



Universidade de Brasília
Departamento de Estatística

Análise de Desempenho no Surf

Tarsila Rodrigues Naves

Relatório final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

Brasília
2022

Tarsila Rodrigues Naves

Análise de Desempenho no Surf

Orientador: Leandro Tavares Correia

Relatório final apresentado para o Departamento de Estatística da Universidade de Brasília como parte dos requisitos necessários para obtenção do grau de Bacharel em Estatística.

**Brasília
2022**

Resumo

O Surf é um esporte que está sendo cada vez mais divulgado e praticado. No Brasil, isso se deve principalmente à vitória do Campeonato Mundial em 2014 pelo Gabriel Medina. Contudo, percebe-se ainda uma necessidade de estudos sobre esta modalidade, que sofre diversas influências como: fatores naturais, hábitos nutricionais, habilidades físicas e domínio técnico e teórico. O presente estudo contém informações de trinta e dois surfistas que fazem treinamento físico funcional específico para o esporte em três centros de treinamentos localizados em Maceió - AL e Marechal Deodoro - AL. Nessa amostra, também foram aplicados quatro testes físicos que abrangem as principais aptidões relacionadas ao Surf. O objetivo desse trabalho é verificar quais variáveis impactam na classificação do surfista como iniciante, intermediário, avançado ou profissional, e obter um modelo que pudesse prever o nível de desempenho do surfista baseado em suas características. Para isso, foi utilizado o método estatístico conhecido como Regressão Logística binária e ordinal, em que a variável resposta é categórica. Ao longo das análises, foi verificado que o tempo e a frequência de Surf; idade do surfista e idade de ingresso no esporte; classificação das manobras: Drop, correr a parede da onda, Cutback, Floater, 360 e Aéreo; e a avaliação de equilíbrio e impulsão horizontal, tiveram um grande impacto na classificação do desempenho do surfista. Também pode-se observar que a distribuição das características do surfista são distintas em relação ao Sexo. Dessa maneira, tornou-se interessante aplicar o método logístico binário não somente para a amostra completa, mas também para subamostras divididas de acordo com o Sexo. Por fim, foi feita a modelagem via regressão logística ordinal, que também apresentou resultados significativos para o estudo.

Palavras-chaves: Surf. Fatores influentes. Regressão Logística. Regressão Ordinal. Modelagem. Predição.

Lista de Tabelas

| | | |
|----|--|----|
| 1 | Medidas resumo da idade do surfista por sexo. | 36 |
| 2 | Medidas resumo da idade inicial do surfista por sexo. | 37 |
| 3 | Medidas resumo do tempo de Surf por nível de desempenho. | 38 |
| 4 | Distribuição da frequência relativa da classificação das manobras Drop e Parede por sexo. | 41 |
| 5 | Distribuição da frequência relativa da classificação das manobras Batida e Rasgada por sexo. | 43 |
| 6 | Distribuição da frequência relativa da classificação das manobras Cutback e Floater por sexo. | 44 |
| 7 | Distribuição da frequência relativa da classificação das manobras 360 e Aéreo por sexo. | 45 |
| 8 | Distribuição de frequência da classificação da manobra tubo por sexo. . . . | 46 |
| 9 | Distribuição conjunta do nível de desempenho por sexo. | 53 |
| 10 | Distribuição da frequência relativa do domínio de execução de manobras complexas por nível de desempenho e valor p do teste de associação. | 53 |
| 11 | Distribuição conjunta da classificação de manobras por nível de desempenho e valor p do teste de independência Qui quadrado. | 54 |
| 12 | Distribuição do domínio teórico por nível de desempenho e valor p do teste de independência Qui quadrado. | 55 |
| 13 | Distribuição conjunta do domínio de execução de manobras complexas por sexo. | 55 |
| 14 | Distribuição conjunta da classificação de manobras por nível de desempenho e valores p do teste de associação Qui quadrado e de Fisher. | 56 |
| 15 | Distribuição conjunta do domínio teórico por sexo e valores p do teste de associação Qui quadrado e de Fisher. | 57 |
| 16 | Medida de desempenho dos modelos binários para a amostra de homens. . . | 59 |
| 17 | Estimativas dos parâmetros do Modelo 1. | 60 |
| 18 | Testes de qualidade do ajuste - Modelo 1 versus Modelo completo. | 61 |

| | | |
|----|---|----|
| 19 | Matriz de confusão para o Modelo 1. | 61 |
| 20 | Medida de desempenho dos modelos binários para a amostra de mulheres. . | 62 |
| 21 | Estimativa do parâmetro dos Modelo 5, 6 e 7. | 63 |
| 22 | Testes de qualidade do ajuste - Modelo 5, 6 e 7 versus Modelo completo. . | 63 |
| 23 | Matriz de confusão para os Modelos 5, 6 e 7. | 64 |
| 24 | Medida de desempenho dos modelos binários para a amostra completa. . . | 65 |
| 25 | Estimativas dos parâmetros do Modelo 11. | 65 |
| 26 | Testes de qualidade do ajuste - Modelo 11 versus Modelo completo. | 66 |
| 27 | Matriz de confusão para o Modelos 11. | 67 |
| 28 | Medidas de desempenho dos modelos ordinais. | 68 |
| 29 | Estimativas dos parâmetros do Modelo 15. | 68 |
| 30 | Matriz de confusão para o Modelos 15. | 70 |

Lista de Figuras

| | | |
|----|--|----|
| 1 | Ilustração do modelo logístico. | 18 |
| 2 | Ilustração de gráfico dos resíduos. | 27 |
| 3 | Ilustração de um problema de separação no modelo logístico. | 28 |
| 4 | Distribuição do nível de desempenho por sexo do surfista. | 35 |
| 5 | Distribuição da idade e idade inicial por sexo e nível de desempenho. | 36 |
| 6 | Distribuição do Tempo de Surf | 37 |
| 7 | Aspectos nutricionais por sexo e nível de desempenho. | 39 |
| 8 | Distribuição da frequência semanal e duração média da sessão por sexo e nível de desempenho. | 40 |
| 9 | Distribuição da prática de outros esportes. | 40 |
| 10 | Distribuição da classificação do Drop e correr a parede da onda por nível de desempenho. | 41 |
| 11 | Distribuição do domínio e tempo dispendido para execução de manobras complexas. | 42 |
| 12 | Distribuição da classificação da Batida e Rasgada por nível de desempenho. | 43 |
| 13 | Distribuição da classificação da Cutback e Floater por nível de desempenho. | 44 |
| 14 | Distribuição da classificação da 360 ^o e Aéreo por nível de desempenho. | 45 |
| 15 | Distribuição da classificação do tubo por nível de desempenho. | 46 |
| 16 | Domínio teórico por sexo e nível de desempenho. | 47 |
| 17 | Resultado dos testes físicos. | 48 |
| 18 | Dispersão entre duas variáveis correlacionadas. | 51 |
| 19 | Gráfico quantil-quantil dos resíduos do Modelo 1. | 61 |
| 20 | Gráfico quantil-quantil dos resíduos dos Modelos 5, 6 e 7. | 63 |
| 21 | Gráfico quantil-quantil dos resíduos do Modelo 11. | 66 |

Sumário

| | |
|--|----|
| 1 Introdução | 9 |
| 2 Revisão Teórica | 15 |
| 2.1 Distribuição de Probabilidade | 15 |
| 2.2 Teste de Hipóteses | 16 |
| 2.3 Regressão Logística | 18 |
| 3 Metodologia | 31 |
| 3.1 Materiais | 31 |
| 3.2 Métodos | 34 |
| 4 Resultados | 35 |
| 4.1 Análise Descritiva | 35 |
| 4.2 Análise de Correlação | 50 |
| 4.2.1 Teste de correlação de Pearson | 50 |
| 4.2.2 Teste de Independência | 52 |
| 4.3 Regressão Logística | 58 |
| 4.3.1 Modelos logísticos binários | 59 |
| 4.3.2 Modelos logísticos ordinais | 67 |
| 5 Conclusão | 71 |
| Referências | 73 |
| Apêndice | 75 |
| A Questionário | 75 |

1 Introdução

De acordo com Farias (2000), o Surf é mundialmente identificado como correr uma onda estando apoiado ou em pé sobre uma prancha. Já Butt e Russel (2005) apud Guimarães (2011), definem o esporte como um ato ilusoriamente simples de apanhar uma onda do oceano em cima de uma prancha e que na realidade, surfar é uma conjunção complexa de forças.

Muito ainda se discute sobre a origem do Surf, segundo Finney e Houston (1995) apud ÁRIAS (2002), os pioneiros a deslizarem sob uma onda foram pescadores nativos ao retornarem do alto mar em seus “caballito” de totora, feitos de fibra de junco. No entanto, há quem defenda que o esporte nasceu no Havaí, no século XI, e que a cultura e a religião do povo havaiano era voltada no Surf.

No Brasil, a prática do esporte iniciou na cidade de Santos - SP, onde o primeiro surfista foi o norte-americano Thomas Ernest Rittcher Jr., sendo sua irmã considerada a primeira mulher a surfar. Contudo, foi somente na cidade do Rio de Janeiro - RJ, que o Surf se consolidou e criou fama, assim como a primeira associação e os primeiros campeonatos brasileiros.

O Surf é caracterizado por ser um esporte radical, e quando não praticado em piscinas de ondas artificiais, é influenciado pelas condições do mar que afetam no desempenho e na intensidade da atividade. O mar está em constante mudança, sofre interferência de diversos fatores tais como a ação dos ventos, correntes marítimas, tipo de fundo oceânico e a ação gravitacional da lua sobre as marés (BRASIL et al., 2001). Essa variabilidade trazem ainda mais dificuldades para os praticantes.

O esporte exige habilidades e métodos específicos para a sua prática, que são utilizados ao sentar na prancha, passar por debaixo da onda (furar a onda), remar, subir e ficar de pé em cima da prancha (chamado de “drop”) e executar manobras mais complexas.

O processo de surfar pode ser dividido em três momentos: o primeiro se dá quando o surfista está entrando no mar, deitado sobre a prancha remando para chegar até o “outside” (região de formação de ondas mais distantes da beira da praia). Ao chegar no outside, o surfista aguarda e seleciona a onda, o segundo momento é dado a partir da tentativa de entrada na onda, onde o surfista desempenha fortes remadas com precisão, com o objetivo de que a prancha adquira velocidade suficiente para estar sobre a onda. Para pegar a onda corretamente a remada é o fator número um (SOUZA, 2004). O corpo deve estar equilibrado no centro da prancha, e a remada precisa ser contínua, evitando

balançar muito o corpo. Ao entrar na onda, o terceiro passo é a execução de manobras, sendo a primeira, o movimento de ficar de pé sobre a prancha, denominado “drop”, que para ser realizado, o surfista deve apoiar as mãos sobre a prancha na altura do peito e fazer uma flexão para se levantar, mantendo o peso centralizado, para não cair. Depois, o surfista pode realizar uma série de manobras, pontuadas por Aranã (2007): cavada, batida, rasgada, slash, cutback, floater, tubo, aéreo, aéreo 180º, tail slide, grab rail, 360º.

Segundo Fleishman e Zaccaro (1992), as capacidades motoras estão associadas ao desempenho motor dos indivíduos, possibilitando maior ou menor desempenho ao executar uma determinada tarefa motora, e são divididas em dois grandes grupos: perceptivo-motoras e proficiência física.

O Surf está associado principalmente a cinco capacidades perceptivas-motoras, que foram definidas através de testes e análises realizadas por Fleishman e Ellison (1962), sendo elas: a coordenação multimembros, caracterizada como a habilidade de coordenar o movimento de vários membros simultaneamente; a precisão de controle, sendo a habilidade de realizar ajustamentos de movimentos altamente controlados; a orientação da resposta, definida por tomar uma escolha rápida dentre numerosas alternativas de movimento; o tempo de reação, dada pela capacidade de reagir rapidamente, e o controle de velocidade, sendo a habilidade de produzir ajustamentos de movimentos antecipatórios e contínuos em resposta a mudanças na velocidade de um alvo ou de um objeto que se move continuamente.

A aptidão física, está relacionada à capacidade funcional do indivíduo realizar alguns tipos de atividades que exijam um determinado empenho muscular. Para Mendez-Villanueva e Bishop (2005), as proficiências físicas que estão mais relacionadas ao esporte são: agilidade, equilíbrio, força muscular e flexibilidade. Os mesmos autores afirmam também que o movimento rápido em resposta a estímulo externo é um importante determinante para as habilidades do surfista.

Farias (2000), analisa o Surf como um esporte de equilíbrio instável, uma vez que o surfista, apoiado sobre uma prancha, desliza sob a onda, uma superfície instável e móvel. Para Dambroz (1988) apud Zeni (2002), o equilíbrio é tido como a habilidade de distribuir o peso do corpo controlando o centro de gravidade.

A força muscular representa a força máxima de esforço que uma pessoa pode reproduzir contra uma resistência, já a resistência muscular, de acordo com Baechele (1994) apud Netto (2005), é definida pela capacidade máxima dos músculos de repetirem um movimento várias vezes. Para Junior (2010), o treinamento específico para a modalidade

juntamente com o fortalecimento da musculatura está associada a um melhor desempenho no esporte e a diminuição do risco de lesões.

Nas performances realizadas no Surf, muitas vezes não há tempo suficiente para atingir a força muscular máxima. Nesses casos, utiliza-se a força explosiva, representada pelo produto da força muscular pela velocidade. Segundo Silva e Clemente (2019), o treinamento de potência para membros inferiores é peça essencial para o desempenho no esporte.

Dantas (1995) define a flexibilidade, como qualidade física responsável pela execução voluntária de um movimento de amplitude angular máxima, por uma articulação ou conjunto de articulações, dentro dos limites morfológicos, sem o risco de provocar lesão. O mesmo autor considera a flexibilidade um dos fatores primordiais no aperfeiçoamento motor e no desenvolvimento da consciência corporal.

Nas últimas décadas, o Surf vem se popularizando e se tornando um dos esportes mais praticados no Brasil e no mundo. De acordo com um documento submetido ao Comitê Olímpico Internacional (COI) em 2012 pela ISA (International Surfing Association), indicou que existiam aproximadamente 35 milhões de surfistas em todo o mundo, cerca de 0,5% da população mundial no mesmo ano. O Brasil se destaca no esporte sendo uma das três maiores potências em conjunto com os EUA e a Austrália (WSL, 2015).

No cenário brasileiro, o esporte passou a ser mais notado após a conquista do Campeonato Mundial de 2014, pelo surfista Gabriel Medina, dando grande visibilidade também a geração de surfistas profissionais brasileiros de grande destaque, que foram nomeados como a tempestade brasileira ou “brazilian storm” pelo ex-presidente da Abrasp (Associação Brasileira de Surf Profissional), Marcelo Andrade. Desta maneira, o Surf brasileiro passou a ser mais difundido, ganhando espaço significativo em tv aberta como aponta estudo (FERREIRA, 2018) que compara a incidência de matérias sobre o surf no programa Esporte Espetacular da Rede Globo de Televisão antes e depois da conquista do Medina.

Além disso, o Surf vem expandindo suas fronteiras entre os mais variados públicos, ganhando admiradores e praticantes de diferentes gêneros, idades, condições sociais e aspectos culturais. A título de exemplo, é observado pela participação feminina no esporte que antes era praticado em sua maioria pelo público masculino, e apesar do grande preconceito que ainda é visto nos dias atuais, o cenário feminino no Surf cresce bastante, incluindo cada vez mais as mulheres dentro das suas etapas do circuito mundial.

O esporte teve a sua estreia como esporte olímpico na Olimpíada de Tóquio de 2020, depois de diversas tentativas da ISA (Associação Internacional do Surfe) de integrar as atividades olímpicas. O Brasil participou desta estreia com a representação de quatro grandes surfistas brasileiros, sendo eles: Gabriel Medina, Ítalo Ferreira, Tatiana Weston-Webb e Silvana Lima. Com o resultado positivo da competição para o Brasil, a vitória de primeiro lugar do Ítalo Ferreira, o Surf se tornou ainda mais divulgado no país.

Apesar do esporte estar passando por uma progressiva popularização nos últimos anos, ainda existem poucos estudos sobre o assunto. Uma pesquisa realizada em 2016 teve como objetivo mapear a produção de artigos sobre surf que foram publicados em revistas científicas brasileiras entre o período de 2000 a 2015, e constatou que apenas 27 artigos foram publicados sobre o tema (GOMES et al., 2017), dos quais 55,55% são relacionados à temas como: treinamento, saúde e aspectos nutricionais. Neste mesmo estudo, os autores enfatizam a necessidade de mais pesquisas que possam efetivamente contribuir com o trabalho daqueles que "fazem a modalidade cotidianamente acontecer".

O Surf vai muito além da arte de deslizar sob as ondas, é o ato de se conectar com o mar e a natureza, ter consciência de cada movimento que é realizado. Também está altamente relacionado ao condicionamento físico do surfista, memória muscular, conhecimento técnico e teórico das manobras, assim como o domínio de leitura do mar e posicionamento na onda. Desta maneira, nota-se que o esporte possui uma complexidade ímpar.

O presente estudo busca agregar valor à literatura científica relacionada ao Surf, de tal maneira que os profissionais da área da Educação Física, professores e praticantes do esporte possam adquirir um maior conhecimento sobre o assunto. Para isso, tem como objetivo ampliar a visão sobre o esporte em suas especificidades, através da identificação e análise das principais características que compõe e influenciam no seu desempenho.

A investigação dos componentes que interferem no desenvolvimento no Surf, denominados como variáveis respostas, será realizada através da aplicação de um questionário, que contará com informações de hábitos alimentares, tempo de prática, frequência semanal, aspectos técnicos e teóricos inerentes ao esporte. Além de medidas de avaliação da aptidão física dos surfistas, com a aplicação de testes físicos propostos por Pollock e Wilmore (1993) e pelo Conselho da Europa (1988). O protocolo de testes EUROFIT é empregado em forma de circuito com testes físicos, motores e um teste de resistência cardiorrespiratória, além das medidas antropométricas, cuja eficiência foi comprovada na Europa (MARTINS, 1998).

Baseado nos dados obtidos será realizada análise descritiva dos dados, assim como avaliar a existência de associação entre as variáveis. Também utilizadas técnicas de regressão para classificar em níveis, o desempenho do surfista, nomeado variável resposta.

A Regressão busca explicar a relação, através de um modelo, entre uma variável dependente ou resposta e um conjunto de variáveis independentes ou explicativas. Uma vez que a variável de interesse do estudo, a performance do surfista, é categórica, a Regressão Logística é mais adequada nesta situação. A distribuição logística, na visão matemática, é extremamente flexível e fácil de ser manuseada, além de ter uma interpretação clinicamente significativa (HOSMER; LEMESHOW, 2000a). Esta ferramenta de modelagem irá atribuir uma probabilidade do surfista pertencer a cada nível de classificação de acordo com os fatores que podem influenciar no seu desempenho esportivo.

Dessa maneira, esse estudo tem como objetivo geral analisar os aspectos técnicos e teóricos, aptidões físicas e condições nutricionais do surfista que influenciam no seu desempenho esportivo. Para isso, os objetivos específicos desse trabalho são: realizar análises descritivas desses componentes e aplicar de técnicas de Regressão Logística (Binomial e Multinomial) para construção de um modelo para predição da classificação de desempenho no Surf.

2 Revisão Teórica

2.1 Distribuição de Probabilidade

Distribuição Bernoulli

A distribuição Bernoulli é uma distribuição de probabilidade discreta de espaço amostral $(0,1)$, cujo o interesse é verificar se o resultado apresenta uma determinada característica, onde 0 representa "fracasso" e 1 "sucesso".

Seja X uma variável aleatória com distribuição bernoulli, então

$$X = \begin{cases} 1 & \text{com probabilidade } p \\ 0, & \text{com probabilidade } 1 - p. \end{cases}$$

A função de probabilidade pode ser expressa como:

$$f(x; p) = p^x(1 - p)^{1-x}, x \in (0, 1). \quad (2.1.1)$$

Distribuição Binomial

A distribuição Binomial consiste em n ensaios independentes da distribuição Bernoulli, para qual a probabilidade de sucesso de cada ensaio é p (MORETTIN; BUSSAB, 2010).

Seja X uma variável aleatória, correspondente ao número de sucessos de um experimento binomial, então sua função de probabilidade é definida por

$$f(x; n; p) = \binom{n}{x} p^x (1 - p)^{n-x}. \quad (2.1.2)$$

Distribuição Multinomial

A distribuição multinomial é uma generalização da binomial, consiste em \mathbf{n} ensaios independentes com mais de dois possíveis resultados, com probabilidade constante de ocorrer um determinado resultado.

A sua função de probabilidade é dada por

$$P(X_1 = n_1, \dots, X_t = n_t; n; p_1, \dots, p_k) = \frac{n!}{n_1! \dots n_t!} p_1^{n_1} \dots p_t^{n_t}. \quad (2.1.3)$$

Distribuição Normal

A distribuição normal (gaussiana) é uma distribuição contínua parametrizada pela média (número real μ) e o desvio-padrão (número positivo σ).

Seja X uma variável aleatória com distribuição normal então a sua função de densidade é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (2.1.4)$$

2.2 Teste de Hipóteses

De acordo com Morettin e Bussab (2010), o teste de hipótese é utilizado para testar se uma determinada afirmação sobre uma população, a partir de resultados experimentais de uma amostra contrariam ou não tal afirmação. A hipótese denominada nula (H_0) é a hipótese a ser testada contra a hipótese alternativa (H_1).

O erro do tipo I, cuja probabilidade é denotada por α , que ocorre quando a hipótese nula é rejeitada sendo verdadeira e o erro do tipo II, cuja probabilidade é denotada por β , que é quando a hipótese nula não é rejeitada sendo falsa.

Teste de independência de Pearson

O Teste de independência para o coeficiente de correlação, é indicado para análise de correlação entre duas variáveis quantitativas, realizado em cinco etapas:

1) Formulação das hipóteses:

$$H_0 : \rho = 0 \text{ vs } H_1 : \rho \neq 0.$$

2) Estatística do Teste:

Para amostras retiradas de uma população para a qual $\rho = 0$, pode-se provar que a estatística:

$$T = r \sqrt{\frac{n-2}{1-r^2}},$$

tem distribuição t-student com $(n-2)$ graus de liberdade, onde r é dado pelo coeficiente de correlação de pearson entre a variável x e y

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

3) Determinar o nível de significância (α) do teste, a região crítica do teste bilateral será

$$RC = (-\infty, -x_c] \cup [+x_c, +\infty),$$

em que x_c é o ponto crítico baseado no quantil da distribuição t-student.

4) Resultado da amostra:

Consiste na substituição do coeficiente r calculado na fórmula do passo 2, da estatística do teste.

5) Conclusão:

Se a estatística do teste pertencer à região crítica, rejeita-se hipótese nula, ou seja, existe dependência entre as variáveis estudadas. Caso contrário, não se rejeita a hipótese nula, desta maneira, existe independência entre as variáveis.

Teste de Independência Qui-quadrado (χ^2)

O Teste Qui-quadrado para independência é indicado para quando as variáveis em análise forem categóricas, seguindo as etapas:

1) Formulação das hipóteses:

$$H_0 : \text{independência presente vs } H_1 : \text{independência ausente.}$$

2) Distribuição da Estatística:

$$\chi^2 = \sum_{k=1}^n \frac{(o_i - e_i)^2}{e_i},$$

com $(c - 1)(d - 1)$ graus de liberdade. Sendo c e d , o número de categorias de cada variável.

Onde,

o_i é o valor observado na posição i na tabela.

e_i é o valor esperado na posição i na tabela.

3) Cálculo do p-valor:

O p-valor permite estabelecer a significância estatística do resultado, isto é, se a amostra é consistente com a hipótese nula a ser testada. E será calculado utilizando a função *chisq.test* do software R.

4) Conclusão:

Definido um nível de significância α , se o p-valor calculado for menor ou igual à α rejeita-se a hipótese nula. Caso contrário, não rejeita-se a hipótese nula.

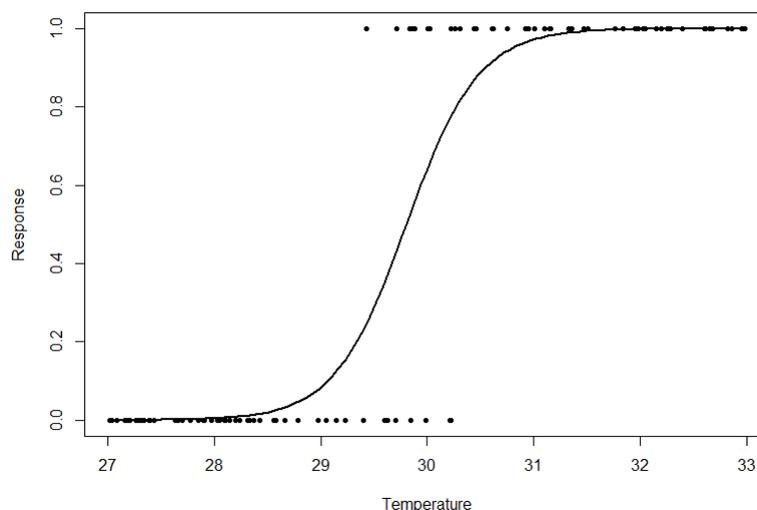
2.3 Regressão Logística

A Regressão Logística é um caso particular de Modelo Linear Generalizado (MLG) cujo objetivo é produzir, a partir de observações prévias, um modelo que permita a predição de valores e estudar a relação entre uma variável dependente categórica (variável resposta) e um conjunto de variáveis independentes (variáveis respostas). Se distingue da Regressão Linear pelo fato da variável resposta Y ser uma variável categórica. Este modelo segue as seguintes suposições:

- observações independentes;
- ausência de multicolinearidade perfeita entre as variáveis explicativas

Além disso, o seu gráfico é definido por uma curva em forma de S, onde os valores esperados estão entre 0 e 1, e a aproximação dos limites acontece de maneira suave (HOSMER; LEMESHOW, 2000b).

Figura 1: Ilustração do modelo logístico.



Fonte: <https://goo.gl/nwec4Q>

Regressão Logística Binária

Seja Y uma variável aleatória dicotômica, i.e, que admite apenas dois valores possíveis. Considerando ainda que a variável resposta pode ser representada por uma indicadora binária de valores 0 e 1, onde 0 representa “fracasso” e 1 “sucesso”, então

$$Y_i = \begin{cases} 1, & \text{com probabilidade } \pi_i \\ 0, & \text{com probabilidade } 1 - \pi_i \end{cases}$$

Suponha que a probabilidade π_i associada aos valores de Y_i dependa de um vetor de covariáveis $X_i = (1, X_{i1}, X_{i2}, \dots, X_{ip})'$. Na Regressão Linear, a esperança condicional de Y_i dado x_i é uma equação linear expressão por

$$E(Y_i|x_i) = \beta X_i,$$

em que $\beta = (\beta_0, \beta_1, \dots, \beta_p)'$.

Contudo, não é possível utilizá-la na Regressão Logística pois a função é não linear, sua esperança condicional varia entre 0 e 1, e a aproximação dos limites é tida de maneira gradual. Devido a isso, é necessário realizar uma transformação de $E(Y_i|x_i)$, conhecida como transformação logito, simplificando a notação $E(Y_i|x_i) = \pi(x_i)$, a transformação é definida, em termos de $\pi(x_i)$ como:

$$\text{logit}(\pi(x_i)) = \log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \alpha + x_i'\beta, \quad i = 1, \dots, n$$

A razão $\frac{\pi(x_i)}{1 - \pi(x_i)}$ é chamada de chance, definida como uma medida de associação que indica a razão entre a chance de sucesso e fracasso, e é justamente a função de ligação que associa os valores esperados da resposta às variáveis explicativas do modelo.

Então, a equação de regressão estimada para o caso logístico é dada por

$$E(Y_i|x_i) = \pi(x_i) = \frac{e^{\alpha + \beta X_i}}{1 + e^{\alpha + \beta X_i}} \quad (2.3.1)$$

Para ajustar um modelo de regressão, é necessário estimar o vetor de interceptos $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_j)$ e o vetor dos parâmetros $\beta = (\beta_1, \dots, \beta_p)$. Para isso, será utilizado o método da máxima verossimilhança, que através dos dados amostrais busca estimar os parâmetros $\hat{\beta}$ que tem a maior probabilidade de replicar o padrão das observações, nos dados da amostra. Este método leva em consideração propriedades importantes, sendo elas:

- **Consistência:** quando o tamanho das observações (n) tende a infinito, o estimador $\hat{\beta}$ converge em probabilidade para β .
- **Propriedade de invariância:** Os estimadores de máxima verossimilhança são invariantes em transformações monotônicas, isto é, se o estimador de máxima verossimilhança de β for $\hat{\beta}$ então o estimador de uma função $g(\beta)$ será $g(\hat{\beta})$.
- **Distribuição assintótica:** Quando n tende a infinito e sob certas condições de regularidade, os estimadores de máxima verossimilhança convergem em distribuição para uma distribuição normal. Isto é,

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(0, I_F^{-1}(\beta))$$

Sendo $I_F^{-1}(\beta)$ a matriz de informação de Fisher, definida pela segunda derivada da função de log-verossimilhança, onde $\beta = \hat{\beta}$.

Sejam Y_i variáveis aleatórias independentes e identicamente distribuídas, tal que, a distribuição de $Y_i|X$ segue uma distribuição Bernoulli (π_i), ou seja,

$$Y_i|X_i \sim \text{Bernoulli}(\pi_i)$$

Então, a distribuição de probabilidade pode ser representada como

$$P(Y_i = y|x_i) = \begin{cases} \pi_i, & \text{se } y = 1 \\ 1 - \pi_i, & \text{se } y = 0 \end{cases}$$

$$P_{Y_i}(y_i) = \pi_i^{y_i}(1 - \pi_i)^{1-y_i}$$

a função verossimilhança é dada por:

$$L(\beta; Y_i; X_i) = \prod_{i=1}^n \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.3.2)$$

Aplicando o resultado (4.2.1) para substituir π_i em (4.2.2)

$$L(\beta; Y_i; X_i) = \prod_{i=1}^n \left(\frac{e^{\beta X_i}}{1 + e^{\beta X_i}} \right)^{y_i} \left(\frac{1}{1 + e^{\beta X_i}} \right)^{1-y_i}$$

a função log-verossimilhança será

$$l(\beta) = \sum_{i=1}^n [y_i \beta X_i - \log(1 + e^{\beta X_i})] \quad (2.3.3)$$

E por fim, derivando (2.3.3) em relação ao p-ésimo coeficiente e igualando a zero:

$$\frac{\partial l(\beta)}{\partial \beta_p} = \sum_{i=1}^n X_i (y_i - \pi_i) = 0 \quad (2.3.4)$$

Os valores que satisfazem a equação (2.3.4) são os estimadores dos coeficientes do modelo, e medem a taxa de variação do logit para uma unidade de variação na variável independente. Ademais, não há uma fórmula fechada para esses valores, para isso, é necessário utilizar métodos numéricos de solução, como por exemplo, o de Newton-Raphson.

Regressão Logística Multinomial

A regressão logística multinomial pode ser vista como uma extensão do modelo logístico binário, em situações nas quais a variável resposta possui três ou mais categorias. Neste caso, o modelo multinomial também pode ser visto com a soma de modelos binários.

Seja Y uma variável aleatória com J categorias, ou seja, com J valores possíveis com probabilidade π_j . A i -ésima observação de Y pode ser representada pela seguinte variável indicadora:

$$y_{ij} = \begin{cases} 1, & \text{se a observação } i \text{ pertence à categoria } j \\ 0, & \text{caso contrário} \end{cases}$$

E a probabilidade da i -ésima observação pertencer a j -ésima categoria é dada por

$$P(Y_i = j) = \pi_{ij} = \frac{e^{\beta_j X_i}}{1 + \sum_{j=1}^{J-1} e^{\beta_j X_i}}$$

sendo $\beta_j = (\beta_{0j}, \beta_{1j}, \dots, \beta_{pj})$ e $X_i = (1, X_{i1}, X_{i2}, \dots, X_{ip})'$, a função de verossimilhança será dada por:

$$L(\beta; Y_i; X_i) = \prod_{i=1}^n [\pi_{i1}^{y_{i1}}, \pi_{i2}^{y_{i2}}, \pi_{i3}^{y_{i3}}, \dots, \pi_{iJ}^{y_{iJ}}]$$

e a função de log-verossimilhança

$$\begin{aligned}
 l(\beta) &= \log \left[\prod_{i=1}^n (\pi_{i1}^{y_{i1}}, \pi_i^{y_{i2}}, \pi_{i3}^{y_{i3}}, \dots, \pi_{iJ}^{y_{iJ}}) \right] \\
 &= \sum_{i=1}^n [\log \pi_{i1}^{y_{i1}} + \log \pi_{i2}^{y_{i2}} + \dots + \log \pi_{iJ}^{y_{iJ}}] \\
 &= \sum_{i=1}^n [y_{i1} \log(\pi_{i1}) + y_{i2} \log(\pi_{i2}) + \dots + y_{iJ} \log(\pi_{iJ})]
 \end{aligned}$$

Usando o fato que $\sum_{i=1}^J \pi_{ij} = 1$ tem-se que:

$$\begin{aligned}
 l(\beta) &= \sum_{i=1}^n \left[\sum_{s=1}^{J-1} y_{is} \log(\pi_{is}) + \left(1 - \sum_{s=1}^{J-1} y_{is} \right) \log \left(1 - \sum_{s=1}^{J-1} \pi_{is} \right) \right] \\
 &= \sum_{i=1}^n \left[\sum_{s=1}^{J-1} y_{is} \log(\pi_{is}) - \sum_{s=1}^{J-1} y_{is} \log \left(1 - \sum_{s=1}^{J-1} \pi_{is} \right) + \log \left(1 - \sum_{s=1}^{J-1} \pi_{is} \right) \right] \\
 &= \sum_{i=1}^n \left[\sum_{s=1}^{J-1} y_{is} \log \left(\frac{\pi_{is}}{1 - \sum_{s=1}^{J-1} \pi_{is}} \right) + \log \left(1 - \sum_{s=1}^{J-1} \pi_{is} \right) \right]
 \end{aligned}$$

como $\pi_{iJ} = 1 - \sum_{s=1}^{J-1} \frac{1}{1 + \sum_{s=1}^{J-1} \exp(\beta_s X_i)}$ então,

$$\begin{aligned}
 l(\beta) &= \sum_{i=1}^n \left[\sum_{s=1}^{J-1} y_{is} \log \left(\frac{\pi_{is}}{\pi_{iJ}} \right) + \log(\pi_{iJ}) \right] \\
 &= \sum_{i=1}^n \left[\sum_{s=1}^{J-1} y_{is} \beta_s X_i - \log \left(1 + \sum_{s=1}^{J-1} \exp(\beta_s X_i) \right) \right] \tag{2.3.5}
 \end{aligned}$$

Para estimar os parâmetros, primeiro deriva-se a equação (4.2.5) em relação a cada categoria e a cada parâmetro deste categoria. De maneira geral, tem-se que:

$$\begin{aligned}
 \frac{\partial l(\beta)}{\partial \beta_{jt}} &= \sum_{i=1}^n \left[y_{ij} X_{it} - \frac{X_{it} \exp(\beta_j X)}{1 + \sum_{s=1}^{J-1} \exp(\beta_s X_i)} \right] \\
 &= \sum_{i=1}^n X_{it} \left[y_{ij} - \frac{\exp(\beta_j X)}{1 + \sum_{s=1}^{J-1} \exp(\beta_s X_i)} \right] \\
 &= \sum_{i=1}^n X_{it} (k_{ij} - \pi_{ij}) = 0,
 \end{aligned}$$

E por fim, a derivada da função log-verossimilhança é igualada à zero:

$$\frac{\partial l(\beta)}{\partial \beta_{jt}} = \sum_{i=1}^n X_{it}(y_{ij} - \pi_{ij}) = 0$$

tal que $j = 1, 2, \dots, J - 1$ e $t = 0, 1, \dots, p$

Regressão Logística Multinomial Ordinal

A estimação dos parâmetros é semelhante à metodologia empregada na situação anterior. No entanto, o modelo ordinal utiliza a probabilidade cumulativa da variável resposta Y assumir valores iguais ou menores que um determinado ponto assim como utiliza interceptos α 's sendo justamente o que difere uma categoria de outra.

$$\begin{aligned} \gamma_{ij} &= P(y_i \leq j|x) = \pi_{i1} + \pi_{i2} + \dots + \pi_{ij} \\ &= \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}, \end{aligned}$$

em que $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_j)$ é o vetor de interceptos, tal que $j = 1, \dots, J$, $\beta = (\beta_1, \beta_2, \dots, \beta_p)$ é o vetor de parâmetros $x = (x_1, x_2, \dots, x_p)$ é o vetor de covariadas regressoras e $\pi_{ij} = P(y_i = j|x)$ representa a probabilidade de ocorrência da j -ésima categoria de resposta para um dado vetor x_i de p covariáveis, com $\sum_{j=1}^J \pi_{ij} = 1$, em que J é a quantidade de categorias de resposta. Desta maneira,

$$\pi_{ij} = \frac{\exp(\alpha_j + \beta x_i)}{1 + \exp(\alpha_j + \beta x_i)} - \frac{\exp(\alpha_{j-1} + \beta x_i)}{1 + \exp(\alpha_{j-1} + \beta x_i)},$$

A função de verossimilhança será dada por,

$$\begin{aligned} L(\beta; \alpha; Y) &= \prod_{i=1}^n P_{Y_i}(y_i) \\ &= \prod_{i=1}^n \left[\prod_{j=1}^J \pi_j(x_i)^{y_{ij}} \right] \\ &= \prod_{i=1}^n \left[\prod_{j=1}^J P(Y \leq j|x_i) - P(Y \leq j-1|x_i) \right]^{y_{ij}}, \end{aligned} \quad (2.3.6)$$

Como,

$$\pi_J(x_i) = \frac{1}{1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \beta_j^T x_i)}, \quad (2.3.7)$$

então, substituindo (2.3.7) em (2.3.6), encontra-se a função de verossimilhança em termos

de $\alpha_1, \dots, \alpha_{J-1}$ e β , dada por:

$$L(\beta; \alpha; Y) = \prod_{i=1}^n \left[\left(\frac{t}{1+t} \right)^{y_{ij}} \left(\prod_{j=2}^{J-1} \left(\frac{u}{1+u} - \frac{v}{1+v} \right)^{y_{ij}} \right) \left(\frac{1}{1+w} \right)^{y_{iJ}} \right],$$

em que $t = \exp(\alpha_1 - x'_i \beta)$, $u = \exp(\alpha_i - x'_i \beta)$, $v = \exp(\alpha_{j-1} - x'_i \beta)$ e $w = \exp(\alpha_{J-1} - x'_i \beta)$.

Então, o log da verossimilhança é dado por:

$$l(\beta, \alpha) = \sum_{j=1}^{J-1} \left[\alpha_j \left(\sum_{i=1}^n y_{ij} \right) + \sum_{k=1}^p \beta_{jk} \left(\sum_{i=1}^n x_{ik} y_{ij} \right) \right] - \sum_{i=1}^n \log \left[1 + \sum_{j=1}^{J-1} \exp(\alpha_j + \beta_j^T x_i) \right]. \quad (2.3.8)$$

Para encontrar os estimadores dos interceptos α_j , deriva-se a log-verossimilhança (4.2.6) em relação ao intercepto na categoria j e igualá-se à zero.

$$\frac{\partial l(\beta)}{\partial \alpha_j} l(\beta, \alpha) = \sum_{i=1}^n \left[(y_{ij} + y_{ij+1}) \left(\frac{\gamma_{ij}(1 - \gamma_{ij})}{\gamma_{ij} - \gamma_{ij-1}} \right) \right] = 0.$$

Analogamente, os estimadores dos parâmetros β_p serão dados por:

$$\frac{\partial l(\beta)}{\partial \beta_{pt}} l(\beta, \alpha) = [y_{ij} X_{it} q + y_{ij+1} X_{it} r] = 0,$$

em que $q = (1 - \gamma_{is} - \gamma_{is-1})$ e $r = (1 - \gamma_{is+1} - \gamma_{is-1})$.

Interpretação dos Parâmetros Estimados

A interpretação dos parâmetros estimados nos modelos de Regressão Logística é feita através do uso da razão de chances (OR) para cada categoria. Se OR é igual à 1 então a variável resposta e a explicativa não estão associadas, quando OR maior que 1 indica que a probabilidade de pertencer a uma dada categoria frente ao nível de referência é grande, sendo considerado um fator de risco, e se OR menor que 1 indica que o sucesso de uma dada categoria frente ao nível de referência é pequeno, sendo considerado um fator de proteção.

A chance de sucesso (Odds) de um modelo de regressão logístico é definida como:

$$odds = \frac{\pi(x_i)}{1 - \pi(x_i)} = e^{\alpha + X'_i \beta}.$$

A razão de chances (Odds Ratio) é dada ao comparar as chances de dois valores x_i e $x_i + 1$ de uma variável preditora X_i , cuja as demais variáveis explicativas são constantes, tem a finalidade de comparar a chance de sucesso em ambos os casos. De maneira geral,

pode ser definido como:

$$OR = \frac{\frac{P(\text{sucesso}|X=x_i+1)}{P(\text{fracasso}|X=x_i+1)}}{\frac{P(\text{fracasso}|X=x_i)}{P(\text{sucesso}|X=x_i)}} = e^{\alpha+\beta p}.$$

Esta relação indica que para $x_i + 1$, a chance de sucesso é multiplicada por e^{β} em relação a chance de sucesso de x_i .

Seleção e Validação do Modelo

A escolha do modelo final será dada pela comparação de todos os modelos, com o intuito de verificar quais os parâmetros e variáveis que são realmente significativos, com o objetivo de definir o modelo que além de melhor descrever o fenômeno em estudo, é parcimonioso. Para isso, será utilizado o Critério de Informação Bayesiano (BIC), o Critério de Informação Akaike (AIC) e o Critério de Informação Akaike Corrigido. Assim como, os testes da Razão de Verossimilhança e de Wald. Já a validação do modelo selecionado será feita através da análise dos seus resíduos.

O Critério de Informação de Akaike (AIC) faz uma ponderação da qualidade do ajuste do modelo com a quantidade de parâmetros estimados. Este método tende a penalizar modelos com muitos parâmetros e considera o modelo com melhor ajuste, aquele com o menor valor AIC. Sendo expresso da seguinte forma:

$$AIC = -2\ln L(\beta) + 2p,$$

em que $L(\beta)$ é a função de verossimilhança e p é o número de parâmetros do modelo. Sendo recomendado apenas quando $\frac{n}{p} \geq 40$ (BURNHAM E ANDERSON, 2002).

Para o caso de pequenas amostras, ou seja, $\frac{n}{p} < 40$, foi desenvolvido critério de AIC corrigido (AICc).

$$AICc = AIC + \frac{2(p+1)(p+2)}{n-p-2}.$$

O Critério de Informação Bayesiano (BIC) faz a comparação entre as verossimilhanças a posteriori, também utiliza como critério de penalização a complexidade do modelo, porém sua regra de penalização é mais forte se comparada ao AIC. É definido como:

$$BIC = -2E[\ln L(\beta)] + p \ln(n).$$

Assim como o AIC, o modelo considerado com um melhor ajuste é aquele cujo valor de BIC é o menor.

Para avaliar a adequação do modelo logístico, utiliza-se os gráficos dos resíduos. Para este, espera-se que haja uma leve suavização dos resíduos dado à probabilidade estimada. Caso contrário, um desvio significativo do comportamento dos resíduos sugere inadequação do modelo.

O Teste da Razão de Verossimilhança faz uma comparação do modelo restrito sob a hipótese nula (com menos parâmetros) com o modelo irrestrito, (com todos os parâmetros livres). Para isso, utiliza a função desvio (D), sendo esta a razão da log-verossimilhança do modelo ajustado sob o modelo saturado, expressa da seguinte forma:

$$D = -2 \sum_{i=1}^n \left[y_i \ln \left(\frac{\hat{\pi}(x_i)}{y_i} \right) + (1 - y_i) \ln \left(\frac{1 - \hat{\pi}(x_i)}{1 - y_i} \right) \right]. \quad (2.3.9)$$

O modelo é dito saturado se contém todas as variáveis, neste caso os valores preditos correspondem aos observados, enquanto o modelo ajustado corresponde ao modelo apenas com as variáveis desejáveis para o estudo. O valor da deviance sempre é positivo, e quanto menor, melhor é o ajuste do modelo. Esta função desempenha o mesmo papel que a soma de quadrados dos resíduos (SSE), na regressão linear.

Sob as hipóteses:

$$H_0 : \beta_p = 0 \text{ vs } H_1 : \beta_p \neq 0.$$

A Estatística do Teste da Razão de Verossimilhança é calculada através da diferença entre a função deviance (4.2.7) de cada modelo:

$$G = 2 \sum_{i=1}^J p_i \ln \left(1 + \frac{p_i - \hat{\pi}_i}{\hat{\pi}_i} \right), \quad (2.3.10)$$

onde $p_i = \frac{n_i}{n}$. A função G possui distribuição assintótica qui-quadrado com J - 2 graus de liberdade.

O Teste de Wald verifica se uma determinada variável explicativa possui relação com a variável resposta. É obtido pela comparação da estimativa de máxima verossimilhança do parâmetro β_p com a estimativa de seu erro padrão.

Sob as hipóteses:

$$H_0 : \beta_p = 0 \text{ vs } H_1 : \beta_p \neq 0.$$

A Estatística do teste é definida abaixo e segue distribuição normal padrão:

$$Z = \frac{\hat{\beta}}{SE(\hat{\beta})}.$$

Analogamente, o Teste de Wald para o caso de múltiplas variáveis explicativas, verifica a hipótese de que cada um dos $p + 1$ coeficientes β é igual a zero, isto é:

$$H_0 : \beta_0 = \beta_1 = \dots = \beta_p = 0 \text{ vs } H_1 : \text{ pelo menos um } \beta \text{ diferente de zero.}$$

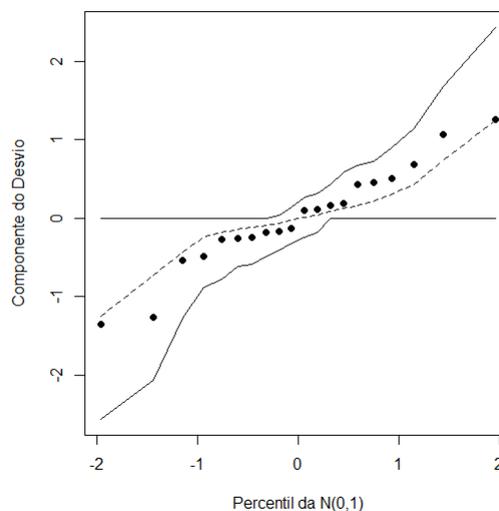
A Estatística de Wald é dado pelo seguinte cálculo matricial:

$$\begin{aligned} W &= \hat{\beta}' [Var(\hat{\beta})^{-1}] \hat{\beta} \\ &= \hat{\beta}' [X' V X] \hat{\beta}, \end{aligned}$$

segue distribuição qui-quadrado com $p + 1$ graus de liberdade

Por fim, para avaliar se o modelo selecionado possui uma boa adequação, utiliza-se o gráfico dos resíduos, já que não são em todas as situações em que a variância será constante.

Figura 2: Ilustração de gráfico dos resíduos.

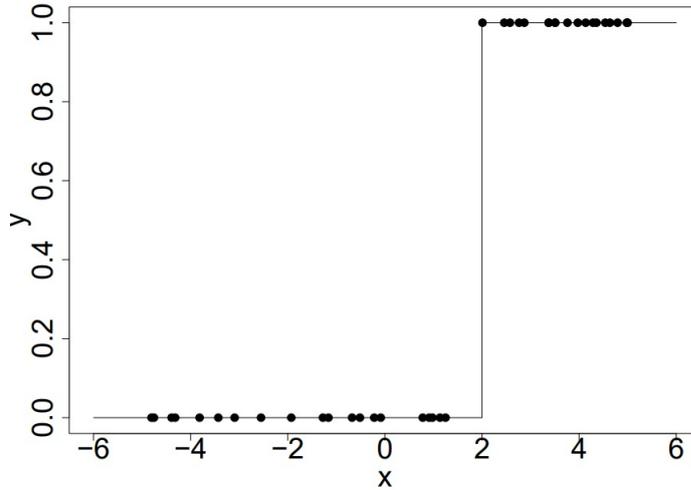


Fonte: De autoria própria.

Problema de separação completa ou quase completa

Na construção do modelo logístico, principalmente quando a variável resposta é binária, podem ocorrer problemas de separação e de não convergência, ilustrados pela Figura 2. Em razão disso, não é possível estimar os parâmetros do modelo pela estimação de máxima verossimilhança.

Figura 3: Ilustração de um problema de separação no modelo logístico.



Fonte: Gelman et al. (2008)

Tal adversidade é dada quando o algoritmo de estimação pela máxima verossimilhança não converge, sendo, de acordo com Allison (2008), na maioria dos casos consequência do padrão dos dados ter completa ou quase completa separação em relação à variável resposta. Isto ocorre quando existe algum vetor de coeficientes β de tal modo que $y_i = 1$ sempre que $\beta x_i > 0$ e $y_i = 0$ sempre que $\beta x_i \leq 0$. Em outras palavras, existe uma função linear de x consegue gerar preditores perfeitos da variável resposta y .

Assim, uma das maneiras de realizar o ajuste no modelo logístico, para evitar os problemas citados, é dado pelo método de penalização da estimação de máxima verossimilhança, proposto por Firth (1993), que visa reduzir o viés da regressão logística em pequenas amostras.

Seja β o vetor de coeficientes regressivos, cujo o estimador de máxima verossimilhança é dado pela solução da equação $\mathbf{U}(\beta) = \frac{\partial \log \mathbf{L}(\beta)}{\partial \beta} = 0$, em que $\mathbf{L}(\beta)$ a sua função de verossimilhança. O método consiste na seguinte modificação da função de verossimilhança

$$\mathbf{L}^*(\beta) = \mathbf{L}(\beta) |\mathbf{I}(\beta)|^{1/2},$$

em que $|\mathbf{I}(\beta)|$ denota o determinante da matriz de informação de Fisher. Sendo agora, o estimador de β denotado por $\hat{\beta}^*$ e obtido pela solução da equação

$$\mathbf{U}^*(\beta) = \frac{\partial}{\partial \beta} \log \mathbf{L}(\beta) + \frac{1}{2} \text{tr} \left\{ \mathbf{I}(\beta)^{-1} \frac{\partial \mathbf{I}(\beta)}{\partial \beta} \right\},$$

Outro procedimento foi sugerido por Gelman et al. (1993), que propôs uma nova

distribuição priori ao modelos clássicos de regressão logística, construídos através de uma padronização das variáveis não binárias para que tenham média igual à 0 e desvio padrão igual à 0,5, e, depois aplicando uma distribuição priori sob os coeficientes. Sendo usualmente utilizada a distribuição Cauchy com locação igual à 0 e escala igual à 2,5.

Esse método faz uma adaptação ao algoritmo padrão da estimação de máxima verossimilhança para obter aproximações inferenciais posteriores para os coeficientes β , sob a forma de uma estimativa $\hat{\beta}$ e uma matriz de covariância V_{β} . A cada iteração, o algoritmo bayesiano determina um pseudo-dado z_i e uma pseudo-variância $(\sigma_i^z)^2$ baseados na linearização da derivada da função log-verossimilhança,

$$\begin{aligned} z_i &= X_i \hat{\beta} + \frac{(1 + e^{X_i \beta})^2}{e^{X_i \beta}} \left(y_i - \frac{e^{X_i \beta}}{1 + e^{X_i \beta}} \right), \\ (\sigma_i^z)^2 &= \frac{1}{n_i} \frac{(1 + e^{X_i \beta})^2}{e^{X_i \beta}}, \end{aligned}$$

e em seguida executa mínimos quadrados ponderados, regredindo z em X com vetor peso $(\sigma_i^z)^2$. A estimativa resultante $\hat{\beta}$ é substituída nas equações acima e a iteração prossegue até convergir.

3 Metodologia

O presente estudo é classificado como descritivo, por ter como objetivo primordial descrever as características de determinada população e estabelecer relações entre variáveis (GIL, 2002). Quanto à abordagem metodológica, é considerada mista. Já em relação ao delineamento, pode ser definida como uma pesquisa de campo, que segundo Fonseca:

A pesquisa de campo caracteriza-se pelas investigações em que, além da pesquisa bibliográfica e/ou documental, se realiza coleta de dados junto a pessoas, com o recurso de diferentes tipos de pesquisa (pesquisa ex-post-facto, pesquisa-ação, pesquisa participante, etc.) (FONSECA, 2022)

3.1 Materiais

A amostra foi composta por 32 (trinta e dois) surfistas de ambos os sexos que praticam treinos específicos fora d'água para o Surf. O número da amostra foi definido baseado na quantidade de alunos de três centros de treinamentos, localizados nas cidades de Maceió- AL e Marechal Deodoro - AL, que realizam treinos funcionais focados no desenvolvimento do esporte.

A coleta de dados foi realizada através de um questionário do tipo semiaberto com 16 (dezesesseis) a 25 (vinte e cinco) perguntas relacionadas à prática do esporte, a depender da habilidade do surfista de realizar manobras complexas, a informação nível de surf de cada aluno por seu respectivo professor e a aplicação de quatro testes físicos, assim como o turno em que foi realizado. Ao final, o banco de dados foi composto por 30 (trinta) covariáveis que, além das recolhidas, foi criada a variável *idade inicial*, subtraindo-se a idade pelo tempo de prática do esporte. Ademais, as covariáveis de classificação das manobras foram parametrizadas como “Realiza com facilidade” se a escala de dificuldade está entre 0 e 2, “Realiza com dificuldade” se entre 3 e 5, e “Não sabe” a não resposta.

O questionário foi elaborado pelo aplicativo de gerenciamento de pesquisas, conhecido como Google Forms, e consolidou informações sobre a idade (em anos), sexo, consumo de álcool, consumo de embutidos e industrializados, consumo de açúcares, tempo (em meses) de surf, frequência (em dias) de prática semanal, média (em minutos) de duração de cada sessão de surf, classificação para realizar as principais manobras, tempo para realização da primeira manobra, se pratica outro esporte e qual, domínio de leitura

de onda, domínio de leitura da previsão de ondas e se consulta a previsão de ondas antes da sessão.

Além disso, os alunos dos centros de treinamentos foram classificados pelos seus respectivos professores de acordo com o grau de desenvolvimento no esporte, que por sua vez será a característica de interesse do estudo, sendo os principais níveis:

- Iniciante: o surfista que está aprendendo a remar, dropar, correr pela parede da onda e ter controle da prancha.
- Intermediário: neste estágio, o surfista já tem domínio dos procedimentos do nível iniciante e está aprendendo a realizar as primeiras manobras básicas como: batida, rasgada e cutback.
- Avançado: neste nível, o surfista já domina manobras básicas e está aprendendo manobras mais complexas.
- Profissional: o surfista possui domínio de manobras básicas e complexas, domina a prancha e surfa ondas de diferentes tamanhos e em diversos lugares.

Com o suporte de três profissionais da Área de Educação Física, que são também os responsáveis de cada centro de treinamento do estudo, foram aplicados quatro testes físicos, por abrangerem as capacidades físicas relacionadas ao esporte, serem considerados de fácil aplicação e relativamente baratos. Sendo eles: o Teste de Flexão de Braços proposto por Pollock e Wilmore (1993), o Teste do Flamingo, o Teste de Sentar e Alcançar e o Teste de Impulsão Horizontal, os últimos três propostos pelo Conselho da Europa (1988).

Teste de Flexão de Braços

Objetivo: Analisar a força do membro superior através da flexão dos braços.

Equipamentos necessários: Nenhum.

Procedimentos: Os homens realizam o teste sem o apoio dos joelhos, já as mulheres, com o apoio dos joelhos.

Protocolo: O avaliado deve se posicionar deitado com a barriga para baixo, com as mãos apoiadas no solo, com uma distância de 10 a 20 cm a partir da linha dos ombros, com os dedos voltados para frente. A posição inicial se dá quando o avaliado está sob o solo com os cotovelos flexionados e a posição final quando os braços estão estendidos. Na posição inicial do movimento, o rosto deve permitir um alinhamento adequado entre o tronco e as pernas, o quadril deve ficar alinhado com a linha do ombro.

Pontuação: Registra-se o número máximo de repetições corretas em um minuto. O teste deve ser interrompido se o avaliado realizar uma execução incorreta ou não completar o movimento de flexão e extensão de cotovelos. A classificação é feita a partir da tabela de valores normativos do teste de flexão de braços em pessoas saudáveis proposta por Pollock e Wilmore (1993).

Teste do Flamingo

Objetivo: Avaliar a capacidade de equilíbrio estático, através da lateralidade dos membros inferiores.

Equipamentos necessários: Um cronômetro e um bosu.

Procedimentos: O participante efetuará o teste com o pé de preferência.

Protocolo: O avaliado deverá colocar um dos pés sobre o eixo longitudinal da trave e, flexionando a perna livre, agarrar no peito do pé com a mão do mesmo lado, imitando a posição de flamingo. Deverá então tentar manter o equilíbrio nesta posição durante um minuto, podendo previamente apoiar-se no antebraço do observador para se colocar na posição correta. O teste começa logo que este apoio cessa. Se o teste é interrompido, uma penalidade é atribuída, o avaliado se reposiciona em apoio novamente e dá-se uma nova partida até que o minuto se esgote.

Pontuação: O número total de quedas ou perda de equilíbrio em 60 segundos é gravado. A tabela de pontuação está disponível no Manual Eurofit (2015), classificada de acordo com a idade e o número de quedas, distribuída em: excelente, muito bom, bom, suficiente, fraco e insuficiente. Se houver mais de 15 quedas nos primeiros 30 segundos, o teste é terminado, não obtendo assim classificação.

Teste de Impulsão Horizontal

Objetivo: Analisar a força de explosão dos membros inferiores.

Equipamentos necessários: Uma trena e uma linha traçada no solo. Procedimentos: Corte um pedaço de 30 cm de fita adesiva e cole em linha reta sob o solo.

Protocolo: O aluno coloca-se imediatamente atrás da linha, com os pés paralelos, ligeiramente afastados, joelhos semi-flexionados, tronco ligeiramente projetado à frente. Ao sinal o aluno deverá saltar a maior distância possível. Serão realizadas duas tentativas, registrando-se o melhor resultado.

Pontuação: A distância do melhor salto é registrada em centímetros.

Teste Sentar e Alcançar

Objetivo: Avaliar a flexibilidade da cadeia posterior, isto é, a parte de trás do corpo, tanto do tronco quanto do quadril.

Equipamentos necessários: Banco de Wells.

Protocolo: Posicione o aluno descalço sentado no chão com os calcanhares sob cada extremidade da fita, oriente o aluno a manter os joelhos estendidos e posicionar uma mão sob a outra, solicite ao aluno que alongue o tronco e após flexione a frente, encostando com os dedos a fita métrica no banco de wells. Repita 2 vezes e anote o melhor resultado em centímetros. Pontuação: De acordo com o alcance do aluno (em centímetros).

3.2 Métodos

No presente estudo, com o auxílio do Software R, será realizado análises descritivas e de correlação das variáveis, com o intuito de investigar o comportamento das mesmas e verificar possíveis tendências. Assim como, será aplicada técnicas de Regressão Logística binária e ordinal. A primeira com a finalidade de averiguar a possibilidade de somar dois ou mais níveis de classificação de desempenho. E a segunda, em caso da análise binária não ser considerada adequada.

Os dados serão analisados através do cálculo de medidas estatísticas, conhecidas como medidas de posição e de dispersão. As primeiras visam resumir as informações em uma única só, posicionada geometricamente entre os valores extremos do conjunto de dados, sendo as principais utilizadas: a média, a mediana, a moda e os quartis da distribuição. E as medidas de dispersão, como a variância e o desvio padrão, serão utilizadas para investigar como os dados estão distribuídos, isto é, avaliar o grau de variabilidade deles em torno da média.

O conjunto de dados será apresentado com o auxílio de recursos gráficos e tabulares. Estes instrumentos facilitam a visualização e a compreensão das informações, visto que as retratam de maneira clara e consolidada.

Para a avaliar o grau de relacionamento entre as variáveis, será calculado o coeficiente de correlação ou associação apropriado a natureza das variáveis em questão. A fim de averiguar a significância do coeficiente calculado, será realizado o teste de hipótese adequado.

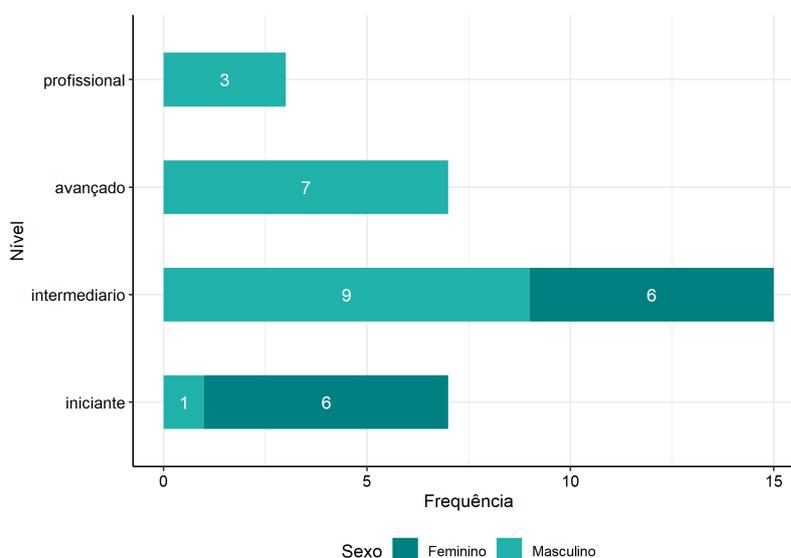
4 Resultados

A amostra contém informações de trinta e dois surfistas, selecionados de maneira aleatória, obtidas através da aplicação do questionário em anexo e dos testes físicos propostos na parte metodológica. A partir disso, é possível analisar o comportamento das variáveis em estudo.

4.1 Análise Descritiva

Os surfistas foram classificados em relação ao seu nível de desempenho no Surf por seus respectivos treinadores físicos como demonstrado na Figura 4.

Figura 4: Distribuição do nível de desempenho por sexo do surfista.



O estudo foi composto por 20 surfistas do sexo masculino, 62,5% do total, e 12 do sexo feminino (37,5%). O número de mulheres que surfam e realizam treinamento específico para o esporte representa apenas 60% da quantidade de homens na amostra. Também percebe-se que as mulheres estão distribuídas nas categorias mais baixas, o que talvez possa ser um reflexo da entrada tardia delas no esporte.

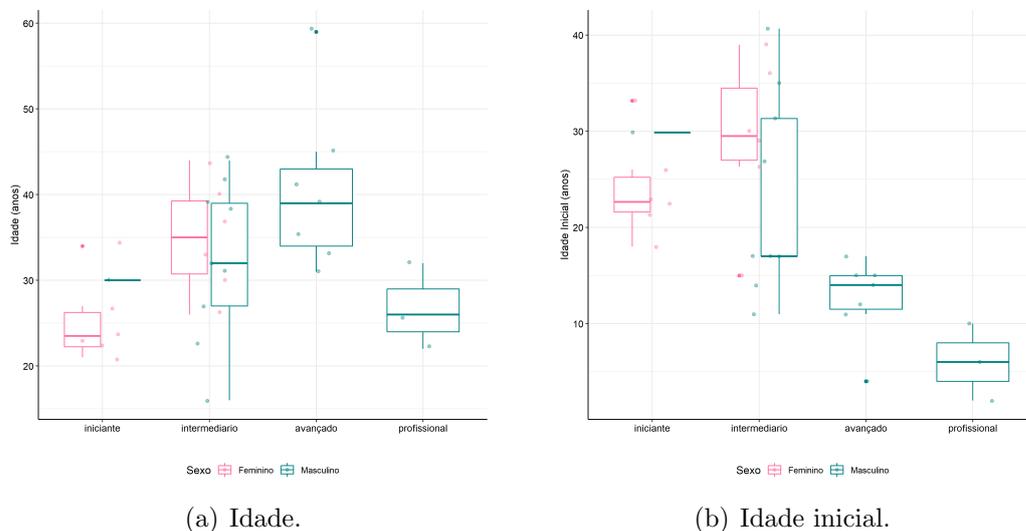
Tabela 1: Medidas resumo da idade do surfista por sexo.

| Sexo | Idade | | | | | |
|----------|-------|------------|------|-------|------------|-----|
| | Mín | 1º quartil | Md | Média | 3º quartil | Máx |
| Mulheres | 21 | 23,75 | 28,5 | 30,08 | 34,75 | 44 |
| Homens | 16 | 29,25 | 32,5 | 34,25 | 39,5 | 59 |
| Total | 16 | 26 | 32 | 32,69 | 39 | 59 |

A participante mais jovem tem 21 anos, por sua vez, o mais jovem tem 16 anos, já a mulher mais velha possui 44 anos e o homem mais velho tem 59 anos. A idade média dos participantes é 32,69 anos, separado por sexo tem-se que a faixa etária média das mulheres é de 30,08 anos e os homens têm em média 34,25 anos.

Nota-se também, pela Tabela 1, uma variação etária maior dos indivíduos do sexo masculino em relação aos do sexo feminino. Os surfistas homens têm idade entre 16 e 59 anos, sendo 44 anos de diferença entre o surfista mais jovem e o mais velho. Já entre as mulheres, a faixa etária varia de 21 a 43 anos, sendo a diferença de 22 anos.

Figura 5: Distribuição da idade e idade inicial por sexo e nível de desempenho.



Depreende-se da Figura 5b que a idade inicial dos homens indica uma predisposição, em que os surfistas que iniciaram mais jovens tendem a estar em categorias de desempenho mais elevadas. Por sua vez, entre as mulheres, ocorre o contrário, as surfistas que iniciaram novas estão na categoria iniciante e as mais velhas no intermediário.

As medidas resumo a seguir demonstram que não somente a média de ingresso diferente entre os sexos, mas que apenas a idade mais antiga a iniciar no esporte é similar entre os dois grupos.

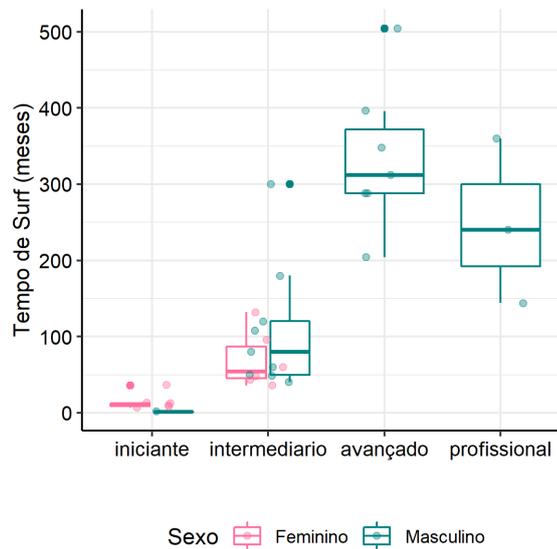
Tabela 2: Medidas resumo da idade inicial do surfista por sexo.

| Sexo | Idade inicial | | | | | |
|----------|---------------|------------|-------|-------|------------|-------|
| | Mín | 1º quartil | Md | Média | 3º quartil | Máx |
| Mulheres | 15 | 22,15 | 26,17 | 26,6 | 30,79 | 39 |
| Homens | 2 | 11 | 15 | 17,29 | 19,46 | 40,67 |
| Total | 2 | 14 | 17,5 | 20,78 | 29,22 | 40,67 |

Essa discrepância das variações tanto da idade quanto da idade inicial dos surfistas de acordo com o sexo, podem ser devidos a discriminação, que a princípio, as mulheres que surfaram, sofriam. Fazendo com que não somente os homens se aprimorassem no esporte antecipadamente, mas também que fossem introduzidos pioneiramente em competições, tivessem mais recursos por patrocínios, oportunidades e visibilidade no esporte.

Adicionalmente, ao analisarmos as Figuras 5a e 5b em paralelo, nota-se também que as mulheres que estão no nível iniciante não só iniciaram no Surf mais novas como também são mais jovens. Isso indica que as mulheres, apesar de iniciarem novas no esporte, não possuem muito tempo de prática. Esta suposição é confirmada na Figura 6, já que claramente observa-se que o sexo feminino desfruta do esporte a menos tempo.

Figura 6: Distribuição do Tempo de Surf



Da Figura 6, tem-se que o intervalo de prática do esporte aparenta aumentar conforme o nível de desempenho do surfista também aumenta. Ademais, ao passo que o nível de desempenho do surfista avança, a variabilidade entre os tempos de prática do Surf também cresce, assim como o tempo médio no esporte, com exceção ao nível profissional.

Tabela 3: Medidas resumo do tempo de Surf por nível de desempenho.

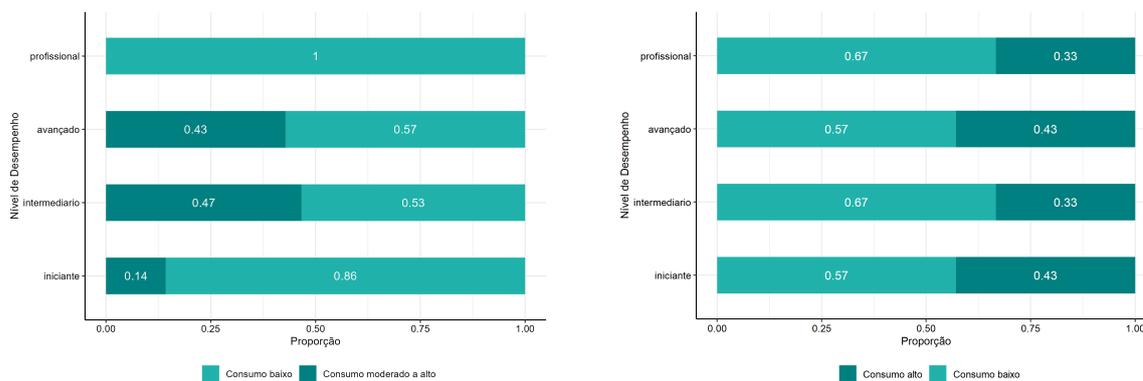
| Nível de desempenho | Tempo de Surf (meses) | | | | | |
|---------------------|-----------------------|------------|-----|-------|------------|-----|
| | Mín | 1º quartil | Md | Média | 3º quartil | Máx |
| Iniciante | 1,5 | 7,5 | 10 | 12,5 | 12,5 | 36 |
| Intermediário | 36 | 48,5 | 60 | 93,47 | 114 | 300 |
| Avançado | 204 | 288 | 312 | 334,3 | 372 | 504 |
| Profissional | 144 | 192 | 240 | 248 | 300 | 360 |

Das medidas antecedentes verifica-se que o participante com menos prática tem 1 mês e meio, enquanto o surfista mais experiente é um homem que surfa há 504 meses, cerca de 42 anos.

Em relação aos aspectos nutricionais dos surfistas, foi recolhido informações relacionadas ao consumo de álcool, de açúcares e de embutidos e industrializados, representados pelas Figuras 7a, 7b e 8, sucessivamente.

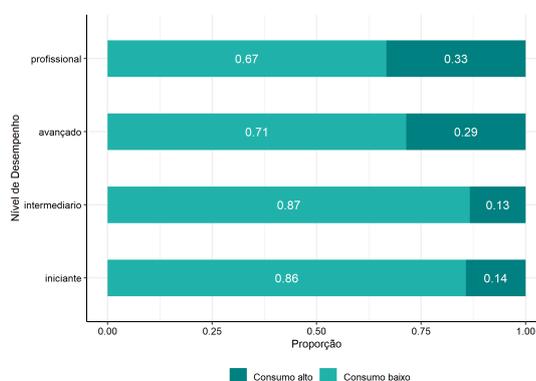
Para avaliar o consumo alcoólico semanal foi considerado que uma lata de cerveja (350ml) corresponde a aproximadamente 2 unidades, um cálice de vinho (90 ml) 1 unidade e uma dose de destilados (35-50ml) 2 unidades. Assim, o consumo abaixo de 10 unidades é indicado como *consumo baixo* e *consumo moderado a alterado*, caso contrário. Já em relação ao consumo diário de açúcares, representado na Figura 7b, foi considerado que uma barra de chocolate possui aproximadamente 50 gramas de açúcar, foi classificado como *alto* se superior a 25g diárias, e *baixo* caso contrário. Por fim, o consumo de embutidos e industrializados foi classificado como *baixo* para aqueles que na escala de consumo responderam entre 0 e 2, os demais foram especificados como *consumo alto*.

Figura 7: Aspectos nutricionais por sexo e nível de desempenho.



(a) Consumo de álcool semanal.

(b) Consumo de açúcar diário.



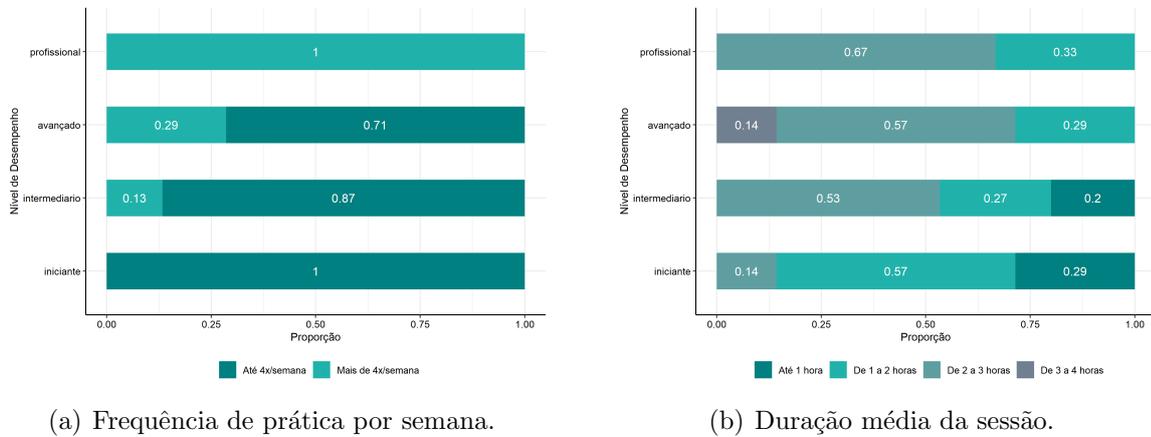
(c) Consumo de embutidos e industrializados diário.

De acordo com a Figura 7a, dentre os surfistas profissionais, esse consumo é baixo, já entre a categoria intermediária e avançada, é dividido quase igualmente entre as duas classificações de ingestão semanal. Da Figura 7b, observa-se apenas que a proporção de surfistas com um alto consumo diário de açúcares é igual para a categoria iniciante e avançada, e também, para o nível intermediário e profissional. Finalmente, ao contrário do esperado, percebe-se, pela Figura 7c, que a proporção dos iniciantes e intermediários que têm um consumo baixo de embutidos e industrializados é superior à dos demais níveis.

Conforme as Figuras 8a e 8b, a maioria dos participantes surfam até 4 vezes por semana e com sessões de duração média entre 2 a 3 horas. Ainda, é esperado que a frequência de prática semanal e a duração da sessão cresça de acordo com a classificação de desempenho do participante.

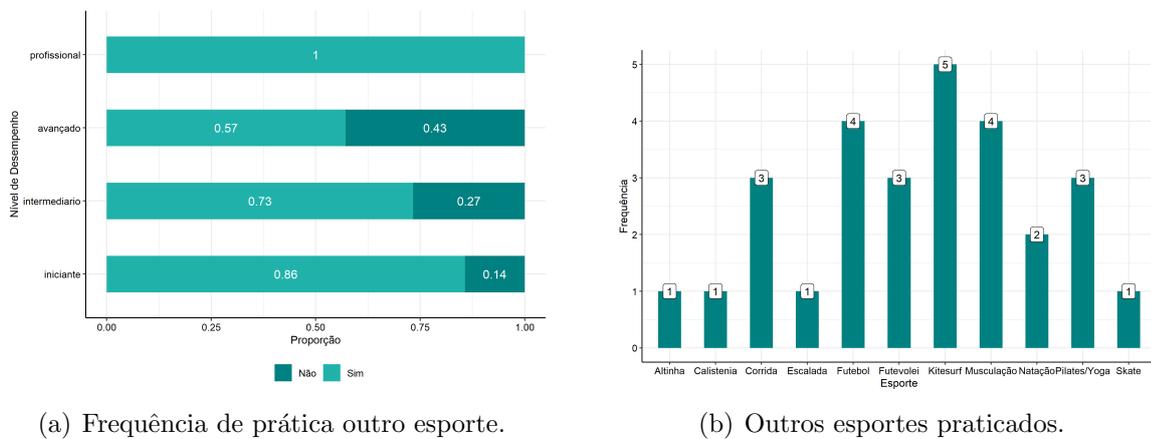
No que tange à duração média de uma sessão de Surf, é possível observar, na Figura 8b, que a maioria dos participantes surfam entre 2 a 3 horas. Os surfistas iniciantes, em sua maioria, relataram que surfam até 2 horas por sessão. A duração da sessão parece aumentar de acordo com o nível de desempenho.

Figura 8: Distribuição da frequência semanal e duração média da sessão por sexo e nível de desempenho.



A partir das Figuras 9a e 9b, conclui-se que maioria dos participantes relataram que também praticam outro esporte além do Surf, sendo os mais citados: kitesurf, musculação e futebol.

Figura 9: Distribuição da prática de outros esportes.



Tendo em vista que as mulheres iniciaram no Surf após os homens, além de ser verificada a distribuição da classificação de execução das manobras do Surf, por nível de desempenho, torna-se interessante também analisar o impacto dessa entrada tardia no esporte.

Relativamente à dificuldade em executar o movimento de ficar em pé na prancha, conhecido como drop, e correr a parede da onda, foi observado que os surfistas iniciantes e intermediários têm mais dificuldade no movimento, enquanto os avançados e profissionais não tem dificuldade.

Figura 10: Distribuição da classificação do Drop e correr a parede da onda por nível de desempenho.

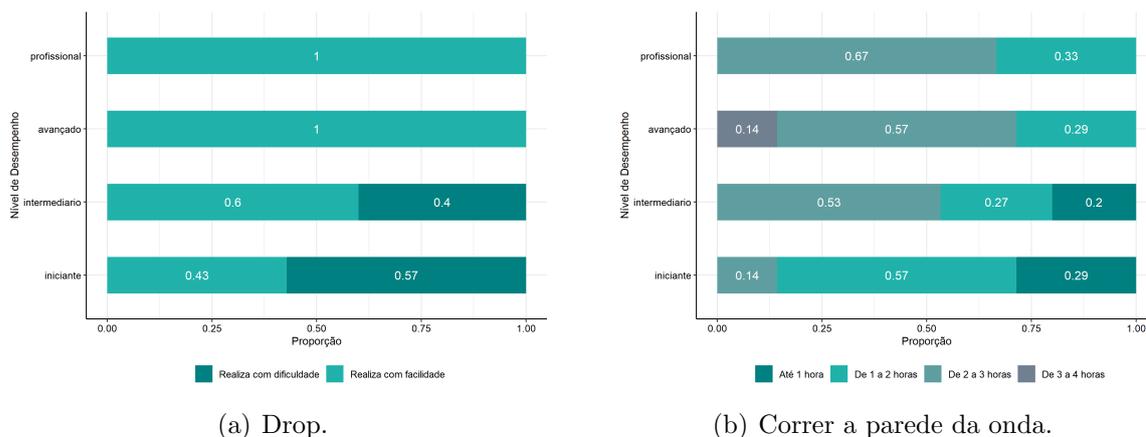


Tabela 4: Distribuição da frequência relativa da classificação das manobras Drop e Parede por sexo.

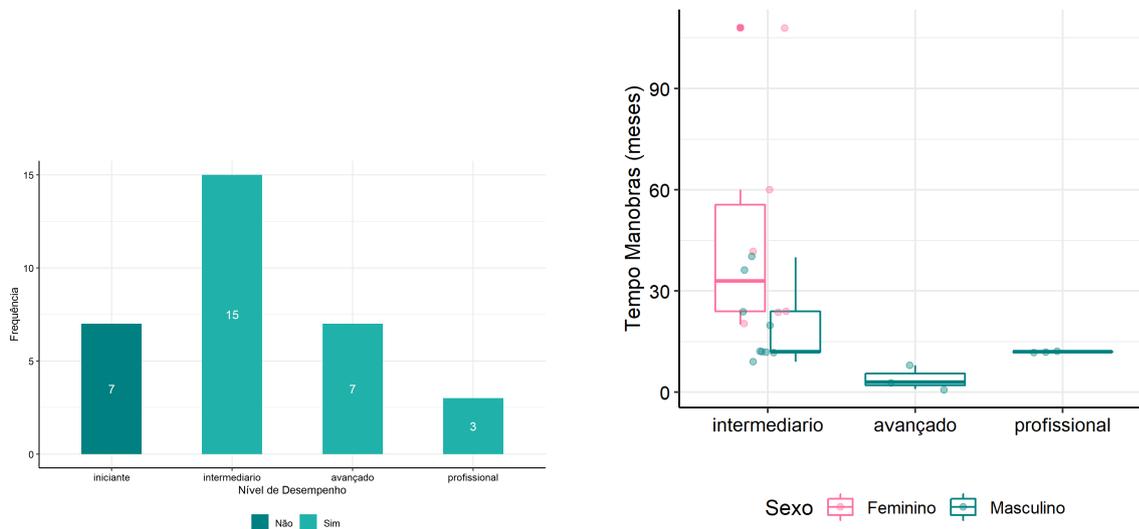
| Manobra | Classificação | Sexo | |
|---------|-------------------------|----------|-----------|
| | | Feminino | Masculino |
| Drop | Realiza com dificuldade | 0,58 | 0,15 |
| | Realiza com facilidade | 0,42 | 0,85 |
| Parede | Realiza com dificuldade | 0,67 | 0,25 |
| | Realiza com facilidade | 0,33 | 0,75 |

Assim, pelas Figuras 10a e 10b, nota-se que os surfistas iniciantes e intermediários possuem mais dificuldade em correr a parede da onda do que realizar o drop. Uma vez que a proporção salta de 57% para 86% dentre os iniciantes, e de 40% para 47% no nível seguinte.

Da Tabela 4, é evidente que os homens consideram mais fáceis os movimentos de subir em pé na prancha e correr a parede da onda em comparação às mulheres. A maioria das surfistas executam esses dois movimentos com dificuldade, o contrário ocorre com os surfistas. Todavia, vale ressaltar que o grupo feminino é composto pelos níveis iniciante e intermediário, fato que, impacta diretamente na classificação desses movimentos.

Os surfistas também foram questionados sobre a sua habilidade em realizar manobras mais complexas como: batida, rasgada, cutback, floater, 360°, aéreo e tubo, assim como o tempo dispendido para executar a primeira manobra complexa. Esses resultados estão expostos nas Figuras 11a e 11b.

Figura 11: Distribuição do domínio e tempo dispendido para execução de manobras complexas.



(a) Domínio de manobras complexas.

(b) Tempo dispendido para a manobra complexa.

De acordo com a Figura 11a, a maioria deles, 78,12%, relatou realizar manobras mais complexas, sendo apenas os iniciantes que não sabem. Em relação a esses, recorda-se que 85,71% deles são do sexo feminino. Dentre os participantes que sabem executar manobras complexas, as mulheres, em geral, levaram mais tempo para realizar esse tipo de movimento, conforme indicado pela Figura 11b.

As duas primeiras manobras complexas avaliadas foram a batida e a rasgada, para ambas nota-se uma tendência de decréscimo da dificuldade em executá-la, na medida em que o nível do surfista avança.

Comparando as Figuras 12a e 12b, conclui-se que os surfistas, em geral, possuem mais dificuldade na manobra Batida do que na Rasgada. Visto que a proporção de surfistas intermediários que têm dificuldade decai de 73% para 53%, dos avançados passa de 3% para 14% e apenas 33% dos surfistas profissionais afirmaram ter dificuldade na Batida, que por sua vez, não possuem na execução das Rasgada.

Figura 12: Distribuição da classificação da Batida e Rasgada por nível de desempenho.

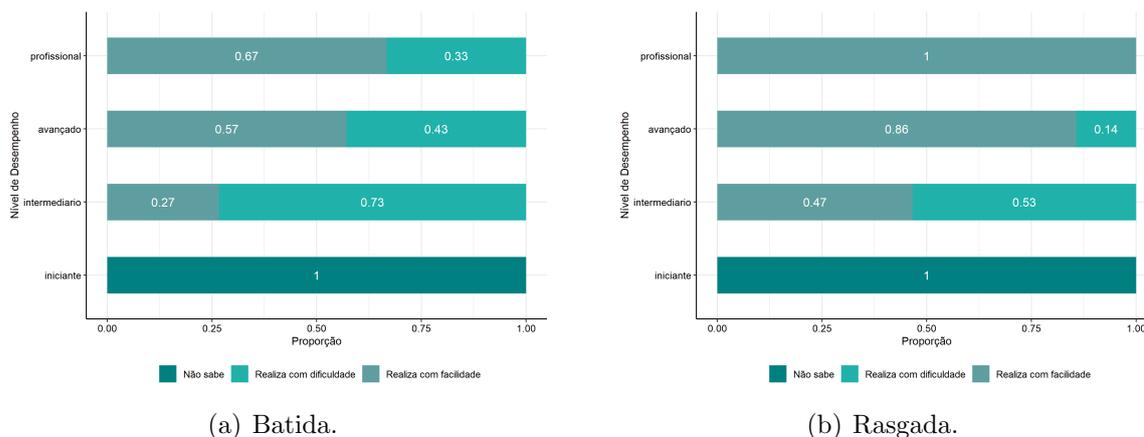


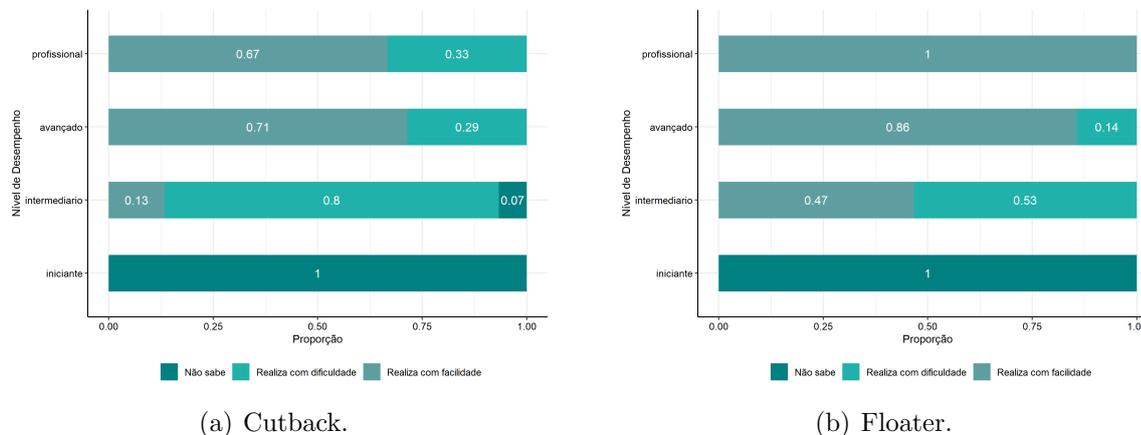
Tabela 5: Distribuição da frequência relativa da classificação das manobras Batida e Rasgada por sexo.

| Manobra | Classificação | Sexo | |
|---------|-------------------------|----------|-----------|
| | | Feminino | Masculino |
| Batida | Não Sabe | 0,5 | 0,05 |
| | Realiza com dificuldade | 0,5 | 0,45 |
| | Realiza com facilidade | 0 | 0,5 |
| Rasgada | Não Sabe | 0,5 | 0,05 |
| | Realiza com dificuldade | 0,3 | 0,25 |
| | Realiza com facilidade | 0,17 | 0,7 |

Também observa-se que 50% das surfistas não sabem realizar nenhuma das duas manobras, ainda assim, a maioria das mulheres que possuem domínio nesses movimentos, o realizam com dificuldade. Enquanto apenas um surfista não sabem executar essas manobras e a maior parte possuem facilidade.

Das seguintes manobras, indicadas nas Figuras 13a e 13b, verifica-se que o número de surfistas intermediários que não possuem domínio nelas cresce para 7% em comparação às duas manobras anteriores.

Figura 13: Distribuição da classificação da Cutback e Floater por nível de desempenho.



Depreende-se que os surfistas consideram mais difícil o Cutback do que o Floater, visto que 80% dos participantes que são intermediários realizam a primeira manobra com dificuldade, esse número cai para 60% na segunda. Enquanto 29% do nível avançado tem dificuldade no Cutback e somente 14% desta categoria afirmou que o Floater é difícil.

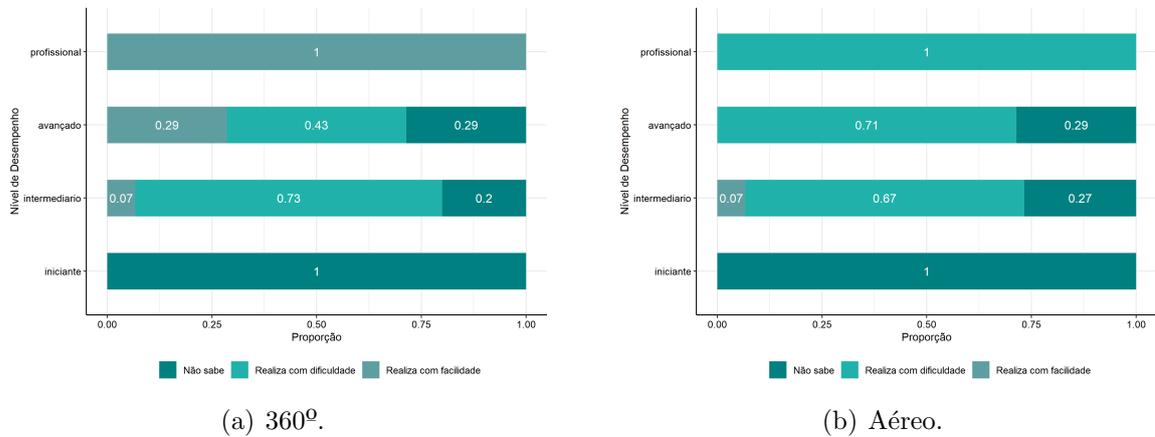
Tabela 6: Distribuição da frequência relativa da classificação das manobras Cutback e Floater por sexo.

| Manobra | Classificação | Sexo | |
|---------|-------------------------|----------|-----------|
| | | Feminino | Masculino |
| Cutback | Não Sabe | 0,5 | 0,1 |
| | Realiza com dificuldade | 0,5 | 0,45 |
| | Realiza com facilidade | 0 | 0,45 |
| Floater | Não Sabe | 0,5 | 0,1 |
| | Realiza com dificuldade | 0,5 | 0,25 |
| | Realiza com facilidade | 0 | 0,65 |

Em relação às duas manobras distribuídas na Tabela 6, tem-se que metade da mulheres não sabem realizá-las, e as que têm domínio afirmaram ter dificuldade. Já em relação ao sexo masculino, apenas dois homens disseram desconhecer a execução dessas manobras, dentre os que sabem, metade realizam o Cutback com dificuldade e a outra parte com facilidade. Enquanto o Floater, em sua maioria, é executada pelos homens com facilidade.

Em relação à execução da manobras 360^o e Aéreo, pelas Figuras 14a e 14b, verifica-se que o número de participantes dos níveis intermediário e avançado que não possuem o domínio necessário para realizá-las aumenta. Ainda assim, observa-se uma maior facilidade no 360^o em comparação ao Aéreo.

Figura 14: Distribuição da classificação da 360º e Aéreo por nível de desempenho.



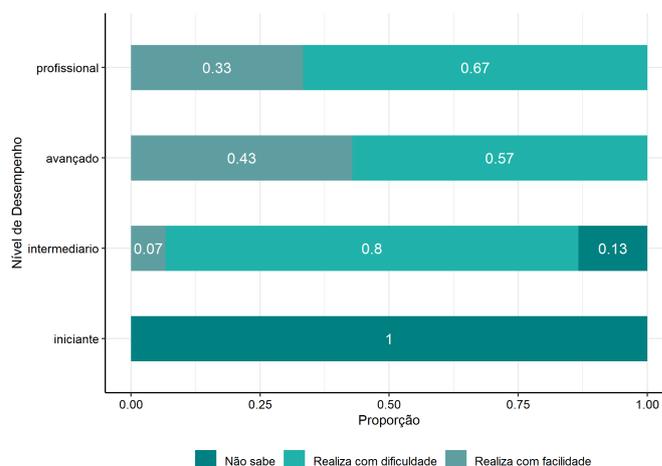
Ademais, apenas um surfista afirmou realizar a manobra Aéreo com facilidade, enquanto os demais ou não sabem ou a executam com dificuldade. Esse cenário se modifica sob o 360º, já que todos os profissionais o executam com dificuldade, 29% dos avançados e 7% da categoria intermediária também. É perceptível a diferença entre os surfistas do sexo masculino e feminino no que tange a execução das manobras: 360º e Aéreo, uma vez que a maioria das mulheres não possuem esse domínio, enquanto, em geral, os homens dominam essas manobras.

Tabela 7: Distribuição da frequência relativa da classificação das manobras 360 e Aéreo por sexo.

| Manobra | Classificação | Sexo | |
|---------|-------------------------|----------|-----------|
| | | Feminino | Masculino |
| 360 | Não Sabe | 0,58 | 0,25 |
| | Realiza com dificuldade | 0,42 | 0,45 |
| | Realiza com facilidade | 0 | 0,3 |
| Aéreo | Não Sabe | 0,5 | 0,5 |
| | Realiza com dificuldade | 0,33 | 0,25 |
| | Realiza com facilidade | 0,17 | 0,7 |

Por fim, relativamente ao Tubo, é perceptível que os surfistas que realizam a manobra, possuem dificuldade na realização. Apenas 13% dos surfistas intermediários afirmaram não ter domínio para executar essa manobra, de acordo com a Figura 15.

Figura 15: Distribuição da classificação do tubo por nível de desempenho.



Também verifica-se que nenhuma mulher possui facilidade nesse movimento, a maioria delas não sabem executá-lo enquanto apenas dois homens não possuem esse domínio e 25% deles afirmaram ter facilidade na manobra tubo.

Tabela 8: Distribuição de frequência da classificação da manobra tubo por sexo.

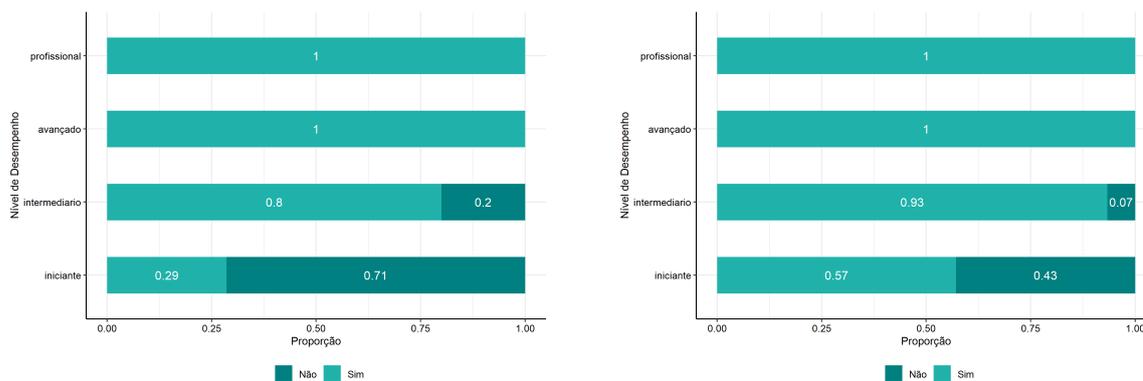
| Sexo | Classificação de dificuldade | | |
|-----------|------------------------------|--------------------|-------------------|
| | Não sabe | Possui dificuldade | Possui facilidade |
| Masculino | 0,1 | 0,65 | 0,25 |
| Feminino | 0,58 | 0,42 | 0 |

A cerca do conhecimento teórico relacionado ao Surf, os surfistas foram questionados à respeito do domínio da leitura da onda e de gráficos de previsão de onda, indicados respectivamente pelas Figuras 16a, 16b e 16c.

A partir das Figuras 16a e 16b, nota-se que a proporção dos participantes que não têm competência teórica na leitura da onda e nos gráficos de previsão de onda decresce à medida que o nível do desempenho aumenta.

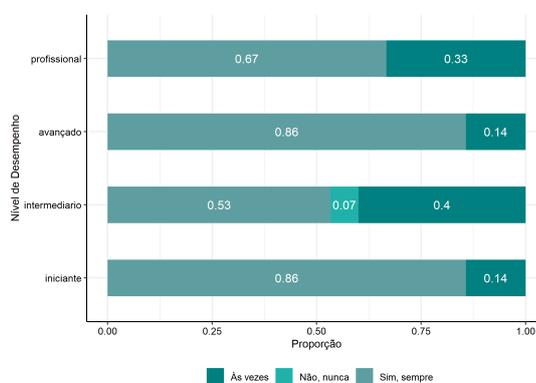
Principalmente o nível iniciante não possui domínio na leitura da onda. Contudo, 20% do grupo intermediário, também se encontra na mesma situação. Esse número decai para 7% em relação aos surfistas intermediários que não sabem fazer a leitura dos gráficos de previsão de onda e para 43% da categoria iniciante.

Figura 16: Domínio teórico por sexo e nível de desempenho.



(a) Leitura de onda.

(b) Leitura de gráficos de previsão de onda.



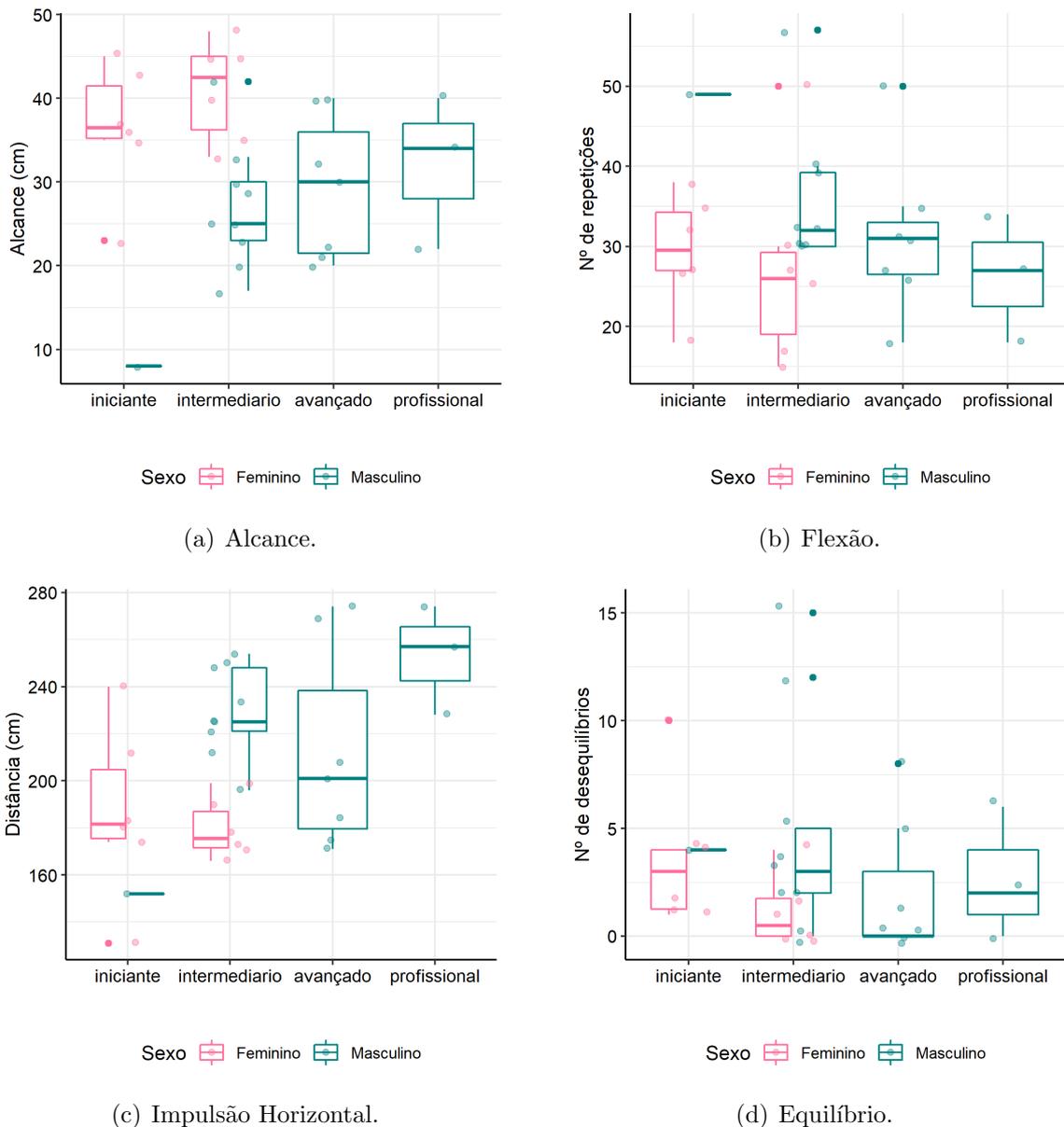
(c) Consulta a gráficos antes da sessão de Surf.

Da Figura 16c, não é evidente nenhum comportamento tendencioso em relação à frequência de consulta à previsão de acordo com o nível do desempenho do surfista. Apesar disso, é perceptível que a grande maioria dos participantes afirmaram sempre consultar a previsão antes de irem surfar.

Para avaliar as características físicas dos surfistas, foram aplicados quatro testes físicos relacionados às principais aptidões inerentes ao Surf. O primeiro e o segundo teste, representados respectivamente pelas Figuras 17a e 17b, tiveram como objetivos, avaliar a flexibilidade do participante através do alcance máximo dos membros superiores sob banco de Wells e avaliar a força dos membros superiores através da repetição máxima do movimento de flexão de braços em um período de um minuto, respectivamente.

O terceiro e o quarto testes, representados respectivamente pelas Figuras 17c e 17d, tiveram como objetivos, analisar a força explosiva dos membros inferiores dos participantes através da realização de um salto horizontal, cuja pontuação foi a distância alcançada e analisar a equilíbrio dos participantes, a pontuação alcançada foi o número de desequilíbrios durante 1 minuto na posição do flamingo, respectivamente.

Figura 17: Resultado dos testes físicos.



Aprende-se da Figura 17a, que as mulheres tiveram um alcance melhor no teste de flexibilidade quando comparado aos homens. Além disso, há uma pequena tendência no aumento do alcance de acordo com o nível do surfista. Todavia, da Figura 17b, ao contrário do que se espera, é perceptível uma queda da distribuição de repetições ao avanço do nível de desempenho do surfista.

O indivíduo que teve o maior alcance sob o banco de Wells, conforme a Figura 17a, chegou a uma distância de 48 centímetros, o menor alcance foi de 8 centímetros, sendo esse um possível outlier no resultado desse teste. Ao contrário do esperado, os surfistas iniciantes e intermediários obtiveram os melhores resultados. Ademais, nota-se que metade dos iniciantes com resultado superior, correspondem a quase 25% dos surfistas

avanzados com melhor alcance.

Relativamente à Figura 17b, apesar da diferença de variação dos resultados por nível de desempenho, observa-se que a maioria dos surfistas se encontram na mesma faixa de performance do teste, em que executaram entre 20 e 40 flexões. Também observa-se similaridades quanto a distribuição da quantidade de repetições dos grupos iniciantes e avanzados. O pior resultado no teste foi 15 flexões, por sua vez, a melhor performance realizou 57 repetições, sendo o número médio de 30 flexões.

A partir do teste representado na Figura 17c, é possível verificar uma tendência geral crescente da proficiência do surfista em relação ao avanço do nível de desempenho. O oposto ocorre ao analisar o comportamento em relação ao gênero do surfista, já que o resultado da categoria feminina decresce do iniciante para o intermediário e o da masculina decai do nível intermediário para o avanzado e aumenta na categoria profissional.

Ainda pela Figura 17c, nota-se que a distância mínima foi de 131 centímetros, realizada por um surfista iniciante, enquanto a máxima foi igual à 275 centímetros, obtida por um surfista avanzado e por um profissional. Percebe-se que a média de alcance foi discrepante entre os níveis de desempenho, assim como os valores obtidos dos três quartis, fato que indica que a distribuição dos resultados nesse teste é distinta segundo o nível de desempenho do surfista.

Da Figura 17d, não observa-se nenhum comportamento tendencioso da quantidade de desequilíbrios pelos níveis de desempenho. Apesar disso, é perceptível alguns resultados que estão distribuídos de maneira distinta em relação à sua categoria, sendo esses, possíveis outliers da distribuição. Apesar da média de desequilíbrios entre os níveis iniciante e intermediário serem muito próximas, o número máximo de desequilíbrios da categoria intermediária ainda representa 50% a mais que o máximo dos iniciantes.

4.2 Análise de Correlação

Nesta seção será avaliado o grau de relacionamento entre as variáveis coletadas através da análise descritiva com cada par de variáveis, também denominada como bi-dimensional. Assim, será aplicado o teste de correlação de Pearson para avaliar a relação entre duas variáveis quantitativas e o teste Qui quadrado para a independência entre duas variáveis qualitativas.

Como o banco de dados contém informações de 30 variáveis, sendo 21 delas classificadas como qualitativas e 9 como quantitativas, nessa parte estudo será abordada de maneira aprofundada somente o par de variáveis com grau de associação ou dependência significativa.

4.2.1 Teste de correlação de Pearson

Sob as hipóteses,

$$H_0 : \rho = 0$$

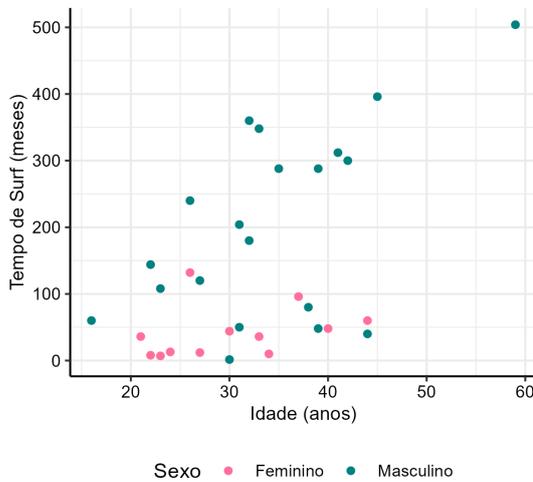
$$H_1 : \rho \neq 0,$$

com um nível de significância $\alpha = 0,10$, quatro pares de variáveis mostraram ser correlacionados entre si, sendo eles: 1º par - Idade x Tempo de Surf, 2º par - Idade Inicial x Alcance no salto horizontal, 3º par - Frequência de prática semanal x Número de flexões e 4º par - Número de desequilíbrios x Alcance no salto horizontal.

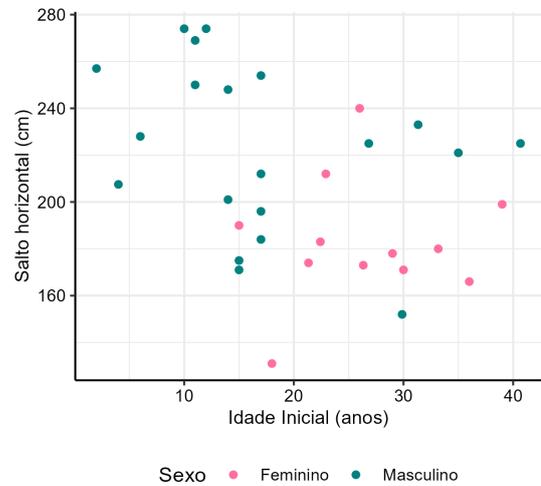
Relativamente a distribuição da idade do surfista pelo tempo em que ele pratica o esporte, avaliado pelo primeiro teste de correlação e demonstrado pela Figura 18a, nota-se que há uma tendência linear crescente. Já o segundo teste, que avaliou a correlação entre a idade inicial e o resultado no salto horizontal, pela Figura 18b, percebe-se que o comportamento é o oposto, nota-se que quanto maior a idade de ingresso no Surf, menor tende a ser o resultado no teste de salto horizontal.

Em relação ao resultado do terceiro teste, sendo este o cruzamento das variáveis número de repetições no teste de flexão e a frequência semanal em que o surfista pratica o Surf, da Figura 18c, é possível verificar que as pessoas que praticam semanalmente mais vezes o esporte, realizaram um número menor de repetições. E, por fim, o quarto teste, pela Figura 18d, percebe-se que à medida que o alcance no teste de impulsão horizontal aumenta, o número de desequilíbrios diminui.

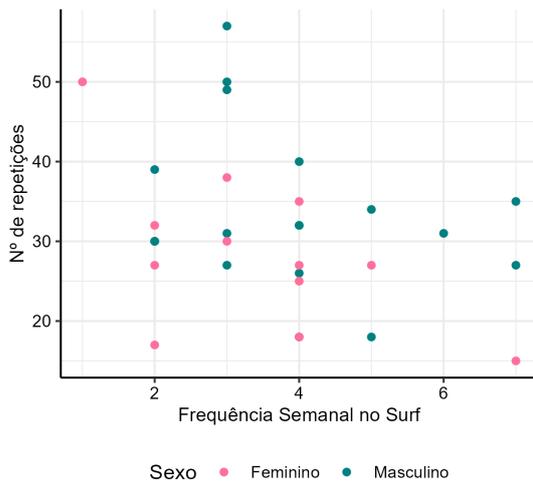
Figura 18: Dispersão entre duas variáveis correlacionadas.



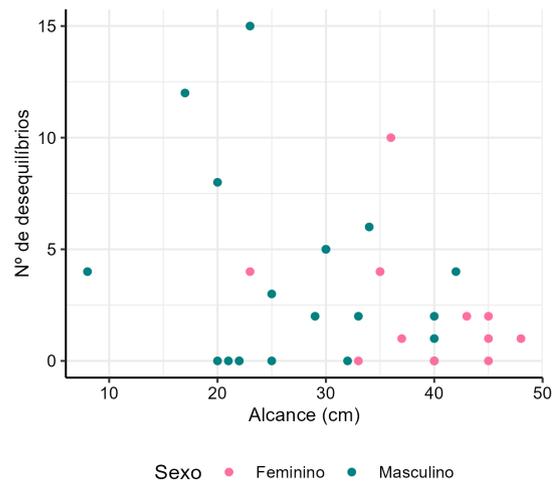
(a) Idade x Tempo de Surf.



(b) Idade Inicial x Alcance no salto horizontal.



(c) Frequência semanal x Nº de Flexões.



(d) Nº de desequilíbrios x Alcance no salto horizontal.

O primeiro teste, cujo a representação gráfica está disposta na Figura 18a, resultou em um valor de p aproximadamente igual à 0,002, sendo este inferior à α , indicando assim a existência de correlação entre as duas variáveis. O coeficiente de correlação de Pearson (ρ) igual à 0,525, sugere uma relação moderada e positiva, isto é, quanto maior a idade do surfista maior o tempo de Surf. Para o grupo de mulheres, o p -valor resultante foi superior ao nível de significância, e por isso, evidencia independência entre as variáveis. Enfim, para os homens, o valor de p igual à 0.012, é possível afirmar um grau de correlação (ρ) = 0,549.

O segundo teste, demonstrado pela Figura 18b, o valor de p obtido foi igual à 0,000036; e, portanto, há evidências para afirmar que existe correlação entre o resultado do teste de impulsão e a idade inicial. O grau de relação é dado pelo coeficiente de Pearson, correspondente à -0,6624, sendo classificado como moderada a forte e negativa. Em outras palavras, quanto menor for a idade inicial no Surf, maior será a distância alcançada no salto horizontal. Novamente analisando separadamente os grupos por sexo, foi observada a inexistência de correlação para o agrupamento de mulheres e dependência (ρ) na proporção de -0,118 para os homens.

Em decorrência ao terceiro teste, representado na Figura 18c, o valor p é dado por 0,07545, evidenciando a dependência entre o número de vezes que os surfistas praticam o esporte e a quantidade de repetições do movimento de flexão de braços. Além disso, como $\rho = -0,33$, essa relação é fraca a moderada e negativa, confirmando a hipótese feita previamente. Para as subamostras, os valores de p iguais à 0,05 e 0,29, para as mulheres e os homens, respectivamente, indicam que há dependência de intensidade -0,56 entre a frequência semanal das mulheres e seu resultado no teste de flexões. Já para os homens, é possível concluir que não há evidências de dependência entre as duas variáveis.

Com a aplicação do teste de correlação para o quarto par de variáveis, cuja sua dispersão está demonstrada na Figura 18d, foi observado o p valor igual à 0,088, sendo este inferior à $\alpha = 0,10$. Então, há evidências para afirmar que essas duas variáveis estão correlacionadas. Além disso, $\rho = -0,305$, confirma a suposição feita, sendo essa associação classificada como fraca a moderada e negativa. Todavia, ao analisar essa relação para as subamostras por sexo, esse teste não retornou valores de p significativos.

4.2.2 Teste de Independência

Tendo em vista que o objetivo do estudo é analisar o desempenho no Surf, e que essa performance é avaliada por níveis, será investigada a independência da variável *nível* em relação às demais variáveis qualitativas. Para isso, será utilizado o teste de independência de Qui quadrado e o teste exato de Fisher.

Além disso, na análise descritiva, foi observado que algumas variáveis se distribuem de maneira heterogênea de acordo com o sexo do surfista. E, por isso, também será avaliada a independência da variável *sexo* com as demais variáveis categóricas, utilizando os mesmos testes.

Sob as hipóteses,

$$H_0 : \text{independência presente vs } H_1 : \text{independência ausente}$$

Com um nível de significância $\alpha = 0,10$, o teste Qui quadrado resultou em treze variáveis que demonstraram associação com o nível de desempenho, e o de Fisher, somente com doze delas.

Antes de tudo, foi analisado o grau de relacionamento entre o nível de desempenho e o sexo do surfista. Da Tabela 9, nota-se que a proporção da distribuição das categorias se distingue de acordo com o sexo.

Tabela 9: Distribuição conjunta do nível de desempenho por sexo.

| Sexo | Nível de desempenho | | | |
|-----------|---------------------|---------------|----------|--------------|
| | Iniciante | Intermediário | Avançado | Profissional |
| Masculino | 0,145 | 0,6 | 1 | 1 |
| Feminino | 0,855 | 0,4 | 0 | 0 |

Também é observado que não há nenhuma mulher nas categorias avançado e profissional, e somente um homem foi classificado como iniciante. O valor p calculado tanto no teste Qui quadrado quanto no teste exato de Fisher foi inferior a 0,1, indicando que existe dependência do nível de desempenho com o sexo do surfista.

Referente ao domínio técnico, os surfistas foram questionados quanto à habilidade em realizar manobras simples e complexas, e foi testando se existe dependência entre nível de desempenho do surfista e a maneira com que eles realizam cada manobra. Primeiramente, a respeito do domínio em executar manobras complexas, sendo elas: batida, rasgada, cutback, floater, 360, aéreo e tubo, foi observada a seguinte distribuição por nível de desempenho do surfista.

Tabela 10: Distribuição da frequência relativa do domínio de execução de manobras complexas por nível de desempenho e valor p do teste de associação.

| Domínio | Nível de desempenho | | | |
|---------|---------------------|---------------|----------|--------------|
| | Iniciante | Intermediário | Avançado | Profissional |
| Sim | 0 | 1 | 1 | 1 |
| Não | 1 | 0 | 0 | 0 |

Nota-se, pela Tabela 10, que nenhum surfista iniciante possui domínio para realizar manobras mais complexas, enquanto os demais grupos afirmaram conseguir executar esse tipo de movimentação. O valor p obtido igual à $0,523^{-6}$, evidencia a relação entre o nível de desempenho do surfista e saber ou não executar manobras complexas. Em conformidade, o teste de Fisher retornou um p-valor de $0,594^{-8}$, sendo este também inferior ao nível de significância $\alpha = 0,1$.

Em seguida foi analisado o domínio do surfista nas manobras indicadas pela Tabela 11, assim como o valor p resultante do teste de associação de como o surfista classifica a manobra e seu respectivo nível de desempenho.

Tabela 11: Distribuição conjunta da classificação de manobras por nível de desempenho e valor p do teste de independência Qui quadrado.

| Manobra | Valor p | Classificação | Nível de desempenho | | | |
|---------|-------------|---------------|---------------------|---------------|----------|--------------|
| | | | Iniciante | Intermediário | Avançado | Profissional |
| Drop | 0,063 | Difícil | 0,57 | 0,40 | 0 | 0 |
| | | Fácil | 0,43 | 0,6 | 1 | 1 |
| Parede | 0,0047 | Difícil | 0,855 | 0,47 | 0 | 0 |
| | | Fácil | 0,155 | 0,53 | 1 | 1 |
| Batida | $3,21^{-6}$ | Difícil | 0 | 0,73 | 0,43 | 0 |
| | | Fácil | 0 | 0,27 | 0,57 | 1 |
| Rasgada | $8,97^{-8}$ | Difícil | 0 | 0,53 | 0,14 | 0 |
| | | Fácil | 0 | 0,47 | 0,86 | 1 |
| Cutback | $1,71^{-6}$ | Não sabe | 0 | 0,07 | 0 | 0 |
| | | Difícil | 0 | 0,8 | 0,285 | 0,33 |
| | | Fácil | 0 | 0,13 | 0,715 | 0,67 |
| Floater | $8,76^{-7}$ | Não sabe | 0 | 0,07 | 0 | 0 |
| | | Difícil | 0 | 0,6 | 0,14 | 0,33 |
| | | Fácil | 0 | 0,33 | 0,86 | 0,67 |
| 360 | $3,06^{-5}$ | Não sabe | 0 | 0,2 | 0,285 | 0 |
| | | Difícil | 0 | 0,73 | 0,43 | 0 |
| | | Fácil | 0 | 0,07 | 0,285 | 1 |
| Aéreo | 0,02 | Não sabe | 0 | 0,27 | 0,285 | 0 |
| | | Difícil | 0 | 0,67 | 0,715 | 1 |
| | | Fácil | 0 | 0,07 | 0 | 0 |
| Tubo | $8,89^{-5}$ | Não sabe | 0 | 0,13 | 0 | 0 |
| | | Difícil | 0 | 0,8 | 0,57 | 0,67 |
| | | Fácil | 0 | 0,07 | 0,43 | 0,33 |

De maneira geral, é possível notar que surfistas mais experientes e de níveis de desempenho mais elevados, tendem a possuir mais domínio nas manobras, classificando-as como fáceis. Os valores p obtidos dos testes de associação Qui quadrado confirmam associação entre o nível de desempenho do surfista e o grau de dificuldade nessas movimentações. Assim, é possível afirmar que o modo com que os surfistas classificam cada manobra tem relação com a sua categoria de desempenho.

Este comportamento também é confirmado pelo teste exato de Fisher, já que quando testado a distribuição do nível de desempenho por cada uma das manobras acima, também retornou um valor-p inferior ao nível de significância.

Tabela 12: Distribuição do domínio teórico por nível de desempenho e valor p do teste de independência Qui quadrado.

| Domínio | Valor p | Classificação | Nível de desempenho | | | |
|------------------|---------|---------------|---------------------|---------------|----------|--------------|
| | | | Iniciante | Intermediário | Avançado | Profissional |
| Leitura da onda | 0,08 | Sim | 0,285 | 0,8 | 1 | 1 |
| | | Não | 0,715 | 0,2 | 0 | 0 |
| Previsão de onda | 0,05 | Sim | 0,57 | 0,93 | 1 | 1 |
| | | Não | 0,43 | 0,07 | 0 | 0 |

Tanto em relação ao domínio na leitura da onda e na leitura de gráficos de previsão de onda, é possível observar na Tabela 12, que a quantidade de surfistas que possuem esses dois domínios aumenta conforme seu nível de desempenho. Sob um nível de significância $\alpha = 0,1$, a suposição de existência de relação entre essas duas variáveis é confirmada pelos valores de p obtidos pelo teste Qui quadrado. Apesar disso, somente a relação das variáveis domínio na leitura de onda e nível de desempenho foi significativa pelo teste exato de Fisher.

Em consequência da associação do nível de desempenho com o sexo do surfista, foi analisado também a existência ou não de dependência da variável *sexo* com as demais variáveis categóricas do estudo. Tanto o teste Qui quadrado quanto o exato de Fisher sinalizou significância de associação do sexo com doze das variáveis em análise.

Tabela 13: Distribuição conjunta do domínio de execução de manobras complexas por sexo.

| Domínio | Sexo | |
|---------|----------|-----------|
| | Feminino | Masculino |
| Não | 0,5 | 0,05 |
| Sim | 0,5 | 0,95 |

Da Tabela 13, nota-se uma discrepância entre a distribuição do domínio de executar manobras mais complexas em relação ao sexo do surfista, já que a maioria dos homens possuem, enquanto as mulheres se dividem entre as que sabem e as que não sabem realizar movimentos complexas. Sendo essa hipótese confirmada a partir da aplicação de ambos os testes estatísticos, que retornaram valores de p inferiores à 0,1.

A Tabela 14 contém o valor p calculado a partir do teste de independência Qui quadrado e exato de Fisher, para aferir o grau de associação entre o sexo do surfista e a classificação de cada manobra, assim como, a distribuição conjunta entre ambas variáveis. Somente na manobra aéreo não foi verificado dependência significativa com o sexo do surfista.

Tabela 14: Distribuição conjunta da classificação de manobras por nível de desempenho e valores p do teste de associação Qui quadrado e de Fisher.

| Manobra | χ^2 | Fisher | Classificação | Sexo | |
|---------|----------|--------|---------------|----------|-----------|
| | | | | Feminino | Masculino |
| Drop | 0,03 | 0,018 | Difícil | 0,58 | 0,15 |
| | | | Fácil | 0,42 | 0,85 |
| Parede | 0,05 | 0,02 | Difícil | 0,67 | 0,25 |
| | | | Fácil | 0,33 | 0,75 |
| Batida | 0,0015 | 0,0008 | Não Sabe | 0,5 | 0,05 |
| | | | Difícil | 0,5 | 0,45 |
| | | | Fácil | 0 | 0,5 |
| Rasgada | 0,0033 | 0,0019 | Não Sabe | 0,5 | 0,05 |
| | | | Difícil | 0,33 | 0,25 |
| | | | Fácil | 0,17 | 0,7 |
| Cutback | 0,005 | 0,004 | Não Sabe | 0,5 | 0,1 |
| | | | Difícil | 0,5 | 0,45 |
| | | | Fácil | 0 | 0,45 |
| Floater | 0,0009 | 0,0002 | Não Sabe | 0,5 | 0,1 |
| | | | Difícil | 0,5 | 0,25 |
| | | | Fácil | 0 | 0,65 |
| 360 | 0,0539 | 0,07 | Não Sabe | 0,58 | 0,25 |
| | | | Difícil | 0,42 | 0,45 |
| | | | Fácil | 0 | 0,3 |
| Tubo | 0,06 | 0,01 | Não Sabe | 0,58 | 0,01 |
| | | | Difícil | 0,42 | 0,65 |
| | | | Fácil | 0 | 0,25 |

O comportamento da variável de classificação das manobras é distinta conforme o sexo do surfista. Geralmente, os surfistas do sexo masculino tendem a afirmar facilidade nas manobras, enquanto as mulheres, em sua maioria, ou não possui domínio de execução ou as classificam como difíceis. Ainda, tanto o teste Qui quadrado quanto o exato de Fisher corroboram com a suposição de associação entre a classificação de todas as manobras expostas na Tabela 14 e o sexo do surfista. Essa relação é observada principalmente nas manobras: Floater e Rasgada, enquanto o Aéreo não indicou significância na relação com o sexo do surfista.

Tabela 15: Distribuição conjunta do domínio teórico por sexo e valores p do teste de associação Qui quadrado e de Fisher.

| Domínio | χ^2 | Fisher | Classificação | Sexo | |
|------------------|----------|--------|---------------|----------|-----------|
| | | | | Feminino | Masculino |
| Leitura da onda | 0,003 | 0,0015 | Sim | 0,42 | 0,95 |
| | | | Não | 0,58 | 0,5 |
| Previsão de onda | 0,027 | 0,013 | Sim | 0,67 | 1 |
| | | | Não | 0,33 | 0 |

Por fim, da Tabela 15, observa-se que, em sua maioria, os homens possuem domínio na leitura de onda e de gráficos de previsão, enquanto boa parte das mulheres ainda afirmam não possuírem conseguirem realizar a leitura corretamente da onda e 33% delas não sabem ler gráficos de previsão de onda. Assim, os valores de p de ambos os testes confirmam a suposição de associação entre o sexo do surfista e cada uma dessas duas variáveis.

4.3 Regressão Logística

Após ser verificado um comportamento distinto das variáveis explanatórias e da variável resposta em relação ao sexo do surfista, optou-se por dividir a amostra, composta por 32 surfistas: 20 homens e 12 mulheres, e a princípio, modelar separadamente. Para isso, utilizou-se o modelo de regressão logístico dicotômico em ambos os casos.

O nível de desempenho da primeira subamostra, composta apenas por homens, foi parametrizada como “0” para os níveis iniciante e intermediário, e como “1” para os níveis avançado e profissional, já a variável resposta referente ao grupo de mulheres foi parametrizada como “0” para o nível iniciante e como “1” intermediário.

Em seguida, foi analisada a amostra completa, sendo inicialmente construído o modelo de regressão logístico binário, cuja a parametrização da variável resposta foi idêntica ao caso da subamostra de surfistas do sexo masculino. E, posteriormente, estruturado o modelo logístico multinomial ordinal para as amostras divididas por sexo e para a amostra completa, parametrizado como “0” para o nível iniciante, “1” para intermediário e “2” para os níveis avançado e profissional.

A construção dos modelos logísticos e as análises de desempenho destes foi realizada utilizando o Software R. A seleção das variáveis explicativas foi feita através da análise de significância de cada uma no modelo, sem a utilização de algoritmos automáticos, já que este retornou modelos com variáveis não significativas. Para isso, utilizou-se um nível de significância de 10%, caso o valor-p do parâmetro adicionado no modelo fosse superior à este, o parâmetro era retirado, até que fossem obtidos modelos compostos apenas por variáveis significantes.

A análise da qualidade do ajuste do modelo foi feita com o auxílio dos testes da Razão de Verossimilhança (RV) e de Wald, comparando o modelo completo ao ajustado. Como o estudo abrangeu informações de 21 variáveis, o modelo completo foi construído somente com parâmetros que indicaram serem significativos no modelo simples, isto é, com apenas uma variável explicativa.

Nos casos em que o modelo de regressão logística tradicional apresentou problemas de não convergência ou de separação, foi utilizado dois ajustes capazes de reverter tais adversidades, sendo a escolha entre eles feita através do menor critério de informação de Akaike (AIC). O primeiro foi proposto por Firth (1993) e implementado no R pelo pacote *logistif*, que realiza uma penalização na estimação de máxima verossimilhança. E o segundo, sugerido por Gelman et al. (1993), realizado com o auxílio do pacote *arm*, em

que é aplicado sob os coeficientes estimados uma distribuição priori de Cauchy.

4.3.1 Modelos logísticos binários

Relativamente ao agrupamento composto somente por surfistas homens foram obtidos um modelo com dois parâmetros e dois modelos com três parâmetros. Pelo método logístico tradicional, foi verificado o problema de não convergência para os três modelos, e assim, para os dois primeiros, a abordagem bayesiana foi utilizada e para o terceiro, foi feito o ajuste de penalização no estimador de máxima verossimilhança. Os modelos selecionados e o saturado foram identificados abaixo e compostos pelas variáveis explicativas descritas.

- **Modelo 1:** tempo de prática de Surf (em meses) e frequência de prática semanal do esporte;
- **Modelo 2:** classificação de correr a parede da onda, resultado do teste de equilíbrio e frequência semanal de Surf;
- **Modelo 3:** classificação do drop, resultado do teste de equilíbrio e duração média da sessão de Surf;
- **Modelo completo:** tempo de prática de Surf, frequência semanal de Surf, idade inicial, classificação dos movimentos: Drop, correr a parede da onda e Cutback e resultado no teste de equilíbrio.

A fim de comparar os modelos obtidos para esse grupo, utilizou-se o critério de Informação de Akaike corrigido (AICc). Nesse caso, por esse critério penalizar modelos com muitos parâmetros, o Modelo 1 foi indicado por possuir uma maior adequabilidade.

Tabela 16: Medida de desempenho dos modelos binários para a amostra de homens.

| Modelo | AICc |
|-----------------|-------|
| Modelo 1 | 13,38 |
| Modelo 2 | 19,43 |
| Modelo 3 | 21,97 |
| Modelo Completo | 34,51 |

A Tabela 17 mostra as estimativas dos parâmetros, respectivos erros padrões e valores-p para o Modelo 1 ajustado.

Tabela 17: Estimativas dos parâmetros do Modelo 1.

| Variável | Parâmetros | Estimativa | Erro padrão | valor-p |
|-----------------|------------|------------|-------------|---------|
| Intercepto | α | -7,752 | 3,053 | 0,011 |
| Tempo Surf | β_1 | 0,016 | 0,006 | 0,013 |
| Frequência Surf | β_2 | 1,158 | 0,6 | 0,053 |

A equação do modelo, em termos dos logitos, é dada por:

$$\log\left(\frac{\pi(x_i)}{1-\pi(x_i)}\right) = -7,752 + 0,016X_1 + 1,158X_2$$

em que X_1 representa o número de meses que o surfista pratica o esporte; X_2 a quantidade de vezes que ele surfa por semana.

A interpretação dos parâmetros estimados é feita pela chance (odds) e expressa a seguir:

- fixado a frequência de surf, para cada acréscimo na variável de tempo de surf, a chance do surfista pertencer ao nível de desempenho parametrizado como “1” é aumentada em $\exp(\beta_1) = 1,016$ vezes;
- fixado o tempo de surf, a chance do surfista ser do nível avançado/profissional aumenta em $\exp(\beta_2) = 3,18$ vezes para cada unidade que é aumentada na frequência semanal de surf.

Assim, pode-se considerar dois surfistas, o primeiro (id_1) com tempo de prática igual à 12 meses e frequência de 5 vezes por semana, e o segundo surfista (id_2) que iniciou no esporte há 1 mês e pratica 2 vezes por semana. A comparação entre esses dois indivíduos é realizada pela razão de chances (odds ratio).

$$OR(id_1, id_2) = \frac{\text{odds}(id_1)}{\text{odds}(id_2)} = e^{\beta(id_1-id_2)} = e^{0,016(12-1)+1,158(5-2)} = 38,74$$

Dessa maneira, infere-se que a chance do primeiro surfista pertencer ao nível de desempenho 1 é 38,74 vezes a chance do segundo ser dessa mesma categoria. Ainda, obteve-se os fatores de inflação da variância dos parâmetros ajustados no modelo, sendo iguais à 1,083 para ambos, indicando ausência de multicolinearidade no ajuste proposto.

A comparação do modelo ajustado com o completo, através dos testes indicados, resultou nos valores das estatísticas, graus de liberdade e valores-p dispostos na Tabela 18. Esses indicam que não há perda significativa ao optar-se pelo Modelo 1, que é considerado

mais parcimonioso em contraste ao modelo completo. Adicionalmente a esse diagnóstico, o comportamento dos resíduos do modelo selecionado, representado na Figura 19, evidencia que este está bem ajustado.

Tabela 18: Testes de qualidade do ajuste - Modelo 1 versus Modelo completo.

| Teste | Estatística | Graus de liberdade | Valor-p |
|-------|-------------|--------------------|---------|
| RV | 3,41 | 6 | 0,8729 |
| Wald | 0,5011 | 6 | 0,7953 |

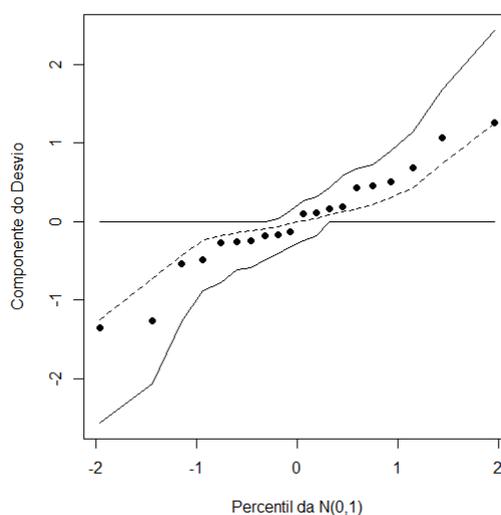


Figura 19: Gráfico quantil-quantil dos resíduos do Modelo 1.

Por fim, através das parâmetros do Modelo, estimou-se o nível de desempenho esperado de acordo com os dados observados, que, na Tabela 19, foi comparado com o nível de desempenho observado.

Tabela 19: Matriz de confusão para o Modelo 1.

| Estimado | Observado | |
|----------|-----------|---|
| | 0 | 1 |
| 0 | 10 | 1 |
| 1 | 0 | 9 |

A matriz de confusão, na Tabela 19, indica que o modelo proposto tem uma acurácia de 95%. Nota-se ainda que o modelo indicou corretamente 90% dos casos do nível parametrizado como 0 e 100% da categoria 1.

Em seguida, realizou-se uma análise similar para a subamostra contendo apenas as surfistas do sexo feminino. Para esse grupo, obteve-se apenas modelos simples, já que ao inserir mais de um parâmetro no modelo somente um deles indicou significância. Dentre os treze significativos, foi observado que os três com menor AIC possuem os mesmos valores do critério e dos parâmetros estimados, e assim, serão analisados conjuntamente. Também foram selecionados os dois modelos seguintes com menor valor de critério.

Nesse caso, também foi necessário realizar ajustes, sendo o escolhido com o menor AIC, devido à problemas de separação da variável resposta, o primeiro modelo foi adequado via penalização do estimador de máxima verossimilhança, e os demais, com a inserção da distribuição priori Cauchy sob os coeficientes estimados. Os modelos selecionados e o saturado estão descritos a seguir.

- **Modelo 4:** formado pela variável explicativa classificação da manobra Rasgada;
- **Modelos 5, 6 e 7:** formados respectivamente pelas variáveis explicativas classificação da manobra Batida, Cutback e Floater;
- **Modelo 8:** formado pela variável explicativa classificação da manobra Cavada;
- **Modelo completo:** formado pelas variáveis explicativas idade, tempo de Surf, domínio de manobras complexas, classificação das manobras: Cavada, Batida, Rasgada, Cutback, Floater, 360°, Aéreo e Tubo.

A fim de comparar os modelos ajustados para a subamostra de mulheres pelo AICc, a Tabela 20 evidencia que os Modelos 5, 6 e 7 são mais adequados, visto que estes possuem um AIC inferior aos demais.

Tabela 20: Medida de desempenho dos modelos binários para a amostra de mulheres.

| Modelo | AICc |
|-----------------|-------|
| Modelo 4 | 8,55 |
| Modelo 5, 6 e 7 | 6,3 |
| Modelo 8 | 9,83 |
| Modelo Completo | 28,65 |

Assim, para o modelo indicado por Modelo 5, 6 e 7, foram observadas as seguintes estimativas dos coeficientes, assim como o seus erros padrões e valores-p.

Tabela 21: Estimativa do parâmetro dos Modelo 5, 6 e 7.

| Variável | Parâmetros | Estimativa | Erro padrão | valor-p |
|------------|------------|------------|-------------|---------|
| Intercepto | α | -3,184 | 1,936 | 0,1 |
| Manobra | β_1 | 6,368 | 0,006 | 0,013 |

A equação do modelo, em termos dos logitos, é dada por:

$$\log\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = -3,184 + 6,368X_1$$

em que $X_1 = 1$ se a surfista classificou a manobra como difícil e $X_1 = 0$ se ela não possui o domínio de execução do movimento.

Assim, a chance da surfista que classificou cada uma das manobras: Batida, Cutback e Floater, como difícil pertencer ao nível intermediário é $\exp(\beta_1) = 583$ vezes a chance da surfista que não sabe executar cada uma dessas manobras.

Ao comparar o modelo ajustado com o completo a fim de verificar se há perdas ao optar por um modelo com menos parâmetros, o teste da Razão de Verossimilhança resultou a favor do modelo ajustado. Assim, pode-se dizer que este está bem adequado. Tal afirmação é confirmada pela Figura 20.

Tabela 22: Testes de qualidade do ajuste - Modelo 5, 6 e 7 versus Modelo completo.

| Teste | Estatística | Graus de liberdade | Valor-p |
|-------|-------------|--------------------|---------|
| RV | 0,3235 | 12 | 1 |

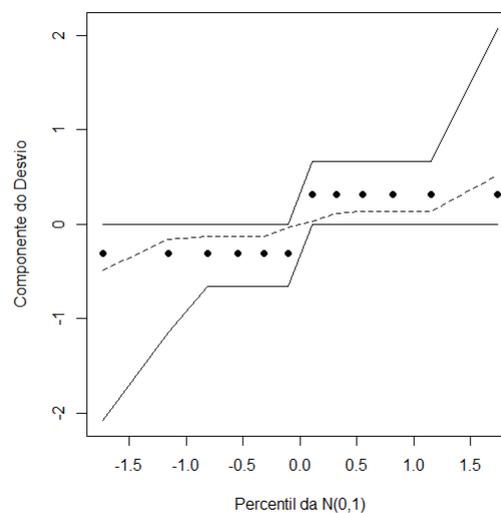


Figura 20: Gráfico quantil-quantil dos resíduos dos Modelos 5, 6 e 7.

Também para os Modelos 5, 6 e 7, os parâmetros estimados foram utilizados para obter a categoria esperada de cada surfista, e, assim, comparar com a que foi observada.

Tabela 23: Matriz de confusão para os Modelos 5, 6 e 7.

| Estimado | Observado | |
|----------|-----------|---|
| | 0 | 1 |
| 0 | 6 | 0 |
| 1 | 0 | 6 |

Apesar da acurácia de 100%, é perceptível que isso ocorre devido ao fato do grupo de mulheres ser composto apenas pelas categorias iniciantes, que afirmaram não possuir domínio nas manobras, e intemediárias, que classificaram as manobras como difícil.

Por fim, para a amostra completa, foram selecionados três modelos significativos também baseados no valor do AIC. Para todos, o melhor ajuste foi dado pela aplicação da distribuição priori Cauchy sob os coeficientes. A composição de cada modelo escolhido, assim como, do modelo irrestrito considerado, é dada a seguir.

- **Modelo 9:** formado pelas variáveis explicativas tempo de Surf e classificação da manobra Drop;
- **Modelo 10:** formado pelas variáveis explicativas tempo de Surf e classificação de correr a parede da onda;
- **Modelo 11:** formado pelas variáveis explicativas tempo de Surf, frequência semanal e idade inicial no esporte;
- **Modelo completo:** formado pelas variáveis explicativas idade, sexo, tempo de Surf, frequência semanal de Surf, domínio da leitura de onda e classificação das manobras: Drop, Parede, Cavada, Batida, Rasgada, Cutback, Floater, 360° e Tubo.

Assim, optou-se por comparar os modelos ajustados através do Critério de Informação de Akaike. Em discordância ao esperado, visto que o AIC tende a penalizar modelos com mais parâmetros, verificou-se o Modelo 11 como o de melhor ajuste.

Tabela 24: Medida de desempenho dos modelos binários para a amostra completa.

| Modelo | AIC |
|-----------------|-------|
| Modelo 9 | 15,1 |
| Modelo 10 | 15,04 |
| Modelo 11 | 14,63 |
| Modelo Completo | 76,38 |

Assim, a Tabela 25 mostra as estimativas dos parâmetros, respectivos erros padrões e valores-p para o Modelo 11 ajustado.

Tabela 25: Estimativas dos parâmetros do Modelo 11.

| Variável | Parâmetros | Estimativa | Erro padrão | valor-p |
|-----------------|------------|------------|-------------|---------|
| Intercepto | α | -3,42 | 3,16 | 0,2783 |
| Tempo Surf | β_1 | 0,014 | 0,006 | 0,0255 |
| Frequência Surf | β_2 | 0,786 | 0,46 | 0,08 |
| Idade Inicial | β_3 | -0,194 | 0,117 | 0,09 |

Sendo o modelo representado, em termos de logito, como:

$$\log \left(\frac{\pi(x_i)}{1 - \pi(x_i)} \right) = -3,42 + 0,014X_1 + 0,786X_2 - 0,194X_3$$

em que X_1 representa o número de meses que o surfista pratica o esporte; X_2 a frequência semanal de prática e X_3 a idade em que o indivíduo começou a surfar.

Quanto à interpretação, tem-se que:

- fixadas a frequência semanal de Surf e a idade inicial, para cada acréscimo na variável de tempo de surf, a chance do surfista pertencer ao nível de desempenho parametrizado como “1” é aumentada em $\exp(\beta_1) = 1,015$ vezes;
- fixados o tempo de surf e a idade inicial, a chance do surfista que pratica o esporte X_2 vezes por semana ser da categoria “1” aumenta em $\exp(\beta_2) = 2,19$ vezes em relação aquele cuja a prática semanal é de $X_2 - 1$;
- fixados o tempo de Surf e a frequência semanal, a chance do surfista que iniciou no esporte com X_3 anos ter desempenho de parâmetro “1” é $\exp(\beta_3) = 0,82$ vezes a chance do que ingressou com idade $X_3 - 1$.

Ainda, para fins comparativos, considere dois surfistas com características distin-

tas. Assim, suponha que o surfista 1 (id_1) tenha iniciado no esporte aos 8 anos, surfe há 10 anos (120 meses) com uma frequência de 4 vezes por semana. Por sua vez, o segundo surfista (id_2) tem 15 anos, surfa há 5 anos (60 meses) e tem frequência semanal igual à 3. Então, a razão das chances de ambos pertencerem ao nível de desempenho 1, é dada por

$$OR(id_1, id_2) = e^{0,014(120-60)+0,786(4-3)-0,194(8-15)} = 19,76$$

A partir disso, pode-se concluir que a chance do surfista 1 ser do nível de desempenho 1 é 19,76 vezes maior que a chance do surfista 2 pertencer à mesma categoria.

Também constatou-se a ausência de multicolinearidade nesse modelo, dado que os fatores de inflação da variância dos seus parâmetros, foram iguais à 1,078 para o tempo de Surf, 1,069 para a frequência de Surf e 1,032 para a idade inicial.

A análise comparativa do modelo ajustado junto ao completo, baseou-se apenas no teste da Razão de Verossimilhança, que não indicou perdas significativas ao se optar pelo Modelo 11, sendo este mais parcimonioso por possuir menos parâmetros.

Tabela 26: Testes de qualidade do ajuste - Modelo 11 versus Modelo completo.

| Teste | Estatística | Graus de liberdade | Valor-p |
|-------|-------------|--------------------|---------|
| RV | 2,768 | 18 | 1 |

Por sua vez, os resíduos do modelo, representados pela Figura 21, apesar de se afastarem da linha dos quantis, ainda é possível confirmar a boa adequação do modelo.

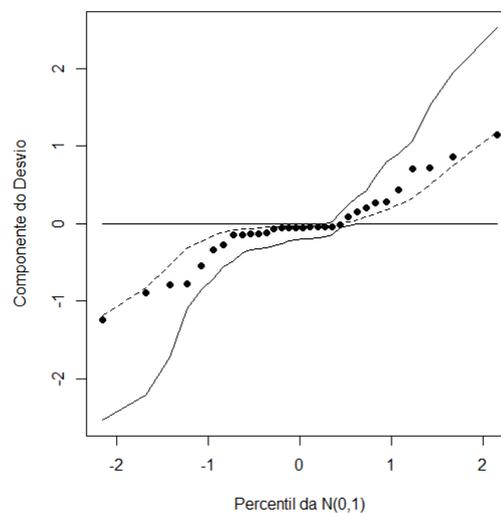


Figura 21: Gráfico quantil-quantil dos resíduos do Modelo 11.

Finalmente, a partir dos dados observados e das estimativas obtidas no Modelo 11, foram gerados níveis de desempenho esperados e comparados com aqueles observados na amostra. Com isso, obteve-se 100% de acurácia.

Tabela 27: Matriz de confusão para o Modelos 11.

| Estimado | Observado | |
|----------|-----------|----|
| | 0 | 1 |
| 0 | 22 | 0 |
| 1 | 0 | 10 |

4.3.2 Modelos logísticos ordinais

No geral, em oposição ao esperado, por se tratar de um estudo com tamanho amostral pequeno, a modelagem utilizando três categorias retornou uma quantidade maior de combinações de parâmetros significativos para compor cada modelo. Além disso, foi observado uma tendência de decréscimo no valor dos critérios de informação AIC, BIC e AICc, a medida em que uma variável era inserida no modelo.

Em contrapartida, principalmente, nos modelos com mais de três variáveis, a verificação do pressuposto de chances proporcionais, retornou uma advertência. A ocorrência dessa situação é dada quando as probabilidades ajustadas para o modelo convergem para 0 ou para 1. Apesar disso, optou-se por também analisar um modelo de 5 parâmetros que retornou esse aviso, por ter tido o menor AIC dentre os demais, e por conter variáveis significativas que trazem muita informação para o estudo como um todo.

Ainda, para alguns dos modelos construídos, observou-se que um alto valor estimado para os parâmetros. A suposição é de que isso tenha ocorrido devido ao pequeno tamanho amostral distribuído por níveis de desempenho de maneira heterogênea.

Nesse caso, não foi possível testar a qualidade dos modelos pelos testes da Razão de Verossimilhança e de Wald, em comparação ao modelo com todos os parâmetros significativos. Já que a hipótese de adequação do modelo de chances proporcionais tende a ser rejeitada em situações de pequenas amostras e muitas variáveis explicativas. Assim, a análise de adequação dos modelos será feita comparando a log verossimilhança e os critérios de informação de cada um deles.

Os modelos selecionados, assim como seus parâmetros, estão descritos a seguir.

- **Modelo 12:** composto pelas variáveis explicativas idade em que o surfista ingressou

no esporte e classificação da manobra Cutback;

- **Modelo 13:** composto pelas variáveis explicativas idade em que o surfista ingressou no esporte e classificação da manobra Floater;
- **Modelo 14:** composto pelas variáveis explicativas idade em que o surfista ingressou no esporte, sexo e classificação da manobra Aéreo;
- **Modelo 15:** composto pelas variáveis explicativas idade do surfista, domínio de manobras complexas, classificação das manobras Drop e 360; e resultado do teste de impulsão horizontal.

Também optou-se por comparar os modelos ajustados através dos critérios de informação e da log-verossimilhança. Com isso, pode-se concluir o Modelo 15 como o de melhor ajuste.

Tabela 28: Medidas de desempenho dos modelos ordinais.

| Modelo | AIC | AICc | BIC | Log-verossimilhança |
|-----------|-------|-------|-------|---------------------|
| Modelo 12 | 31,15 | 32,01 | 38,48 | -12,20 |
| Modelo 13 | 34,4 | 35,25 | 41,72 | -10,57 |
| Modelo 14 | 47,67 | 49,15 | 56,47 | -17,83 |
| Modelo 15 | 24,38 | 27,74 | 36,11 | -4,19 |

Assim, à respeito do Modelo 15, os parâmetros estimados, seus respectivos erros padrões e valores-p calculados são apresentados na Tabela 29.

Tabela 29: Estimativas dos parâmetros do Modelo 15.

| Variável | Parâmetros | Estimativa | Erro padrão | valor-p |
|----------------|------------|------------|-------------|---------|
| Intercepto 1 | α_1 | 33,487 | 18,9868 | 0,077 |
| Intercepto 2 | α_2 | 74,694 | 14,038 | < 0,001 |
| Idade | β_1 | 0,575 | 0,395 | 0,145 |
| Drop (Fácil) | β_2 | 23,6 | 13,7 | 0,085 |
| Manobras (Sim) | β_3 | 43,41 | 6,54 | < 0,001 |
| 360 (Difícil) | β_4 | -5,73 | 4,37 | 0,189 |
| 360 (Fácil) | β_5 | 12,32 | 9,134 | 0,185 |
| Impulsão | β_6 | -0,047 | 0,032 | 0,14 |

Sendo o modelo representado, em termos de logito, como:

$$\log \left[\frac{\hat{P}(Y_i \leq j|x_i)}{1 - \hat{P}(Y_i \leq j|x_i)} \right] = \alpha_j - 0,575X_1 - 23,6X_2 - 43,41X_3 + 5,73X_4 - 12,32X_5 + 0,047X_6$$

em que X_1 representa a idade do surfista; $X_2 = 1$ se o surfista classifica o Drop como difícil e $X_2 = 0$, caso contrário; $X_3 = 1$ se o surfista possui domínio em manobras complexas e $X_3 = 0$, caso contrário; $X_4 = 1$ se o surfista classifica o 360 como difícil e $X_4 = 0$, caso contrário; $X_5 = 1$ se o surfista classifica o 360 como fácil e $X_5 = 0$, caso contrário; e X_6 representa o alcance que o surfista obteve no teste de impulsão horizontal.

Quanto à interpretação, tem-se que:

- fixados o domínio em manobras complexas, a classificação do Drop e do 360 e o resultado no teste de impulsão, a chance do surfista de um surfista com X_1 anos passar do nível de desempenho iniciante para o intermediário ou o avançado aumenta em $\exp(\beta_1) = 1,77$ vezes em relação a um surfista de idade $X_1 - 1$;
- fixados a idade do surfista, o domínio em manobras complexas, a classificação da manobra 360 e o resultado no teste de impulsão, a chance de um surfista que classificou o Drop como fácil de subir de nível é $\exp(\beta_2) = 1,77 \times 10^{10}$ vezes a chance de um indivíduo que classifica essa manobra como difícil;
- fixados a idade do surfista, a classificação das manobras Drop e 360, e o resultado do teste de impulsão horizontal, a chance de um surfista que possui domínio em manobras complexas de avançar de nível é $\exp(\beta_3) = 7,16 \times 10^{18}$ vezes a chance daquele que não sabe executar esse tipo de manobra;
- fixados a idade do surfista, o domínio em manobras complexas, a classificação da manobra Drop e o resultado no teste de impulsão, a chance de um surfista que classificou o 360 como difícil de subir de nível é $\exp(\beta_3) = 0,0032$ vezes a chance de um indivíduo que não possui domínio nessa manobra;
- fixados a idade do surfista, o domínio em manobras complexas, a classificação da manobra Drop e o resultado no teste de impulsão, a chance de um surfista que classificou o 360 como fácil de subir de nível é $\exp(\beta_4) = 2,05 \times 10^6$ vezes a chance de um indivíduo que não possui domínio nessa manobra;
- fixados a Idade, o domínio em manobras complexas, e classificação do Drop e do

360, a chance de um surfista que obteve o resultado X_5 no teste de impulsão subir de nível é $\exp(\beta_5) = 0,953$ vezes a chance daquele que alcançou $X_5 - 1$.

Ainda, pode-se supor dois surfistas, id_1 e id_2 , sendo o primeiro com 30 anos, que classifica o Drop como fácil, possui domínio em manobras complexas, considera o 360 como fácil e obteve o alcance de 274 cm no teste de impulsão, e o segundo de 20 anos, considera o Drop difícil, não possui domínio em manobras complexas e no 360 e teve um alcance de 131 cm. Sendo possível então, compará-los pela razão de chances (OR), de tal maneira que, conforme a equação (5.3.6), o primeiro surfista tem $1,507 \times 10^{15}$ vezes mais a chance de avançar de nível que a segunda surfista.

$$OR(id_1, id_2) = e^{0,575(30-20)+23,6+43,41+12,32-0,047(274-131)} = 1,073 \times 10^{34} \quad (4.3.1)$$

Também constatou-se a presença de multicolinearidade nesse modelo. Apesar disso, como os coeficientes estimados demonstraram significância, a multicolinearidade não parece inflacionar a variância das estimativas.

Tabela 30: Matriz de confusão para o Modelos 15.

| Estimado | Observado | | |
|----------|-----------|----|---|
| | 0 | 1 | 2 |
| 0 | 7 | 0 | 0 |
| 1 | 0 | 14 | 1 |
| 2 | 0 | 1 | 9 |

Por fim, para o Modelo 15, foi utilizado o modelo ajustado para indicar o nível de desempenho esperado do surfista, baseado nas informações da amostra. E, assim, foi comparado o nível esperado com observado nos dados, através da matriz de confusão representada pela Tabela 30, que evidenciou uma acuarica de 93,75%.

5 Conclusão

Durante todo o estudo foi possível perceber que os surfistas homens e mulheres se comportam de maneira distinta, não somente no nível de desempenho em que foram classificados, mas também na experiência, habilidades físicas e domínio técnico. Sendo isso fruto da marginalização que as mulheres sofreram, e ainda sofrem, no esporte. Devido à isso, optou-se por ajustar um modelo de regressão binária para cada grupo, e posteriormente, construir outro para a amostra completa.

Para a amostra de homens, o modelo que demonstrou uma maior adequabilidade teve o critério de informação de Akaike (AIC) igual à 11,88. Este modelo foi composto pelas variáveis explicativas: tempo de prática de Surf (em meses) e frequência de prática semanal. Conclui-se que a chance de um surfista ser classificado como nível “1” aumenta conforme o tempo de prática e a frequência do surfista também aumentam.

Devido a distribuição das surfistas serem idênticas no que tange a classificação das manobras: Batida, Cutback e Floater, o modelo simples construído para estas manobras apresentou estimativas iguais. E, por apresentar o menor AIC dentre todos os modelos contruídos, foi tido como o de melhor ajuste. Apesar da acurácia obtida ser de 100%, é perceptível que isso é consequência do grupo de mulheres ser composto apenas por iniciantes, que não sabem realizar a manobra, e intermediárias, que a classificaram como difícil. Além disso, é possível concluir que a chance da surfista pertencer ao nível parametrizado “1” é maior para aquelas que classificaram a manobra como difícil em relação as que não possuem esse domínio.

O modelo binário com a amostra completa com melhor ajuste, cujo o AIC foi igual à 13,157, é formado pelas variáveis tempo de prática de Surf, frequência semanal e idade de ingresso no esporte. Dessa maneira, é possível concluir que a cada unidade aumentada no tempo de Surf, a chance do surfista pertencer ao nível “1” é elevada conforme o tempo de prática e a frequência do surfista também aumentam e diminuí na medida em que a idade de ingresso do indivíduo aumenta.

Ainda, a modelagem via regressão ordinal, ao contrário do esperado, retornou um maior número de modelos com combinações de variáveis significativas. Apesar disso, devido ao pequeno tamanho amostral, os limites de estimação retornaram valores elevados para os coeficientes, e assim, as razões de chances obtidas também são consideradas altas. Também, em oposição ao esperado, o modelo ordinal que atingiu o menor AIC, foi aquele com o maior número de variáveis explicativas.

Assim, o modelo ordinal selecionado foi composto pelas variáveis explicativas: idade, classificação do Drop, domínio de manobras complexas, classificação do 360 e resultado do teste de impulsão horizontal. O critério de informação de Akaike foi igual à 24,38.

Para este, é possível concluir que a chance de um surfista passar do nível de desempenho iniciante para o intermediário ou o avançado aumenta conforme a sua idade. Essa chance também é elevada se o surfista classifica o Drop como fácil, em relação aos que consideram difícil, se o indivíduo possui domínio em manobras complexas e se ele considera a manobra 360 como fácil ou difícil, em relação aos surfistas que não possui domínio nela. Por último, o alcance obtido no teste de impulsão horizontal, indicou um decréscimo na chance do surfista progredir de nível.

A avaliação dos hábitos nutricionais, ainda que utilizado um questionário criado pela OMS, não retornou resultados significativos, podendo ser consequência da subjetividade das perguntas, uma vez que o indivíduo responde conforme o que se lembra de seus hábitos em relação aos últimos dias.

Os testes físicos, apesar de abrangerem e medirem as principais aptidões que envolvem o esporte, no geral, não retornaram resultados significativos ou esperados, já que o provável seria que os surfistas com os melhores resultados pertencessem à categoria mais elevada. Entretanto, observou-se que a pontuação atingida nessas avaliações, sofrem o impacto de um preparo físico que não está somente ligado a prática do Surf, sendo decorrente de toda a trajetória esportiva do indivíduo, além de poder ter interferência de aspectos genéticos.

Para trabalhos futuros, uma análise confirmatória com uma amostra maior e bem distribuída em relação aos níveis de classificação e ao sexo dos surfistas tenderá a trazer resultados mais precisos, além de poder indicar outras variáveis como significativas para o modelo logístico de predição.

Referências

- ALLISON, P. D. Convergence failures in logistic regression. *SAS Global Forum* University of Pennsylvania, Philadelphia, PA, 2008.
- ARANã, C. *Aprenda a surfar: esporte, lazer, saúde e integração com a natureza*. [S.l.]: Apostila, 2007.
- BRASIL, F. et al. Frequência cardíaca e tempo de movimento durante o surfe recreacional - estudo piloto. *Revista Brasil Ciência e Movimento* 9 (4): 65-75, 2001.
- DANTAS, E. H. M. *Surf: conteúdos para a prática*. [S.l.]: Porto Alegre, RS: Aktmed, 1995.
- FARIAS, S. F. *Surf: conteúdos para a prática*. [S.l.]: Editora UFSC, 2000.
- FERREIRA, A. C. M. O surf e a tv aberta - uma análise midiática dos anos de 2014 e 2015. *Trabalho de Conclusão de Curso (Bacharelado em Educação Física) Universidade Tecnológica Federal do Paraná, Curitiba*, 2018.
- FIRTH, D. "bias reduction of maximum likelihood estimates.". *Biometrika* 80: 27-38, 1993.
- FLEISHMAN, E. A.; ZACCARO, S. J. *Toward a taxonomy of team performance functions*. [S.l.]: In R. W. Swezey E. Salas (Eds.), 1992.
- FONSECA, J. J. S. *Metodologia da pesquisa científica*. [S.l.]: Fortaleza: UEC, 2002. Apostila, 2022.
- GELMAN, A. et al. "a weakly informative default prior distribution for logistic other regression models". *Abb. Appl. Stat.*, 2,4, 1993.
- GELMAN, A. et al. "bayesian generalized linear models and an appropriate default prior.". *Columbia University*, 2008.
- GIL, A. C. *Como Elaborar Projeto de Pesquisa*. [S.l.]: Editora Atlas, 2002.
- GOMES, L. do C. et al. Nas ondas da ciência: Perfil da produção do conhecimento sobre o surfem periódicos brasileiros. *LICERE - Revista Do Programa De Pós-graduação Interdisciplinar Em Estudos Do Lazer*, 20(4), 285-311, 2017.
- GUIMARãES, R. E. Análise do contributo do surf para o estilo de vida de seus praticantes. *Dissertação de Mestrado em Ciências do Desporto, Universidade do Porto*, 2011.
- HOSMER, D. W.; LEMESHOW, S. *Applied logistic regression*. [S.l.]: Wiley New York, 2000.
- HOSMER, D. W.; LEMESHOW, S. *Applied Logistic Regression*. [S.l.: s.n.], 2000.
- JUNIOR, N. N. Exercícios funcionais e proprioceptivos: sua importância para a prevenção de lesões no surf. *Monografia de Graduação em Educação Física. Centro de Desportos. Universidade Federal de Santa Catarina, Florianópolis*, 2010.

- MARTINS, F. O. S. Avaliação das qualidades físicas de crianças através do eurofit. *Dissertação de Mestrado. Rio de Janeiro, RJ. Universidade Castelo Branco*, 1998.
- MENDEZ-VILLANUEVA, A.; BISHOP, D. Physiological aspects of surfboard riding performance. *Sports MED*, 2005.
- MORETTIN, P. A.; BUSSAB, W. de O. *Estatística Básica*. [S.l.]: Editora Saraiva, 2010.
- NETTO, J. S. M. Diagnóstico do nível de aptidão física em amostra de crianças de 7 a 14 anos em Curitiba, utilizando o protocolo do “eurofit”. *Trabalho de Conclusão de Curso (Bacharelado em Educação Física) Universidade Tecnológica Federal do Paraná, Curitiba*, 2005.
- POLLOCK, M.; WILMORE, J. *Exercícios na Saúde e na Doença: Avaliação e Prescrição para Prevenção e Reabilitação*. [S.l.]: Editora Médica e Científica Ltda, 1993.
- SILVA, B.; CLEMENTE, F. M. Physical performance characteristics between male and female youth surfing athletes. *Journal Sports Medicine Physical Fitness*, 2019.
- SOUZA, R. *Boas ondas: surfando com Rico de Souza*. [S.l.]: Rio de Janeiro: Ediouro, 2004.
- ZENI, A. L. Caracterização das capacidades físicas do surf e fundamentos para a prática. *Monografia – Escola de Educação Física e Esportes*, 2002.
- ÁRIAS, M. *Surf Gênese Antropologia do Surf*. [S.l.]: Encarte Revista Alma Surf, n. 8, cap. I, 2002.

Apêndice

A Questionário

Análise de Desempenho no Surf

Formulário destinado à coleta de dados do TCC

***Obrigatório**

1. Nome *

2. Idade *

3. Sexo *

Marcar apenas uma oval.

Feminino

Masculino

4. Consumo de álcool - Sabendo que uma lata de cerveja (350ml) corresponde a *
aproximadamente 2 unidades, um cálice de vinho (90 ml) 1 unidade e uma
dose de destilados (35-50ml) 2 unidades, quantas unidades semanais de álcool
você consome?

Marcar apenas uma oval.

0 a 10 unidades

10 a 21 unidades

Acima de 21 unidades

5. Consumo de açúcares - Sabendo que uma barra de chocolate possui aproximadamente 10 colheres de sopa de açúcar, ou 50 gramas de açúcar, quantas colheres de açúcar você estima que consome diariamente? *

Marcar apenas uma oval.

- Não consome
- 0 a 3 Colheres de Açúcar - (0 a 15g)
- 3 a 5 Colheres de Açúcar - (15 a 25g)
- Acima de 5 Colheres de Açúcar - (Acima de 25g)

6. Consumo de embutidos e industrializados - Com relação a sua dieta, qual a quantidade que consome de alimentos industrializados e embutidos (biscoitos, salgadinhos, comida pronta, alimentos processados...)? *

Marcar apenas uma oval.

| | 0 | 1 | 2 | 3 | 4 | 5 | |
|------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------|
| Nunca come | <input type="radio"/> | Come em Excesso |

7. Surfa a quanto tempo? (Em meses) *

8. Com qual frequência (em dias) você surfa semanalmente? *

Marcar apenas uma oval.

- 1
- 2
- 3
- 4
- 5
- 6
- 7

21. Com qual dificuldade você realiza a manobra 360°? (entende-se por facilidade: realizar a manobra 80% das últimas sessões de surf)

Marcar apenas uma oval.

| | | | | | | | |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | |
| Com muita facilidade | <input type="radio"/> | Com muita dificuldade |

22. Com qual dificuldade você realiza a manobra aéreo? (entende-se por facilidade: realizar a manobra 80% das últimas sessões de surf)

Marcar apenas uma oval.

| | | | | | | | |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | |
| Com muita facilidade | <input type="radio"/> | Com muita dificuldade |

23. Com qual dificuldade você realiza a manobra tubo? (entende-se por facilidade: realizar a manobra 80% das últimas sessões de surf)

Marcar apenas uma oval.

| | | | | | | | |
|----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | |
| Com muita facilidade | <input type="radio"/> | Com muita dificuldade |

Domínio técnico e teórico

24. Possui domínio na leitura da onda? *

Marcar apenas uma oval.

- Sim
 Não

25. Possui domínio na leitura de previsão da onda? *

Marcar apenas uma oval.

Sim

Não

26. Consulta a previsão das ondas antes da sessão de surf? *

Marcar apenas uma oval.

Sim, sempre

Às vezes

Não, nunca

Este conteúdo não foi criado nem aprovado pelo Google.

Google Formulários