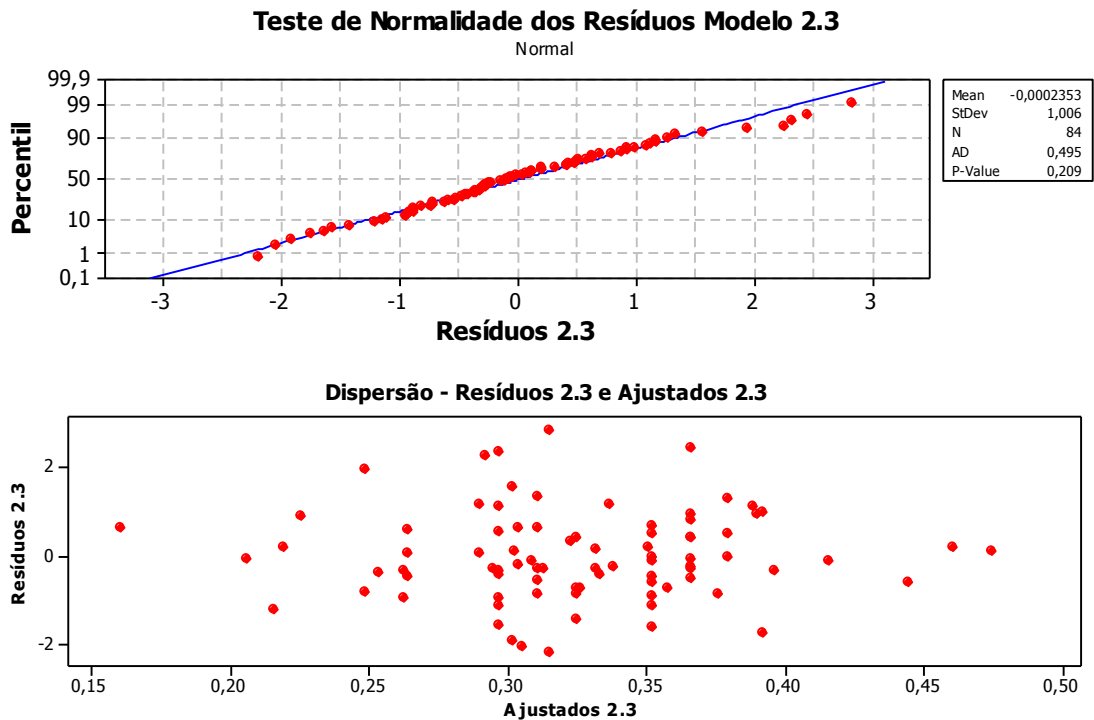
Elaboração Própria

O modelo 2.3 apresenta todas as variáveis estatisticamente significantes, os VIF's estão todos abaixo de cinco e não há contradição entre os resultados dos testes F e dos testes T individuais para cada variável explicativa. O modelo não contou com as variáveis Motorização e Nº de Itens de Segurança. Passamos para a análise dos resíduos deste modelo.

O gráfico 15 contém o gráfico de probabilidade normal dos resíduos, assim como o teste de Anderson Darling e um diagrama de dispersão dos resíduos contra os valores ajustados pelo modelo 2.3. Notamos que, pelo p-valor do teste de Anderson Darling aplicado aos resíduos, não temos evidências para rejeitar a hipótese de normalidade dos erros. Desta forma o modelo 2.3, elaborado após retiramos os pontos discrepantes, atende ao pressuposto de normalidade. Quanto ao pressuposto de linearidade da relação entre a variável resposta e as explicativas, conseguimos validar a suposição de linearidade do modelo pelo diagrama de

dispersão, uma vez que não existe correlação entre os resíduos e os valores ajustados do modelo e o teste de falta de ajuste da ANOVA não rejeita a hipótese que o modelo linear está bem ajustado.

Gráfico 15 – Análise de Resíduos do modelo 2.3 (banco de dados de dois anos)



Elaboração Própria

Em relação à homocedasticidade dos erros, os pontos estão bem distribuídos nos gráfico e não apresentam nenhuma tendência de se disporem em padrões do tipo “funil”. Desta forma, validamos a suposição de homogeneidade da variância dos erros. Fizemos a análise dos resíduos do modelo verificando o gráfico dos resíduos contra as variáveis preditoras e também não identificamos nenhum padrão de “funil”. (Apêndice B)

Podemos concluir então que o modelo 2.3 atende as hipóteses básicas para o ajuste de um modelo de regressão e, desta forma, ele é o modelo final do banco de dados de dois anos.

A equação do modelo final de regressão do banco de dados de dois anos é:

$$\begin{aligned} \text{VRV.02} = & 0,283 + 0,0138 \text{ N}^\circ \text{Itens_Conforto.02} - 0,0328 \text{ M-Toyota.02} \\ & - 0,0432 \text{ M-Hyundai.02} + 0,0398 \text{ M-Renault.02} \\ & + 0,0500 \text{ M-Nissan.02} - 0,150 \text{ M-Honda.02} \\ & - 0,0481 \text{ Seg-Hatch Pequeno.02} \\ & + 0,0945 \text{ Seg-Sedan Grande.02} + 0,0788 \text{ Seg-Monocab.02} \\ & - 0,0642 \text{ Pot-> 190} \end{aligned}$$

A categoria de referência do modelo 2.3 é um veículo que pertença as Montadoras Ford, GM, Mercedes Benz, Audi, Volkswagen, Fiat, Peugeot ou BMW, além de pertencer a uma dessas categorias de segmento Entrada, Sedan Médio, SW, SUV, Sedan Compacto, Hatch Médio, Sedan Pequeno, tem a potência menor ou igual a 190cv.

O valor médio para a desvalorização de um modelo de veículo em dois anos, caso estivermos avaliando um modelo de veículo que pertença à categoria de referência de todas as variáveis explicativas é 28,3%, valor do intercepto da equação do modelo.

Vamos tomar como exemplo as variáveis, Montadora, Segmento e N° Itens de Conforto para explicar a interpretação dos coeficientes do modelo 2.3 como a seguir.

- M-Honda.02: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a montadora Honda, sua desvalorização média em dois anos será reduzida em 15,0 pontos percentuais em relação a um veículo das montadoras de referência. Logo percebemos que a montadora Honda tem o efeito de reduzir a depreciação de um modelo de veículo em relação à depreciação das montadoras de referência.
- Seg-Monocab: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer ao segmento Monocab, sua desvalorização média em dois anos será acrescida de 7,88 pontos percentuais em relação a um veículo do segmento de referência. Desta forma notamos que o fato de um modelo veículo pertencer ao segmento Monocab tem o efeito de aumentar a depreciação deste

modelo de veículo em relação à depreciação dos segmentos de referência.

- Nº Itens de Conforto: para cada item de conforto que estiver presente no modelo de veículo, serão acrescidos 1,38 pontos percentuais em sua desvalorização média após dois anos, considerando todas as outras variáveis constantes.

Vale destacar que, de forma similar ao comportamento da variável “BMW” no modelo 1.2, a variável “Pot->190” apresentou também um comportamento estranho considerando ao que sabe do problema estudado, mostrando o efeito de reduzir a depreciação média do veículo. O que seria esperado é que, mantendo todas as demais variáveis constantes, o fato de um modelo de veículo ter potência superior a 190CV iria aumentar a desvalorização deste veículo. Mantivemos a variável no modelo, pois, entendemos que seu comportamento pode estar sendo influenciado pelo baixo número de observações no banco de dados.

Como exemplo para utilização do modelo final para o banco de dados de dois anos, vamos considerar um modelo de veículo com as seguintes características: Montadora Hyundai; Segmento Hatch Pequeno; Motorização igual a 1.0; Potência menor ou igual a 130; 2 itens de segurança e 2 itens de conforto.

Teríamos a seguinte equação para este modelo:

$$VRV.02 = 0,283 + 0,0138 * (2) - 0,0328 * (0) - 0,0432 * (1) + 0,0398 (0) + 0,0500 * (0) - 0,150 * (0) - 0,0481 * (1) + 0,0945 (0) + 0,0788 (0) - 0,0642 * (0)$$

$$VRV.02 = 0,283 + 0,0276 - 0,0432 - 0,0481 = 21,93\%$$

A desvalorização média para um veículo com essas características após a utilização por dois anos é de 21,93%

4.2.3 Modelo de regressão para o banco de dados de três anos

O Quadro 11 apresenta o modelo inicial para o banco de dados de três anos considerando todas as variáveis explicativas. Conforme exposto na análise descritiva, neste banco a variável Segmento foi agrupada da seguinte maneira: Entrada, Hatch Pequeno e Sedan Pequeno formam apenas uma categoria e passam a ser o grupo de referência; as categorias Sedan Médio e Sedan Grande foram unidas e estão representadas no modelo pela variável “Seg- Sedan Médio- Sedan Grande.03”. Quanto à Motorização, a categoria de referência passa a ser “ ≤ 1.8 ”. Na variável Potência, a categoria de referência passa a ser “ ≤ 130 ” e a categoria “ >130 e ≤ 190 ” está representada pela variável “Pot->130 \leq 190.03”. Observamos no Quadro 9 que temos ausência das seguintes variáveis: montadora Ford, Segmento de Entrada, Hatch Pequeno e Sedan Pequeno, Motorização ≤ 1.8 ” e Potência ≤ 130 ”. Sendo assim, um veículo que se encaixe nessas categorias é o veículo de referência e terá sua desvalorização média representadas pelo intercepto do modelo.

Analisando os VIFs das variáveis preditoras do modelo 3.1, notamos que quatro variáveis têm um VIF superior a cinco, sendo que “Nº Itens Segurança.03” se destaca com um VIF de 10,59. Após retirá-la do modelo, a variável “Pot->130 \leq 190.02” e “Seg-Sedan Médio-Sedan Grande.03” ainda se mantiverem com um VIF acima de 5. Sendo assim, a variável “Pot->130 \leq 190.02” foi retirada e, após este passo, não tivemos mais nenhuma variável com VIF superior a cinco. Não foi observada contradição entre os resultados dos testes F da tabela ANOVA e os testes T das variáveis preditoras. Sendo assim, uma vez garantida a ausência de multicolinearidade, passamos ao refinamento do modelo utilizando o critério de saída uma-a-uma considerando aqueles que obtiveram um maior p-valor da estatística do teste T.

Como pode ser notado no Quadro 11, inicialmente o modelo 3.1 contou com 23 variáveis candidatas a preditoras. Contudo duas foram retiradas por critério de VIF e mais onze foram retiradas considerando a eliminação devido a serem estatisticamente nulas.

Quadro 11 – Modelo de regressão (3.1) da variável resposta VRV.03 (banco de dados de três anos)

```

VRV.03 = 0,301 + 0,00705 N°_Itens_Segurança.03 + 0,00360 N°_Itens_Conforto.03
+ 0,0331 M-Volkswagen.03 + 0,0380 M-Toyota.03 + 0,0416 M-Fiat.03
- 0,0279 M-Hyundai.03 - 0,0003 M-GM.03 + 0,0828 M-Renault.03
+ 0,0609 M-Nissan.03 + 0,0310 M-Citroen.03 + 0,0782 M-Peugeot.03
+ 0,0242 M-Audi.03 - 0,0049 M-Honda.03 + 0,0334 M-Mercedes Benz.03
+ 0,0186 Seg-Hatch Médio.03 + 0,0764 Seg-Sedan Compacto.03
+ 0,0094 Seg-Sedan Médio-Sedan Grande.03 + 0,0524 Seg-SpaceWagon.03
+ 0,0367 Seg-Monocab.03 + 0,0176 Seg-SUV.03 + 0,0089 Mot->= 2.0.03
+ 0,0365 Pot-> 130 < = 190.02 + 0,0151 Pot-> 190

Predictor          Coef      SE Coef      T          P          VIF
Constant           0,30110    0,02034     14,80     0,000
N°_Itens_Segurança.03 0,007048  0,004941     1,43     0,161    10,590
N°_Itens_Conforto.03 0,003596  0,004062     0,89     0,381     4,828
M-Volkswagen.03    0,03314    0,01895     1,75     0,087     2,747
M-Toyota.03        0,03800    0,02345     1,62     0,112     2,025
M-Fiat.03           0,04164    0,01965     2,12     0,040     3,229
M-Hyundai.03       -0,02792   0,02075    -1,35     0,185     1,951
M-GM.03            -0,00027   0,01860    -0,01     0,988     2,892
M-Renault.03       0,08278    0,02136     3,88     0,000     2,443
M-Nissan.03        0,06095    0,02106     2,89     0,006     2,374
M-Citroen.03       0,03096    0,02693     1,15     0,256     1,376
M-Peugeot.03       0,07821    0,02447     3,20     0,003     2,206
M-Audi.03           0,02419    0,03154     0,77     0,447     1,888
M-Honda.03         -0,00490   0,04208    -0,12     0,908     1,705
M-Mercedes Benz.03 0,03343    0,04723     0,71     0,483     2,149
Seg-Hatch Médio.03 0,01858    0,02989     0,62     0,537     6,176
Seg-Sedan Compacto.03 0,07637   0,02696     2,83     0,007     1,379
Seg-Sedan Médio-Sedan Grande.03 0,00945  0,02956     0,32     0,751     7,305
Seg-SpaceWagon.03 0,05243    0,02825     1,86     0,070     1,515
Seg-Monocab.03     0,03668    0,02923     1,25     0,216     2,397
Seg-SUV.03         0,01762    0,02256     0,78     0,439     2,726
Mot->= 2.0.03      0,00890    0,01983     0,45     0,656     4,053
Pot-> 130 < = 190.02 0,03648    0,02409     1,51     0,137     8,057
Pot-> 190          0,01513    0,03588     0,42     0,675     4,742

S = 0,0319857   R-Sq = 79,8%   R-Sq(adj) = 69,4%

Analysis of Variance

Source      DF      SS      MS      F      P
Regression  23    0,181417  0,007888  7,71  0,000
Residual Error  45    0,046039  0,001023
  Lack of Fit  36    0,031131  0,000865  0,52  0,919
  Pure Error   9     0,014908  0,001656
Total       68    0,227455
    
```

Elaboração Própria

Chegamos ao modelo 3.2, apresentado no Quadro 12, após refazer o modelo treze vezes retirando as variáveis na seguinte ordem:

1. N° Itens de Segurança.03

2. Pot $>130 \leq 190$ (Após a retirada desta segunda variável eliminamos a multicolinearidade do modelo, nenhuma variável apresentou VIF maior que 5)
3. M-GM.03 (A partir desta variável fomos eliminando pelo critério do maior p-valor)
4. M-Honda.03
5. M-Audi.03
6. Pot- >190 .03
7. M-Citroen.03
8. Mot- ≥ 2.0
9. M-Mercedes Benz.03
10. M-Toyota.03
11. M.Volkswagen.03
12. M.Fiat.03
13. Seg-Sedan Compacto.03

O modelo 3.2 apresenta todas as variáveis estatisticamente significantes e sem indícios de multicolinearidade. O coeficiente de determinação ajustado deste modelo nos evidencia que 65,7% da variância da desvalorização do veículo após três anos é explicada pelas variáveis preditoras, coeficiente este que, de forma similar aos dois bancos anteriores, atende as necessidades deste trabalho.

Quadro 12 – Modelo de regressão (3.2) da variável resposta VRV.03 (banco de dados de três anos)

$$\begin{aligned} \text{VRV.03} = & 0,333 + 0,00834 \text{ N}^\circ \text{ Itens Conforto.03} - 0,0553 \text{ M-Hyundai.03} \\ & + 0,0467 \text{ M-Renault.03} + 0,0338 \text{ M-Nissan.03} + 0,0456 \text{ M-Peugeot.03} \\ & + 0,0543 \text{ Seg-Hatch Médio.03} + 0,0457 \text{ Seg-Sedan Médio-Sedan Grande.03} \\ & + 0,0872 \text{ Seg-SpaceWagon.03} + 0,0866 \text{ Seg-Monocab.03} + 0,0441 \text{ Seg-SUV.03} \end{aligned}$$

Preditoras	Coef	SE Coef	T	P	VIF
Constant	0,333096	0,008553	38,95	0,000	
Nº Itens Conforto.03	0,008339	0,002809	2,97	0,004	2,057
M-Hyundai.03	-0,05530	0,01643	-3,37	0,001	1,091
M-Renault.03	0,04667	0,01555	3,00	0,004	1,153
M-Nissan.03	0,03383	0,01504	2,25	0,028	1,080
M-Peugeot.03	0,04565	0,01932	2,36	0,022	1,225
Seg-Hatch Médio.03	0,05429	0,01595	3,40	0,001	1,568
Seg-Sedan Médio-Sedan Grande.03	0,04570	0,01578	2,90	0,005	1,855
Seg-SpaceWagon.03	0,08724	0,02490	3,50	0,001	1,049
Seg-Monocab.03	0,08657	0,02070	4,18	0,000	1,072
Seg-SUV.03	0,04406	0,01640	2,69	0,009	1,283

S = 0,0338817 R-Sq = 70,7% R-Sq(adj) = 65,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	10	0,160873	0,016087	14,01	0,000
Residual Error	58	0,066582	0,001148		
Lack of Fit	29	0,023513	0,000811	0,55	0,946
Pure Error	29	0,043069	0,001485		
Total	68	0,227455			

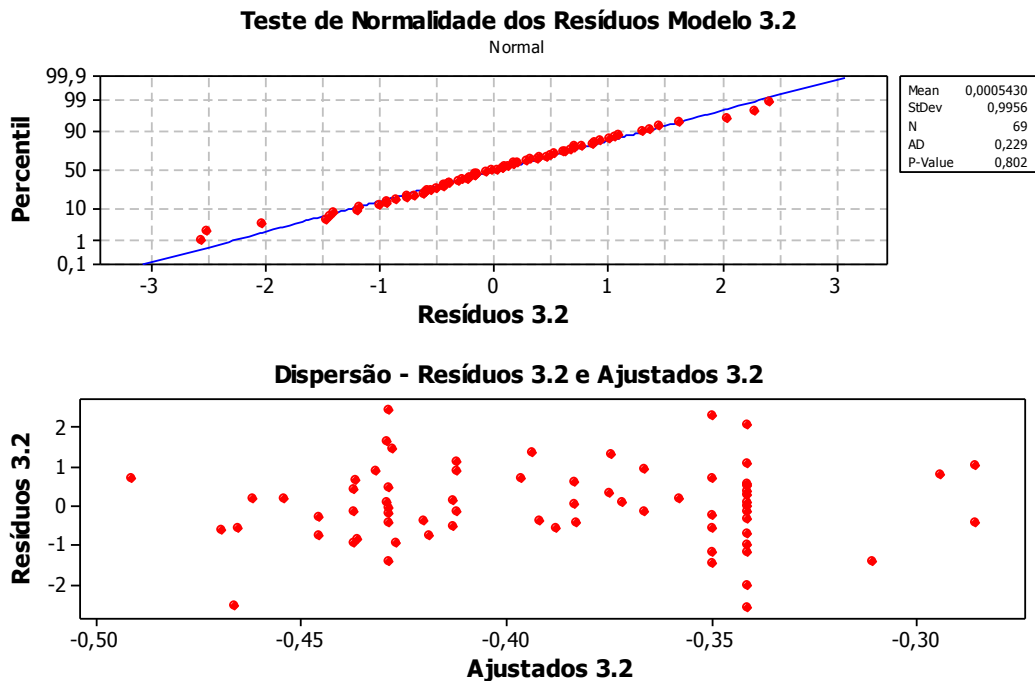
Elaboração Própria

Para elegermos o modelo 3.2 como nosso modelo final, precisamos analisar seus resíduos e investigar os pressupostos de linearidade, homocedasticidade e normalidade. O gráfico 16 apresenta o gráfico de probabilidade normal, juntamente com o teste de normalidade de Anderson-Darling aplicado aos resíduos do modelo. Pelo p-valor, notamos que não temos evidências suficientes para rejeitar a hipótese de normalidade dos erros e ainda notamos que os pontos estão bem distribuídos sobre a linha de referência e com isso reforçamos a suposição de normalidade dos erros. Observamos também que não temos a presença da variável Nº de Itens de Segurança, nem nenhuma categoria de Motorização ou Potência neste modelo.

O gráfico 16 apresenta também o diagrama de dispersão dos resíduos contra os valores ajustados do modelo 3.2. Este diagrama apresenta pontos que estão distribuídos ao longo de uma faixa constante em torno do valor zero para o resíduo e não apresentam nenhum padrão de correlação com os valores ajustados. Isto indica

que as suposições de homocedasticidade dos erros e de linearidade do modelo podem ser consideradas válidas. Observando os resíduos contra as variáveis preditoras (Apêndice C), não temos nenhuma tendência, desta forma reforçamos a ideia de homocedasticidade dos erros.

Gráfico 16 – Análise de Resíduos do modelo 3.2 (banco de dados de três anos)



Elaboração Própria

Analisando a presença de pontos de alavanca ou influência, por meio do exame valores de alavanca Hi e a distância de Cook, não identificamos nenhum ponto influente. Desta forma, elegemos o modelo 3.2 como o modelo final do banco de dados de três anos.

A equação final do modelo é:

$$\begin{aligned}
 \text{VRV.03} = & 0,333 + 0,00834 \text{ N}^\circ \text{ Itens_Conforto.03} - 0,0553 \text{ M-Hyundai.03} \\
 & + 0,0467 \text{ M-Renault.03} + 0,0338 \text{ M-Nissan.03} \\
 & + 0,0456 \text{ M-Peugeot.03} + 0,0543 \text{ Seg-HatchMédio.03} \\
 & + 0,0457 \text{ Seg-SedanMédio-Sedan Grande.03} \\
 & + 0,0872 \text{ Seg-SpaceWagon.03} + 0,0866 \text{ Seg-Monocab.03} \\
 & + 0,0441 \text{ Seg-SUV.03}
 \end{aligned}$$

A categoria de referência no modelo 3.2 é um veículo das Montadoras Ford, GM, Honda, Audi, Mercedes Benz, Toyota, Volkswagen ou Fiat, do Segmento de Entrada, Hatch Pequeno, Sedans pequenos ou compactos.

No modelo 3.2, o intercepto representa o valor médio da desvalorização em três anos de um veículo pertencente à categoria de referência. Então, para um modelo de veículo que foi utilizado por três anos e pertencente à categoria de referência, a depreciação média será de 33,3% no período de tempo citado.

Quanto ao efeito do número de itens de conforto na desvalorização do veículo em três anos, para cada item de conforto que estiver presente no modelo de veículo, será acrescido, em média, 0,834 pontos percentuais na desvalorização do veículo, considerando todas as outras variáveis constantes. Tomando como exemplo a variável Montadora e Segmento, os coeficientes do modelo 3.2 podem ser interpretados como a seguir.

- M.Renault.03: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a montadora Renault, sua desvalorização média em três anos será acrescida de 4,67 pontos percentuais em relação a um modelo de veículo da categoria de referência. Desta forma, notamos que o fato de o veículo ser da montadora Renault tem o efeito de aumentar a depreciação de um modelo de veículo em relação à depreciação das montadoras de referência.
- Seg-Sedan Médio-Sedan Grande.03: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a categoria de segmento Sedan Médio ou Sedan Grande, espera-se que a desvalorização média aumente em 4,57 pontos percentuais em relação ao segmento de referência.

Como exemplo de utilização do modelo final para o banco de dados de três anos, vamos considerar um modelo de veículo com as seguintes características: Montadora Volkswagen; Segmento Sedan Grande; Motorização superior ou igual a 2.0; Potência maior ou igual a 190; 5 itens de segurança e 6 itens de Conforto.

Teríamos a seguinte equação para este modelo:

$$\text{VRV.03} = 0,333 + 0,00834 * (6) - 0,0553 * (0) + 0,0467 (0) + 0,0338 (0) + 0,0456 (0) + 0,0543 (0) + 0,0457 (1) + 0,0872 (0) + 0,0866 (0) + 0,0441 (0)$$

$$\text{VRV.03} = 0,333 + 0,05004 + 0,0457 = 42,88\%$$

A desvalorização média para um veículo com essas características após a utilização por três anos é de 42,88%.

4.2.4 Modelo de regressão para o banco de dados de quatro anos

O Quadro 13 apresenta o modelo inicial para o banco de dados de quatro anos, considerando todas as variáveis explicativas. Neste banco de dados as categorias do variável segmento foram agrupadas da seguinte forma: Entrada, Hatch Pequeno e Sedan Pequeno formam apenas uma categoria e passam a pertencer ao grupo de referência. Motorização passa a ter como categoria de referência “≤1.4”. Para a variável potência a categoria de referência passa a ser “≤130”, a categoria “>220” foi agrupada com “>190 e ≤220”, e agora são representadas por “>190”.

Analisando o modelo 4.1, presente no Quadro 13, notamos que 11 variáveis preditoras apresentam um VIF superior a cinco. Ao retirarmos em sequência as que tiveram os maiores valores (“Nº Itens de Segurança.04”, “Mot-> 2.0.04”, “Pot->130≤160” e “M.Fiat.04”), não tivemos mais nenhuma variável com valor de VIF superior a cinco. Se observarmos o p-valor da estatística do teste F na tabela ANOVA do modelo 4.1, notamos que não existe contradição entre os resultados do teste F e dos testes T individuais para cada variável explicativa, desta forma afastamos a hipótese de existência de multicolinearidade entre as variáveis deste modelo.

Após retirarmos do modelo quatro variáveis para atender ao pressuposto da ausência de multicolinearidade, iniciamos o procedimento de retirada uma-a-uma das variáveis preditoras que são estatisticamente nulas. Como podemos notar o modelo 4.1, conta com 24 variáveis candidatas a preditoras. Contudo, após os procedimentos mencionados chegamos ao modelo 4.2, que é apresentado no Quadro 14, com apenas 9 variáveis preditoras.

Quadro 13 – Modelo de regressão (4.1) da variável resposta VRV.04 (banco de dados de quatro anos)

VRV.04 = 0,353 + 0,00491 N°_Itens_Segurança.04 + 0,0120 N°_Itens_Conforto.04
 + 0,0269 M-Volkswagen.04 + 0,0431 M-Toyota.04 + 0,0204 M-Fiat.04
 - 0,0599 M-Hyundai.04 - 0,0304 M-GM.04 + 0,0677 M-Renault.04
 + 0,0437 M-Nissan.04 - 0,0140 M-Citroen.04 + 0,154 M-Peugeot.04
 - 0,0390 M-Honda.04 - 0,0346 M-Mercedes Benz.04
 + 0,0170 Seg-Hatch Médio.04 + 0,0926 Seg-Sedan Compacto.04
 - 0,0007 Seg-Sedan Médio.04 + 0,0450 Seg-SW-Sedan Grande.04
 + 0,0494 Seg-Monocab.04 + 0,0198 Seg-SUV.04
 - 0,0005 Mot-> 1.4 <= 1.8.04 - 0,0339 Mot->= 2.0.04
 + 0,0397 Pot-> 130 < = 160.04 + 0,0340 Pot-> 160 < = 190.04
 + 0,0068 Pot-> 190

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,35345	0,02968	11,91	0,000	
N°_Itens_Segurança.04	0,004910	0,005759	0,85	0,402	11,712
N°_Itens_Conforto.04	0,011996	0,005387	2,23	0,035	6,810
M-Volkswagen.04	0,02690	0,02644	1,02	0,318	5,063
M-Toyota.04	0,04312	0,03186	1,35	0,188	3,077
M-Fiat.04	0,02043	0,02765	0,74	0,467	6,595
M-Hyundai.04	-0,05988	0,03059	-1,96	0,061	2,836
M-GM.04	-0,03041	0,02912	-1,04	0,306	7,315
M-Renault.04	0,06773	0,02928	2,31	0,029	3,393
M-Nissan.04	0,04370	0,03271	1,34	0,193	4,233
M-Citroen.04	-0,01401	0,03400	-0,41	0,684	2,384
M-Peugeot.04	0,15415	0,06323	2,44	0,022	4,207
M-Honda.04	-0,03905	0,04592	-0,85	0,403	2,219
M-Mercedes Benz.04	-0,03461	0,05434	-0,64	0,530	3,108
Seg-Hatch Médio.04	0,01704	0,04284	0,40	0,694	5,562
Seg-Sedan Compacto.04	0,09264	0,02626	3,53	0,002	1,422
Seg-Sedan Médio.04	-0,00074	0,03442	-0,02	0,983	6,731
Seg-SW-Sedan Grande.04	0,04502	0,02638	1,71	0,100	2,109
Seg-Monocab.04	0,04944	0,03033	1,63	0,115	2,789
Seg-SUV.04	0,01978	0,02332	0,85	0,404	3,092
Mot-> 1.4 <= 1.8.04	-0,00047	0,01440	-0,03	0,974	2,829
Mot->= 2.0.04	-0,03389	0,04666	-0,73	0,474	17,321
Pot-> 130 < = 160.04	0,03972	0,03098	1,28	0,211	8,890
Pot-> 160 < = 190.04	0,03404	0,05335	0,64	0,529	5,871
Pot-> 190	0,00681	0,05482	0,12	0,902	11,889

S = 0,0305228 R-Sq = 86,0% R-Sq(adj) = 73,0%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	24	0,1485165	0,0061882	6,64	0,000
Residual Error	26	0,0242226	0,0009316		
Lack of Fit	23	0,0210593	0,0009156	0,87	0,650
Pure Error	3	0,0031633	0,0010544		
Total	50	0,1727392			

Elaboração Própria

Chegamos ao modelo 4.2, após refazer o modelo dezessete vezes retirando as variáveis na seguinte ordem:

1. Nº Itens de Segurança.04
2. Mot- \geq 2.0.04
3. Pot $>130\leq$ 160.04
4. M-Fiat.04 (Após a retirada desta quarta variável, eliminamos a multicolinearidade do modelo, nenhuma variável apresentou VIF maior que 5)
5. M- \rightarrow 1.4 \leq 1.8.04 (A partir desta variável, fomos eliminando pelo critério do maior p-valor)
6. M-Mercedez Benz.04
7. M-Volkswagen.04
8. Pot- \rightarrow 160 \leq 190.04
9. M-Nissan.04
10. M-Toyota.04
11. Seg-Sedan Médio04
12. M-Honda.04
13. Seg-SUV.04
14. Pot- \rightarrow 190.04
15. M-Citroen.04

Quadro 14 – Modelo de regressão (4.2) da variável resposta VRV.04 (banco de dados de quatro anos)

The regression equation is

$$VRV.04 = 0,383 + 0,0140 N^{\circ}_Itens_Conforto.04 - 0,0789 M-Hyundai.04 - 0,0382 M-GM.04 + 0,0519 M-Renault.04 + 0,0943 M-Peugeot.04 + 0,0600 Seg-Hatch\ M\u00e9dio.04 + 0,0731 Seg-Sedan\ Compacto.04 + 0,0572 Seg-SW-Sedan\ Grande.04 + 0,0659 Seg-Monocab.04$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,382678	0,008085	47,33	0,000	
N ^o _Itens_Conforto.04	0,014047	0,002162	6,50	0,000	1,086
M-Hyundai.04	-0,07895	0,01869	-4,22	0,000	1,049
M-GM.04	-0,03821	0,01388	-2,75	0,009	1,646
M-Renault.04	0,05186	0,01643	3,16	0,003	1,058
M-Peugeot.04	0,09431	0,04004	2,36	0,023	1,670
Seg-Hatch M\u00e9dio.04	0,06003	0,02525	2,38	0,022	1,914
Seg-Sedan Compacto.04	0,07306	0,02525	2,89	0,006	1,303
Seg-SW-Sedan Grande.04	0,05716	0,01893	3,02	0,004	1,075
Seg-Monocab.04	0,06592	0,01867	3,53	0,001	1,046

S = 0,0306726 R-Sq = 77,7% R-Sq(adj) = 72,8%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	9	0,134166	0,014907	15,85	0,000
Residual Error	41	0,038573	0,000941		
Lack of Fit	20	0,016954	0,000848	0,82	0,666
Pure Error	21	0,021619	0,001029		
Total	50	0,172739			

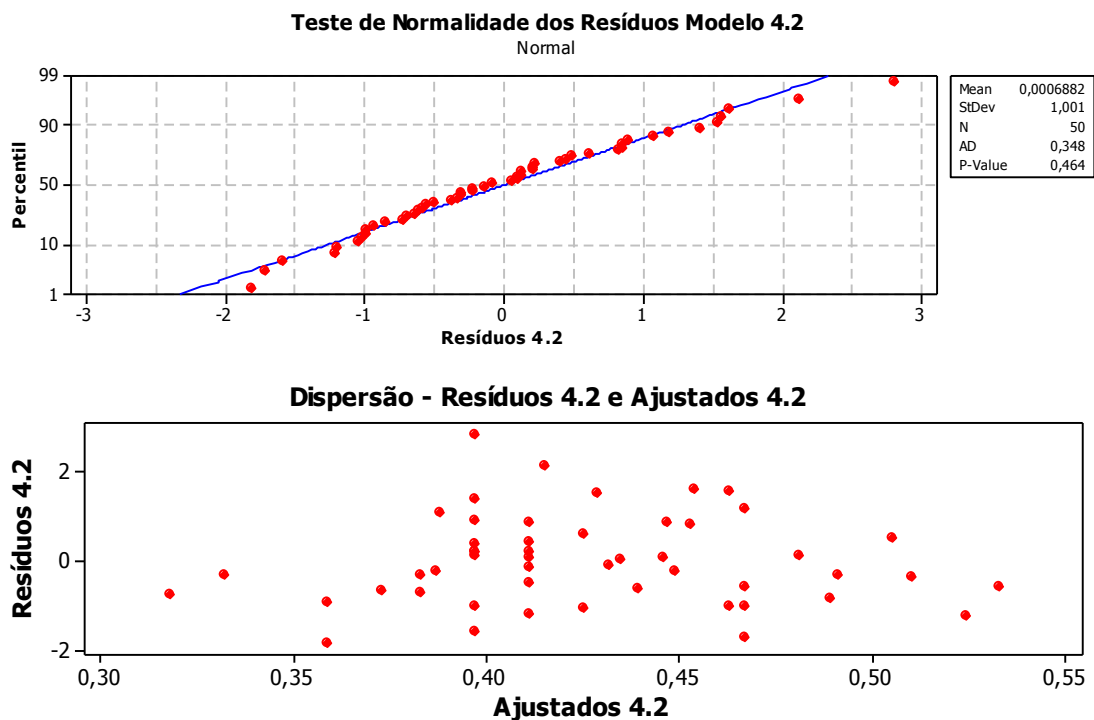
Elaboração Pr\u00f3pria

O Modelo 4.2 apresenta todas as vari\u00e1veis estatisticamente significantes e sem a presen\u00e7a de multicolinearidade. N\u00e3o temos contradi\u00e7\u00e3o entre o teste F da ANOVA e dos testes T das vari\u00e1veis preditoras, logo, refor\u00e7amos a aus\u00eancia de multicolinearidade. As vari\u00e1veis preditoras conseguem explicar 72,8% da vari\u00e2ncia da desvaloriza\u00e7\u00e3o do ve\u00edculo ap\u00f3s quatro anos e este coeficiente de determina\u00e7\u00e3o ajustado est\u00e1 em linha com o valor apresentado nos modelos anteriores e \u00e9 satisfat\u00f3rio para este estudo. Notamos que o modelo n\u00e3o contou com a presen\u00e7a da vari\u00e1vel N^o Itens de Seguran\u00e7a e nenhuma categoria das vari\u00e1veis motoriza\u00e7\u00e3o e pot\u00eancia.

Precisamos avaliar os demais pressupostos para o modelo 4.2. O gr\u00e1fico 17 apresenta o gr\u00e1fico de probabilidade normal dos res\u00edduos junto a um teste de normalidade de Anderson-Darling, conforme seu p-valor e os pontos plotados no gr\u00e1fico, notamos que os res\u00edduos possuem normalidade refor\u00e7ando a suposi\u00e7\u00e3o de

normalidade dos erros. Pelo diagrama de dispersão dos resíduos contra os valores ajustados pelo modelo, notamos que não temos correlação entre os resíduos do modelo e os valores ajustados, além de estarem bem distribuídos não apresentando nenhuma tendência que indique variância não constante dos erros. Desta forma validamos as suposições de linearidade do modelo e de homocedasticidade dos erros. E reforçamos essa suposição ao analisar o gráfico de resíduos contra as variáveis preditoras do modelo (Apêndice D), onde notamos que os pontos estão bem distribuídos e não apresentam tendência.

Gráfico 17 – Análise de Resíduos do modelo 4.2 (banco de dados de quatro anos)



Elaboração Própria

Ao analisarmos a presença de pontos de alavanca ou influência, por meio do exame dos valores de alavanca H_i e a distância de Cook, identificamos alguns pontos que poderiam ser considerados valores discrepantes, contudo, após realizamos alguns testes não vimos nenhum ganho que justificasse a retirada dessas observações do modelo. Sendo assim o modelo final para o banco de dados de quatro anos é o modelo 4.2.

A equação final do modelo é:

$$\begin{aligned} \text{VRV.04} = & 0,383 + 0,0140 \text{ N}^\circ \text{ Itens_Conforto.04} - 0,0789 \text{ M-Hyundai.04} \\ & - 0,0382 \text{ M-GM.04} + 0,0519 \text{ M-Renault.04} + 0,0943 \text{ M-Peugeot.04} \\ & + 0,0600 \text{ Seg-Hatch Médio.04} + 0,0731 \text{ Seg-Sedan Compacto.04} \\ & + 0,0572 \text{ Seg-SW-Sedan Grande.04} + 0,0659 \text{ Seg-Monocab.04} \end{aligned}$$

O intercepto do modelo 4.2 representa o valor médio da desvalorização em quatro anos de um veículo que pertença a categoria de referência, que, neste modelo de regressão, é composta pelas montadoras Ford, Fiat, Mercedes Benz, Volkswagen, Toyota, Honda e Citroen, os segmentos de Entrada, Hatch Pequeno, Sedan Pequeno, Sedan Médio e SUV. Espera-se então que um modelo de veículo utilizado por quatro anos e que pertença à categoria de referência tenha uma desvalorização média de 38,3%.

A interpretação dos coeficientes do modelo 4.2, pode ser feita conforme a seguir.

- N° Itens de Conforto: considerando todas as outras variáveis constantes, para cada item de conforto que estiver presente no modelo de veículo, será acrescido em sua desvalorização média de quatro anos 1,40 p.p.
- M-Renault.04: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a montadora Renault, sua desvalorização média em quatro anos será acrescida em 5,19 pontos percentuais em relação a um modelo de veículo das montadoras de referência.
- Seg-SW-Sedan Grande.04: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer ao segmento SW ou Sedan Grande, sua desvalorização média em quatro anos será acrescida em 5,72 pontos percentuais em relação a um veículo dos segmentos de referência.

Para exemplificar a utilização do modelo final do banco de dados de quatro anos, vamos considerar um modelo de veículo com as seguintes características: Montadora Peugeot; Segmento Hatch Médio; Potência maior que 130 e menor ou igual a 160; Motorização maior ou igual a 2.0; 3 itens de segurança e 4 de conforto.

Teríamos a seguinte equação para este modelo:

$$VRV.04 = 0,383 + 0,0140 * (4) - 0,0789 * (0) - 0,0382 * (0) + 0,0519 * (0) + 0,0943 * (1) + 0,0600 * (1) + 0,0731 * (0) + 0,0572 * (0) + 0,0659 * (0)$$

$$VRV.04 = 0,383 + 0,0560 + 0,0943 + 0,0731 = 60,64\%$$

A desvalorização média para um veículo com essas características após a utilização por quatro anos é 60,64%

4.2.5 Modelo de regressão para o banco de dados de cinco anos

O quadro 15 apresenta o modelo 5.1 contendo todas as variáveis candidatas a preditoras para o banco de dados de cinco anos. Neste banco de dados as categorias de segmento Entrada e Hatch Pequeno integram o grupo de referência; as categorias Sedan Pequeno e Sedan Compacto agora são representadas por “Seg-Sedan Pequeno-Sedan Compacto.05”. Para Motorização a categoria de referência passou a ser “ ≤ 1.4 ”. Em quanto a Potência, a referência passa a ser “ ≤ 130 ” e as categorias “ >160 e ≤ 190 ”, “ >190 e ≤ 220 ” e “ >220 ” foram agrupadas e agora são representadas por “Pot->160”.

O Modelo 5.1 apresentou praticamente todas as suas variáveis preditoras com VIF maior que 5, com exceção de “Mot->1.4 \leq 1.8”, o próprio software estatístico utilizado retirou a variável “Pot->160” por ser altamente correlacionada com outras variáveis do modelo. Tivemos variáveis com elevados valores de VIF, como “Nº Itens de Segurança”, cujo valor é 141,89. Para atingirmos um modelo com ausência de multicolinearidade, foi necessário retirar cinco variáveis do modelo, sendo elas retiradas nesta ordem: “Pot->160”, “Nº Itens de conforto.05”, “Mot- \geq 2.0”, “Pot $>130 \leq 160$ ” e “Hatch Médio”. Após esse procedimento, chegamos a um modelo que não apresentou mais nenhuma variável com VIF superior a cinco, contudo ainda tínhamos várias variáveis estatisticamente nulas no modelo.

Uma vez retiradas as variáveis que estavam violando o pressuposto de ausência de multicolinearidade, retiramos uma-a-uma aqueles que foram consideradas estatisticamente nulas. Sendo assim partimos do modelo 5.1 com 16

variáveis preditoras no modelo e chegamos ao modelo 5.2, presente no quadro 16, com 4 variáveis preditoras.

Quadro 15 – Modelo de regressão (5.1) da variável resposta VRV.05 (banco de dados de cinco anos)

```

VRV.05 = 0,501 + 0,0498 N°_Itens_Segurança.05 - 0,0250 N°_Itens_Conforto.05
- 0,134 M-Volkswagen.05 - 0,124 M-Fiat.05 - 0,0053 M-GM.05
- 0,142 M-Hyundai.05 - 0,0129 M-Renault.05 + 0,244 M-Peugeot.05
+ 0,0300 M-Honda.05 - 0,270 Seg-Hatch Médio.05
- 0,0420 Seg-Sedan Pequeno-Sedan Compact - 0,213 Seg-Sedan Médio.05
- 0,0525 Seg-SUV.05 + 0,0259 Mot-> 1.4 <= 1.8.05 + 0,144 Mot->= 2.0.05
- 0,0097 Pot-> 130 < = 160.05

Predictor           Coef   SE Coef   T       P       VIF
Constant            0,50055 0,02633  19,01  0,000
N°_Itens_Segurança.05 0,04983 0,02778   1,79  0,116 141,894
N°_Itens_Conforto.05 -0,02495 0,01403  -1,78  0,119  23,614
M-Volkswagen.05     -0,13427 0,06199  -2,17  0,067  13,347
M-Fiat.05            -0,12414 0,05995  -2,07  0,077  15,851
M-GM.05              -0,00526 0,03779  -0,14  0,893   8,501
M-Hyundai.05        -0,14235 0,08408  -1,69  0,134   8,962
M-Renault.05        -0,01287 0,03598  -0,36  0,731   5,710
M-Peugeot.05        0,24377 0,07946   3,07  0,018   8,004
M-Honda.05           0,02999 0,07451   0,40  0,699   7,039
Seg-Hatch Médio.05  -0,2698  0,1381  -1,95  0,092  66,184
Seg-Sedan Pequeno-Sedan Compact -0,04200 0,03377  -1,24  0,254   5,029
Seg-Sedan Médio.05  -0,2126  0,1320  -1,61  0,151  76,784
Seg-SUV.05          -0,05249 0,04069  -1,29  0,238   5,749
Mot-> 1.4 <= 1.8.05  0,02592 0,02398   1,08  0,315   4,563
Mot->= 2.0.05        0,14446 0,07063   2,05  0,080  22,001
Pot-> 130 < = 160.05 -0,00972 0,03951  -0,25  0,813  10,240

S = 0,0274934   R-Sq = 91,1%   R-Sq(adj) = 70,7%

Analysis of Variance

Source           DF          SS          MS          F          P
Regression       16  0,0541270  0,0033829  4,48  0,026
Residual Error   7  0,0052912  0,0007559
  Lack of Fit     6  0,0048957  0,0008159  2,06  0,488
  Pure Error      1  0,0003955  0,0003955
Total            23  0,0594182
    
```

Elaboração Própria

Chegamos ao modelo 5.2, após refazer o modelo treze vezes retirando as variáveis na seguinte ordem:

1. Pot-≥160.05
2. N° Itens de Segurança.05
3. Mot->2.0.05
4. Pot->130≤160.05

5. Seg-Hatch Medio.05 (Após a retirada desta segunda variável eliminamos a multicolinearidade do modelo, nenhuma variável apresentou VIF maior que 5)
6. Seg-Sedan Pequeno-Sedan Compacto (A partir desta variável fomos eliminando pelo critério do maior p-valor)
7. Seg-SUV.05
8. Mot->1.4≤1.8.05
9. M-GM.05
10. N° Itens de Conforto.05
11. M-Hyundai.05
12. M-Volkswagen.05
13. M-Fiat05

Quadro 16 – Modelo de regressão (5.2) da variável resposta VRV.05 (banco de dados de cinco anos)

The regression equation is

$$VRV.05 = 0,458 + 0,0498 \text{ M-Renault.05} + 0,201 \text{ M-Peugeot.05} - 0,0724 \text{ M-Honda.05} + 0,0479 \text{ Seg-Sedan Médio.05}$$

Predictor	Coef	SE Coef	T	P	VIF
Constant	0,458164	0,006729	68,08	0,000	
M-Renault.05	0,04983	0,01467	3,40	0,003	1,056
M-Peugeot.05	0,20083	0,02692	7,46	0,000	1,022
M-Honda.05	-0,07241	0,03010	-2,41	0,026	1,278
Seg-Sedan Médio.05	0,04794	0,01648	2,91	0,009	1,333

S = 0,0260632 R-Sq = 78,3% R-Sq(adj) = 73,7%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	0,046512	0,011628	17,12	0,000
Residual Error	19	0,012907	0,000679		
Total	23	0,059418			

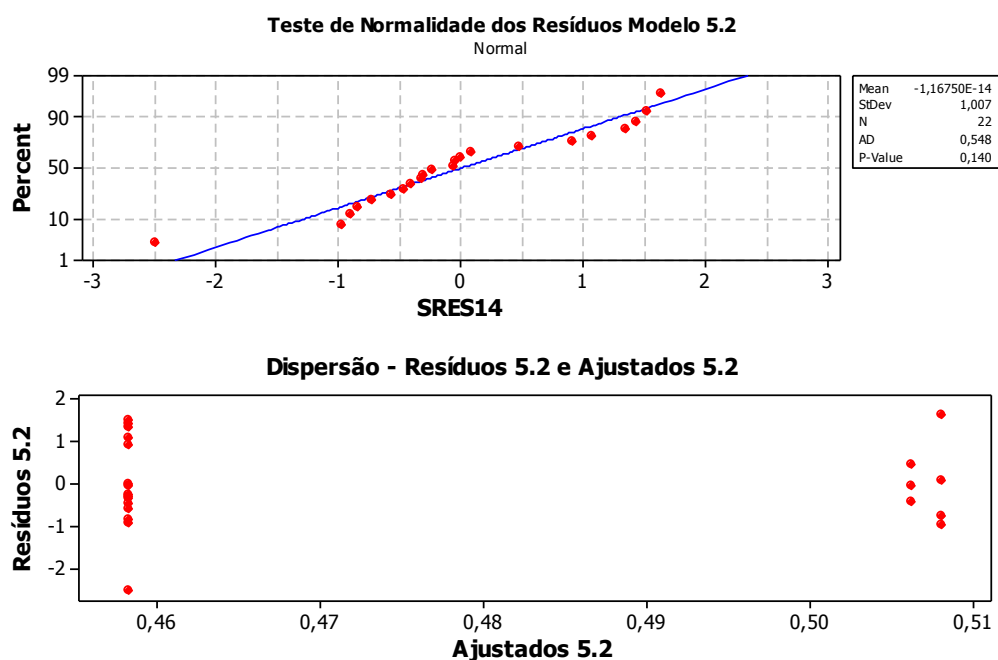
Elaboração Própria

O modelo 5.2 apresenta todas as variáveis estatisticamente significantes, os valores dos VIF menores que cinco e não há contradição entre os resultados do teste F da tabela ANOVA e dos testes T das variáveis explicativas, o que reforça a hipótese de que não existe multicolinearidade entre as variáveis deste modelo.

Neste modelo não temos a presença das variáveis N^o de Itens de Conforto, N^o de Itens de Segurança, Motorização e Potência. O coeficiente de determinação ajustado nos demonstra que as quatro variáveis preditoras são capazes de explicar 73,7% da variância do valor residual dos veículos com cinco anos de uso. Este percentual é superior aos modelos anteriores e é considerado satisfatório.

Necessitamos passar agora para a análise de resíduos a fim de validar os demais pressupostos para o modelo 5.2. O gráfico 17 avalia a suposição de normalidade dos erros do modelo por meio do gráfico de probabilidade normal dos resíduos padronizados do modelo 5.2 e também apresenta o teste de Anderson Darling aplicado aos resíduos. Observando o p-valor deste teste, podemos concluir que não temos evidências para rejeitar a hipótese de normalidade dos erros. O gráfico 17 apresenta também um diagrama de dispersão dos resíduos contra os valores ajustados pelo modelo e, neste gráfico, notamos que não existe nenhuma correlação entre os resíduos e os valores preditos pelo modelo, além da distribuição dos pontos não mostrar evidência de nenhum comportamento tendencialmente decrescente ou crescente. Com o exposto, conseguimos validar as suposições de normalidade dos erros, linearidade do modelo e homocedasticidade dos erros.

Gráfico 18 – Análise de Resíduos do modelo 5.2 (banco de dados de cinco anos)



O gráfico dos resíduos contra as variáveis preditoras do modelo (Apêndice E) corrobora com a validação das suposições de homocedasticidade e linearidade, uma vez que não apresenta nenhum padrão de correlação e nenhuma tendência ou padrão tipo “funil”.

Fizemos uma análise de possíveis pontos discrepantes, utilizando os valores de alavanca Hi e distância de Cook. De forma similar ao ocorrido no banco de dados de quatro anos, fizemos vários testes em relação aos pontos que tiveram valores elevados de Hi e distância de Cook, mas, ao retirá-los, não estávamos observando ganhos que justificassem a saída dos dados. Desta forma o modelo final para o banco de dados de cinco anos é:

$$\text{VRV.05} = 0,458 + 0,0498 \text{ M-Renault.05} + 0,201 \text{ M-Peugeot.05} \\ - 0,0724 \text{ M-Honda.05} + 0,0479 \text{ Seg-Sedan Médio.05}$$

A categoria de referência para o modelo final do banco de dados de cinco anos é composta pelas montadoras Nissan, Fiat, Volkswagen e Hyundai os segmentos de Entrada, Hatch Pequeno, Hatch Médio, SUV, Sedan Pequeno e Sedan Compacto. E, caso um modelo de veículo pertença a categoria de referência, é esperado que sua desvalorização média após a utilização por um período de cinco anos seja de 45,8%, valor representado pelo intercepto do modelo.

A interpretação dos coeficientes do modelo 5.2 pode ser feita conforme a seguir.

- Seg-Sedan Médio.05: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer ao segmento Sedan Médio sua desvalorização média em cinco anos será acrescida em 4,79 pontos percentuais em relação a um veículo dos segmentos de referência.
- M-Honda.05: considerando todas as demais variáveis constantes, se um modelo de veículo pertencer a montadora Honda, sua desvalorização média em cinco anos será reduzida em 7,24 pontos percentuais em relação a um modelo de veículo das montadoras de referência.

Para demonstrar a utilização do modelo final do banco de dados de cinco anos, vamos considerar um modelo de veículo com as seguintes características: Montadora Renault; Segmento Suv; Potência maior que 130 e menor ou igual a 160; Motorização maior ou igual a 2.0; 2 itens de segurança e 5 de conforto.

Teríamos a seguinte equação para este modelo:

$$VRV.05 = 0,458 + 0,0498 * (1) + 0,201 * (0) - 0,0724 * (0) + 0,0479 * (0)$$

$$VRV.05 = 0,458 + 0,0498 = 50,78\%$$

A desvalorização média para um veículo com essas características após a utilização por cinco anos é 50,78%

5. Conclusão

Esse trabalho teve como objetivo construir cinco modelos de regressão que conseguissem estimar o valor residual de um modelo de veículo, após sua utilização por um, dois, três, quatro e cinco anos. Foram construídos cinco bancos de dados, um para cada período de desvalorização estudado e, ao todo, foram considerados 49 modelos de veículos, que se desdobraram em 312 observações divididas entre os cinco bancos de dados.

Seis características dos modelos de veículos foram escolhidas como possíveis variáveis preditoras, sendo quatro categóricas (Montadora, Segmento, Motorização e Potência), e duas quantitativas discretas (Nº de Itens de Segurança e Nº de Itens de Conforto). Dada a natureza das variáveis categóricas, elas poderiam se desdobrar em 31 variáveis indicadoras. Logo poderiam ser inseridas 33 variáveis em cada modelo, considerando as variáveis quantitativas discretas.

Na análise exploratória de dados, analisamos o comportamento de cada variável em relação a variável resposta de cada banco de dados e, em alguns momentos, optamos por agrupar algumas categorias das variáveis categóricas, dado sua similaridade de suas respostas, a fim de construirmos modelos mais parcimoniosos.

Chegou-se, então, a cinco modelos de regressão, que conseguiram atender a todos os pressupostos para um bom ajuste de modelo e que apresentaram um coeficiente de determinação ajustado satisfatório. Os modelos são capazes de explicar a influência das características de um veículo em sua desvalorização ao longo do tempo. Conseguem também estimar qual é a desvalorização média esperada para um veículo após um período de utilização, dado algumas de suas características.

Os modelos de regressão compartilham de alguns pontos em comum: em nenhum deles, as variáveis candidatas a preditoras “Nº de Itens de Segurança” e Motorização conseguiram explicar a variância da variável resposta. Logo, essas variáveis não estão presentes em nenhum modelo e conclui-se que, com o conjunto de dados estudado, não foi possível detectar a importância dessas variáveis para explicar a desvalorização de um veículo ao longo do tempo. Montadora e Segmento são as variáveis mais relevantes para explicar a desvalorização de um veículo ao longo do tempo, pois estão presentes em todos os modelos de regressão com ao menos uma de suas categorias. “Nº de Itens de Conforto” é uma variável que está presente nos modelos de regressão que consideram o período de um a quatro anos de uso, mas não foi considerada importante para explicar a depreciação do veículo depois de cinco anos de uso. A partir do período de três anos de utilização, de acordo com os dados analisados, a variável Potência perde sua capacidade de explicar a desvalorização de um modelo de veículo, pois essa variável está presente apenas nos dois primeiros modelos. Contudo, no modelo para o período de dois anos, a variável Potência ainda apresentou um comportamento diferente do que era esperado considerando-se o conhecimento prévio sobre a desvalorização de veículos. Apenas uma categoria de variável esteve presente em todos os modelos de regressão: a categoria da variável Montadora. Em todos os modelos de regressão, o fato de um modelo de veículo pertencer à montadora Renault faz com que sua desvalorização aumente em relação às outras montadoras presentes nas categorias de referência de cada modelo de regressão.

As variáveis “Pot->130≤160”, “Pot->160≤190”, “M-Fiat” e “M-BMW” contribuíram para explicar parte da variância da desvalorização de um modelo de

veículo após um ano e não contribuíram em mais nenhum outro modelo. Já a variável “Seg-Hatch Pequeno” só conseguiu explicar parte da variância da desvalorização de um modelo de veículo no período de dois anos. A variável “Seg-SUV” auxiliou na explicação de parte da variância da desvalorização de um veículo com três anos e não teve participação em nenhum outro modelo. E as variáveis “M-GM” e “Seg-Sedan Compacto” tiveram participação apenas no modelo de regressão para o banco de dados de quatro anos.

O modelo de regressão final para o quinto ano de utilização de um modelo de veículo apresentou um número de variáveis preditoras bem distinto dos demais modelos. Isto pode ser devido ao fato que o banco de dados para cinco anos tem um número restrito de observações. Enquanto os modelos de regressão para um, dois, três e quatro anos contaram respectivamente com 11, 10, 10 e 9 variáveis preditoras, respectivamente, e o modelo de cinco anos contou com 4 variáveis preditoras. Para o banco de dados de cinco anos, as variáveis “M-Honda.05” e “M-Peugeot.05”, da categoria Montadora, contam apenas com uma observação no banco de dados e que, quando essas variáveis são retiradas do modelo, reduzem consideravelmente o coeficiente de determinação ajustado para este modelo. Desta forma, cria-se uma incerteza em relação a confiabilidade deste modelo de regressão para cinco anos, pois, mesmo que ele atenda todos os requisitos necessários a um bom ajuste, existe uma dependência grande em relação a apenas dois tipos de veículos.

Além desta limitação mencionada em relação ao banco de dados de cinco anos, este estudo também apresentou outros fatores que limitaram nossa análise. O banco de dados como um todo não contou com o grande número de observações, dadas as dificuldades existentes em considerar um mesmo modelo de veículo ao longo do tempo, pois, como mencionado no início deste trabalho, os veículos sofrem reestilizações ou facelifts que podem alterar (ou não) drasticamente sua estética ou posicionamento de preço. Desta forma, não tivemos como separar uma parte do banco de dados para trabalhar com um grupo de validação, o que nos impediu de comparar a qualidade das estimativas feitas pelos modelos de regressão com dados reais. Outra limitação refere-se a fato de que trabalhamos com apenas seis

características de um veículo, pois poderíamos ter testado um número maior de variáveis candidatas a preditoras, caso tivéssemos mais características de um veículo nos bancos de dados disponíveis *on-line*.

Por fim, entendemos que este trabalho alcançou seu objetivo na medida em que os modelos de regressão elaborados conseguiram atingir os objetivos propostos e, como sugestão para estudos futuros, recomenda-se construir modelos de regressão considerando um maior número de modelos de veículos e que se incorpore outras variáveis como, cor do modelo de veículo, volume do porta malas, comprimento do modelo de veículo, entre outras.

6. Referências Bibliográficas

CARROS NA WEB. Disponível em <<https://www.carrosnaweb.com.br>>. Acesso em: Novembro 2017

FENABRAVE. Disponível em <<http://www3.fenabreve.org.br:8082/plus/>>. Acesso em: Novembro 2017

GUJARATI, D.; PORTER, D. Econometria Básica. Quinta Edição. São Paulo: McGraw-Hill 2008.

ICARROS. Disponível em <<https://www.icarros.com.br>>. Acesso em: Novembro 2017

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. São Paulo: LTC Livros Técnicos e Científicos, 2009.

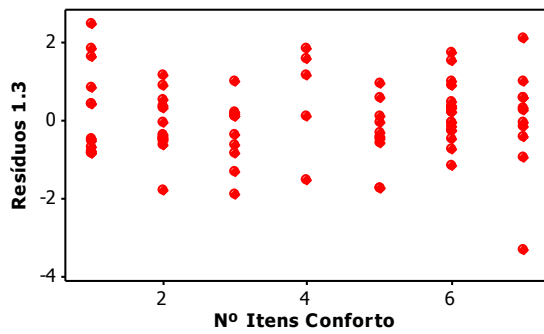
PAGANO, M.; GAUVREAU, K. Princípios de bioestatística. São Paulo: Pioneira Thompson Learning, 2004

WEBMOTORS. Disponível em <<https://www.webmotors.com.br>>. Acesso em: Novembro 2017

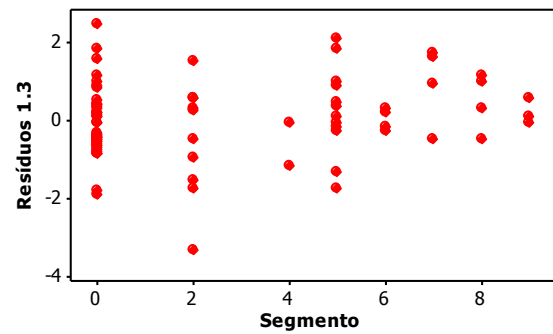
7. Apêndices

Apêndice A – Gráfico de Resíduos e Preditoras banco de dados de um ano

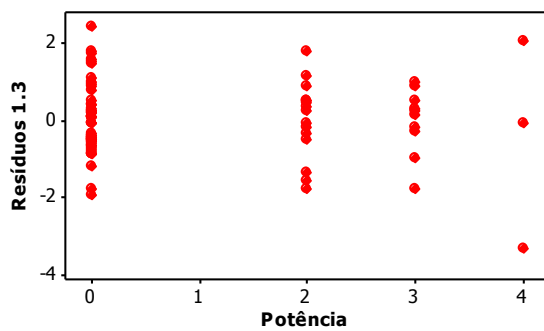
Dispersão Resíduos e Preditora Nº Itens Conforto



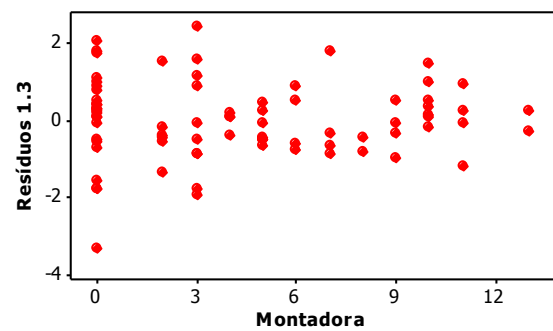
Dispersão Resíduos e Preditora Segmento



Dispersão Resíduos e Preditora Potência



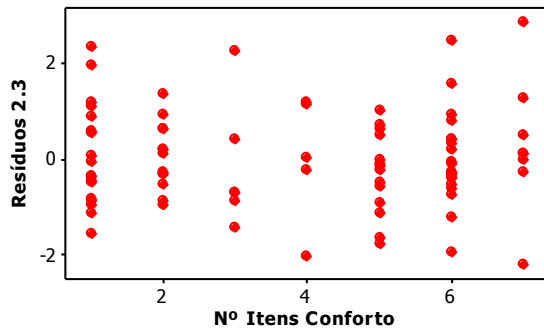
Dispersão Resíduos e Preditora Montadora



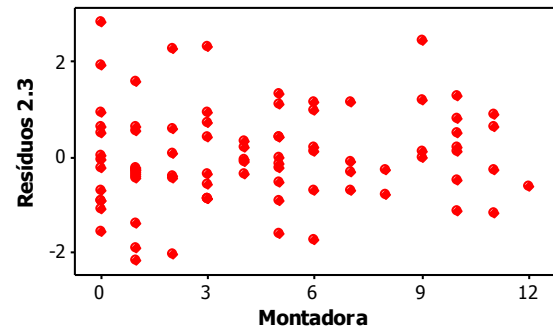
Elaboração Própria

Apêndice B – Gráfico de Resíduos e Predictoras banco de dados de dois anos

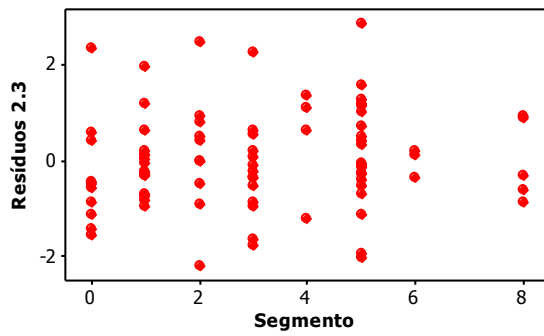
Dispersão Resíduos e Preditora N° Itens de Conforto



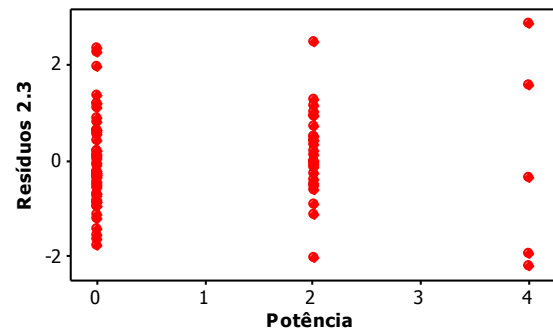
Dispersão Resíduos e Preditora Montadora



Dispersão Resíduos e Preditora Segmento

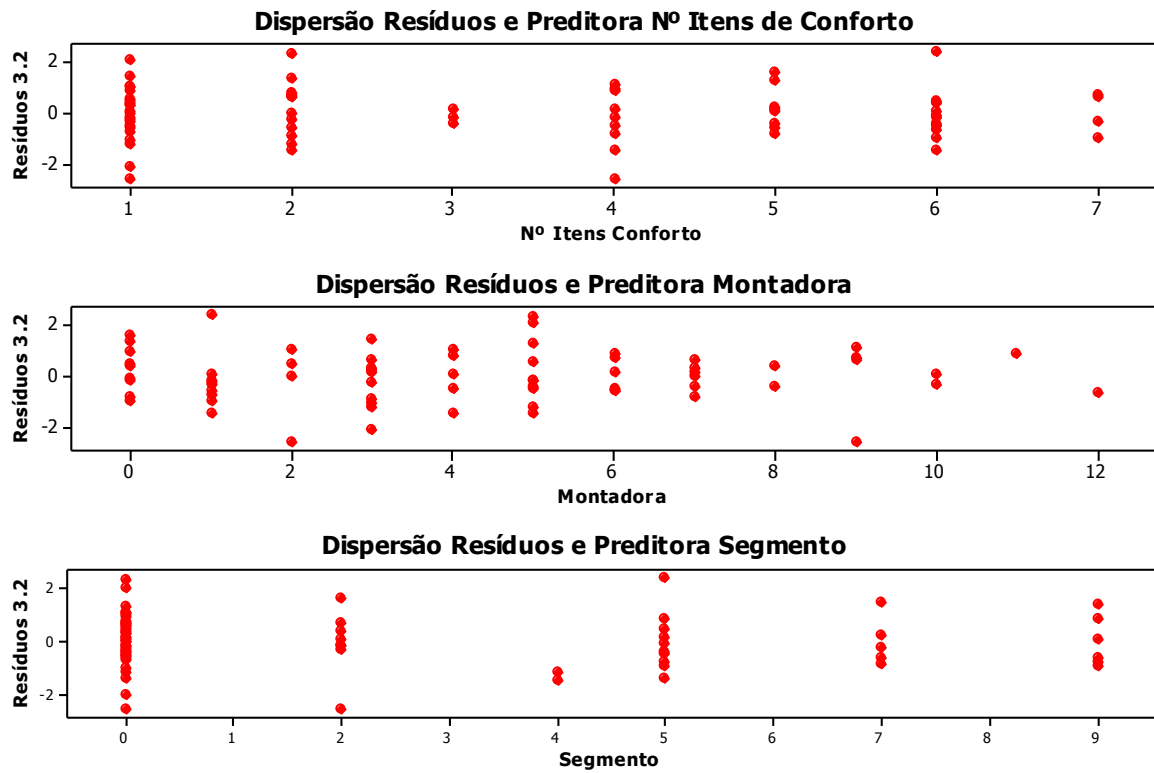


Dispersão Resíduos e Preditora Montadora



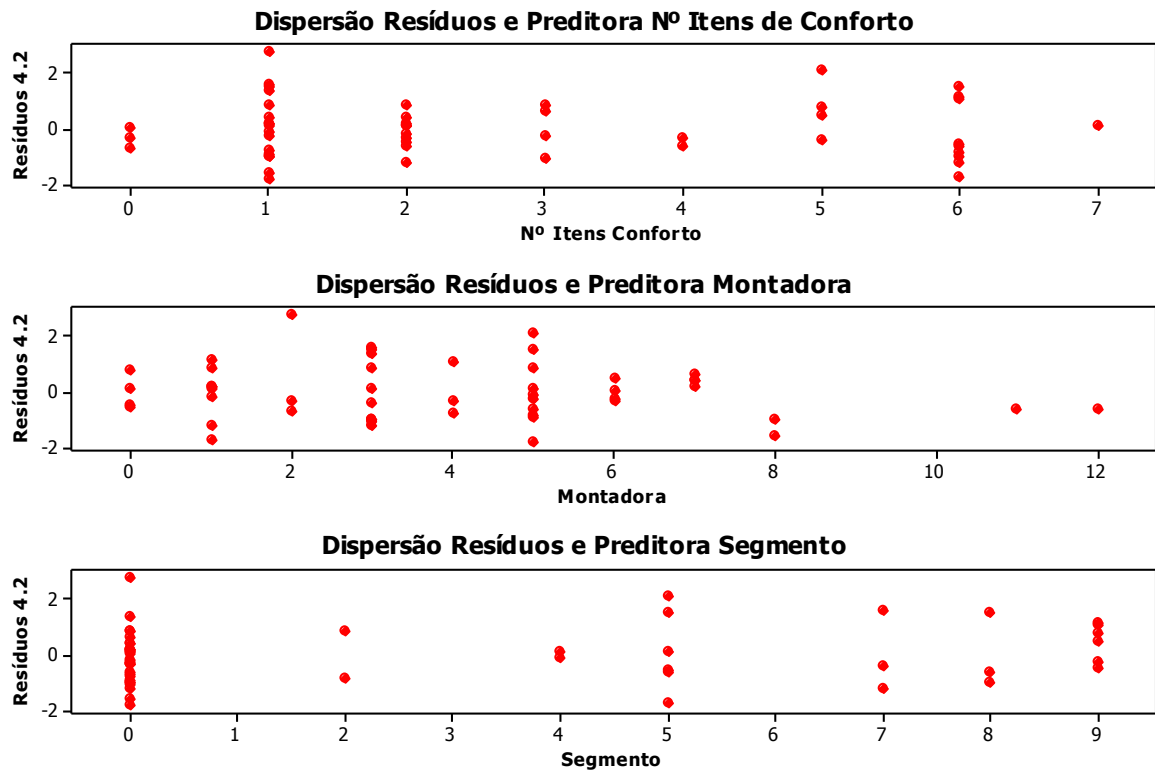
Elaboração Própria

Apêndice C – Gráfico de Resíduos e Preditoras banco de dados de três anos



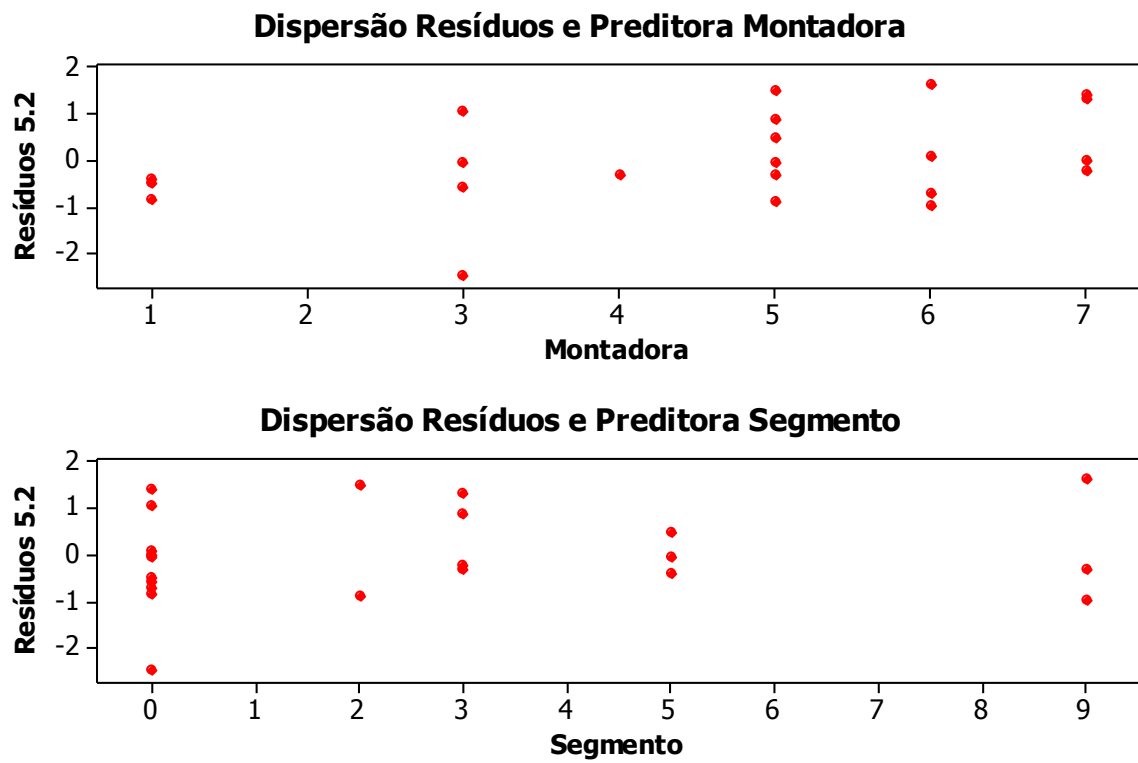
Elaboração Própria

Apêndice D – Gráfico de Resíduos e Preditoras banco de dados de quatro anos



Elaboração Própria

Apêndice E – Gráfico de Resíduos e Preditoras banco de dados de cinco anos



Elaboração Própria