

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE MÉTODOS ESTATÍSTICOS

Gustavo Arantes Silva
Luan Coelho Vieira da Silva

MODELOS LINEARES MISTOS COM EFEITOS
ALEATÓRIOS CRUZADOS PARA ANÁLISE DE DADOS DA
MODALIDADE AUTOMOBILÍSTICA DE FÓRMULA 1

RIO DE JANEIRO
2023

GUSTAVO ARANTES SILVA
LUAN COELHO VIEIRA DA SILVA

MODELOS LINARES MISTOS COM EFEITOS
ALEATÓRIOS CRUZADOS PARA ANÁLISE DE DADOS DA
MODALIDADE AUTOMOBILÍSTICA DE FÓRMULA 1

**Projeto Final de Curso submetido ao
Departamento de Métodos Estatísticos
do Instituto de Matemática da Univer-
sidade Federal do Rio de Janeiro como
requisito para obtenção do grau de Ba-
charel em Estatística.**

Rio de Janeiro, março de 2023

UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

**GUSTAVO ARANTES SILVA
LUAN COELHO VIEIRA DA SILVA**

Projeto Final de Curso submetido ao Departamento de Métodos Estatísticos do Instituto de Matemática da Universidade Federal do Rio de Janeiro como requisito para obtenção do grau de Bacharel em Estatística.

Orientadora: Professora Marina S. Paez -
UFRJ, D.Sc.

Prof. - UFRJ

Prof. - UFRJ

Prof. - UFRJ

Rio de Janeiro, 10 de março de 2023

RESUMO

Inserir resumo

Palavras-chave:.

ABSTRACT

Resumo em Inglês

Keywords: Palavras Chaves.

LISTA DE ILUSTRAÇÕES

Figura 1 – Boxplot dos tempos de volta por Grande Prêmio	8
Figura 2 – Visualização das trocas de marcha nas voltas mais rápidas de cada Grande Prêmio	10
Figura 3 – Boxplot dos tempos de volta por piloto	11
Figura 4 – Comportamento ao longo das voltas de cada piloto no Grande Prêmio da Holanda	12
Figura 5 – Gráfico dos tempos de volta	13
Figura 6 – Gráfico dos tempos de volta com cores específicas para cada circuito . .	14
Figura 7 – Gráfico dos tempos de volta por Grande Prêmio	15
Figura 8 – Modelo com um efeito fixo	18
Figura 9 – Modelo com um efeito fixo e efeito aleatório no intercepto	19
Figura 10 – Modelo com efeito aleatório no intercepto e no coeficiente angular da variável explicativa	20
Figura 11 – Exemplo de voltas aninhadas por circuito e piloto de maneira cruzada .	21
Figura 12 – Exemplo de voltas aninhadas por piloto e circuito de maneira cruzada .	22

LISTA DE TABELAS

Tabela 1 – Tabela de pontuação do Campeonato de Fórmula 1 de 2022	2
Tabela 2 – Tabela de Grandes Prêmios do Campeonato de Fórmula 1 de 2022 . .	3
Tabela 3 – Resultados por equipe do Campeonato de Fórmula 1 de 2022	4
Tabela 4 – Resultados por piloto do Campeonato de Fórmula 1 de 2022	4
Tabela 5 – Características dos circuito	9
Tabela 6 – Tabela de pontuação do Campeonato de Fórmula 1 de 2022	26

LISTA DE ABREVIATURAS E SIGLAS

F1	Fórmula 1
GP	Grande Prêmio (de Fórmula 1)

LISTA DE SÍMBOLOS

Ω	Espaço amostral
ω	Evento pertencente ao espaço amostral Ω
\mathbb{R}	Conjunto dos números reais

SUMÁRIO

1	INTRODUÇÃO	1
1.1	A Fórmula 1	1
1.1.1	Campeonato Mundial de Fórmula 1 de 2022	2
1.2	Objetivo	5
1.2.1	Objetivo Geral	5
1.2.2	Objetivos Específicos	5
1.3	Tratamento dos Dados	6
1.4	Análise Exploratória	7
1.4.1	<i>Grid Position</i>	7
1.4.2	Pneus	7
1.4.3	Circuitos	8
1.4.4	Pilotos	10
1.4.5	Voltas	10
1.4.6	Fatores Externos	14
2	REVISÃO DE LITERATURA	16
2.1	Análise de Regressão	16
2.2	Modelos Lineares	17
2.2.1	Método de Mínimos Quadrados	17
2.3	Modelos Lineares Mistos	17
2.3.1	Dados aninhados	17
2.3.2	Dados cruzados	17
2.3.3	Efeitos mistos	17
2.4	Estatística Bayesiana	18
3	MODELOS PROPOSTOS	21
3.1	Estrutura dos dados	21
3.2	Modelos	23
3.3	Procedimento de Inferência	24
3.4	CrITÉRIOS de Comparação	25
3.4.1	DIC	25
3.4.2	Erro Quadrático Médio	25
4	RESULTADOS	26
4.1	Comparação de Modelos	26
4.2	Análises dos resultados	26

5	CONSIDERAÇÕES FINAIS	27
	REFERÊNCIAS	28
	APÊNDICES	29
	APÊNDICE A – APÊNDICE 1	30

1 INTRODUÇÃO

1.1 A Fórmula 1

Fórmula 1 (F1, de forma abreviada) é a modalidade de automobilismo mais popular do mundo. O "Campeonato Mundial de Pilotos", torneio da modalidade que premia pilotos e equipes que conseguirem maior pontuação na temporada, existe desde 1950. Durante estas mais de sete décadas atraiu a atenção dos apaixonados por carros e o interesse das maiores montadoras do planeta. A história da Fórmula 1 é inseparável da história de algumas gigantes do automobilismo como Ferrari, Williams, Renault e McLaren.

O alto valor de mercado das empresas que veem na modalidade uma forma de consolidar ainda mais suas marcas eleva o nível de competitividade da modalidade. Com isso, há investimentos milionários feitos pelas equipes todos os anos. Tais investimentos são baseados em pesquisas avançadas buscando o máximo desempenho nas pistas. Fatores como aerodinâmica, tração dos pneus, entre outros são estudados com afincos a fim de maximizar a performance.

[Tremayne 2006, 6] descreve o carro de Fórmula 1 como uma confecção das mais atualizadas técnicas de desenvolvimento de produto e pesquisa criada dentro de uma escala de tempo que é inevitavelmente muito curta, para ser enviada para uma competição acirrada da qual não há trégua durante 'a temporada'.

Não somente as marcas, mas também diversos pilotos ganharam notoriedade na história do esporte. Dentre eles o alemão Michael Schumacher, vencedor de sete temporadas, e o inglês Lewis Hamilton, também detentor de sete títulos e recordista de vitórias em corridas - com 103 conquistas. Há também ídolos brasileiros como Nelson Piquet - vencedor de três temporadas, Emerson Fittipaldi - duas vezes campeão, Rubens Barrichello - detentor do recorde de corridas disputadas - 322, e o maior deles, Ayrton Senna - também com três títulos.

Uma temporada de Fórmula 1 consiste em corridas - conhecidas com Grandes Prêmios (GPs). Os GPs são realizados em diversos locais do mundo, em pistas construídas especialmente para a prática da modalidade.

Falar de qualificatória bla bla

bla

bla

1.1.1 Campeonato Mundial de Fórmula 1 de 2022

Em 2022, o prêmio de melhores piloto e montadora foram dadas para o condutor e equipe com maior número de pontos ao longo de 22 Grandes Prêmios durante o ano. Disputaram o título 10 equipes. Cada equipe participa com 2 carros e, consequentemente, 2 pilotos.

Tabela 1 – Tabela de pontuação do Campeonato de Fórmula 1 de 2022

Corrida		Corrida Sprint	
Posição	Pontuação	Posição	Pontuação
1	25	1	8
2	18	2	7
3	15	3	6
4	12	4	5
5	10	5	4
6	8	6	3
7	6	7	2
8	4	8	1
9	2	Volta mais rápida da corrida	
10	1	1	1

Conforme dados da Tabela 1, por Grande Prêmio, de modo geral, é possível pontuar entre 0 e 26 pontos - 25 na corrida, e 1 de bonificação de volta mais rápida da corrida. No ano de 2022, os Grandes prêmios de Emília-Romanha, Áustria e São Paulo contaram também com a corrida sprint, possibilitando até mais 8 pontos no evento.

Como não é escopo do trabalho falar sobre cada piloto individualmente, e a fim de evitar referenciá-los de formas distintas ao longo deste projeto, serão utilizados os sobrenomes dos pilotos na forma como são apresentados na transmissão televisiva globalmente.

Os pilotos podem ser substituídos no decorrer do campeonato. Não é comum, e não é uma estratégia interessante pois cada piloto pontua individualmente e quando um piloto novo entra na disputa ele não herda os pontos do seu antecessor. O piloto Albon - da equipe Williams teve um problema de saúde e foi substituído por de Vries no Grande Prêmio da Itália. Ainda, Vettel - da equipe Aston Martin foi diagnosticado com COVID no começo da temporada e não disputou as duas primeiras provas do ano - Grande Prêmio do Bahrein e Grande Prêmio da Arábia Saudita, sendo substituído por Hulkenberg.

A Tabela 2 mostra quais foram os eventos da temporada de 2022, com a data de realização e o piloto vencedor. Trata-se de um campeonato longo, com duração de 8 meses e corridas em circuitos espalhados pelo mundo. As equipes precisam levar isso em consideração ao fazer o planejamento, e adaptar-se conforme o andamento do campeonato.

Tabela 2 – Tabela de Grandes Prêmios do Campeonato de Fórmula 1 de 2022

Nome do Evento	Data do Evento	Piloto Vencedor
Grande Prêmio do Bahrein	20/3/2022	Leclerc - Ferrari
Grande Prêmio da Arábia Saudita	27/3/2022	Verstappen - Red Bull Racing
Grande Prêmio da Austrália	10/4/2022	Leclerc - Ferrari
Grande Prêmio Emília-Romanha	24/4/2022	Verstappen - Red Bull Racing
Grande Prêmio de Miami	8/5/2022	Verstappen - Red Bull Racing
Grande Prêmio da Espanha	22/5/2022	Verstappen - Red Bull Racing
Grande Prêmio de Mônaco	29/5/2022	Perez - Red Bull Racing
Grande Prêmio do Azerbaijão	12/6/2022	Verstappen - Red Bull Racing
Grande Prêmio do Canadá	19/6/2022	Verstappen - Red Bull Racing
Grande Prêmio da Inglaterra	3/7/2022	Sainz - Ferrari
Grande Prêmio da Áustria	10/7/2022	Leclerc - Ferrari
Grande Prêmio da França	24/7/2022	Verstappen - Red Bull Racing
Grande Prêmio da Hungria	31/7/2022	Verstappen - Red Bull Racing
Grande Prêmio da Bélgica	28/8/2022	Verstappen - Red Bull Racing
Grande Prêmio da Holanda	4/9/2022	Verstappen - Red Bull Racing
Grande Prêmio da Itália	11/9/2022	Verstappen - Red Bull Racing
Grande Prêmio de Singapura	2/10/2022	Perez - Red Bull Racing
Grande Prêmio do Japão	9/10/2022	Verstappen - Red Bull Racing
Grande Prêmio dos EUA	23/10/2022	Verstappen - Red Bull Racing
Grande Prêmio da Cidade do México	30/10/2022	Verstappen - Red Bull Racing
Grande Prêmio de São Paulo	13/11/2022	Russel - Mercedes
Grande Prêmio de Abu Dhabi	20/11/2022	Verstappen - Red Bull Racing

Tabela 3 – Resultados por equipe do Campeonato de Fórmula 1 de 2022

Posição	Equipe	Vitórias	Pontos
1	Red Bull Racing	17	759
2	Ferrari	4	554
3	Mercedes	1	515
4	Alpine	0	173
5	McLaren	0	159
6	Alfa Romeo	0	55
7	Aston Martin	0	55
8	Haas F1 Team	0	37
9	AlphaTauri	0	35
10	Williams	0	8

Tabela 4 – Resultados por piloto do Campeonato de Fórmula 1 de 2022

Posição	Piloto	Equipe	Vitórias	Pontos
1	Verstappen	Red Bull Racing	15	454
2	Leclerc	Ferrari	3	308
3	Perez	Red Bull Racing	2	305
4	Russel	Mercedes	1	275
5	Sainz	Ferrari	1	246
6	Hamilton	Mercedes	0	240
7	Norris	McLaren	0	122
8	Ocon	Alpine	0	92
9	Alonso	Alpine	0	81
10	Bottas	Alfa Romeo	0	49
11	Vettel	Aston Martin	0	37
12	Ricciardo	McLaren	0	37
13	Magnussen	Haas F1 Team	0	25
14	Gasly	AlphaTauri	0	23
15	Stroll	Aston Martin	0	18
16	Tsunoda	AlphaTauri	0	12
17	Schumacher	Haas F1 Team	0	12
18	Zhou	Alfa Romeo	0	6
19	Albon	Williams	0	4
20	Latifi	Williams	0	2
21	de Vries	Williams	0	2
22	Hulkenberg	Aston Martin	0	0

As tabelas 3 e 4 mostram que a temporada de 2022 foi dominada pela equipe Red Bull Racing, vencedora de 17 dos 22 Grandes Prêmios. Ferrari e Mercedes foram as outras a conseguir vitórias em eventos - 4 e 1 -, respectivamente. Especificamente, o piloto Verstappen foi o responsável por 15 das 17 vitórias de sua equipe.

EXPLICAR SAFETY CAR, VIRTUAL SAFETY CAR, GRID...

1.2 Objetivo

1.2.1 Objetivo Geral

Este trabalho tem como objetivo estudar as variáveis que impactam o tempo necessário para completar uma volta em uma corrida de Fórmula 1.

1.2.2 Objetivos Específicos

Fundamentados no objetivo geral apresentado, destacam-se:

- Explorar os dados do Campeonato Mundial de Fórmula 1 de 2022 e apresentá-los de forma concisa;
- Enumerar as principais variáveis que influenciam na nossa variável de interesse - o tempo para completar uma volta;
- Apresentar um modelo linear misto sob estatística bayesiana que consiga prever nossa variável de interesse com alto grau de credibilidade;

1.3 Tratamento dos Dados

Para o presente trabalho foram utilizados dados da API para desenvolvedor Ergast, de uso livre para fins não comerciais. Esta ferramenta disponibiliza dados desde o primeiro Campeonato Mundial de Fórmula 1, em 1950 e está em constante evolução. Estes dados são referência na comunidade de desenvolvedores do ecossistema da Fórmula 1 - servindo de base para sites e aplicativos.

Iremos considerar apenas 20 pilotos. Os pilotos de Vries e Hulkenberg terão seus dados agregados aos de seus companheiros de equipe Albon e Vettel, que disputaram 21 e 20 das 22 corridas, respectivamente. Para não gerar dúvidas, serão utilizados Albon* e Vettel* no restante do trabalho.

Ainda, foram desconsiderados da análise deste trabalho os Grandes Prêmios de: Emília-Romanha, Japão, Mônaco e Singapura. A razão para isso foi pelo fato de haver chuva durante as corridas. Embora tenhamos dados de presença ou ausência de chuva, estes são insuficientes para medir o impacto na corrida. Para fazê-lo de forma eficiente seria necessário dados pluviométrico durante cada volta além de informação sobre a capacidade de absorção do asfalto.

Desta forma, o objeto deste trabalho serão as demais 18 corridas realizadas no ano de 2022. Ainda, para garantir a confiabilidade das análises foram:

- retiradas as voltas nas quais o piloto entra ou sai do pit stop
- retiradas as voltas com e imediatamente após *Safety Car*/ *Virtual Safety Car*.
- mantidas apenas voltas sob bandeiras verde e amarela

A saber, o *Safety Car*... A bandeira verde está presente quando não há perigo na pista. Já a bandeira amarela significa que há perigo, porém a corrida, a princípio, não será interrompida. Como foram retiradas as voltas com *Safety Car*, as bandeiras amarelas restantes tendem a ser não problemáticas.

1.4 Análise Exploratória

Nesta seção serão explicadas as principais variáveis utilizadas neste trabalho a fim de explicar o tempo de uma volta em corrida. Estas variáveis sabidamente impactam na nossa variável de interesse, e são levadas em consideração nas análises das equipes de Fórmula 1.

1.4.1 *Grid Position*

A cada Grande Prêmio são realizadas 3 provas qualificatórias - Q1, Q2 e Q3, no mesmo circuito da corrida. Essas provas definem a posição inicial do piloto na corrida, usando como critério seu melhor tempo de volta. Isso dá aos pilotos que largam nas primeiras posições uma enorme vantagem, pois a ultrapassagem é uma das tarefas mais difíceis na disputa automobilística.

O termo *grid position* é utilizado para definir esta colocação, e o primeiro colocado nas qualificatórias larga na *pole position*, isto é, primeira posição. O *grid position* é um ótimo indicativo da corrida pois é consequência de um teste do piloto e seu carro no local da corrida.

1.4.2 Pneus

Por corrida, as equipes podem usar 3 tipos de pneus anunciados pela organização do evento dentre 5 possíveis. Em caso de chuva, outros 2 são disponibilizados. Como os eventos com chuva foram desconsiderados desta análise, iremos nos restringir ao primeiro caso. As informações abaixo foram retiradas do site da Pirelli, empresa responsável pelo fornecimento de pneus do campeonato. [Pirelli]

- C1: o pneu mais duro, escolhido para circuitos que desgastam mais os pneus. Desenvolvido para atingir máxima resistência ao calor, é capaz de percorrer distâncias maiores com baixa diminuição de performance;
- C2: o segundo mais duro, porém um pouco mais versátil que o C1. Funciona para pistas mais quentes e rápidas como o C1;
- C3: está presente em todas as corridas. Pode ser nominado pela organização como pneu duro, pneu médio ou macio, e portanto é extremamente versátil;
- C4: desenvolvido para performar bem em pistas apertadas e com muitas curvas. O desgaste deste pneu é alto e ele atinge o pico de performance rapidamente;
- C5: o pneu mais macio, ideal para pistas mais lentas que desgastam pouco o pneu e requer que o piloto manobre o carro constantemente nas curvas.

Em outras palavras, quanto mais macio for o pneu mais fácil será para o piloto realizar as curvas e manobrar o veículo. No entanto, o desgaste também será maior, e em poucas voltas a aderência do pneu ao asfalto do circuito pode comprometer-se. Analogamente, quanto mais duro for o pneu maior será sua durabilidade e resistência ao calor, sendo esta característica ideal para pistas com retas longas e altas velocidades.

1.4.3 Circuitos

Cada circuito tem suas características únicas, demandando adaptabilidade dos pilotos e equipes ao longo do campeonato. Esta variabilidade acarreta em diferenças no tempo de volta em cada pista, tanto na medida de valor central - mediana, como na concentração dos tempos do evento, conforme mostrado na figura abaixo.

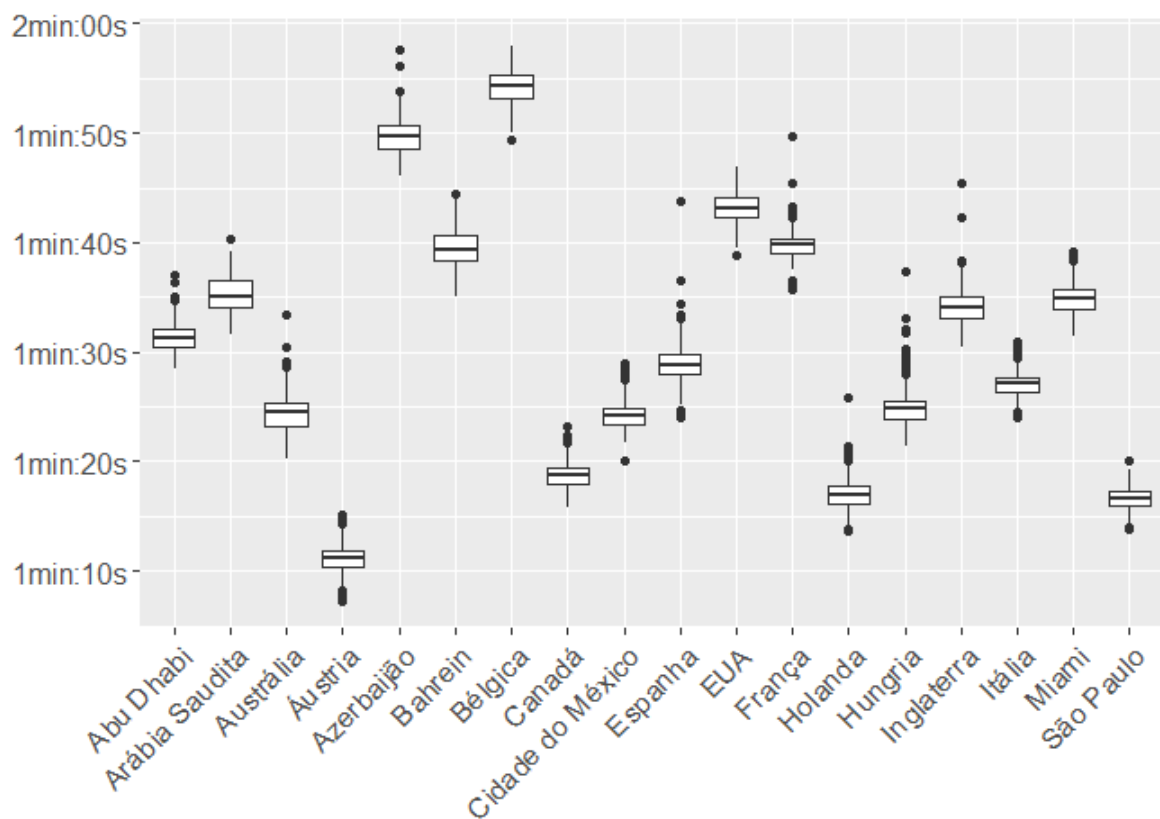


Figura 1 – Boxplot dos tempos de volta por Grande Prêmio

No GP da Áustria a maior parte das voltas foram completadas em pouco mais de 1 minuto e 10 segundos. Já no GP da Bélgica, para que o mesmo ocorresse, foi necessário cerca de 40 segundos a mais. Isto nos mostra que tratar as voltas de diferentes pistas como dados independentes não é uma boa estratégia para traçar um modelo que explique o tempo de volta. A seguir serão listados alguns dados que corroboram esta afirmação.

Tabela 5 – Características dos circuito

Local do Evento	Pneu mácio	Pneu médio	Pneu duro	Comprimento do circuito (km)	Voltas
Bahreïn	C3	C2	C1	5,412	57
Arábia Saudita	C4	C3	C2	6,174	50
Austrália	C5	C3	C2	5,303	58
Miami	C4	C3	C2	5,412	57
Espanha	C3	C2	C1	4,675	66
Azerbaijão	C5	C4	C3	6,003	51
Canadá	C5	C4	C3	4,361	70
Inglaterra	C3	C2	C1	5,891	52
Áustria	C5	C4	C3	4,318	71
França	C4	C3	C2	5,842	53
Hungria	C4	C3	C2	4,381	70
Bélgica	C4	C3	C2	7,004	44
Holanda	C3	C2	C1	4,259	72
Itália	C4	C3	C2	5,793	53
EUA	C4	C3	C2	5,513	56
Cidade do México	C4	C3	C2	4,304	71
São Paulo	C4	C3	C2	4,309	71
Abu Dhabi	C5	C4	C3	5,281	58

Aqui vale observar que o circuito do GP da Áustria possui 4,318km de comprimento, enquanto o circuito do GP da Bélgica possui 7,004km. O comportamento visto no boxplot ocorre, em grande parte, por este fator. Outro fator relevante é o conjunto de pneus adotados, definidos conforme características específicas da pista.

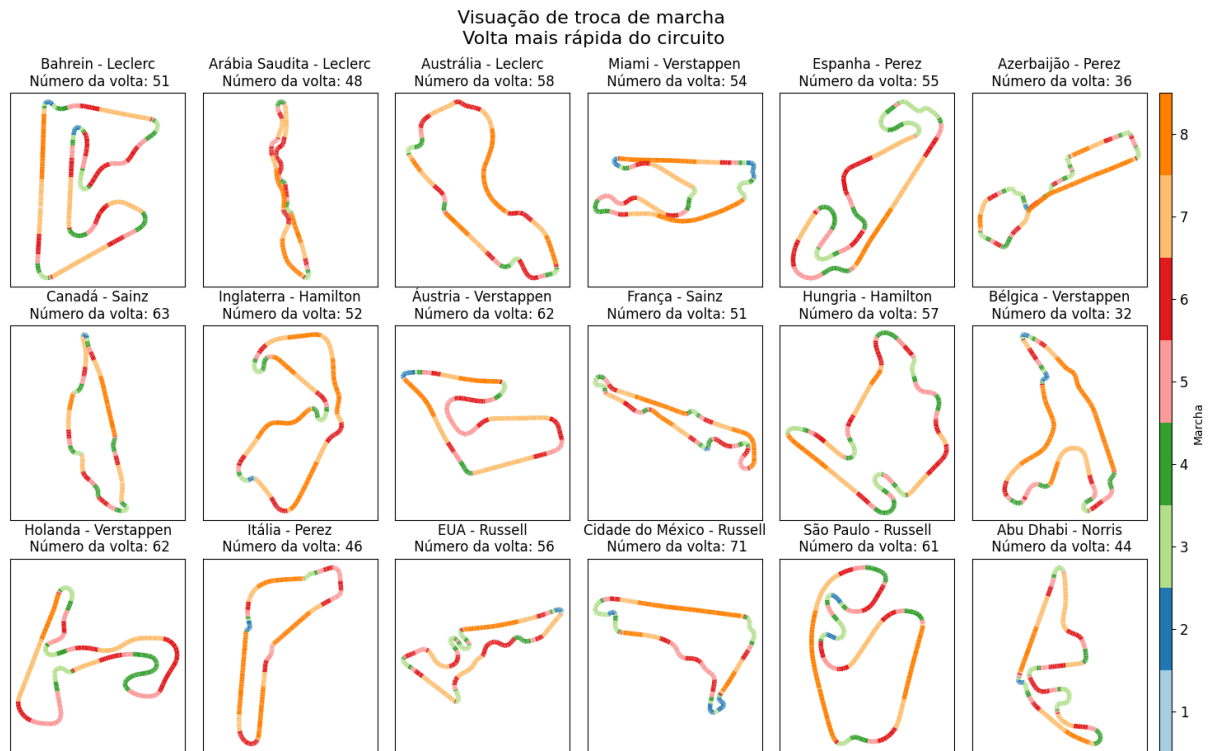


Figura 2 – Visualização das trocas de marcha nas voltas mais rápidas de cada Grande Prêmio

Vou fazer um gráfico com todos os circuitos, ainda falta ajustar algumas configurações.

1.4.4 Pilotos

1.4.5 Voltas

A figura 3 mostra os tempos de volta dos pilotos no GP da Holanda com diferentes pneus. As linhas tracejadas em vermelho delimitam os *stints*. Toda vez que um piloto faz uma parada na área da pista que se encontra sua equipe - mecânicos e engenheiros - geralmente para troca de pneus, inicia-se um novo *stint*. É possível notar alguns comportamentos:

Primeiro, os tempos tendem a serem menores conforme o número de voltas aumenta. Isso se dá pois a quantidade de combustível cai conforme é consumido pelo veículo, fazendo com que ele fique mais leve. Ainda, a cada *stint*, costuma ocorrer uma redução imediata no tempo devido a substituição de um pneu desgastado, e conforme o novo pneu é exigido, o tempo volta a subir.

Espera-se ainda que o desgaste ocorra mais rapidamente nos pneus mais macios, e mais lentamente nos pneus duros. No caso acima, o pneu mais duro é o C1, representado em vermelho, e o mais macio é o C3 representado em azul, concordando com a Tabela 5

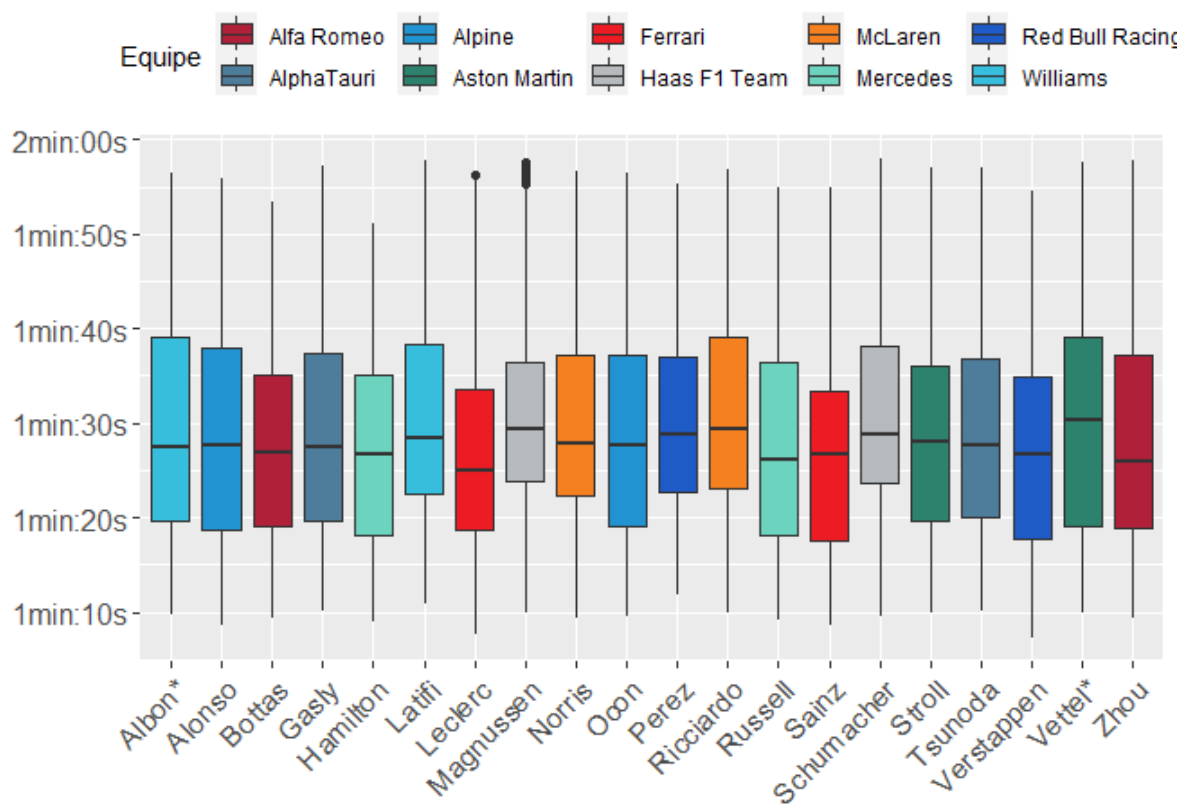


Figura 3 – Boxplot dos tempos de volta por piloto

apresentada anteriormente.

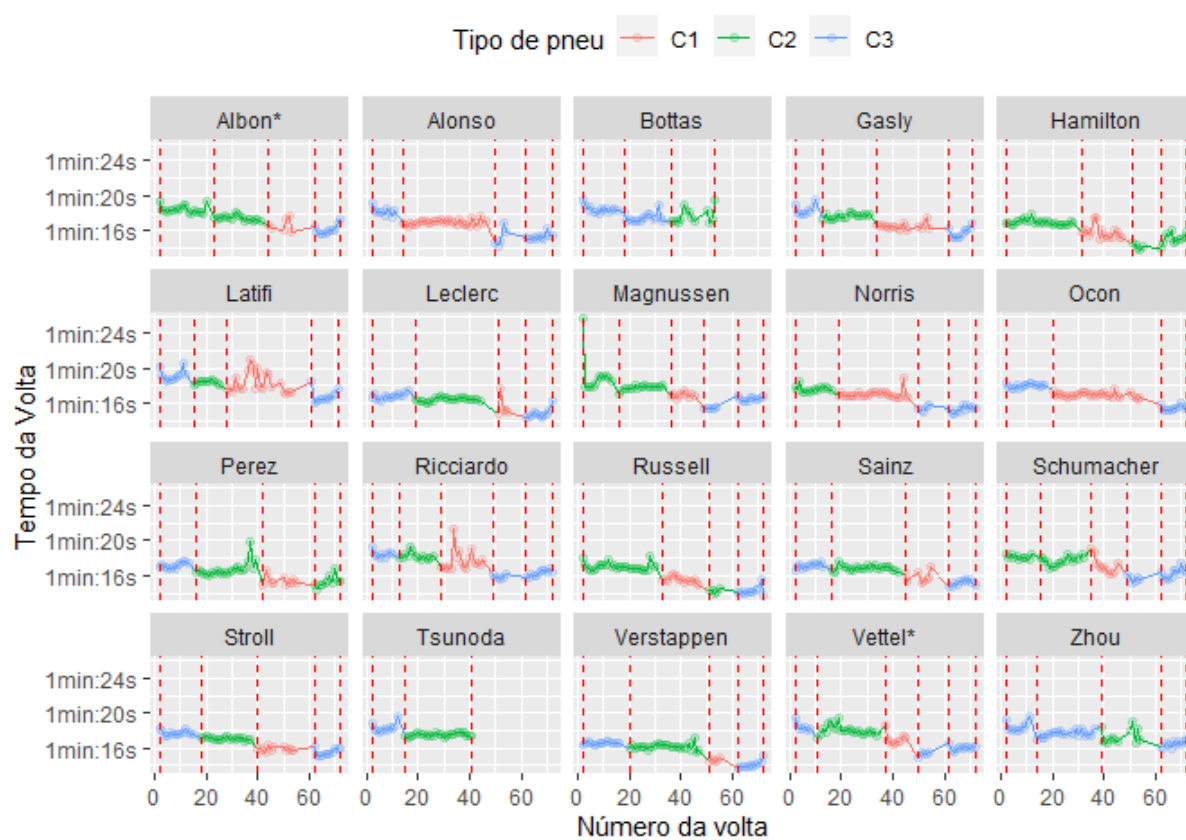


Figura 4 – Comportamento ao longo das voltas de cada piloto no Grande Prêmio da Holanda

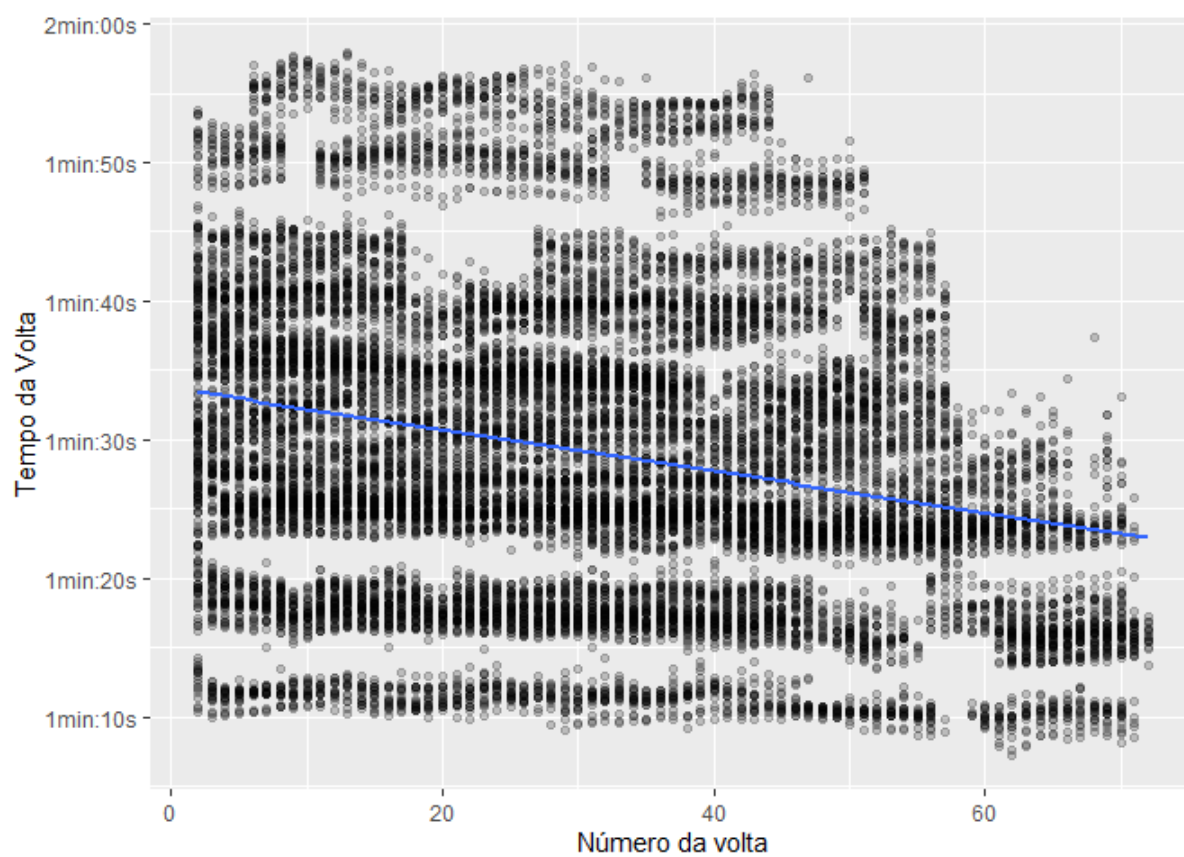


Figura 5 – Gráfico dos tempos de volta

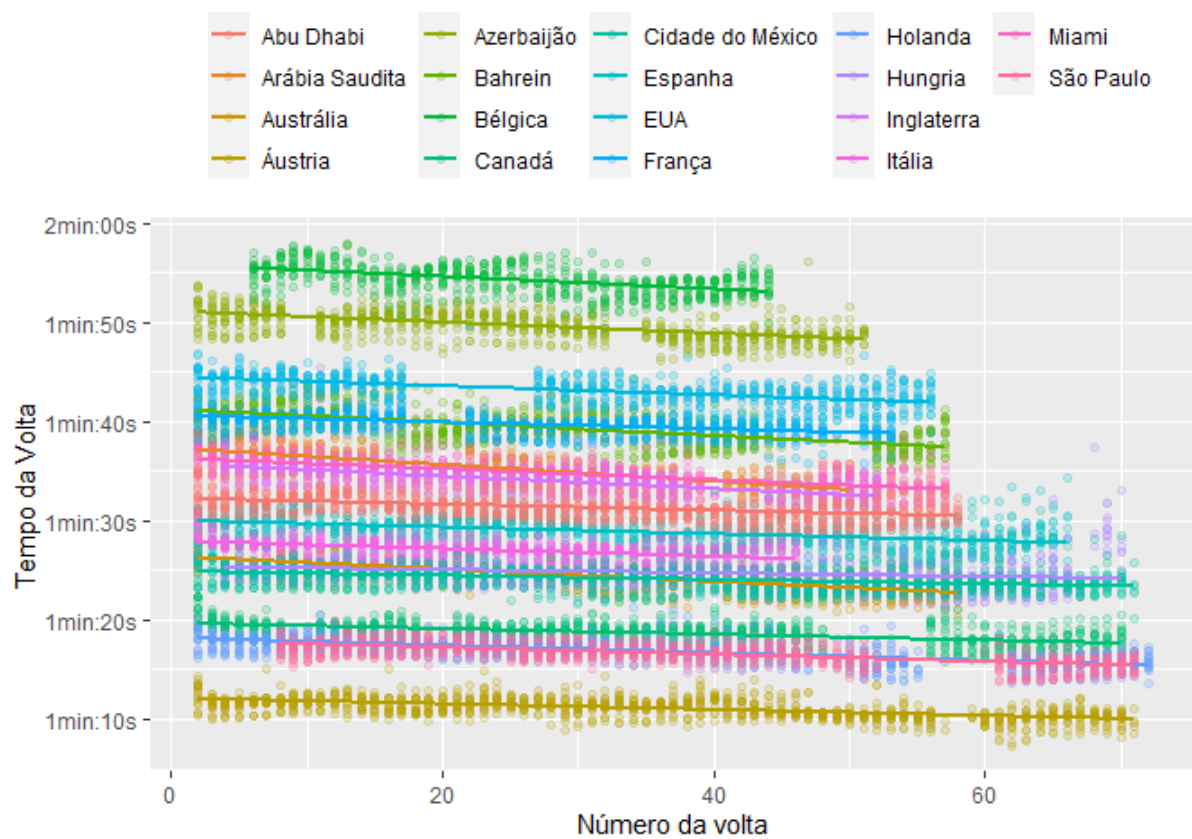


Figura 6 – Gráfico dos tempos de volta com cores específicas para cada circuito

Para que o comportamento da figura 3 ocorra é preciso, porém, especificar um evento, e individualizar os pilotos. A figura 4 mostra inicialmente como seria se desconsiderássemos estes fatores e explorasse diretamente a relação de números de voltas e tempo de volta. Os dados ficam muito dispersos, e embora nas últimas voltas ainda seja possível notar uma diminuição nos tempos esta relação é muito fraca. O gráfico ao lado, especificando por cores cada Grande Prêmios ilustra o impacto do evento na nossa variável de interesse, sendo possível notar claros agrupamentos nos dados. Já na figura 5, com os gráficos por evento representados individualmente nota-se que os tempo ficam muito mais concentrados.

1.4.6 Fatores Externos

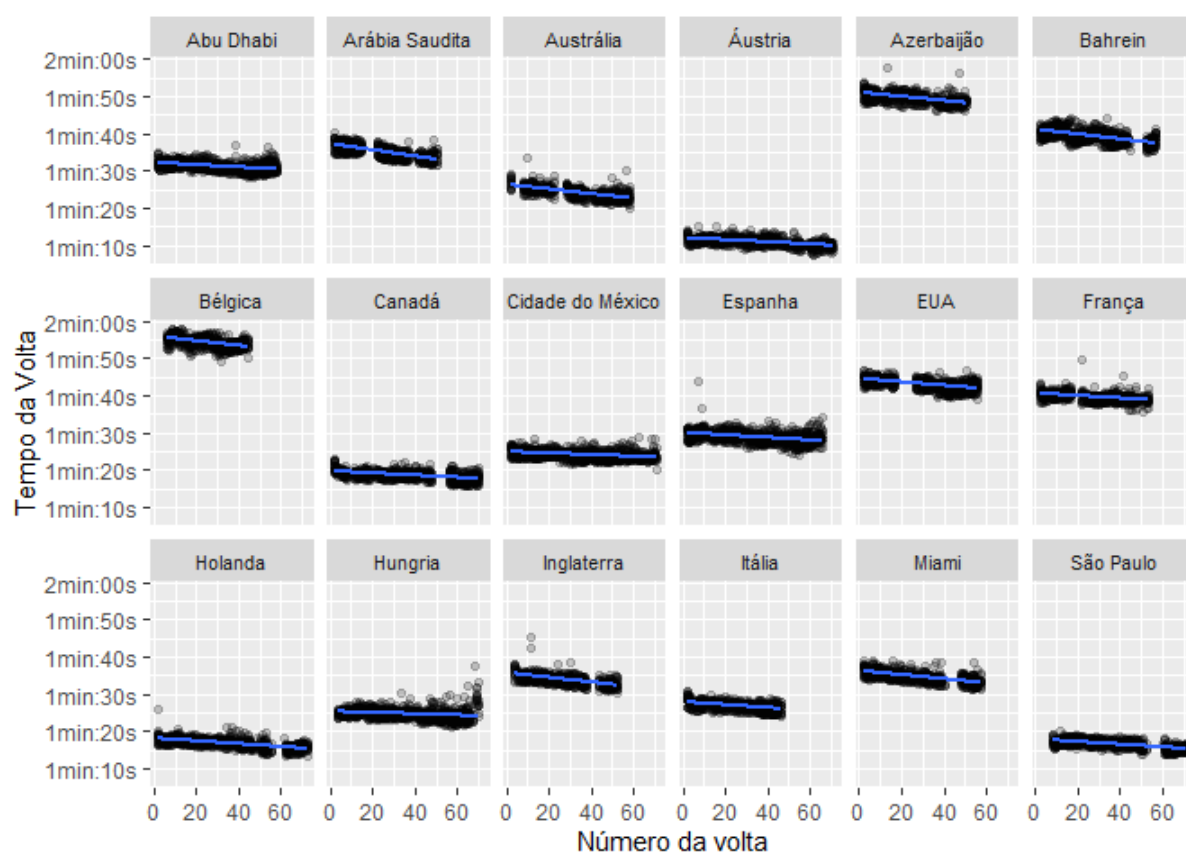


Figura 7 – Gráfico dos tempos de volta por Grande Prêmio

2 REVISÃO DE LITERATURA

Neste capítulo serão discutidos conceitos empregados neste trabalho. Dentre os tópicos abordados teremos: Análise de Regressão, Modelos Lineares, Modelos Lineares Mistos e Estatística Bayesiana.

2.1 Análise de Regressão

A Análise de Regressão é uma ferramenta estatística utilizada quando é razoável pensar que certas quantidades se relacionam, de maneira que seja possível prever um valor de interesse através das quantidades explicativas.

Exemplo 1 Supondo que um carro esteja na posição inicial S_0 , e desloca-se com aceleração constante por um período de tempo t , medido em segundos. Sejam S_f , v_0 e a : a posição após t segundos, a velocidade inicial no instante $t = 0$ e a aceleração, respectivamente.

Sabe-se que $S_f = S_0 + v_0 t + \frac{at^2}{2}$. Podemos escrever $S_f = f(S_0, v_0, t, a)$, ou seja, tratamos a variável S_f como variável dependente, e explicitamos uma relação com as variáveis independentes S_0, v_0, t e a . O exemplo acima é de uma relação determinística. A análise de regressão é utilizada em eventos estocásticos, quando há incerteza e não é possível precisar o valor que assumirá nossa quantidade de interesse, sendo essa sempre representada numericamente.

Exemplo 2 Consideremos a variável de interesse seja o tempo que um carro leva para, saindo do repouso, atingir a velocidade de 100 km/h .

Faz sentido pensar que o custo do motor do veículo influencie neste tempo. Outra variável que pode explicar este tempo é o peso do veículo. No entanto, não conseguimos uma fórmula direta para esta relação, como no exemplo anterior. Existem diversos outros fatores que influenciam na nossa variável de interesse de maneira que não conseguimos precisar, como a capacidade real do motor, uma vez que o preço é apenas um indicativo, o destreza do piloto, e a aderência dos pneus à pista são alguns desses.

Trata-se de um evento aleatório, por possuir influência indeterminadas de variáveis conhecidas e variáveis desconhecidas. Ainda, nossa variável de interesse, o tempo para atingir a velocidade de 100 km/h é numérica, e especificamente, contínua. Este é o tipo de caso tratado na análise de regressão. Por meio de métodos estatísticos, conforme exemplificado na seção a seguir, busca-se estimar os efeitos de covariáveis usando dados amostrados de eventos observados.

2.2 Modelos Lineares

2.2.1 Método de Mínimos Quadrados

2.3 Modelos Lineares Mistos

Modelos Lineares Mistos são extensões dos modelos lineares mais básicos, utilizada quando a estrutura dos dados analisados possui aninhamentos.

2.3.1 Dados aninhados

Estrutura de dados aninhadas são encontradas comumente em diversos estudos. Trata-se de dados nos quais cada observação pertence a um grupo, e acredita-se que este grupo possua um efeito importante na variável resposta. Trata-se de uma estrutura hierárquica.

2.3.2 Dados cruzados

Dados cruzados são referentes a estruturas nas quais dois ou mais grupos são agrupados de forma a conter todas as possíveis combinações. Dados cruzados não são uma estrutura hierárquica, em num modelo com cruzamento em um nível hierárquico ocorre o agrupamento desses fatores, sendo a combinação deles um nível hierárquico. Este cruzamento é utilizado quando assume-se que há independência nas unidades estudadas.

2.3.3 Efeitos mistos

Os modelos lineares mistos possuem este nome por possuírem parâmetros de efeitos fixos e parâmetros de efeitos aleatórios, sendo portanto caracterizados como modelos lineares de efeitos mistos.

Efeitos fixos e aleatórios dizem respeito à natureza das variáveis que estão sendo modeladas. O primeiro assume que o efeito da variável é fixo para todas as unidades estudadas, e são utilizados para inferir sobre uma variável de interesse.

Por outro lado, os efeitos aleatórios são característicos de unidades que permite-se variar aleatoriamente. Isso ocorre para explicar variações desconhecidas e não explicadas pelo modelo, quando acredita-se que os dados não sejam independentes.

Exemplo 3 Deseja-se estimar quantas voltas um piloto de Fórmula 1 irá completar antes de parar para troca de pneus.

Podemos considerar um modelo com um efeito fixo representado pela variável que representa o tempo de vida do pneu em uso. No entanto, sabe-se que cada pista possui

uma interação do asfalto com o pneu diferente, exigindo o pneu de formas diferentes e difíceis de prever.

Um efeito aleatório no intercepto para cada circuito pode ser interpretado como diferentes médias do tempo de duração para cada pista. Pode-se ainda assumir um efeito aleatório no coeficiente angular do tempo de vida do pneu em uso no modelo. Este efeito dá a entender que o tempo de vida do pneu impacta marginalmente de forma diferente na nossa variável de interesse.

Figuras exemplificativas, alterar com exemplo próprio.

1.png

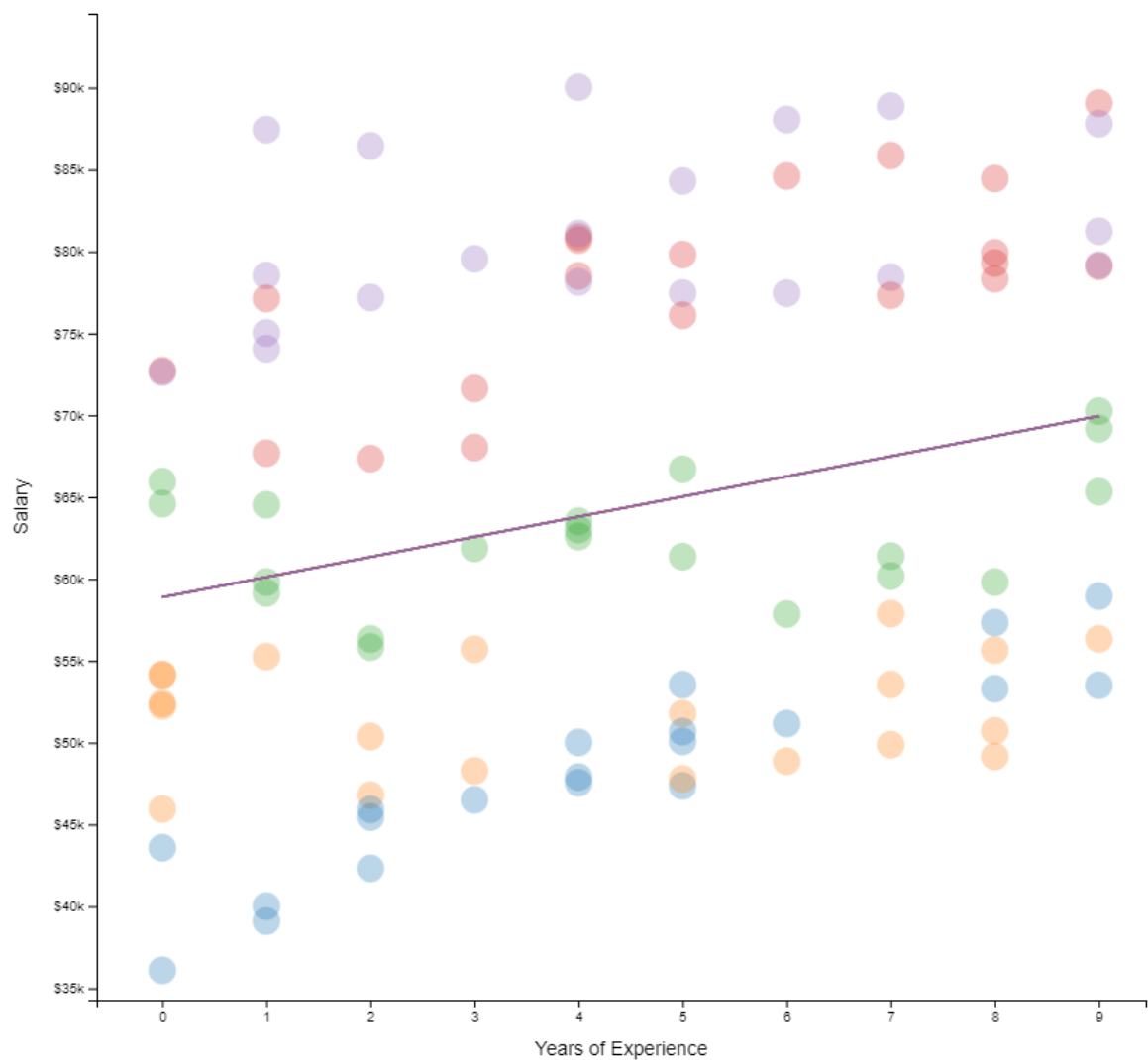


Figura 8 – Modelo com um efeito fixo

2.4 Estatística Bayesiana

[Gelman e Hill 2006]

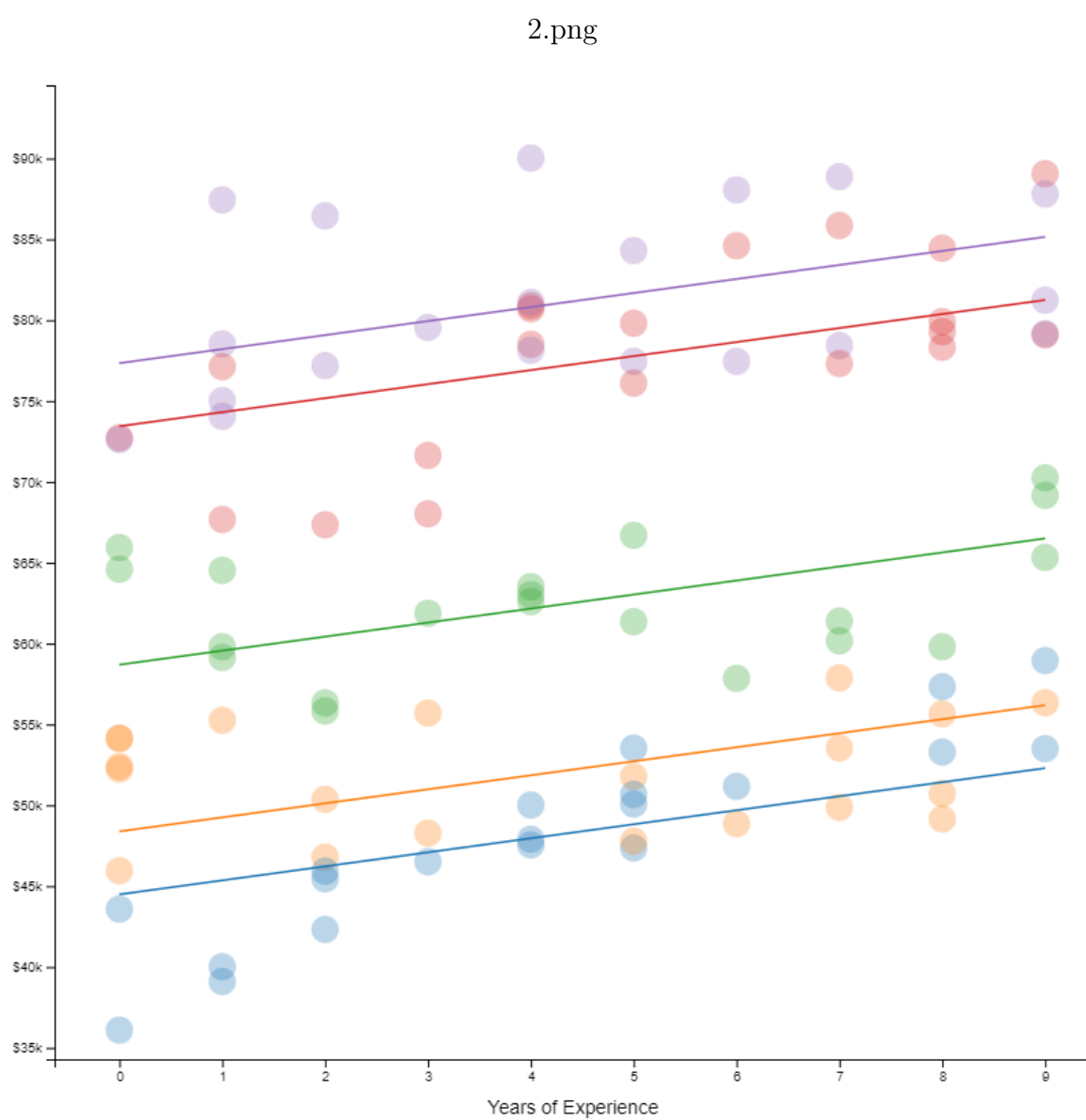


Figura 9 – Modelo com um efeito fixo e efeito aleatório no intercepto

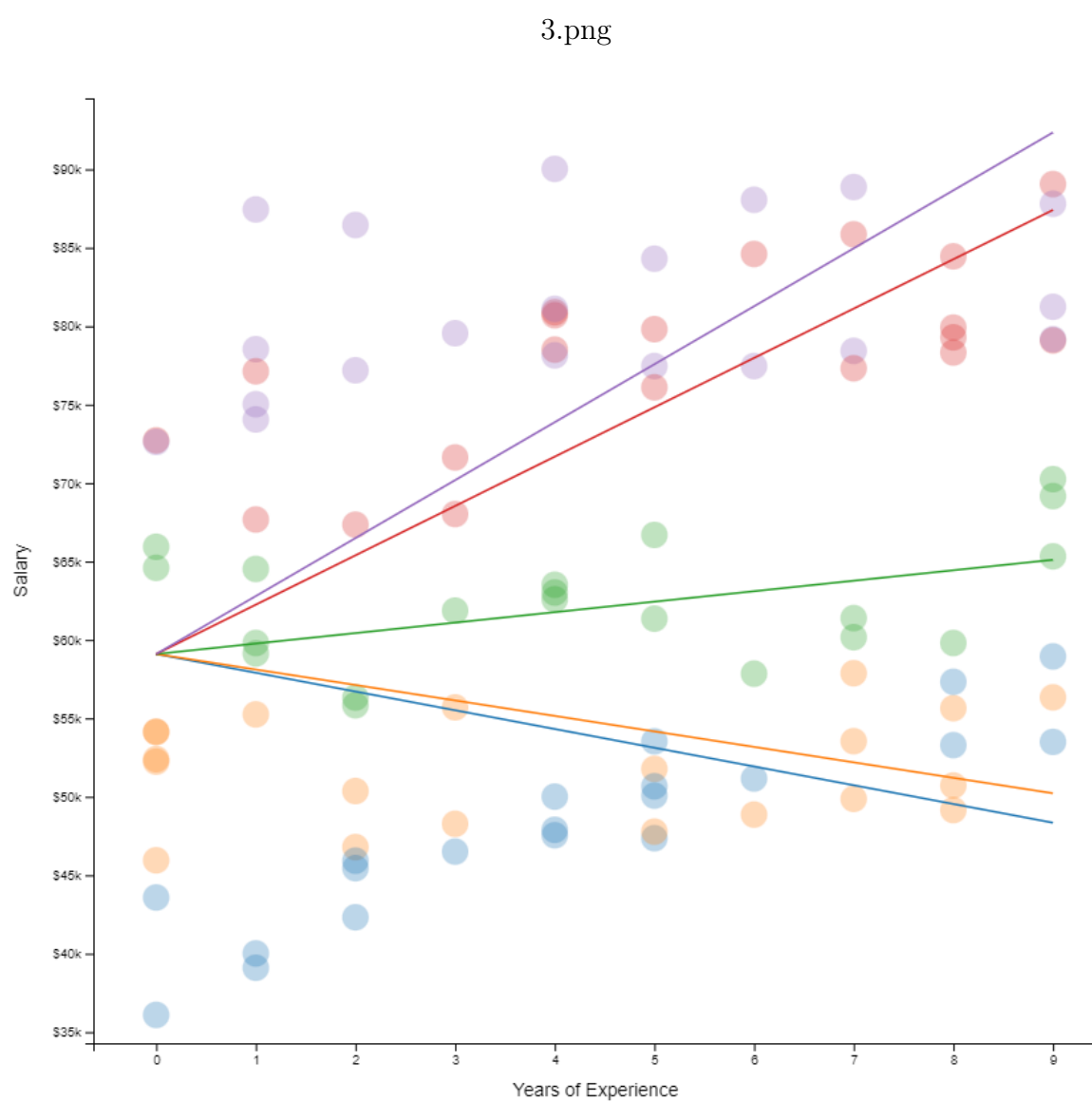


Figura 10 – Modelo com efeito aleatório no intercepto e no coeficiente angular da variável explicativa

3 MODELOS PROPOSTOS

Neste capítulo será discutida a metodologia para modelagem dos tempos de volta no Campeonato Mundial de Fórmula 1 de 2022.

3.1 Estrutura dos dados

As voltas de uma corrida de Fórmula 1 podem ser agrupadas por piloto, assim como as voltas de cada piloto podem ser agrupadas por circuito (Grande Prêmio).

De maneira análoga, as voltas de um conjunto de corridas pode ser agrupada por circuito, e as voltas em cada circuitos podem ser agrupados por piloto.

Acontece que apesar de existirem 2 grupos não há 3 níveis de hierarquia. Isso se daria apenas se os pilotos pudessem ser aninhados por circuitos ou vice-versa, o que não é o caso neste estudo.

Ao considerarmos um modelo linear misto levando em conta as influências dos pilotos e dos circuitos, temos um modelo com efeitos aleatórios cruzados, conforme mostrado abaixo.

Vale ressaltar que esta estrutura é inerante aos dados, e independe do modelo adotado.

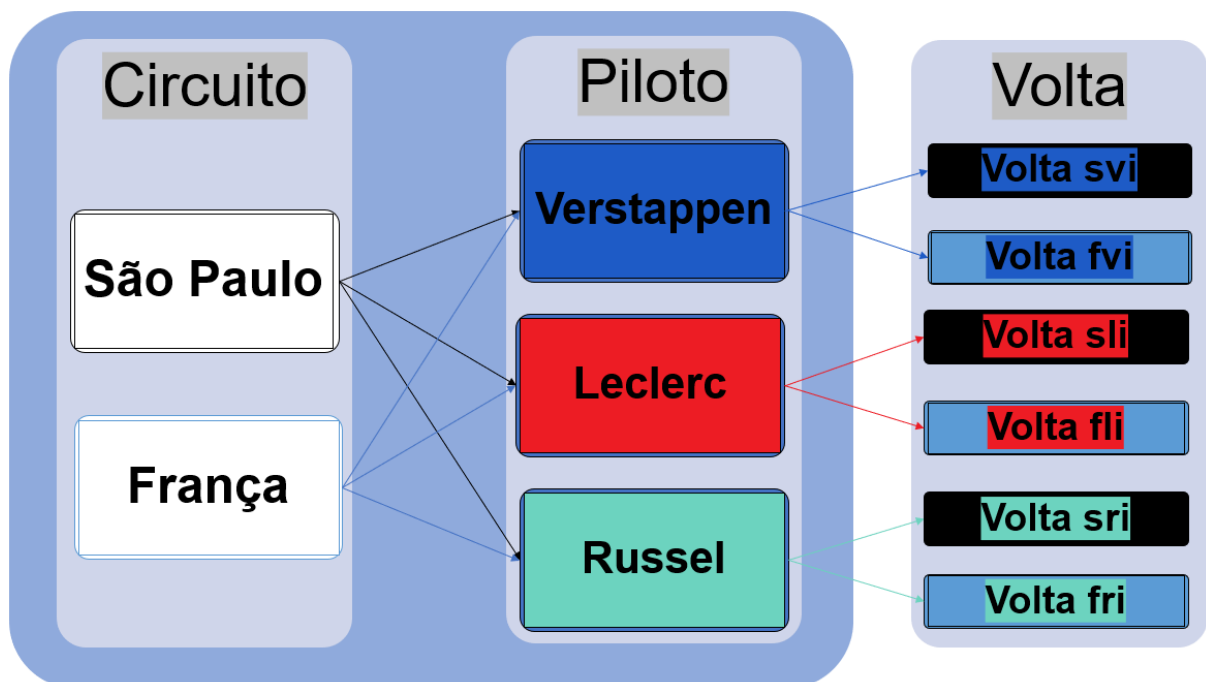


Figura 11 – Exemplo de voltas aninhadas por circuito e piloto de maneira cruzada

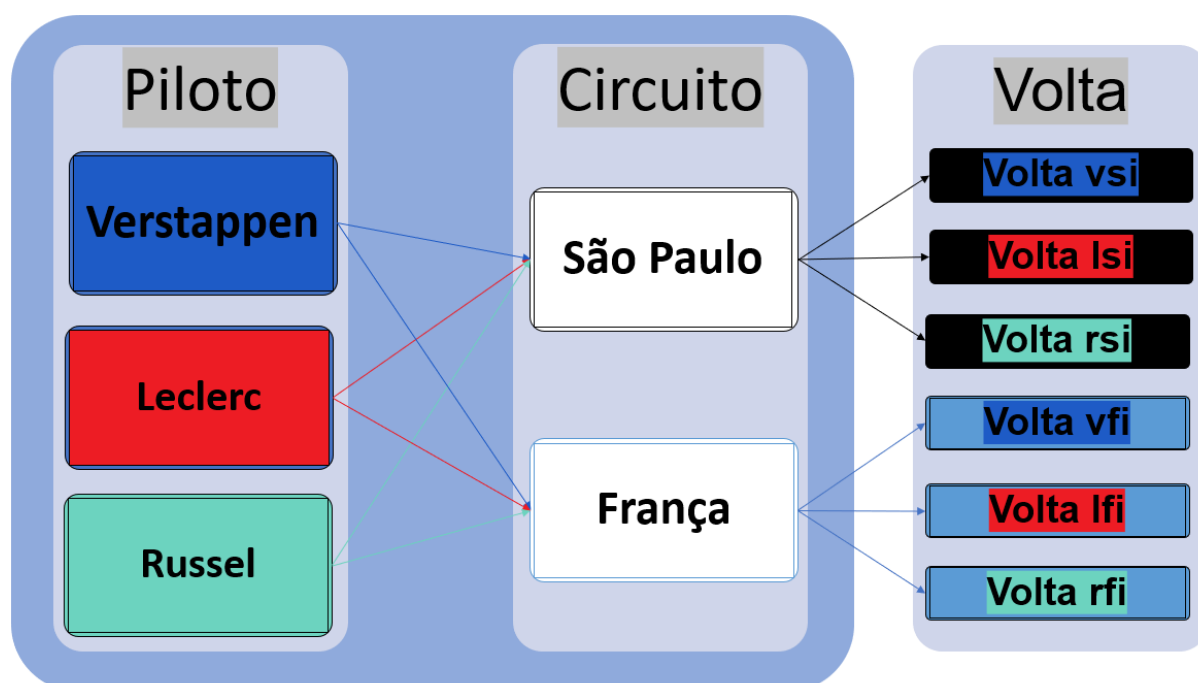


Figura 12 – Exemplo de voltas aninhadas por piloto e circuito de maneira cruzada

3.2 Modelos

Modelo 1 Assume-se que o tempo para completar a volta é dado em função de variáveis dependentes sem considerar efeitos específicos do circuito e do piloto. Trata-se, portanto, de um modelo linear múltiplo com variáveis preditivas apenas em um nível. A variável y_{ijc} representa a o tempo necessário para completar a i -ésima volta do piloto j no circuito c . As variáveis independentes consideradas foram (x_1) número da volta; (x_2) quantidade de voltas dadas pelo pneu em uso; (x_3) tipo de pneu em uso; e (x_4) posição no grid de largada. Desta forma, temos que $x_{1ijc} = i$, e, portanto:

$$y_{ijc} = \beta_0 + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad \epsilon_{ijc} \sim N(0, \sigma^2) \quad (3.1.1)$$

Neste modelo, temos que β_0 é o intercepto do modelo, $\beta_1, \beta_2, \beta_3$ e β_4 são os coeficientes angulares para as variáveis $(x_1), (x_2), (x_3)$ e (x_4) , respectivamente, dados pelo Método dos Mínimos Quadrados. Já ϵ_{ijc} representa o erro do modelo.

Modelo 2 Assume-se um modelo linear misto permitindo que o intercepto varie por circuito em função de suas características específicas conhecidas. A saber, considerou-se a variável (x_5) tamanho do circuito, medido em km .

$$y_{ijc} = \beta_{0c} + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad \epsilon_{ijc} \sim N(0, \sigma^2) \quad (3.2.1)$$

$$\beta_{0c} = \beta_0 + \beta_5 x_{5c}$$

$$y_{ijc} = (\beta_0 + \beta_5 x_{5c}) + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad (3.2.2)$$

Modelo 3 Assume-se um modelo linear misto assim como em **M2**, porém permite-se um efeito aleatório no nível do circuito. Este efeito explica as influências dadas por características do circuito não observadas. O modelo é dado por:

$$y_{ijc} = \beta_{0c} + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad \epsilon_{ijc} \sim N(0, \sigma^2) \quad (3.3.1)$$

$$\beta_{0c} = \beta_0 + \beta_5 x_{5c} + u_c \quad u_c \sim N(0, \sigma_u^2)$$

$$y_{ijc} = (\beta_0 + \beta_5 x_{5c} + u_c) + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad (3.3.2)$$

Modelo 4 Adiciona-se ao **Modelo 3** outro efeito aleatório, desta vez no nível do piloto. Este efeito explica as influências dadas por características não observadas do piloto.

$$y_{ijc} = \beta_{0jc} + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad \epsilon_{ijc} \sim N(0, \sigma^2) \quad (3.4.1)$$

$$\beta_{0jc} = \beta_{0c} + v_j \quad v_j \sim N(0, \sigma_v^2); \quad \beta_{0c} = \beta_0 + \beta_5 x_{5c} + u_c \quad u_c \sim N(0, \sigma_u^2)$$

$$y_{ijc} = (\beta_0 + \beta_5 x_{5c} + v_j + u_c) + \beta_1 i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad (3.4.2)$$

Modelo 5 Adiciona-se ao **Modelo 4** um efeito aleatório no coeficiente angular da variável (x_1) que representa o número de voltas.

Tal efeito foi escolhido pelo fato de cada circuito ter um número específico de voltas. Com isso, a preparação para as corridas varia, principalmente na quantidade de combustível em cada veículo. Por exemplo, estar na 44^a volta no Grande Prêmio da Bélgica significa estar terminando a corrida, e o carro encontra-se quase sem gasolina e muito mais leve. No GP do Brasil, com 71 voltas para completar a corrida, esta relação é diferente.

$$\begin{aligned} y_{ijc} &= \beta_{0c} + \beta_{1c}i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} & \epsilon_{ijc} &\sim N(0, \sigma^2) \\ \beta_{0jc} &= \beta_{0c} + v_j & v_j &\sim N(0, \sigma_v^2); & \beta_{0c} &= \beta_0 + \beta_5 x_{5c} + u_c & u_c &\sim N(0, \sigma_u^2) \\ \beta_{1c} &= \beta_1 + w_c & w_c &\sim N(0, \sigma_w^2) \end{aligned} \quad (3.5.1)$$

$$y_{ijc} = (\beta_0 + \beta_5 x_{5c} + u_c + v_j) + (\beta_1 + w_c)i + \beta_2 x_{2ijc} + \beta_3 x_{3ijc} + \beta_4 x_{4ijc} + \epsilon_{ijc} \quad (3.5.2)$$

Observação: Não foi testado modelo com covariáveis no nível do piloto uma vez que a competição de Fórmula 1 é muito dinâmica e decidida muitas vezes por detalhes tecnológicos. Ex-campeões como Alonso e Vettel, com muitos anos de experiência, no ano de 2022 não brigavam pelo título por estarem em equipes que não foram capazes de brigar pelo título.

Por questões de custo computacional (tempo de execução e convergência) definiu-se que os efeitos aleatórios acima serão não correlacionados.

3.3 Procedimento de Inferência

Inferência realizada sobre a metodologia bayesiana. CONTINUAR

Obtem-se a distribuição a posteriori, que não possui forma fechada. Utilizou-se então métodos de Monte Carlo via Cadeia de Markov para amostrar da distribuição a posteriori.

Utilizou-se o software JAGS [Plummer 2003] por meio de um pacote da linguagem de programação R [R Core Team 2022].

3.4 Critérios de Comparação

3.4.1 DIC

O Deviance Information Criterion (DIC) é uma medida de seleção de modelo que é amplamente utilizada em modelos Bayesianos. Ele foi proposto por Spiegelhalter et al [Spiegelhalter et al. 2002] e é definido como a diferença entre o deviance médio da amostra observada e o deviance médio esperado pelo modelo.

O deviance é uma medida de dissimilaridade entre o modelo e os dados observados, que é análoga à distância quadrática entre os dois. Ele é dado por:

$$D = -2 \log L$$

onde L é a verossimilhança do modelo, que é a probabilidade de observar uma determinada amostra da população, sendo no caso a amostra representada pelos dados e a população sob distribuição do modelo ajustado. Quanto maior o valor de L , melhor o ajuste aos dados observados.

O DIC é dado por:

$$DIC = D_{mean} + p_D$$

onde D_{mean} é o deviance médio da amostra observada e p_D é o valor esperado para o deviance médio sob o modelo. Assim, o DIC é a diferença entre o deviance médio da amostra observada e o deviance médio esperado pelo modelo. Quanto menor o valor do DIC, melhor o ajuste aos dados observados.

Além disso, o DIC também fornece um intervalo de confiança para a média do deviance, ajudando na avaliação da incerteza do modelo. Ele é muito utilizado sob o paradigma bayesiano, pois permite avaliar o modelo sem depender de um conjunto específico de parâmetros.

3.4.2 Erro Quadrático Médio

$$msLapTime_i \sim N(\mu, \sigma^2)$$

$$\mu = \alpha_{j[i],k[i]} + \beta_{1k[i]}(LapNumber) + \beta_2(TyreType_{C2}) + \beta_3(TyreType_{C3}) + \beta_4(TyreType_{C4}) + \beta_5(TyreType_{C5}) + \beta_6(TyreLife)$$

$$\alpha_j \sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(GridPosition), \sigma_{\alpha_j}^2\right), \text{ for Driver } j = 1, \dots, J$$

$$\begin{pmatrix} \alpha_k \\ \beta_{1k} \end{pmatrix} \sim N\left(\begin{pmatrix} \gamma_0^\alpha + \gamma_1^\alpha(CircuitLength) \\ \mu_{\beta_{1k}} \end{pmatrix}, \begin{pmatrix} \sigma_{\alpha_k}^2 & 0 \\ 0 & \sigma_{\beta_{1k}}^2 \end{pmatrix}\right), \text{ for EventName } k = 1, \dots, K$$

4 RESULTADOS

4.1 Comparação de Modelos

Tabela 6 – Tabela de pontuação do Campeonato de Fórmula 1 de 2022

Modelo	ν	D	p_D (Temporariamente AIC)	DIC	EPD	EQM
Modelo 1			347696	426772		
Modelo 2			328132	425146		
Modelo 3			275042	275266		
Modelo 4			272821	273774		
Modelo 5			271487	272547		

4.2 Análises dos resultados

Falar sobre as distribuições a posteriori.

5 CONSIDERAÇÕES FINAIS

REFERÊNCIAS

GELMAN, A.; HILL, J. *Data analysis using regression and multilevel/hierarchical models*. [S.l.]: Cambridge university press, 2006. Citado na página 18.

PIRELLI. *F1 tires: details and technical data*. <<http://web.archive.org/web/20221020080408/https://www.pirelli.com/tires/en-us/motorsport/f1/tires>>. Accessed: 2022-12-18. Citado na página 7.

PLUMMER, M. *JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling*. 2003. Citado na página 24.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: <<https://www.R-project.org/>>. Citado na página 24.

SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 64, n. 4, p. 583–639, 2002. Citado na página 25.

TREMAYNE, D. *The science of formula 1 design: expert analysis of the anatomy of the modern grand prix car*. [S.l.]: Haynes North America, 2006. Citado na página 1.

Apêndices

APÊNDICE A – APÊNDICE 1

Voltas removidas - motivo