



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DEPARTAMENTO DE ESTATÍSTICA



Estimativa da Probabilidade de Vitória em Lutas de MMA

Autor: Daniel Sganzerla
Orientadora: Professora Dra. Suzi Alves Camey

Porto Alegre, 08 de agosto de 2016.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Estimativa da Probabilidade de Vitória em Lutas de MMA

Autor: Daniel Sganzerla

Trabalho de Conclusão de Curso
apresentado para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dra. Suzi Alves Camey
Professor Dr. Álvaro Vigo

Porto Alegre, 08 de agosto de 2016.

Dedico este trabalho aos meus pais e a minha irmã, por tudo.

“Deixe sua bússola moral indicar o caminho através do mundo das ‘Pessoas Grandes’, e não comprometa a essência de quem você é”.

A Sabedoria do Condado

Agradecimentos

Agradeço imensamente aos meus pais, Darci e Salete, a minha irmã Daniela e ao meu cunhado Marcos, por nunca deixarem que me faltasse nada.

Agradeço a Professora Suzi pelos conselhos, pela paciência e por todo o muito tempo que me auxiliou durante o trabalho. Ao Professor Álvaro, pelos atendimentos fora do horário, a Professora Jandyra por ser uma das pessoas mais brilhantes que já conheci e a Bianca Miarka por me incentivar e tornar este trabalho possível, sem me conhecer. Ao amigo e chefe Rodrigo Coster, por me fazer gostar de R e ter a paciência de me ensinar.

Agradeço a minha namorada Sabrina, por me apoiar em tudo. Agradeço aos meus amigos de longa data, por me apoiarem longe de casa. Agradeço aos amigos e colegas de faculdade, especialmente ao Maik e ao time Random Walk FC.

Em especial, meu agradecimento à empresa Borrachas VIPAL e ao Programa Bom Aluno, por fazerem parte da formação do meu caráter, por acreditarem em mim, pelo apoio e suporte por muitos anos.

Agradeço a Deus por me guiar.

Resumo

A análise de lutas de MMA permite observar diversos lutadores competindo mais de uma vez. Dessa forma, os estudos envolvendo a estimação da probabilidade de vitória e do comportamento das variáveis de interesse necessita que se utilize alguma metodologia que considere a correlação existente entre as unidades amostrais. Em acréscimo, cada luta é formada por dois lutadores, de modo que também não exista independência entre ambos.

Uma das formas de se considerar a existência de dependência entre as unidades amostrais é com as Equações de Estimação Generalizadas (GEE). Essa metodologia, proposta por Zeger e Liang (1986) se diferencia dos Modelos Lineares Generalizados por incorporar uma matriz de correlação de trabalho que modela a dependência entre as unidades amostrais. Este trabalho busca, também, comparar os resultados produzidos por diferentes estruturas de matrizes de correlação ao considerar a estimação da probabilidade de vitória das lutas, no contexto explicativo e preditivo.

Foram calculados escores para representar diversas situações de combate. Com a metodologia GEE, verificou-se que nas 1111 lutas analisadas o escore correspondente à efetividade de ataque de golpes se mostrou mais relacionado com a vitória, sendo uma prática a ser considerada nos treinamentos. Neste caso, a consideração da estrutura de dependência que existe quando os lutadores são avaliados mais de uma vez se mostrou mais eficiente do que considerar a luta como um agrupamento, mesmo que a soma da probabilidade de vitória não seja igual a 1. Neste caso, as matrizes apresentaram comportamentos similares.

Ao considerar a estimação de um modelo preditivo, a variável ODDS, um escore da banca de apostas, apresentou um maior percentual de acerto do que os modelos só com os escores. Os modelos com os escores e a ODDS apresentaram percentual de acerto maior que 60%. Ao se calcular um ganho fictício em relação às apostas, os modelos com os escores apresentaram ganhos, enquanto que os modelos com a ODDS resultaram em prejuízo, mostrando que a modelagem com os escores foi mais eficiente para este fim.

Sumário

1) Introdução	7
2) Metodologia	10
2.1) Modelos Lineares Generalizados – GLM.....	10
2.2) Modelo para dados binários.....	12
2.3) Modelo para dados correlacionados	13
2.4) Modelo GEE	13
2.4.1) Matriz de Correlação de trabalho	15
2.5) Escores de Desempenho e Dicionário de Variáveis.....	18
3) Resultados	23
3.1) Modelos Explicativos	25
3.2) Modelos Preditivos	27
4) Discussão	31
Referências Bibliográficas	33

1) Introdução

As artes marciais mistas (MMA – *Mixed Martial Arts*) têm se tornado um dos esportes que mais cresce no mundo na última década [4]. O *Ultimate Fighting Championship* (UFC) é a maior e mais bem-sucedida organização desta movimentada indústria. Para mostrar a magnitude deste evento, é possível utilizar como exemplo o UFC 128, realizado em Curitiba, no dia 14 de maio de 2016, onde 15 mil espectadores presenciaram a pesagem dos lutadores, representando o maior público em uma pesagem oficial do evento até então. No dia da luta, mais de 45 mil pessoas compareceram à Arena da Baixada para contemplarem o evento de Artes Marciais Mistas.

Um aspecto estatístico desses combates é que no decorrer das lutas são contabilizados os totais de chutes, socos, quedas e diversas outras situações e momentos do combate, refletindo a importância que é dada aos números e análises de dados neste evento. O FightMetric (<http://www.fightmetric.com>), provedor oficial de dados do UFC e do banco de dados aqui utilizado, é a companhia responsável pela contabilização e validação destes resultados, utilizando definições restritas, e sendo a referência em diversos estudos [2,4,8].

A análise destes resultados para prover um *feedback* a atletas e treinadores não é uma novidade nos esportes de combate [2], assim como a busca por fatores que contribuam para vitória ou derrota de um lutador. Além disso, também se tem comparações sobre o tempo em movimentos e análises técnico-táticas em relação à vitória ou derrota [2].

O interesse na probabilidade de vitória em esportes também não é uma busca nova. Existem diversos trabalhos que aplicam diferentes metodologias visando à estimação da probabilidade de vitória [3,7]. Em outro aspecto, existem diversas casas de apostas que possuem profissionais para calcular uma variável conhecida como ODDS. Esse indicador é utilizado para representar a

probabilidade de determinado evento ocorrer, vitória ou derrota, e quanto a casa de apostas paga ao apostador caso ele acerte.

Nos casos supracitados, deseja-se estudar a relação entre a variável resposta (vitória ou derrota) e as variáveis independentes, onde são utilizados modelos de regressão. Inicialmente, os modelos se restringiam a variável resposta contínua e com distribuição normal. Em 1972, Nelder e Wedderburn escreveram os Modelos Lineares Generalizados (GLM - *Generalized Linear Models*) de forma unificada, onde a natureza da variável resposta foi estendida (binárias, de contagem, categóricas) e com outras distribuições de probabilidade (pertencentes à família exponencial).

No GLM, uma das suposições é de que as unidades amostrais observadas sejam independentes [1]. No caso das lutas de MMA, isto significa supor que o resultado (vitória/derrota) de um lutador em uma luta não influencia nas próximas lutas. Como cada lutador foi avaliado em diversas lutas, é possível que esta hipótese não esteja correta.

Quando um lutador é avaliado mais de uma vez, espera-se, naturalmente, que suas medidas estejam mais correlacionadas entre si do que com outros lutadores. Ou seja, existe um grupo de medidas para este mesmo lutador. Este fenômeno, de dependência entre as unidades amostrais, pode ser analisado utilizando Modelos Mistos e Equações de Estimação Generalizadas (GEE – *Generalized Estimating Equation*).

No caso das lutas de MMA, além da correlação entre lutas de um mesmo lutador, existe a dependência dentro da mesma luta, uma vez que para cada lutador vitorioso, existe um lutador que é derrotado. Assim, surge a questão de como considerar as estruturas de correlação e qual seria a melhor abordagem.

Em um contexto de estimar a probabilidade de vitória, é possível que se opte por construir um modelo em que se observem os acontecimentos em uma luta e se relacione com o seu resultado de vitória ou de derrota, tendo um modelo explicativo. Por exemplo, o número maior de socos de um lutador deve estar associado com a vitória, mas não se sabe o número de socos antes da luta

ocorrer. Por outro lado, pode-se tentar, com base em variáveis conhecidas à priori da luta, estimar a probabilidade de vitória ou derrota, optando por um modelo preditivo. Ao se considerar a construção de um modelo, pode-se utilizar variáveis que representem o histórico do lutador como variáveis explicativas, usualmente renomeadas como escores [8].

Neste trabalho optou-se pelas Equações de Estimação Generalizadas (GEE) para um desfecho binário, de vitória ou de derrota, para modelar tanto o modelo explicativo quanto o preditivo. O objetivo do trabalho é aplicar esses modelos ao conjunto de resultados de lutas de MMA para verificar qual estratégia de luta está mais associada com a vitória e se há uma predição melhor de resultado da luta do que o uso da ODDS.

2) Metodologia

A estimação da probabilidade de vitória ou derrota para cada luta que será abordada neste trabalho considera a existência da dependência entre as unidades amostrais de duas maneiras distintas. A primeira delas consiste em considerar que cada lutador é avaliado mais de uma vez. A segunda leva em consideração que cada luta é formada por dois lutadores, uma vez que para cada vitória há uma derrota, configurando uma estrutura de correlação.

Assim, o banco de dados está estruturado de maneira que cada luta corresponda a duas linhas, uma para cada lutador. Para considerar que o resultado de uma luta pode ser influenciado tanto pelos escores do lutador como de seu oponente, as variáveis de cada lutador também estão contidas na linha do seu oponente.

Alguns trabalhos, que procuram estimar a probabilidade de vitória no MMA [4,8], optam por selecionar um dos lutadores da determinada luta, calcular a diferença entre cada variável explicativa dos lutadores e utilizar a variável resultante como possível preditora na modelagem, relacionando-a com o resultado do lutador selecionado. Com essa metodologia perde-se informação de um dos lutadores.

Como citado na introdução, há pelo menos duas classes de modelos que lidam com dados correlacionados, os modelos mistos e equações de estimação generalizadas (GEE). Aqui, sem discussão de mérito, optou-se pelo GEE.

Antes de abordar este modelo, se faz útil uma introdução ao Modelo Linear Generalizado (GLM) e das relações entre eles. Após isso, pode se ter um melhor entendimento de determinados aspectos do modelo GEE.

2.1) Modelos Lineares Generalizados – GLM

O modelo linear normal desenvolvido por McCullagh e Nelder [11] considera um vetor \mathbf{y} de n observações independentes, $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, que representa a variável resposta e uma matriz X de p variáveis preditoras. Nesta

definição, supõe-se que \mathbf{y} segue distribuição normal com média $E(\mathbf{y})$, e variância, $V(\mathbf{y})$, definidos como:

$$E(\mathbf{y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \text{ e } V(\mathbf{y}) = \boldsymbol{\sigma}^2 \quad (1)$$

onde $\boldsymbol{\mu}$ é um vetor $n \times 1$ de médias, \mathbf{X} é uma matriz $n \times p$ de variáveis independentes, $\boldsymbol{\beta}$ é um vetor $p \times 1$ de parâmetros e $\boldsymbol{\sigma}^2$ é uma matriz diagonal $n \times n$ de variâncias

O avanço tecnológico ocorrido na década de 70 permitiu o desenvolvimento de diversas técnicas estatísticas [12]. Em 1972, Nelder e Wedderburn escreveram os Modelos Lineares Generalizados de forma unificada, onde a natureza da variável resposta \mathbf{y} foi estendida (binárias, de contagem, categóricas) e com outras distribuições de probabilidade para a variável resposta (pertencentes à família exponencial).

Para tanto, é preciso que se definam três importantes partes que caracterizam o modelo linear generalizado:

1. **Componente Aleatório:** É a especificação de que a variável resposta \mathbf{y} , segue determinada distribuição de probabilidade pertencente à família exponencial e apresenta independência entre as observações.

2. **Componente Sistemático (preditor linear):** Formado pelas variáveis explicativas que entram no modelo de regressão na forma de uma soma linear de seus efeitos.

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} \text{ ou } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \quad (2)$$

3. **Função de Ligação:** É a função que associa o componente aleatório ao componente sistemático do modelo. A função é denotada por $g(\cdot)$ e deve ser real, monótona e diferenciável, tendo o papel de linearizar a relação entre os componentes. Mais detalhes podem ser vistos no livro de McCullagh e Nelder (1989) [11].

2.2) Modelo para dados binários

Para os casos onde a variável resposta é binária, é possível definir como distribuição de probabilidade a distribuição Binomial. Neste caso, existem algumas funções de ligações usuais, que são apresentadas no quadro abaixo, destacando a função canônica logit.

Quadro 1: Funções de Ligação usuais para desfechos binários.

Função	Fórmula
Logit	$g(\pi) = \log\left(\frac{\pi}{1-\pi}\right)$
Probit	$g(\pi) = \Phi^{-1}(\pi)$
Complementar Log Log	$g(\pi) = \log(-\log(1-\pi))$

O modelo logístico é um caso particular do GLM, onde a média, que não é linear, é ligada aos parâmetros do modelo pela função de ligação logit, especificado da seguinte maneira:

$$\text{logit}(\pi_i) = \log\left[\frac{\pi_i(x)}{1-\pi_i(x)}\right] = g(\pi_i) = x_i^t \beta = \eta_i \quad (3)$$

na qual x_i é o vetor $p \times 1$ de covariáveis referente a i -ésima observação; π_i representa a probabilidade da resposta da i -ésima observação ser igual a 1, dado o conjunto de covariáveis ($\pi_i = P[Y_i = 1 | \mathbf{X}_i]$).

Assim, para o caso de estimar a probabilidade de vitória do lutador em cada luta, definem-se os termos do trinômio descrito anteriormente como:

- i) Distribuição da variável resposta: Binomial.
- ii) Variáveis explicativas: Envergadura, Escores de desempenho, ODDS, Lado dominante.
- iii) Função de ligação: logit

Relembrando que uma das pressuposições do modelo é de que as unidades amostrais sejam independentes, tem-se que o modelo GLM não é adequado, já que um lutador foi avaliado em mais de uma luta. Quando o lutador compete mais de uma vez, existe uma estrutura de correlação que deve ser considerada. Ou ainda, pode-se pensar que cada luta é formada por dois lutadores, um ganhador e um perdedor, ou seja, o resultado de um lutador define o resultado do outro.

2.3) Modelo para dados correlacionados

A necessidade e importância de se utilizar uma modelagem que leve em consideração a dependência entre as unidades observadas pode ser verificada em diversos trabalhos [1, 5, 6, 9, 10].

O modelo de GEE, proposto por Zeger e Liang (1986) considera no modelo uma estrutura de correlação de trabalho entre as observações, obtendo-se estimativas consistentes e não viciadas para os parâmetros do modelo que considera o grupo de medidas correlacionadas.

Neste trabalho será considerada a modelagem através do GEE, onde o grupo de medidas se refere ao conjunto de lutas em que o mesmo lutador foi avaliado. Também será utilizado o GEE para se correlacionar os indivíduos dentro de uma mesma luta, configurando-o como um grupo formado por duas medidas, e comparando estas duas estruturas de dados correlacionados.

2.4) Modelo GEE

De maneira simplificada, o GEE é uma extensão do GLM com o acréscimo de uma matriz de correlação de trabalho na variância da variável resposta [1,9,10]. No GLM considera-se que as unidades amostrais são independentes. No GEE, é possível se pensar que cada unidade amostral faz parte de um grupo de unidades amostrais que não são independentes.

Suponha que existam n grupos de unidades amostrais semelhantes, onde y_{ij} é a variável resposta de interesse para a j -ésima unidade amostral do i -ésimo grupo e \mathbf{X}_{ij} sendo um vetor de $p \times 1$ de covariáveis desta determinada unidade

amostral, onde $i = 1, \dots, n$ e $j = 1, \dots, m_i$, sendo m_i o tamanho do grupo i . Neste caso, os n grupos podem ser possuir tamanhos distintos.

Percebe-se que para o i -ésimo grupo, existe um vetor $m_i \times 1$ de respostas, $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{im_i})'$ e uma matriz de covariáveis $m_i \times p$, $\mathbf{X}_i = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{im_i})'$.

O modelo GEE supõe que:

a) A relação entre a média da variável resposta $\boldsymbol{\mu}_i$, e as variáveis explicativas pode ser expressa sob forma linear através da função de ligação definida, de forma que:

$$g(\boldsymbol{\mu}_i) = \mathbf{X}_i \boldsymbol{\beta} \quad (4)$$

b) A variância da variável resposta é uma função conhecida da média, da matriz de correlação de trabalho e do parâmetro de escala da respectiva distribuição. O quadro 2 traz a expressão da variância no caso GEE e no GLM.

Quadro 2: Relação entre a média e variância nos modelos GLM e GEE.

	GLM	GEE
Relação entre Média e a Variância	$V_i = \frac{f(\boldsymbol{\mu}_i)}{\phi}$	$V_i = \frac{A_i^{\frac{1}{2}} R_i(\alpha) A_i^{\frac{1}{2}}}{\phi}$

Neste quadro, é possível ver que no GLM, a variância é uma razão entre uma função da média, $f(\boldsymbol{\mu}_i)$ e do o parâmetro de escala ϕ . No GEE, o numerador considera A_i , que é uma matriz diagonal $m_i \times m_i$ tendo $f(\boldsymbol{\mu}_i)$ como elementos da diagonal principal e $R_i(\alpha)$, que é a matriz de correlação de trabalho correspondente ao i -ésimo grupo de observações correlacionadas de dimensão $m_i \times m_i$.

Liang e Zeger incorporaram uma matriz de correlação de trabalho no termo de variância, transformando V_i em uma matriz de covariância para o grupo i .

Quando m_i for igual a 1, considerando um indivíduo, o estimador do GEE equivale ao do GLM.

A estimativa dos parâmetros de regressão, definidos por Liang e Zeger [9], é dada pela solução do sistema de equações diferenciais quasi-escore dado por:

$$U_k(\boldsymbol{\beta}) = \sum_{i=1}^n D_i V_i^{-1} S_i = 0, \quad k = 1, \dots, p \quad (5)$$

onde, $D_i = \frac{\partial u_i}{\partial \beta_k}$ e $S_i = (y_i - u_i)$.

Para obter $\hat{\boldsymbol{\beta}}$, Liang e Zeger [9] descrevem um processo iterativo baseado no escore de Fisher [1] que é detalhado no trabalho original.

Considerando as duas formas de agrupamento do caso das lutas, tem-se que no caso em que cada indivíduo é um grupo e suas medidas são formadas pelas vezes em que ele é avaliado, existe a pressuposição de que os demais agrupamentos (todos os outros indivíduos) formados sejam independentes. Ao se considerar essa dependência existente do lutador, perde-se a indexação de que uma luta é formada por dois lutadores, implicando que a soma das probabilidades estimadas não resulte em 1. Nesse caso, o lutador vitorioso considerado pelos modelos estimados será o que apresentar maior probabilidade estimada da referida luta.

Por outro lado, quando se assume que cada luta é um grupo, supõe-se que os resultados das lutas sejam independentes. Neste caso, perde-se a informação de que um lutador é avaliado mais de uma vez.

2.4.1) Matriz de Correlação de trabalho

Uma decisão ao lidar com o GEE é qual matriz de correlação de trabalho ($R_i(\alpha)$) é adequada ao problema de pesquisa. A especificação correta aumenta a eficiência dos parâmetros do modelo de regressão [5]. Apesar disso, Liang e

Zeger [9] afirmam que o modelo é robusto, uma vez essas estimativas permanecem consistentes.

A matriz de correlação de trabalho $R_i(\alpha)$ representa a correlação entre os indivíduos do mesmo grupo e possui dimensão de acordo com o tamanho do grupo. Essa matriz é definida no intervalo de medida da correlação, entre -1 e 1 e pode possuir diferentes estruturas, onde se destacam as matrizes definidas no quadro 3:

Quadro 3: Matrizes de Correlação de trabalho.

Estrutura	Definição	Exemplo (m=3) $R(\alpha) =$	Número de Parâmetros
Independente	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1, se j = k \\ 0, se j \neq k \end{cases}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	0
Permutável	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1, se j = k \\ \alpha, se j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha & \alpha \\ \alpha & 1 & \alpha \\ \alpha & \alpha & 1 \end{pmatrix}$	1
AR(1)	$Corr(Y_{ij}, Y_{ij+t}) = \{\alpha^t, \quad t=0,1,2\}$	$\begin{pmatrix} 1 & \alpha & \alpha^2 \\ \alpha & 1 & \alpha \\ \alpha^2 & \alpha & 1 \end{pmatrix}$	1
M-dependente	$Corr(Y_{ij}, Y_{ij+t}) = \begin{cases} 1, se t = 0 \\ \alpha_t, se t = 1,2, \dots, M \\ 0, se t > M \end{cases}$ sendo M o número máximo de dependência entre as observações.	$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ \alpha_1 & 1 & \alpha_1 \\ \alpha_2 & \alpha_1 & 1 \end{pmatrix}$ M=2	M
Não estruturada	$Corr(Y_{ij}, Y_{ik}) = \begin{cases} 1, se j = k \\ \alpha_{jk}, se j \neq k \end{cases}$	$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 \\ \alpha_1 & 1 & \alpha_3 \\ \alpha_2 & \alpha_3 & 1 \end{pmatrix}$	$\frac{m(m-1)}{2}$ Sendo m o número de medidas.

No caso de observações independentes dentro do mesmo grupo, a matriz utilizada é a independente. Quando se supõe que as medidas possuem uma correlação de primeira ordem, com uma estrutura de temporalidade, se utiliza a matriz AR(1). Em um caso geral, onde se supõe que as medidas do grupo possuem correlações distintas entre elas, utiliza-se a matriz não-estruturada. A

matriz permutável é utilizada quando se supõe que a correlação entre as medidas do grupo é a mesma entre cada medida.

Para mostrar o efeito da escolha da matriz de trabalho foram utilizadas as quatro estruturas de correlação aqui apresentadas. Os modelos serão comparados de acordo com os critérios QIC e QIC_c , propostos por Pan[14] que indicam um melhor modelo para valores menores. Além disso, os modelos serão comparados segundo o percentual de acerto de cada um, sendo que para verificar qual lutador é considerado vitorioso para cada modelo, compara-se a probabilidade estimada de cada lutador na luta e o que apresentar maior valor é o definido.

2.5) Escores de Desempenho e Dicionário de Variáveis

A criação de escores de desempenho para os lutadores de MMA pode ser realizada de diversas maneiras, uma vez que cada analista pode julgar determinado componente da luta como fator importante a ser levado em consideração. Outros trabalhos também utilizam, através de seus critérios, escores para descrever o comportamento dos lutadores [8].

Neste trabalho, inicialmente, as ações do lutador são computadas através de simples contagem dentro de cada *round*. A partir disso foram calculados os totais de cada lutador em cada luta para cada variável. O lutador teve seu histórico de lutas avaliado de janeiro de 2012 até setembro de 2014, tendo como fonte de dados o FightMetric.

Os escores também foram computados para representar situações distintas dentro da luta que são descritas abaixo:

Striking ou Trocação: Representa a troca de chutes, socos, cotoveladas e joelhadas entre os lutadores, e se inicia, na maioria das vezes, com os lutadores em pé, mas que pode ocorrer na luta de chão. Pode-se destacar que esta é uma das situações que empolga o público e representa alguns estilos como boxe, *karatê*, *taekowndo*, *kickboxing* e *muaythai*.

Neste caso, as ações descritas acima foram agrupadas e nomeadas como “Golpes” e contabiliza três escores para cada lutador em cada luta.

$$\text{Efetividade de Golpe} = \frac{\text{Golpes Efetivos}}{\text{Golpes Desferidos}} \quad (6)$$

Representa a efetividade no ataque, sendo contabilizada pelos golpes em que o lutador acerta o oponente (Golpes Efetivos) e pelo total de tentativas (Golpes Desferidos).

$$\text{Defesa de Golpe} = \frac{\text{Golpes Defendidos}}{\text{Total de Golpes Recebidos}} \quad (7)$$

Representa a efetividade na defesa dos golpes, sendo contabilizada pelos golpes em que o lutador não é acertado (Golpes Defendidos) e pelo total de tentativas do seu oponente (Total de Golpes Recebidos), incluindo as que o oponente acertou e errou.

Dentro de uma luta, a estimativa de ataque de um lutador é complementar à defesa de seu adversário neste escore. Isso não acontece quando é considerado o histórico de lutas prévias.

Quedas: Quantifica o ato de derrubar o adversário por meio de golpes na luta agarrada ou arremessando o oponente. Cabe a ressalva de que esta queda não é a mesma de quando um adversário cai em consequência de um golpe de trocação, como um soco. Este escore representa o *wrestling*, luta olímpica, sumô, judô, *aikido*, luta olímpica, sambô, jiu-jitsu, dentre outros, e é quantificado pelo número de quedas que o lutador conseguiu aplicar, pelo número de tentativas.

Assim como existem diversas maneiras de se aplicar a queda, existem diversas maneiras de se esquivar ou de se defender. Neste caso, não foi feita nenhuma discriminação entre elas. O escore de defesa de quedas foi contabilizado através do número de quedas que o oponente não conseguiu concluir dividido pelo número de tentativas.

$$\text{Efetividade de Queda} = \frac{\text{Quedas Efetivas}}{\text{Tentativas de Quedas}} \quad (8)$$

$$\text{Defesa de Queda} = \frac{\text{Quedas Defendidas}}{\text{Total de Quedas Recebidas}} \quad (9)$$

Uma vez que este escore apresenta grande quantidade de lutas onde o lutador não tenta a queda, foi realizada a seguinte categorização:

$$\text{Efetividade de Queda} = \begin{cases} 0, & \text{Sem tentativas ou sem quedas efetuadas} \\ 1, & \text{Pelo menos uma queda efetuada} \end{cases} \quad (10)$$

$$\text{Defesa de Queda} = \begin{cases} 0, \text{ Sem tentativa- Oponente não tentou nenhuma queda} \\ 1, \text{ Baixa Efetividade-Escore menor ou igual a } 0,5 \\ 2, \text{ Alta Efetividade-Escore maior que } 0,5 \end{cases} \quad (11)$$

Atividade no Solo: Quantifica as ações do lutador na luta de solo. Também representa as artes marciais do grupo de quedas, onde um dos objetivos é obter a vitória através de uma técnica de finalização. Neste caso, foi verificada a quantidade de movimentos que o lutador conseguiu realizar por queda efetiva ou por *knockdown*, mostrando determinado domínio no solo, sem considerar diretamente o número de finalizações. Os *Knockdowns* também foram considerados neste caso, pois por diversas vezes o lutador se aproveita que o oponente está caído para aplicar as técnicas e buscar a finalização.

Os movimentos no solo considerados foram: Avanço para meia-guarda, Avanço para o lado, Avanço para a montada, Avanço para trás e “*Sweeps*” (Passagem de uma posição neutra ou inferior para uma posição dominante), relativos ao número de quedas da luta. Neste caso, foram consideradas as quedas da luta, uma vez que mesmo nas quedas sofridas o lutador pode obter movimentos favoráveis no solo.

$$\text{Atividade no Solo} = \frac{\text{Movimentos}}{\text{Total de Quedas da Luta}} \quad (12)$$

***Knockdowns* por Golpe Efetivo:** Existem situações em que o lutador aplica golpes com menos frequência e pode possuir uma “mão pesada”, característica atribuída ao lutador que costuma derrubar o adversário com um golpe, representando, de certa forma, a força do golpe. Para quantificar essa situação, optou-se por verificar o número de *knockdowns* por golpe que o lutador acerta no oponente. Este escore só será utilizado no modelo preditivo.

$$\text{Taxa de Knockdowns} = \frac{\text{Knockdowns}}{\text{Golpes Efetivos}} \quad (13)$$

Estes escores serão utilizados de maneiras distintas nos modelos explicativos e preditivos. No primeiro caso, os escores correspondentes às próprias lutas serão padronizados para ser possível comparar a importância de cada escore. Por outro lado, ao definir-se a estimação de um modelo preditivo, é necessário que se conheça o valor das variáveis explicativas antes da luta ocorrer. Assim sendo, é necessária uma modificação no banco de dados para os modelos preditivos.

A primeira luta de cada lutador passa a ser utilizada apenas para o cálculo dos escores de desempenho, servindo como estimativa dos escores da segunda luta deste lutador. Assim, quando um lutador for lutar pela terceira vez, seu escore antes da terceira luta é a média do escore calculado das ações ocorridas na luta 1 e na luta 2 deste lutador. Por fim, só serão consideradas nessa estimação as lutas onde os dois lutadores já haviam competido pelo menos uma vez.

ODDS: Calculada pela casa de apostas, esta variável indica quanto a casa irá retornar em percentual ao apostador caso o resultado da aposta seja correto. A metodologia de cálculo pode variar de acordo com a banca apostadora e quase nunca é apresentada. Na ODDS utilizada neste trabalho, quanto maior a ODDS, maior o lucro e mais a banca considera que este lutador irá ser derrotado. Ou seja, quanto maior essa ODDS, menor a chance do lutador vencer.

O formato da ODDS aqui utilizada é decimal, mais utilizada na Europa. O retorno com a aposta é calculado da seguinte maneira:

1. Multiplica-se o valor apostado pelo valor da ODDS e tem-se quanto de unidades monetárias será retornado.
2. Diminuindo essa quantidade do valor apostado, tem-se o lucro.

Por exemplo, uma aposta de 100\$ em uma ODDS de 3.00. Caso o palpite seja correto, o apostador terá $100\$ \times 3 = 300\$$, correspondendo a $300\$ - 100\$ = 200\$$ de lucro. Ou uma aposta de 100\$ a uma ODDS de 1.40. Caso o palpite seja

correto, o apostador terá $100\$ \times 1,4 = 140\$$, correspondendo a $140\$ - 100\$ = 40\$$ de lucro.

Lado Dominante: Corresponde ao lado em que o lutador apresenta maior facilidade de combate e sua importância é dada pela mudança da base em que o lutador mantém no octógono. Existem 3 categorias possíveis: Destro, canhoto ou ambidestro.

Envergadura: É a distância entre o dedo médio de uma mão a outra, estando os braços abertos horizontalmente.

3) Resultados

Foram avaliadas 1136 lutas, de ambos os sexos, no período de janeiro de 2012 a novembro de 2014. Destas, foram excluídas 25 lutas cujo resultado foi empate (4), ou ‘Sem Resultado’ (21), que representa o fim da luta por circunstâncias imprevistas, como, por exemplo, golpes ilegais. Assim, o número de lutas resultante foi de 1111 ou 2222 observações.

Dentre as 1111 lutas, houve 739 lutadores distintos. A Tabela 1 mostra que há uma variação no número de lutas de cada lutador, entre 1 até 9 disputas. Apenas 26,7% dos lutadores competiram uma única vez, tornando necessário que se leve em consideração a dependência entre as observações.

Tabela 1: Distribuição do número de lutas

Número de lutas	Número de Lutadores	%
1	197	26,7
2	172	23,3
3	118	16,0
4	94	12,7
5	58	7,8
6	57	7,7
7	27	3,6
8	14	1,9
9	2	0,3
Total	739	100

Cada um destes lutadores apresenta características tempo-independentes como ENVERGADURA e LADO DOMINANTE, e tempo-dependentes como os escores de desempenho.

É possível verificar, pela Tabela 2, que todos os escores de desempenho observados apresentaram uma média maior para o lutador que obteve o resultado de vitória. Também é possível observar que os escores de queda apresentam muitas situações em que não houve tentativa, resultando na impossibilidade de calcular a efetividade e diminuindo o número de observações, mostrando a necessidade da categorização descrita na metodologia.

Tabela 2: Descritivas dos escores de desempenho por resultado

Escore de Efetividade	Resultado	N	Mínimo	Máximo	Média	Desvio-Padrão
Atividade Solo	Derrota	1029	0	6	0,250	0,560
	Vitória	1029	0	8	0,715	0,979
Ataque de Golpe	Derrota	1109	0	1	0,469	0,186
	Vitória	1108	0	1	0,576	0,162
Defesa de Golpe	Derrota	1108	0	1	0,424	0,162
	Vitória	1109	0	1	0,531	0,186
Knockdown	Derrota	1085	0	0,13	0,002	0,009
	Vitória	1106	0	1	0,019	0,054
Ataque de Queda	Derrota	712	0	1	0,271	0,326
	Vitória	806	0	1	0,509	0,345
Defesa de Queda	Derrota	806	0	1	0,491	0,345
	Vitória	712	0	1	0,729	0,326
ENVERGADURA	Derrota	1059	1,57	2,13	1,830	0,097
	Vitória	1101	1,55	2,13	1,835	0,101

A Figura 1 mostra que para as categorias de queda, nos casos em que o lutador efetuou pelo menos uma queda, houve 62,3% de vitórias. Em contrapartida, quando o lutador não conseguiu realizar nenhuma queda o percentual de vitórias é de apenas 38,5, mostrando a importância de ter se conseguido efetuar a ação. Nos casos de defesa de queda, lutadores com baixa efetividade foram derrotados em 70,2% das lutas. O percentual de vitórias quando a efetividade de queda é alta ou não houve necessidade de defender-se é de 59,6 e 56,7, respectivamente.

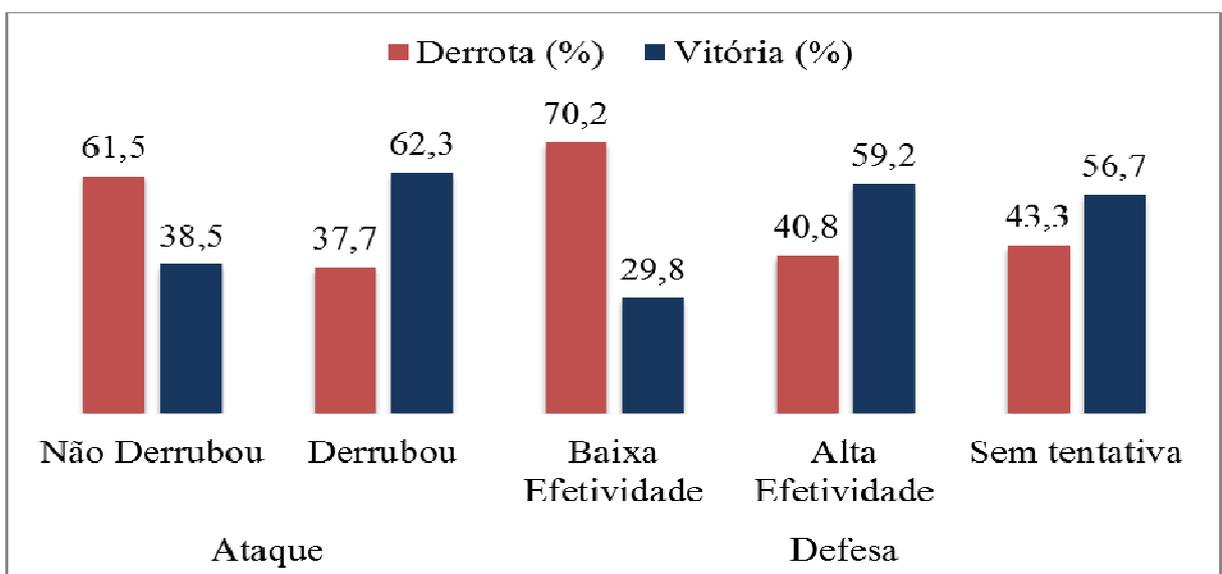


Figura 1: Distribuição do resultado da luta com os escores de queda

De acordo com a Tabela 3, verifica-se que quando destros lutaram contra canhotos houve um percentual de vitória bastante semelhante para cada um dos dois, com um percentual de vitórias um pouco superior para o atleta canhoto. Os ambidestros venceram mais tanto os destros quanto os canhotos, sendo que o percentual de vitória contra canhotos é de 70%.

Tabela 3: Percentual de vitórias de acordo com o lado dominante do lutador.

% de Vitórias	Lutas	Destros	Canhotos	Ambidestros
Destros x Canhotos	318	48,1	51,9	-
Destros x Ambidestros	30	46,7	-	53,3
Canhotos x Ambidestros	10	-	30	70

3.1) Modelos Explicativos

É intuitivo que os escores de desempenho apresentados estejam relacionados com o resultado da luta, uma vez que devem colaborar para a pontuação dos juízes ou para que se consiga o nocaute. Essa hipótese foi verificada através da construção de dois grupos de modelos utilizando os escores da própria luta como variáveis explicativas, de modo que se observe quais das situações de combate mais impactaram na probabilidade de vitória e servindo para programação de treinamentos.

A tabela 4 mostra que os modelos com diferentes matrizes de correlação de trabalho são bastante semelhantes entre si em relação às estimativas dos parâmetros e seus erros-padrões. No modelo onde foi considerado a correlação dos lutadores dentro de cada luta, observa-se que os escores de queda não foram significativos ($p > 0,05$). Neste caso, os escores de ataque e atividade no solo apresentaram valores iguais em módulo.

Tabela 4 – Coeficientes (erro-padrão) do modelo explicativo com diferentes estruturas de matrizes de correlação de trabalho e de dependência entre as observações.

Efetividade	Dependência entre lutadores				Dependência dentro da luta
	Independente	AR (1)	Permutável	Não-Estruturada	
Ataque Queda¹	0,308 (0,1115)	0,308 (0,1115)	0,308 (0,1115)	0,344 (0,1103)	0,001 ³ (0,0006)
Defesa Queda²	0,485 (0,1454)	0,485 (0,1453)	0,485 (0,1454)	0,484 (0,1429)	0,00006 ³ (0,0008)
Defesa Queda²	0,791 (0,126)	0,791 (0,126)	0,792 (0,1261)	0,698 (0,1252)	0,00009 ³ (0,0004)
Ataque de Golpe	0,882 (0,0756)	0,882 (0,0756)	0,884 (0,0756)	0,875 (0,0747)	0,556 (0,0702)
Defesa de Golpe	0,867 (0,0787)	0,866 (0,0787)	0,867 (0,0787)	0,845 (0,0798)	0,556 (0,0702)
Atividade no Solo	0,690 (0,1021)	0,690 (0,1021)	0,690 (0,1021)	0,673 (0,1007)	0,926 (0,1203)
Atividade no Solo - Adversário	-0,697 (0,1084)	-0,697 (0,1084)	-0,697 (0,1084)	-0,676 (0,1079)	-0,927 (0,1204)

¹Sem Queda é a categoria de Referência. ²Baixa Efetividade é a categoria e referência. ³p>0,05.

De acordo com as estimativas, o escore de ataque de golpe é o que mais tem impacto no resultado da luta, sendo seguido pelo escore de defesa de golpe. Ao compararem-se os escores de atividade no solo, o escore do oponente foi levemente superior (em módulo), mostrando que é preferível não deixar o adversário configurar movimentos no solo a movimentar-se. Para a efetividade de ataque de quedas, a chance de vitória o lutador que efetuou pelo menos uma queda é 1,36 vezes maior ($e^{0,308} = 1,36$) do que os lutadores que não a fizeram.

Sobre a efetividade na defesa de queda, a chance de vitória dos lutadores que apresentaram uma alta efetividade é de 2,21 vezes maior ($e^{0,792} = 2,21$) dos que apresentaram uma baixa efetividade. Além disso, ter uma baixa efetividade de defesa de queda é pior que não ter sido posto a prova (sem quedas).

A Tabela 5 mostra que os critérios de comparação de modelos apresentaram valores bastante semelhantes ao considerar-se a dependência dos lutadores, sendo que os 4 modelos apresentaram melhor desempenho do que o que considera a dependência dentro das lutas, em todos os critérios. A utilidade

de mostrar que os modelos apresentaram um considerável percentual de acerto é de que as estimativas produzidas possam apresentar validade ao proposto. Pode-se verificar que embora o modelo que considera a matriz de correlação de trabalho “Não-Estruturada” apresentar um percentual de acerto maior, obteve os critérios QIC e QIC_c também maiores.

Tabela 5: Comparação entre os modelos explicativos.

Critério	Dependência entre lutadores				Dependência dentro da luta
	Independente	AR (1)	Permutável	Não-Estruturada	
QIC	2057,34	2057,34	2057,35	2057,87	2135,34
QIC_c	2054,36	2054,36	2054,36	2055,31	2141,26
Acertos (%)	77,04	77,04	77,04	77,24	76,26

3.2) Modelos Preditivos

O número de lutas desta etapa da análise foi de 625, com 503 lutadores distintos, uma vez que as primeiras lutas de cada lutador serviram apenas para o cálculo dos escores de desempenho. Dentre os lutadores, há uma variação entre uma luta (que anteriormente era sua segunda luta), até 8 lutas.

Analisando a Tabela 6, é possível verificar que os escores são bastante semelhantes em relação à vitória ou derrota. O escore de efetividade de ataque de golpe apresentou valores médios maiores para a derrota, de forma contraintuitiva.

Tabela 6: Descritivas dos lutadores de acordo com o resultado da luta– Modelo Preditivo

Efetividade	Resultado	N	Mínimo	Máximo	Média	Desvio-Padrão
Atividade Solo	Derrota	607	0	4	0,528	0,708
	Vitória	595	0	4	0,569	0,700
Ataque de Golpe	Derrota	625	0	1	0,550	0,146
	Vitória	625	0	1	0,538	0,136
Defesa de Golpe	Derrota	625	0	1	0,478	0,146
	Vitória	624	0	1	0,506	0,142
Knockdown	Derrota	623	0	0,5	0,013	0,039
	Vitória	623	0	0,33	0,012	0,027
Ataque de Queda	Derrota	524	0	1	0,423	0,297
	Vitória	530	0	1	0,428	0,288
Defesa de Queda	Derrota	529	0	1	0,630	0,294
	Vitória	532	0	1	0,673	0,288

A envergadura média dos lutadores vitoriosos foi de 1,835 (DP= 0,095), enquanto que para os lutadores derrotados foi de 1,832 (DP=0,101). A semelhança entre as medidas se deve que os lutadores, na grande maioria das vezes, fazem parte da mesma categoria de peso e acabam por apresentarem características antropométricas semelhantes.

Verificou-se também, antes da modelagem, se for considerado o lutador como vitorioso o que possuir uma maior envergadura ou um dos escores maior, resulta em uma maneira efetiva de classificação, os valores são exibidos pela Tabela 7.

Assim, com as estimativas computadas, o percentual de acerto do vencedor da luta é maior que 50% em quase todos os escores e para a envergadura. Para estas lutas, o oponente que apresentava um valor de efetividade de ataque menor venceu mais vezes, o que é contraintuitivo. É importante ressaltar que os escores calculados não levam em consideração o histórico anterior a 2012.

Tabela 7: Percentual de acerto escolhendo o maior valor.

Medida	Quantidade de Lutas	Acerto Percentual
Envergadura	545	52,00
Efetividade Golpe Ataque	624	46,40
Efetividade Golpe Defesa	624	56,60
Efetividade Queda Ataque	428	50,93
Efetividade Queda Defesa	424	57,31
Escore Atividade solo	512	50,39
Escore KND	380	54,47

Foram estimados 4 modelos distintos para a probabilidade de vitória, considerando em todos eles a matriz de correlação de trabalho Permutável e a dependência devido cada lutador ter sido avaliado mais de uma vez, dado pela tabela 8.

O modelo 1 considera somente a ODDS. Para o modelo 2, adicionou-se as características do lutador e de seu oponente, juntamente com os escores de desempenho descritos. O modelo 3 desconsidera a ODDS e utiliza apenas os escores de desempenho e características com $p < 0,05$. O modelo 4 considera apenas os escores de Efetividade de Defesa de Golpe.

Tabela 8: Variáveis dos modelos preditivos.

Modelo	Variáveis
1	ODDS.
2	Envergadura ¹ , Lado Dominante ¹ , Escores de Efetividade ¹ ; ODDS.
3	Efetividade de Defesa de Golpe ¹ , Efetividade de Ataque de Queda ¹ , Efetividade de Defesa de Queda ¹ .
4	Efetividade de Defesa de Golpe ¹²

¹Variáveis do oponente também no modelo.

A Tabela 9 mostra que nenhum dos modelos apresentou percentual de acerto maior que o modelo só com a ODDS. Além disso, ao acrescentar no modelo 1 os escores de desempenho e características do lutador e oponente, o modelo 2 tem-se uma piora nesse percentual de acerto. Os modelos que não consideraram a utilização da ODDS apresentaram um percentual de acerto inferior a 60%.

Entretanto, caso fosse escolhido apostar no lutador que apresentou maior probabilidade estimada na luta, os modelos que consideram a ODDS teriam um retorno negativo, configurando prejuízo ao apostador. Por outro lado, os modelos sem a ODDS apresentaram um retorno positivo, representando lucro ao apostador. Este retorno é calculado da maneira descrita anteriormente, com uma aposta fictícia de 100\$ em cada luta.

Por exemplo, foram realizadas 604 apostas fictícias segundo o modelo 4, 100\$ em cada luta. No fim do período, o apostador teria 2583,8\$ a mais, de lucro. No total, o apostador investiu 60.400\$, tendo como retorno 62983,8\$ de retorno, que implica em um lucro de quase 5%. No caso do modelo 1, o apostador teria um prejuízo total de 103,55\$.

Tabela 9: Comparação entre os modelos estimados.

Critério/Modelo	1 - ODDS	2 - Completo	3 - Lutas	4 - Defesa de Golpe
<i>QIC</i>	1506,767	1355,488	1479,808	1489,679
<i>QIC_c</i>	1505,546	1357,655	1479,31	1489,246
Acerto (%)	68,33	65,12	56,43	56,52
Lucro	-103,55	-1102,85	148	2583,8

4) **Discussão**

A estimação dos modelos explicativos com as variáveis padronizadas pode auxiliar nas estratégias de treinamento, uma vez que os lutadores podem priorizar o aprimoramento de determinada técnica de acordo com a importância da ação. Neste parece que aumentar a efetividade dos golpes é o mais importante. Também se verificou a importância de se conseguir efetuar pelo menos uma queda na luta e, quando na atividade de solo, é melhor evitar que o adversário faça os movimentos do que se expor a fazê-lo. A interpretação desses escores e das estimativas deve ser balanceada ao programar o treinamento como que cada lutador possui de especialidade, de modo que devem ser analisados conjuntamente.

Os resultados mostraram que a consideração da estrutura de dependência entre os lutadores foi mais eficiente desde que se verifique a probabilidade de vitória estimada, de modo que se escolha um lutador como vencedor, sendo aqui a sugerida. Também se mostrou eficiente a consideração das variáveis do oponente na modelagem.

Em acréscimo, as matrizes de correlação mostraram resultados bastante semelhantes e é preciso avaliar qual faz mais sentido no estudo. Neste caso, a estrutura independente não seria considerada, uma vez que as lutas não são independentes. Por mais que uma luta ocorra temporalmente, uma após a outra, a estrutura AR (1) pode não ser a melhor das alternativas, pois os dias entre as lutas variam muito. É possível assumir que a estrutura de correlação entre as medidas do mesmo lutador seja a mesma para todos os tempos em que ele lute, ou estimar estas medidas pelo próprio modelo, sugerindo a utilização da matriz não estruturada ou a permutável.

Ao considerar a estimação de um modelo preditivo, não foi possível aumentar o percentual de acerto ao comparar com um modelo somente com a ODDS. Além disso, ao utilizarem-se somente os escores, os percentuais de acerto foram baixos. Por outro lado, é possível comparar os resultados com o retorno das apostas, onde os modelos com os escores apresentaram melhor

desempenho. De maneira intuitiva, a ODDS consegue acertar mais resultados, mas é justamente nesses casos que o retorno é menor. Essa relação pode ser feita da seguinte maneira: quando um lutador é favorito para vencer, o retorno das apostas, a ODDS, é menor, ou seja, a aposta retorna menos unidades monetárias, assim, os modelos sem a ODDS acabam por acertar o resultado com retornos maiores. A indicação é estimar um modelo que vise maximizar o retorno e não somente o percentual de acerto, se o objetivo for este.

Algumas situações de combate não foram consideradas neste trabalho e podem ser abordadas, como o tempo que o lutador fica em determinadas ações da luta. Além disso, outras variáveis como “Dias sem lutar”, se o lutador estava se recuperando de lesão ou não e as especialidades de cada lutador podem colaborar para aumentar o acerto do modelo preditivo.

Referências Bibliográficas

- [1] Agranonik M . Equações de Estimção Generalizadas (GEE): aplicação em estudo sobre mortalidade neonatal em gemelares de Porto Alegre, RS (1995-2007) [Dissertação de Mestrado]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2009.
- [2] Miarka, B. et al. A comparison of time-motion performance between age groups in judo matches. *Journal of Sports Sciences*, London, v. 30, n. 9, p. 899-905, 2012.
- [3] Chumacero, R.A. (2009), “Altitude or Hot Air?,” *Journal of Sports Economics*, 10(6), 619-638.
- [4] Collier T, Johnson AL, Ruggiero J. "Aggression in Mixed Martial Arts: An Analysis of the Likelihood of Winning a Decision". In: *Violence and Aggression in Sporting Contests: Economics, History, and Policy*. Vol4. Springer; 2011. p. 97-109.
- [5] Fitzmaurice GM. A caveat concerning independence estimating equation with multivariate binary data. *Biometrics*. 1995;51(1): 309-17.
- [6] Fitzmaurice GM. Clustered data. *Nutrition*. 2001 Jun; 17(6): 487-8
- [7] Gilsdorf, K.F. Sukhatme, V.A. (2008), “Testing Rosen’s Sequential Elimination Tournament Model: Incentives and Player Performance in Professional Tennis,” *Journal of Sports Economics*, 9(3), 287-303.
- [8] Johnson JD. “Predicting outcomes of mixed martial arts fights with novel fight variables” [TCC]. Athens (Georgia): University of Georgia; 2012 .
- [9] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;4273(1): 13-22.
- [10] Twisk JW. Longitudinal data analysis. A comparison between generalized estimating equations and random coefficient analysis. *Eur J Epidemiol*. 2001 Ago; 19(8): 769-76
- [11] McCullagh P. *Generalized Linear Models*: Chapman and Hall 1983.
- [12] Paula, G. A. (2004). *Modelos de regressão com apoio computacional*, São Paulo: IME/USP.

[13] Pan W. Information criterion in generalized estimating equation, *Biometrics* 2001; 57(1): 120-5