

Universidade Federal de Minas Gerais  
Instituto de Ciências Exatas  
Departamento de Estatística

CRISTIANO GIORDANO XAVIER

**PREVISÃO DE RESULTADOS DE JOGOS DE VOLEIBOL UTILIZANDO  
REGRESSÃO LOGÍSTICA**

Belo Horizonte – 2016

CRISTIANO GIORDANO XAVIER

**PREVISÃO DE RESULTADOS DE JOGOS DE VOLEIBOL UTILIZANDO  
REGRESSÃO LOGÍSTICA**

Monografia apresentada ao curso de  
Especialização em Estatística do Departamento  
de Estatística do Instituto de Ciências Exatas da  
UFMG como requisito para a obtenção do título  
de Especialista em Estatística.

Professor Orientador: Edna Afonso Reis

Belo Horizonte - 2016

## RESUMO

O objetivo do presente estudo foi encontrar modelos de regressão logística para estimar a probabilidade de vitória de um time de voleibol em função de fatores como local do jogo (em casa, fora ou neutro), posição na tabela de classificação, diferença de pontos nos *sets* e ter vencido dois *sets* primeiro foram estudados. Foram analisados jogos de duas competições distintas: 132 jogos da Superliga Nacional 2015/2016, que envolve times de clubes, e 82 jogos da Liga Mundial de Vôlei 2016, disputados entre seleções de países. No caso dos jogos entre times de clubes, dois casos foram analisados: a probabilidade de vitória no jogo do time que ganhou dois sets primeiro e a probabilidade de vitória do time que jogou em casa. No terceiro caso, também foi estudada a probabilidade de vitória no jogo do time que ganhou dois sets primeiro, mas considerando apenas os jogos de seleções que foram realizados em campo neutro. Os resultados mostraram que a diferença de pontos no primeiro *set* ganho não foi significativa em nenhum dos três casos estudados, mas que a diferença de pontos no segundo *set* ganho foi significativa no segundo caso. Verificou-se, também, que estar melhor colocado que o adversário na tabela de classificação influencia na probabilidade de vitória dos clubes, mas não é significativo no caso de seleções. Observou-se que ganhar dois *sets* primeiro aumenta a probabilidade de vitória, que é muito parecida quando o time joga em casa (84%) ou fora (83%). Concluiu-se que as variáveis utilizadas nesse estudo não são suficientes para determinar a probabilidade de um time vencer o jogo, sugerindo que novos fatores sejam adicionados a um futuro estudo. Entretanto, os resultados relacionados a ganhar dois *sets* primeiro e à diferença de posições na tabela podem ser vistos como informações valiosas para tomar decisões antes ou durante um jogo.

**Palavras-chave:** Jogar em casa. Jogar fora. Probabilidade de vitória. *Sets*.

## ABSTRACT

The aim of the present study was to find logistic regression models to predict the probability of a volleyball team winning a match in terms of factors such as match location (home, away or neutral), position in the standings, score difference in the sets and winning two sets first were studied. Matches from two distinct competitions were assessed: 132 from Superliga Nacional de Voleibol, which comprises teams from clubs and 82 from the Volleyball World League 2016, comprising international squads. Regarding the matches among clubs, two cases were analysed: the victory probability of a team which won two sets first and the victory probability when a team is playing home games. In the third case, the victory probability of a team which won two sets first was also studied, but in this case, only games which took place in a neutral venue regarding international squads were taken into account. The results showed that the difference in points in the first set won was not significant in any of the three analysed cases. However, the difference in points in the second set won came across as a significant factor in the second case. Evidences were also found that holding a better place in the standings than the opponent influences in the winning probability of the team clubs. However, the same factor is not significant in international squad's matches. It was observed that winning two *sets* first increases the probability of winning a match and that the probability is very similar when these two sets were won by the home team(84%) or the away team(83%). It was possible to conclude that the variables utilised in this study are not enough to clearly determine the odds of winning a match suggesting that other factors should be added to future studies. However, the results related to winning two *sets* first and the difference in the standings between the teams might be seen as valuable information to take decisions before or during a match.

**Keywords:** Home match. Away match. Probability of victory. *Sets*.

## **LISTA DE SIGLAS**

CBV	Confederação Brasileira de Voleibol
FIVB	<i>Fédération Internationale de Volleyball</i>
LMVM	Liga Mundial De Voleibol Masculino
SLVM	Super Liga De Voleibol Masculino

## SUMÁRIO

RESUMO .....	III
ABSTRACT .....	IV
LISTA DE SIGLAS .....	V
1. INTRODUÇÃO.....	1
2. JUSTIFICATIVA .....	2
3. OBJETIVOS.....	2
4. BANCOS DE DADOS.....	2
5. MÉTODOS.....	3
5.1. Modelo de Regressão Logística.....	3
5.1.1. Chance e Razão de Chances.....	4
5.1.2. Teste dos coeficientes do modelo.....	5
5.1.3 Teste de ajuste do modelo .....	5
5.1.4 Sensibilidade, Especificidade e Curva ROC.....	6
6. RESULTADOS .....	7
6.1. Primeiro Caso .....	7
6.1.1. Análises Descritivas .....	7
6.1.2. Modelos de Regressão Logística.....	11
6.2. Segundo Caso .....	14
6.2.1. Análises Descritivas .....	14
6.2.2. Modelos de Regressão Logística.....	17
6.3. Terceiro Caso.....	24
6.3.1. Análises Descritivas .....	24
6.3.2. Modelos de Regressão Logística.....	27
7. CONCLUSÕES.....	28
7.1 Conclusões do Primeiro Caso .....	28
7.2 Conclusões do Segundo Caso .....	28
7.3 Conclusões do Terceiro Caso .....	29
8. CONSIDERAÇÕES FINAIS .....	30
9. REFERÊNCIAS .....	31
APÊNDICES .....	33
A - Primeiro Caso .....	33
B - Segundo Caso .....	35

C - Terceiro Caso .....	46
-------------------------	----

## 1. INTRODUÇÃO

O voleibol *indoor* (jogado em quadra dentro de local coberto) ganhou seu espaço entre os brasileiros depois de resultados significativos nas Olimpíadas de 1984, em Los Angeles (Estados Unidos) e de 1988, em Seul (Coréia do Sul), onde conquistou a medalha de prata e o quarto lugar respectivamente, mas, principalmente, após os jogos de 1992, em Barcelona (Espanha), onde chegou ao seu primeiro ouro olímpico. Hoje em dia, o Brasil figura em um grupo restrito da elite deste esporte. Além disso, o voleibol é o segundo esporte mais praticado no país (Ministério do Esporte) e continua em constante crescimento e aperfeiçoamento.

Ao longo do tempo, as regras do jogo foram sendo modificadas, com o intuito de tornar as partidas de voleibol mais atrativas e mais dinâmicas. Após as mudanças realizadas nas regras do voleibol em 1999, o jogo passou por uma renovação completa no sentido técnico-tático (Zadražnik, M., Marelić, N., & Rešetar, T., 2009).

A evolução dos treinos tem sido refletida no crescimento da homogeneidade das características dos atletas de alto nível (Sheppard et al., 2009). Times de ponta são similares em altura média dos jogadores e no desempenho físico e técnico (Gabbett et al., 2007). Desta forma, os jogos entre os melhores times são, geralmente, bem disputados.

Nos dias de hoje, um jogo de voleibol de alto nível, jogado nas principais ligas do mundo, não é simplesmente definido pela habilidade dos jogadores em quadra, mas pelo trabalho deles em conjunto com as comissões técnica e analítica. Estas comissões fazem parte do chamado *staff*, que em sintonia quase que imediata com o técnico da equipe, passam informações valiosas, objetivando a minimização dos erros dentro do jogo.

Em um jogo de vôlei, vencer ou perder depende de um grande número de fatores. Um desses fatores é a eficiência necessária para solucionar situações técnicas-táticas em fases do jogo (Zadražnik, M., Marelić, N., & Rešetar, T., 2009). O dinamismo do jogo obriga cada vez mais as equipes a minimizarem seus erros, tornando cada informação extremamente importante. Isto faz com que o estatístico, que é parte do *staff*, tenha papel fundamental na obtenção e na análise dessas informações.

A Estatística e o vôlei brasileiro começaram sua relação em 1986. José Carlos Brunoro, na época assistente técnico da seleção, notou que o Brasil precisava de algo a mais para se tornar uma potência no vôlei. Com a ajuda de Sandra Caldeira, ex-jogadora e formada em Estatística, surgiu a ideia de coletar dados estatísticos dos times e analisá-los, buscando entender melhor o comportamento das equipes em quadra e tentando prever a próxima ação



do oponente através dos dados coletados. Hoje em dia, a Estatística é parte fundamental em qualquer grande clube de ponta.

No voleibol *indoor* é declarada vencedora a equipe que conquistou três *sets* primeiro, independentemente da sequência. Vence o *set* a equipe que chegar primeiro aos 25 pontos, com uma diferença mínima de pelo menos dois pontos em relação ao seu oponente. Caso tal diferença não seja alcançada no limite de 25 pontos, o *set* segue até que um dos times consiga abrir dois pontos de diferença para o outro time. Se, durante o jogo, ocorrer um empate no qual cada equipe tenha vencido dois *sets*, realiza-se um *set* de desempate, no qual os pontos necessários para a vitória somam 15, mantendo-se a regra de diferença de dois pontos para que um time seja vencedor.

Pode-se afirmar que um jogo de vôlei *indoor* é composto por três, quatro ou cinco “micro-jogos” separados, onde cada time inicia cada “micro-jogo” com uma probabilidade de sucesso de 50%, quando fatores como estar jogando em casa ou não, desempenho no *set* anterior e posição na tabela de classificação, por exemplo, são desconsiderados.

Uma vez que em cada um desses *sets* existe sempre uma equipe vencedora, poder-se-á afirmar que o jogo de VP é constituído por dois ou três “micro-jogos”. Sendo estes “micro-jogos” ou *sets*, unidades independentes\*, é possível que os desempenhos das equipes sejam diferenciados no decurso do jogo. Deste modo, não só importa estudar a possível variação dos indicadores fisiológicos e funcionais no decurso do jogo como também a *performance* tática, tendo como referência os diferentes *sets* que o integram (Marcelino R, Mesquita I, Sampaio J, 2009).

O resultado em cada um dos *sets*, principalmente no caso de fracasso, pode modificar totalmente a tática inicial estabelecida pelo técnico da equipe, obrigando-o a passar novas orientações ao time, dado o que aconteceu nos *sets* anteriores. Estas orientações podem ter sua eficiência elevada caso o técnico tenha um conhecimento probabilístico do que acontece em tais situações.

\*independente não no sentido estatístico

## 2. JUSTIFICATIVA

Tendo em vista a suposição de que a probabilidade de vitória de um time depende de diversas variáveis, é importante identificar quais são estas variáveis e como elas afetam o resultado. Juntamente com todo o aparato tecnológico utilizado pelas equipes hoje em dia, tal análise pode ser uma ferramenta a mais para ajudar a identificar problemas táticos/técnicos, buscando ajudar o técnico a efetuar modificações caso esteja perdendo o jogo. Além disso, busca-se encontrar respostas para algumas crenças como: o time da casa tem mais chance de vencer o jogo.

## 3. OBJETIVOS

O estudo foi dividido em três partes. Na primeira parte, o objetivo é encontrar um modelo de regressão logística que estime a probabilidade do time que ganhou dois *sets* primeiro ganhar o jogo, em função de seu desempenho nos dois *sets* que ganhou primeiro, da sua posição na tabela de classificação em relação ao adversário e do fato de estar jogando ou não em casa.

Na segunda parte do estudo, busca-se verificar o efeito, no resultado final (venceu ou não o jogo) do time da casa, dos fatores: diferença de posição entre os adversários na tabela de classificação; diferença de pontos no primeiro *set* do time que ganhou dois *sets* primeiro; diferença de pontos no segundo *set* do time que ganhou dois *sets* primeiro; e se o time que ganhou dois *sets* primeiro estava jogando em casa.

Na terceira parte, o objetivo é entender se a neutralidade do lugar do jogo (jogar em um ginásio onde nenhuma das equipes estão familiarizadas com as dimensões e pontos de referência do local) interfere ou não na probabilidade do time que ganhou dois *sets* primeiro ganhar o jogo, em função de seu desempenho nos *sets* ganhos e da diferença de posições entre os times.

## 4. BANCOS DE DADOS

Foram utilizados dois conjuntos de dados para as análises deste estudo. O primeiro conjunto de dados (*Banco 1*) foi coletado na página da internet da *Confederação Brasileira de Voleibol* (CBV). Esse banco apresenta os resultados dos 132 jogos disputados na temporada

2015/2016 da *Superliga Masculina de Voleibol (SLMV)*, competição mais importante da elite do voleibol nacional. A SLMV é composta por 12 equipes brasileiras e todas se enfrentam em um sistema de turno e retorno. Cada equipe joga 11 partidas “em casa” e 11 partidas “fora de casa”. O Banco 1 foi utilizado para fazer as análises das hipóteses dos casos 1 e 2 deste estudo.

O segundo conjunto de dados (*Banco 2*), coletado na página da internet da *Fédération Internationale de Volleyball (FIVB)*, são os jogos da *Liga Mundial de Vôlei (LMV)* de 2016. As seleções de 36 países realizaram 152 partidas em um sistema de grupos, semi-finais e finais. Dentre tais partidas, 82 delas foram realizadas em território neutro, que foram as partidas utilizadas no terceiro objetivo deste estudo.

Com as partidas realizadas em campo neutro, o objetivo é analisar se ganhar dois *sets* primeiro aumenta a probabilidade de um time ganhar o jogo. Diferente do primeiro caso, agora não temos o fator “jogar em casa”. Além disso, as diferenças de posições dos times podem ser muito maiores, não somente pelo fato de termos mais times, mas também por ser utilizado o *ranking* da FIVB para definir a diferença de posições entre as equipes. A LMV não conta exatamente com os 36 selecionados que possuem as melhores posições no *ranking* da FIVB. Sendo assim, a diferença de posições pode ser maior que 35. No caso dos dados analisados, temos a seleção com o melhor *ranking* ocupando o 1º posto e a seleção de pior *ranking* ocupando o posto 50. O *ranking* mostra a posição de cada selecionado internacional em outubro de 2015, logo após as últimas competições internacionais daquele ano.

## 5. MÉTODOS

Os modelos de regressão tornaram-se um componente essencial em qualquer análise de dados preocupada em descrever a relação entre uma variável resposta e uma ou mais variáveis explicativas (Hosmer-Lemeshow, 2000). Para cada tipo de variável resposta que estamos tentando descrever o resultado, tem-se uma classe de modelos a ser utilizada.

### 5.1. Modelo de Regressão Logística

No modelo de regressão logística, a variável resposta  $Y$  é qualitativa dicotômica, ou seja, tem apenas duas categorias, chamadas de “sucesso” de um evento ( $Y=1$ ) e “fracasso” deste evento ( $Y=0$ ). Por exemplo, o evento pode ser o voto em um determinado candidato, a

liberação de crédito bancário para um cliente, ou a vitória de uma equipe em uma atividade esportiva.

O modelo de regressão logística é a equação que define a probabilidade do sucesso do evento, ou seja,  $P(Y=1)$ , em função dos valores das variáveis explicativas  $X_1, X_2, \dots, X_k$ :

$$P(Y = 1) = \frac{e^{B_0 + B_1X_1 + B_2X_2 + \dots + B_nX_k}}{1 + e^{B_0 + B_1X_1 + B_2X_2 + B_nX_n}}$$

na qual  $B_0, B_1, B_2, \dots, B_n$  são os coeficientes de regressão a serem estimados e  $e$  é a base do logaritmo natural.

Todos os quadros presentes no estudo com os testes de significância dos modelos, dos termos, de capacidade preditiva dos modelos e as equações de regressão logística foram feitos usando o *software Minitab 17*.

### 5.1.1. Chance e Razão de Chances

Tudo começa com o conceito de probabilidade. Digamos que a probabilidade de sucesso de um evento seja 0.8. Logo, a probabilidade desse evento não ocorrer (fracasso) é de  $1 - 0.8 = 0.2$ . A *chance* de sucesso é definida pela razão entre a probabilidade de sucesso do evento sobre a probabilidade de fracasso do evento. No exemplo, a chance de sucesso seria definida por  $\frac{0.8}{0.2} = 4$ . Isso quer dizer que a chance de sucesso é de 4 para 1, indicando que a probabilidade de sucesso é quatro vezes maior que a probabilidade de fracasso. Para uma interpretação mais simples, digamos que um evento tem probabilidade de sucesso = 0.5, sua chance de sucesso seria de 1 para 1. Ou seja, tanto sucesso quanto fracasso têm a mesma probabilidade de acontecer.

A *razão de chances* (RC) é uma medida que permite comparar a chance de sucesso de um evento entre dois grupos ou duas situações. Por exemplo, se a chance de um evento é 2 no Grupo A e igual a 5 em no Grupo B, dizemos que a razão de chances do evento do Grupo B em relação ao Grupo A é dada por  $RC = \frac{5}{2} = 2.5$ . Ou seja, a chance do evento ocorrer é duas vezes e meia maior no Grupo B quando comparada ao Grupo A.

No modelo de regressão logística, o coeficiente estimado  $b_j$  de uma variável explicativa  $X_j$  fornece uma estimativa do logaritmo natural da razão de chances de sucesso na variável resposta ( $Y=1$ ), comparando as situações quando  $X_j=x$  e  $X_j=x+1$ , e mantendo-se

constantes os valores das demais variáveis explicativas. A estimativa da razão de chances é obtida, portanto, através da exponenciação do coeficiente:  $RC = e^{bj}$ . Um coeficiente positivo leva a  $RC > 1$ , ou seja, há um aumento na chance ao se passar de  $X_j = x$  para  $X_j = x + I$ ; por outro lado, um valor negativo do coeficiente leva a  $RC < 1$ , ou seja, a chance de sucesso é reduzida ao se passar de  $X_j = x$  para  $X_j = x + I$ ; e, quando o coeficiente é nulo,  $RC = 1$ , ou seja, a chance de sucesso é igual para  $X_j = x$  para  $X_j = x + I$ .

### 5.1.2. Teste dos coeficientes do modelo

Para saber se o coeficiente  $B_i$  (da variável explicativa  $X_i$ ) é significativo, ou seja, diferente de zero, comparamos o valor de  $D$  (*deviance*) com e sem a variável no modelo, sendo  $D$  expresso por:

$$D = -2 \ln \left[ \frac{(\text{verossimilhança do modelo ajustado})}{(\text{verossimilhança do modelo saturado})} \right]$$

O modelo saturado corresponde ao modelo com todas as variáveis e o modelo ajustado corresponde ao modelo com as variáveis sendo testadas.

A significância da variável é dada pela estatística  $G$  que compara as mudanças no valor de  $D$  com e sem a variável no modelo. Esse valor pode ser obtido por:

$$G = D(\text{modelo sem a variável}) - D(\text{modelo com a variável})$$

Sob a hipótese nula  $H_0 : B_i = 0$ , a estatística  $G$  tem distribuição assintótica Qui-quadrado. O valor-p é definido como a probabilidade de se observar valores de  $G$  tão extremos quanto o valor calculado quando a hipótese nula é verdadeira. Para valores-p baixos, menores que  $\alpha$ , rejeitamos a hipótese nula, e concluímos que a variável explicativa  $X_i$  tem relação significativa com a variável resposta, ou seja, que  $B_i \neq 0$ .

### 5.1.3 Teste de ajuste do modelo

O teste de ajuste do modelo, para saber se o modelo ajustado tem um bom poder de previsão, confronta as hipóteses:

$$H_0: \text{O modelo está bem ajustado} \quad \text{vs} \quad H_1: \text{O modelo não está bem ajustado.}$$

Serão utilizados os testes Pearson, da Deviance e de Hosmer-Lemeshow. Um valor-p alto, que leva à não rejeição da  $H_0$ , indica que não existem diferenças significativas entre as probabilidades previstas pelo modelo e os valores observados no banco de dados. Neste estudo, vamos considerar que há um bom ajuste do modelo se os três testes apresentarem valor-p acima de 0.05.

#### 5.1.4 Sensibilidade, Especificidade e Curva ROC

A probabilidade de sucesso estimada pelo modelo de regressão logística pode ser usada como uma previsão do sucesso ou fracasso do evento: se ela for alta, prevemos um sucesso; se ela for baixa, prevemos um fracasso. A porcentagem de sucessos previstos pelo modelo dentre os sucessos reais dos dados é chamada de *sensibilidade*; e a porcentagem de fracassos previstos pelo modelo dentre os fracassos reais do banco de dados é chamada de *especificidade*.

Mas a partir de que valor uma probabilidade de sucesso estimada pelo modelo pode ser considerada alta? Este valor é chamado de *ponto de corte*. Usualmente é tomado como 0.5, mas não precisa necessariamente ser este valor. Quanto maior (mais próximo de 1) é o ponto de corte, maior é a especificidade do modelo, mas menor é a sua sensibilidade. Assim, na escolha do ponto de corte, levamos em consideração a intenção do modelo como critério de classificação.

A **Curva ROC** é uma ferramenta gráfica usada para avaliar a qualidade do ajuste de um modelo logístico em função dos valores de sensibilidade e especificidade para diversos pontos de corte. Nela, os valores de sensibilidade e 1-especificidade, para cada ponto de corte, são dispostos um contra o outro em uma curva. É esperado que a curva chegue o mais próximo possível do canto esquerdo superior do quadrado, indicando um bom ajuste do modelo. Avalia-se a qualidade do modelo com a curva ROC observando a área abaixo da curva dentro do quadrado unitário: quanto mais próxima de 1 (distante de 0.5) for esta área, melhor o ajuste.

## 6. RESULTADOS

### 6.1. Primeiro Caso

Para cada um dos 132 jogos da SLMV, a variável resposta  $Y$  indica se o time que ganhou dois *sets* primeiro venceu o jogo ( $Y=1$ ) ou não ( $Y=0$ ).

A base de dados apresenta quatro potenciais variáveis explicativas:

$X_1$ : diferença de pontos no primeiro *set* ganho pelo time que ganhou dois *sets* primeiro (valores inteiros maiores ou iguais a 2);

$X_2$ : diferença de pontos no segundo *set* ganho pelo time que ganhou dois *sets* primeiro (valores inteiros maiores ou iguais a 2);

$X_3$ : se o time que ganhou dois *sets* primeiro jogava “em casa”(=1) ou “fora de casa” (=0);

$X_4$ : diferença de posição, na tabela de classificação, do time que ganhou os dois *sets* primeiro (valores inteiros de -11 a -1 e de 1 a 11).

Sobre as variáveis, em  $X_1$  e  $X_2$  busca-se entender, através da diferença de pontos, como o desempenho em cada um desses *sets* modifica (ou não) a probabilidade de vitória do time que venceu dois *sets* primeiro. Em relação a  $X_3$ , o objetivo é entender o quanto ganhar dois *sets* primeiro jogando em casa ou fora pode influenciar (ou não) a vitória do time. Em  $X_4$ , o objetivo é entender se estar acima ou abaixo do adversário quando se ganha dois *sets* primeiro tem influência no resultado final do jogo.

#### 6.1.1. Análises Descritivas

Dentre os 132 jogos realizados, em 111 (84%) o time que ganhou dois *sets* primeiro ganhou o jogo. Já era esperado que a porcentagem de jogos vencidos pelo time que ganhou dois *sets* primeiro fosse maior que 50%, pois este é um resultado conhecido no meio.

A Tabela 1 mostra que a porcentagem de jogos vencidos pelo time que ganhou dois *sets* primeiro é um pouco maior se ele jogou em casa (86%) do que se ele jogou fora de casa (82%). O teste do Qui-quadrado, que compara essas proporções, apresenta valor-p igual a 0.296, indicando que não existe evidência para se dizer que há diferença significativa entre as proporções, ao nível de significância de 5%. Ou seja, a variável “jogar em casa” não parece ser significativa para explicar o resultado no jogo do time que ganhou dois *sets* primeiro.

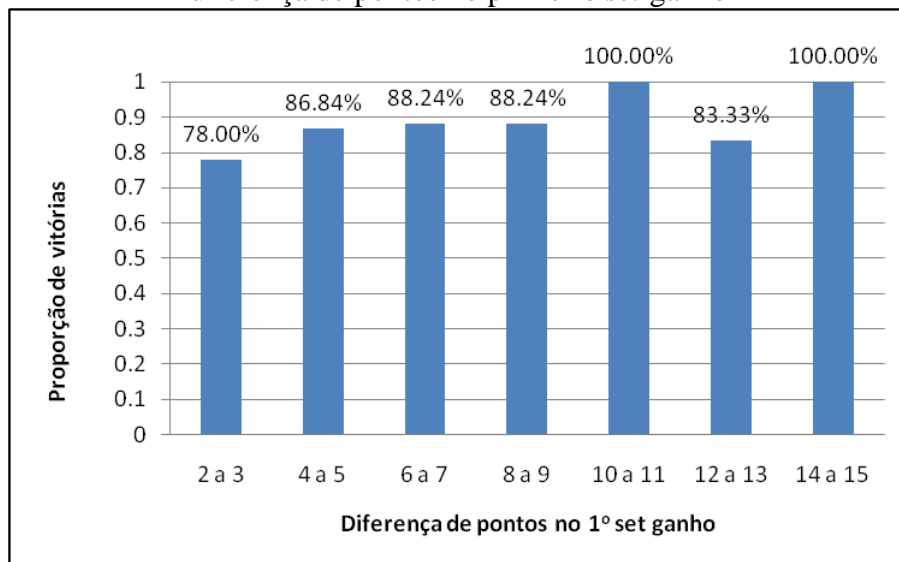
**Tabela 1.** Número de jogos segundo resultado final no jogo do time que ganhou dois *sets* primeiro, estratificados por local do jogo (em casa ou fora de casa)

O time que ganhou dois <i>sets</i> primeiro jogava em casa?	Resultado do time		Total
	Perdeu (Y=0)	Venceu (Y=1)	
Não ( $X_3=0$ )	11 (18%)	50 (82%)	61 (100%)
Sim ( $X_3=1$ )	10 (14%)	61 (86%)	71 (100%)
<b>Total</b>	21	111	132

FONTE: Dados de pesquisa, 2016.

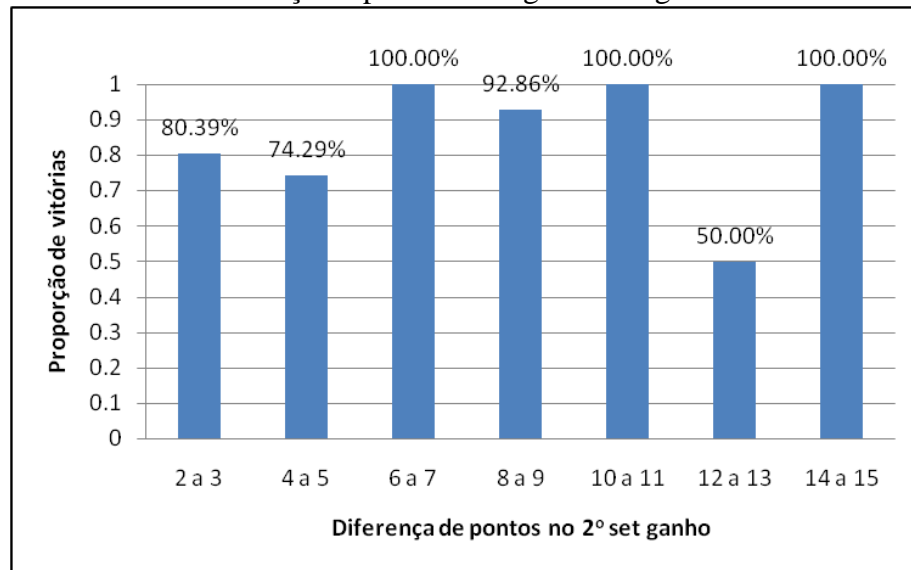
As figuras 1 e 2 mostram a proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores diferença de pontos, respectivamente, no primeiro *set* ganho ( $X_1$ ) e no segundo *set* ganho ( $X_2$ ). Nota-se que há uma leve tendência de aumento da proporção de vitórias com o aumento nas diferenças de pontos, mas os valores são muito parecidos, indicando que estas variáveis não são importantes para explicar a probabilidade de vitória no jogo do time que ganhou dois *sets* primeiro.

**Figura 1.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores da diferença de pontos no primeiro *set* ganho





**Figura 2.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores da diferença de pontos no segundo *set* ganho



As tabelas 2 e 3 representam as frequências das diferenças de pontos no primeiro e no segundo *set* ganhos pelo time que ganhou dois *sets* primeiro, respectivamente. Nota-se que a maioria dos *sets* são vencidos por 2 ou 3 pontos de diferença, indicando vitórias apertadas em ambos. Além disso, as baixas frequências para diferenças de pontos iguais ou maiores que 10 indicam que poucos *sets* são vencidos com uma ampla vantagem. Nesse caso, seria melhor considerarmos somente as proporções derivadas de frequências mais altas para uma análise mais precisa.

**Tabela 2 e Tabela 3.** Frequência das diferenças de pontos em  $X_1$ (Tabela2) e  $X_2$ (Tabela3)

X1	Frequency	Wins	Propor
2 a 3	50	39	0.78
4 a 5	38	33	0.868421
6 a 7	17	15	0.882353
8 a 9	17	15	0.882353
10 a 11	3	3	1
12 a 13	6	5	0.833333
14 a 15	1	1	1
	132	111	

X2	Frequency	Wins	Propor
2 a 3	51	41	0.803922
4 a 5	35	26	0.742857
6 a 7	23	23	1
8 a 9	14	13	0.928571
10 a 11	6	6	1
12 a 13	2	1	0.5
14 a 15	1	1	1
	132	111	

Em relação à posição dos times na tabela de classificação ( $X_4$ ), na Figura 3 observamos que a proporção de vitórias do time que ganhou dois *sets* primeiro parece ser

maior quando a diferença nas posições é maior, mesmo quando este time está abaixo na tabela.

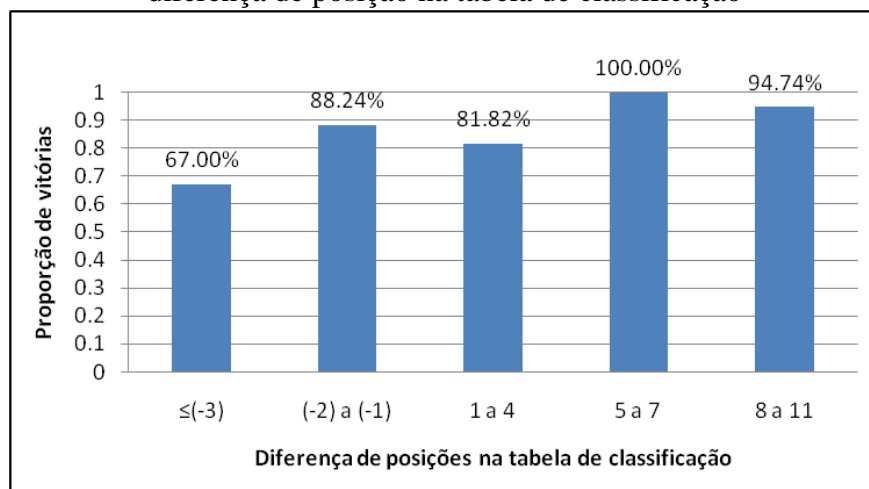
Entretanto, na Tabela 4 vemos que, quando os times que ganharam dois *sets* primeiro estavam acima de seus adversários, a porcentagem de vitória no jogo foi muito maior (90%) do que quando estes times estavam abaixo (73%). Nesse caso, o teste do Qui-quadrado comparando estas duas proporções apresentou valor-p igual a 0.000, o que leva a concluir que as proporções são significativamente diferentes.

**Tabela 4.** Número de jogos segundo resultado final do time que ganhou dois *sets* primeiro, estratificados por posição na tabela de classificação

Posição na tabela de classificação*	Resultado do time		Total
	Perdeu (Y=0)	Venceu (Y=1)	
Abaixo ( $X_4 < 0$ )	12 (27%)	32 (73%)	44 (100%)
Acima ( $X_4 > 0$ )	9 (10%)	79 (90%)	88 (100%)
<b>Total</b>	21	111	132

\*Do time que venceu dois *sets* primeiro. FONTE: Dados de pesquisa, 2016.

**Figura 3.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo a diferença de posição na tabela de classificação



A tabela 5 apresenta com que frequências os times venceram dois *sets* primeiro segundo suas posições na tabela de classificação.

**Tabela 5.** Frequência dos times que ganharam dois sets primeiro segundo posição na tabela

X4	Frequency	Wins	Propor
≤(-3)	27	17	0.67
(-2) a (-1)	17	15	0.882353
1 a 4	44	36	0.818182
5 a 7	25	25	1
8 a 11	19	18	0.947368
	132	111	

### 6.1.2. Modelos de Regressão Logística

A Tabela 3 mostra os valores dos coeficientes estimados no modelo de regressão logística para as quatro variáveis explicativas separadamente.

Nos modelos individuais foi adotado  $\alpha=0.1$  e apenas as variáveis  $X_2$  e  $X_4$  foram consideradas significativas.

Para verificar o efeito conjunto das quatro variáveis explicativas, foi ajustado o modelo de regressão logística múltiplo com todas elas (Tabela 6; Apêndice: Quadro 1). Foi confirmada a significância de  $X_4$ , mas  $X_2$  deixou de ser importante. A fim de verificar uma possível influência conjunta de  $X_4$  e  $X_2$ , um modelo foi ajustado somente com as duas variáveis (Tabela 6; Apêndice: Quadro 2) e foi confirmado que apenas a variável  $X_4$  é significativa. Nesse caso, o  $\alpha$  adotado foi de 0.05.

**Tabela 6.** Valores dos coeficientes estimados nos modelos de regressão logística para as quatro variáveis explicativas

Variável Explicativa	Modelos Individuais			Modelo Múltiplo 1 (X1, X2, X3 e X4)				Modelo Múltiplo 2 (X2 e X4)			
	Coef	Valor-p	Erro	Coef	Valor-p	EP*	Vif	Coef	Valor-p	EP*	VIF
<b>X1</b>	0.085	0.328	0.091	0.024	0.802	0.096	1.04	-	-	-	-
<b>X2</b>	0.179	0.077	0.109	0.128	0.220	0.110	1.04	0.128	0.217	0.110	1.03
<b>X3</b>	0.294	0.537	0.048	0.264	0.598	0.501	1.01	-	-	-	-
<b>X4</b>	0.160	0.003	0.056	0.143	0.008	0.057	1.04	0.146	0.006	0.056	1.03

X1 : diferença de pontos no primeiro *set* ganho pelo time que ganhou dois *sets* primeiro.

X2 : diferença de pontos no segundo *set* ganho pelo time que ganhou dois *sets* primeiro.

X3 : se o time que ganhou dois *sets* primeiro jogava “em casa”(=1) ou “fora de casa” (=0).

X4 : diferença de posição, na tabela de classificação, do time que ganhou os dois *sets* primeiro.

O modelo final foi ajustado apenas com a variável  $X_4$  (Quadro 1). Os três testes de qualidade de ajuste do modelo apresentam valor-p maior que 0.05, indicando um bom ajuste deste modelo segundo estes testes. Entretanto, a curva ROC (Figura 4) mostra que o modelo não tem uma boa capacidade preditiva, uma vez que a curva está distante do canto superior esquerdo do quadrado unitário.

A probabilidade do time que ganhou dois *sets* primeiro vencer o jogo é estimada por:

$$P(Y = 1) = \frac{e^{1.518 + 0.1598 X_4}}{1 + e^{1.518 + 0.1598 X_4}}$$

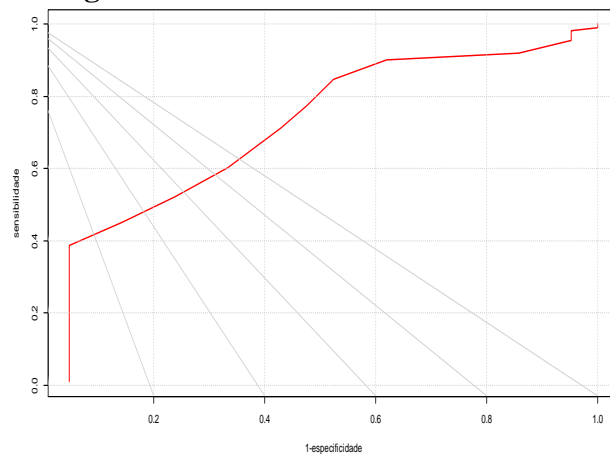
Como o coeficiente de  $X_4$  é positivo, conforme a diferença de posições na tabela aumenta, a probabilidade do time que ganhou dois *sets* primeiro conseguir ganhar a partida

também aumenta. A Tabela 7 e a Figura 5 apresentam as estimativas desta probabilidade para diferentes valores de  $X_4$ . Por exemplo: se o time que ganhou dois *sets* primeiro está 10 posições acima do adversário, podemos dizer, com 95% de confiança, que ele tem uma probabilidade entre 87.0% e 98.7% de ganhar o jogo. Na interpretação da razão de chances, a cada aumento de uma unidade na diferença de posições, a chance do time que ganhou dois *sets* primeiro vencer o jogo aumenta em 17.3% (pois  $\exp(0.1598)=1.173$ ).

#### Quadro 1: Resultados do ajuste do Modelo Final

Link function	Logit					
Rows used	132					
Response Information						
Variable	Value	Count				
Y	1	111	(Event)			
	0	21				
	Total	132				
Deviance Table						
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value	
Regression	1	9.010	9.0100	9.01	0.003	
X4	1	9.010	9.0100	9.01	0.003	
Error	130	106.664	0.8205			
Total	131	115.674				
Model Summary						
Deviance	R-Sq		R-Sq(adj)		AIC	
	7.79%	6.92%	110.66			
Coefficients						
Term	Coef	SE Coef	VIF			
Constant	1.518	0.247				
X4	0.1598	0.0562	1.00			
Odds Ratios for Continuous Predictors						
	Odds Ratio	95% CI				
X4	1.1732	(1.0508, 1.3099)				
Goodness-of-Fit Tests						
Test	DF	Chi-Square	P-Value			
Deviance	130	106.66	0.933			
Pearson	130	133.47	0.400			
Hosmer-Lemeshow	6	7.49	0.278			

**Figura 4: Curva ROC do Modelo Final**



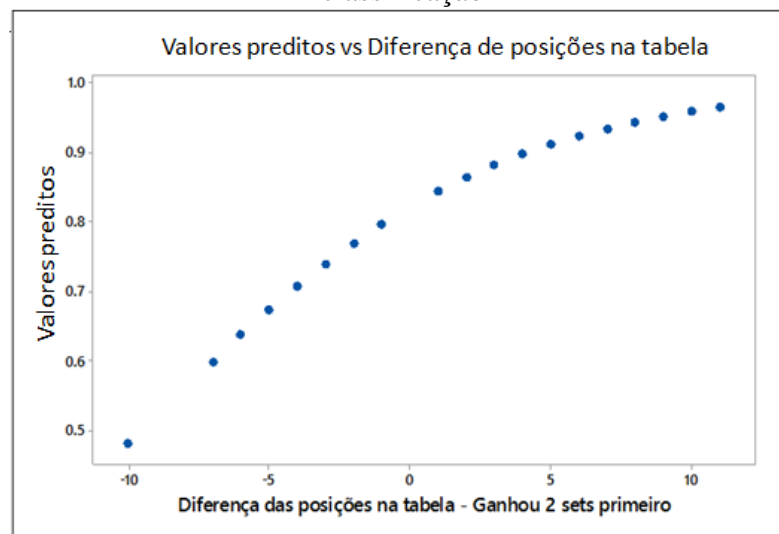
**Tabela 7.** Estimativas da probabilidade do time que ganhou dois *sets* primeiro vencer o jogo segundo a diferença de posições na tabela de classificação

$X_4$	Estimativa Pontual	Erro Padrão	IC 95%
10	0.957	0.0252	(0.8698, 0.9870)
5	0.910	0.0310	(0.8282, 0.9552)
2	0.863	0.0325	(0.7858, 0.9149)
-2	0.768	0.0478	(0.6619, 0.8487)
-5	0.672	0.0813	(0.4989, 0.8088)
-10	0.480	0.1517	(0.2191, 0.7524)

$X_4$ : diferença de posição na tabela de classificação. IC 95%: Intervalo de Confiança de 95%.

FONTE: Dados de pesquisa, 2016.

**Figura 5.** Valores preditos para a probabilidade de vitória em relação a posição na tabela de classificação



## 6.2. Segundo Caso

Neste segundo caso, a variável resposta  $Y$  indica se o *time da casa* venceu o jogo ( $Y=1$ ) ou não ( $Y=0$ ), considerando os 132 jogos da SLMV.

As variáveis explicativas a serem investigadas são:

- $X_1$ : diferença de posição, na tabela de classificação, do time da casa em relação ao adversário (valores inteiros de -11 a -1 e de 1 a 11);
- $X_2$ : diferença de pontos do time da casa em relação ao adversário no primeiro *set* ganho pelo time que ganhou dois *sets* primeiro (valores inteiros  $\leq -2$  ou  $\geq 2$ );
- $X_3$ : diferença de pontos do time da casa em relação ao adversário no segundo *set* ganho pelo time que ganhou dois *sets* primeiro (valores inteiros  $\leq -2$  ou  $\geq 2$ );
- $X_4$ : se o time da casa ganhou dois *sets* primeiro, sim (=1) ou não (=0).

No que diz respeito às variáveis explicativas, em  $X_1$  o objetivo é entender se estar acima ou abaixo do adversário influencia ou não no resultado final do time da casa. Sobre as variáveis  $X_2$  e  $X_3$ , busca-se entender através, da diferença de pontos, como o desempenho em cada um dos *sets* modifica (ou não) a probabilidade de vitória do time da casa. Em  $X_4$ , o objetivo é entender o quanto ganhar dois *sets* primeiro jogando em casa pode influenciar (ou não) a vitória do time da casa.

### 6.2.1. Análises Descritivas

Dentre os 132 jogos realizados, o time da casa venceu o jogo em 71 (54%). Era esperado que a porcentagem de jogos vencidos pelo time da casa fosse bem maior que 50%, pois há uma crença no meio esportivo de que o time da casa tem chance maior de vencer.

A Tabela 8 mostra que a porcentagem de jogos vencidos pelo time da casa é muito maior se ele ganhou dois *sets* primeiro (84%) do que quando não ganhou dois *sets* primeiro (17%). Esta diferença já era esperada pelos resultados da Seção 6.1, na qual verificou-se que a percentual de vitórias do time que ganhou dois *sets* primeiro era de 84% (valor-p =0.000)

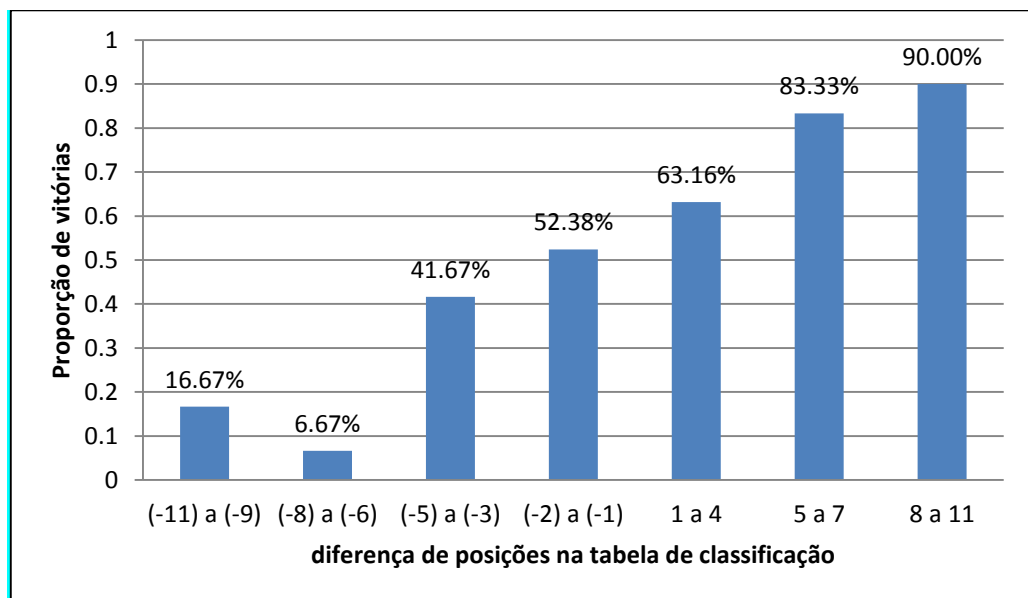
**Tabela 8.** Número de jogos segundo resultado final do time da casa, estratificados pelo fato do time da casa ter vencido ou não dois *sets* primeiro

Time da casa ganhou dois <i>sets</i> primeiro	Resultado do time da casa		Total
	Perdeu (Y=0)	Venceu (Y=1)	
Não ( $X_4=0$ )	50 (83%)	<b>10 (17%)</b>	60 (100%)
Sim ( $X_4=1$ )	11 (16%)	<b>61 (84%)</b>	72 (100%)
<b>Total</b>	61 (46%)	<b>71 (54%)</b>	132 (100%)

FONTE: Dados de pesquisa, 2016.

A Figura 6 apresenta a proporção de vitórias do time da casa segundo a diferença de posições na tabela ( $X_1$ ). Observa-se claramente que há uma tendência de crescimento desta proporção conforme aumenta a diferença nas posições, indicando que esta variável é importante para explicar a probabilidade de vitória do time da casa no jogo.

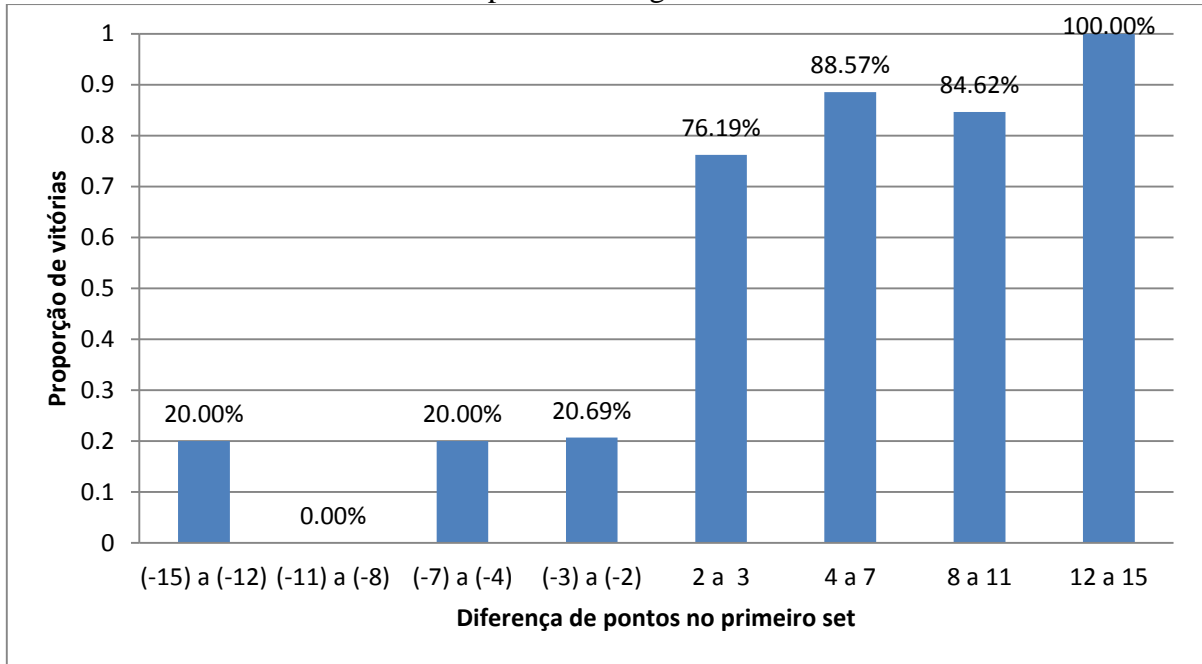
**Figura 6.** Proporção de vitórias do time da casa segundo a diferença de posições na tabela



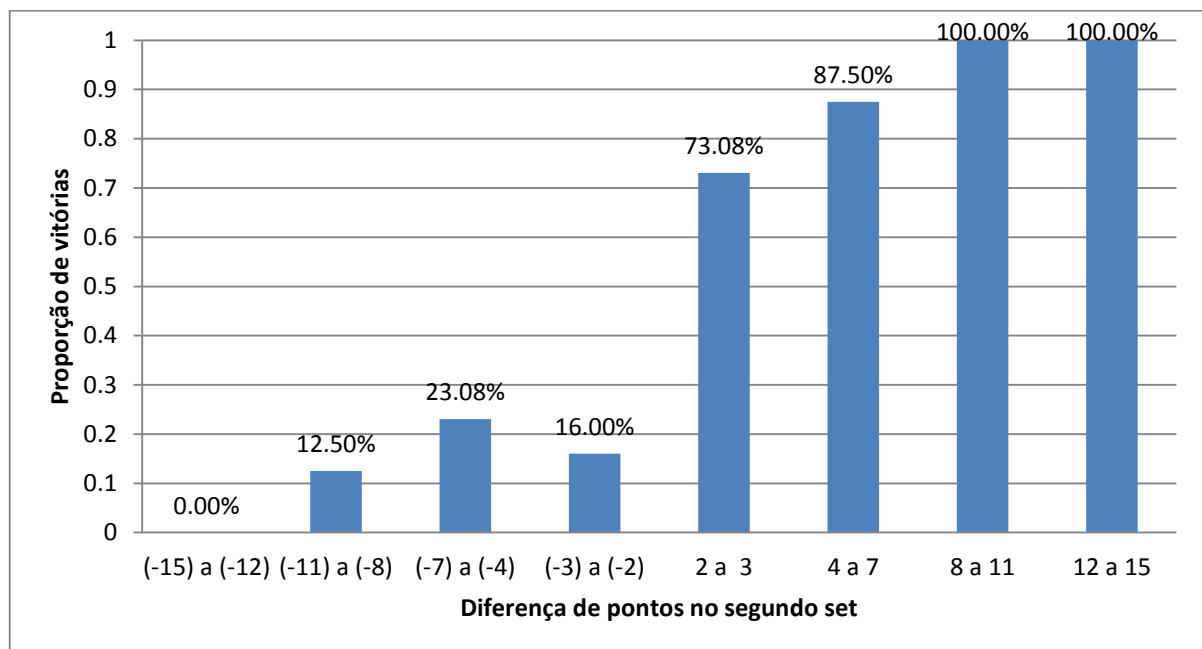
As figuras 7 e 8 mostram a proporção de vitórias do time da casa em função da sua diferença de pontos em relação ao adversário, no primeiro ( $X_2$ ) e no segundo ( $X_3$ ) *set* ganhos primeiro. Observa-se que, tanto em  $X_2$  quanto em  $X_3$ , quando a diferença de pontos é negativa (quando o time da casa perde o *set*), a proporção de vitórias do time da casa é muito baixa. Por outro lado, quando a diferença é positiva (o time da casa venceu o *set*), em ambos os *sets* a proporção de vitórias é muito maior. Nota-se tanto em  $X_2$  quanto em  $X_3$  uma leve tendência de aumento na proporção de vitórias conforme a diferença de pontos aumenta, indicando que vencer os *sets* por diferenças maiores pode aumentar a probabilidade de vencer o jogo.

Entretanto, nota-se que o mais importante é ganhar os *sets*, independente de uma diferença baixa ou alta.

**Figura 7.** Proporção de vitórias do time da casa segundo os valores diferença de pontos no primeiro *set* ganho



**Figura 8.** Proporção de vitórias do time da casa segundo os valores diferença de pontos no primeiro *set* ganho





## 6.2.2. Modelos de Regressão Logística

A Tabela 9 mostra os valores dos coeficientes estimados no modelo de regressão logística para as quatro variáveis explicativas separadamente. Todas as variáveis são significativas a 5%.

Para verificar o efeito conjunto das quatro variáveis explicativas, foi ajustado o modelo de regressão logística múltiplo com todas as variáveis (Tabela 9; Apêndice: Quadro 3). Apenas  $X_1$  e  $X_3$  tiveram a significância confirmada. Note-se que  $X_2$ ,  $X_3$  e  $X_4$  apresentam valores de VIF próximos de 5 (4.06 e 4.61) e valores acima de 5 (6.8), indicando correlações moderada e alta, respectivamente, entre as variáveis, o que afetou os resultados e fez com que apenas uma delas ficasse significativa na presenças das outras duas. Para confirmar se havia de fato correlação entre as variáveis, foram feitos os gráficos de correlação entre elas (Apêndice: Figura 1) e verificou-se que de fato todas as variáveis estão correlacionadas.

Foram ajustados modelos utilizando todas as combinações possíveis entre as variáveis predictoras (Apêndice: Quadros de 4 a 12). Entretanto, comparando o AIC e os testes de qualidade de ajuste, o modelo final com  $X_1$  e  $X_3$  foi o que apresentou os melhores resultados.

**Tabela 9.** Valores dos coeficientes estimados nos modelos de regressão logística para as quatro variáveis explicativas

Variável Explicativa	Modelos Individuais			Modelo Múltiplo 1 ( $X_1$ , $X_2$ , $X_3$ e $X_4$ )				Modelo Final ( $X_1$ e $X_3$ )			
	Coef	Valor-p	Erro	Coef	Valor-p	EP*	Vif	Coef	Valor-p	EP*	VIF
<b>X1</b>	0.221	0.000	0.091	0.150	0.006	0.058	1.01	0.199	0.000	0.062	1.00
<b>X2</b>	0.319	0.000	0.109	0.035	0.722	0.098	4.06	-	-	-	-
<b>X3</b>	0.379	0.000	0.048	0.256	0.023	0.127	4.61	0.342	0.005	0.064	1.00
<b>X4</b>	3.211	0.000	0.056	0.430	0.741	1.3	6.80	-	-	-	-

$X_1$  : diferença de posição, na tabela de classificação, do time da casa em relação ao adversário

$X_2$  : diferença de pontos do time da casa em relação ao adversário no primeiro *set* ganho pelo time que ganhou dois *sets* primeiro.

$X_3$  : diferença de pontos do time da casa em relação ao adversário no segundo *set* ganho pelo time que ganhou dois *sets* primeiro.

$X_4$  : se o time da casa ganhou dois *sets* primeiro, sim (=1) ou não (=0).

O modelo final foi ajustado somente com as duas variáveis significativas no Modelo Múltiplo 1 (Quadro 2) e verificou-se que as duas continuam significativas. Os três testes de qualidade de ajuste do modelo apresentam valor-p maior que 0.05, indicando um bom ajuste deste modelo.

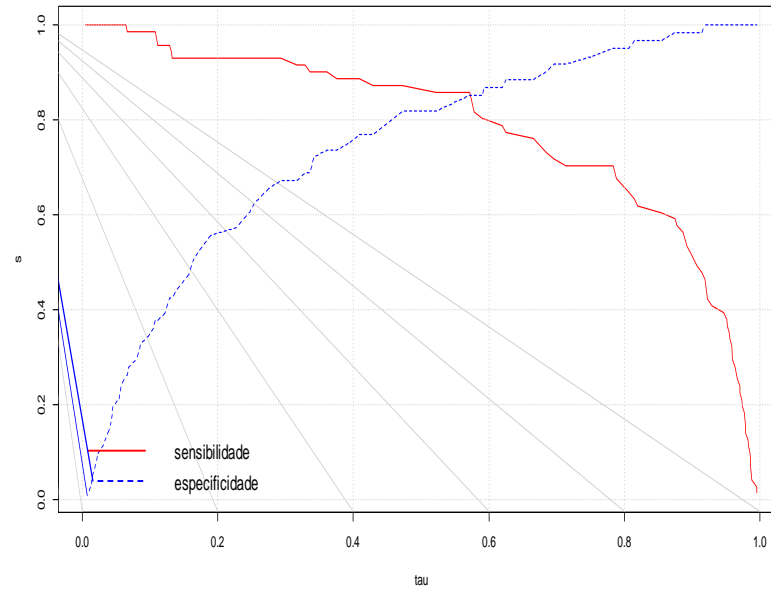
A Figura 9(b) mostra a curva ROC do modelo final. Visualmente, nota-se que a curva está muito mais próxima do canto superior esquerdo do quadrado do que da linha central. As

curvas de sensibilidade e especificidade (Figura 9(a)) mostram que, para um ponto de corte de 0.57 que, o modelo é capaz de acertar que o time da casa time vai ganhar em 85.9% das vezes e que vai perder em 85.2% da vezes.

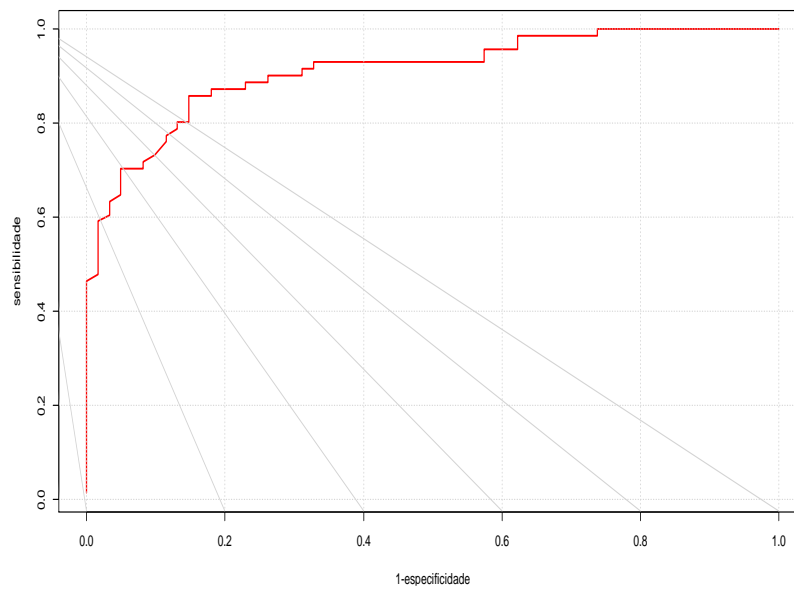
**Quadro 2:** Resultados do ajuste do modelo final.

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	84.57	42.2844	84.57	0.000
X1	1	12.41	12.4056	12.41	0.000
X3	1	45.76	45.7582	45.76	0.000
Error	129	97.66	0.7571		
Total	131	182.23			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
46.41%	45.31%	103.66			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	0.202	0.257			
X1	0.1994	0.0623	1.00		
X3	0.3425	0.0642	1.00		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2206	(1.0803, 1.3792)			
X3	1.4084	(1.2420, 1.5972)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = 0.202 + 0.1994 X1 + 0.3425 X3					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	97.66	0.982		
Pearson	129	110.69	0.876		
Hosmer-Lemeshow	8	7.06	0.530		

**Figura 9:** (a) Sensibilidade e especificidade do modelo 2 em função do ponto de corte.



(b) Curva ROC do modelo Final.



A probabilidade do time da casa vencer o jogo é estimada por:

$$P(Y = 1) = \frac{e^{0.202 + 0.1994 x_1 + 0.3425 x_3}}{1 + e^{0.202 + 0.1994 x_1 + 0.3425 x_3}}$$

Quanto à interpretação dos coeficientes do modelo, podemos dizer que: a cada aumento de uma unidade na diferença de posições, a chance do time da casa vencer o jogo aumenta em 22% considerando a posição constante; e a cada aumento em uma unidade na diferença na pontuação do segundo *set* ganho, a chance do time da casa de vencer o jogo aumenta em 40%.

A Tabela 10 apresenta as estimativas da probabilidade de vitória do time da casa segundo sua posição na tabela de classificação quando vence seu segundo *set* por 2 pontos de diferença ( $X_3 = 2$ ). Por exemplo: quando o time da casa está 11 posições acima do adversário e ganha seu segundo *set* com uma diferença de 2 pontos, podemos dizer, com 95% de confiança, que ele tem uma probabilidade entre 82.7% e 99% de ganhar o jogo. Conforme a diferença de posições na tabela diminui, a probabilidade de ganhar o jogo também diminui. Entretanto, para os valores positivos, quando o time da casa está acima na tabela, essa probabilidade não diminui de maneira drástica (Figura 10).

Quando o time da casa está 11 posições abaixo e vence seu segundo *set* por 2 pontos de diferença, o modelo indica que sua probabilidade de vitória está entre 0.06% e 52.4%, com 95% de confiança. Isto mostra que, quando o time da casa está muito abaixo na tabela e vence seu segundo *set* por uma diferença pequena, a probabilidade de vitória desse time é baixa.

Quando o time da casa está abaixo na tabela ( $X_1 < 0$ ), mesmo que a diferença de posições diminua, a probabilidade de vitória desse time estará no máximo entre 57.7% e 78%. O aumento da probabilidade em função das diferenças de posições quando o time da casa vence seu segundo *set* por uma diferença de 2 pontos é mostrado na Figura 10.

**Tabela 10-** Previsões das probabilidades de vitória para o time da casa segundo a posição na tabela de classificação considerando uma diferença de 2 pontos no segundo set ganho

X1	X3	Estimativa Pontual	Erro Padrão	IC 95%
11	2	0.956	0.033	(0.827 , 0.990)
10	2	0.947	0.036	(0.814 , 0.986)
9	2	0.936	0.040	(0.800 , 0.982)
8	2	0.923	0.043	(0.785 , 0.975)
7	2	0.907	0.046	(0.769 , 0.967)
6	2	0.889	0.049	(0.751 , 0.955)
5	2	0.868	0.052	(0.731 , 0.941)
4	2	0.844	0.054	(0.709 , 0.923)
3	2	0.815	0.055	(0.683 , 0.900)
2	2	0.783	0.056	(0.654 , 0.874)
1	2	0.748	0.058	(0.618 , 0.844)
-1	2	0.665	0.066	(0.527 , 0.780)
-2	2	0.620	0.073	(0.471 , 0.749)
-3	2	0.572	0.081	(0.411 , 0.719)
-4	2	0.522	0.091	(0.348 , 0.691)
-5	2	0.473	0.101	(0.289 , 0.665)
-6	2	0.423	0.110	(0.233 , 0.640)
-7	2	0.376	0.117	(0.184 , 0.616)
-8	2	0.330	0.122	(0.143 , 0.592)
-9	2	0.288	0.124	(0.110 , 0.570)
-10	2	0.249	0.123	(0.083 , 0.547)
-11	2	0.213	0.120	(0.062 , 0.525)

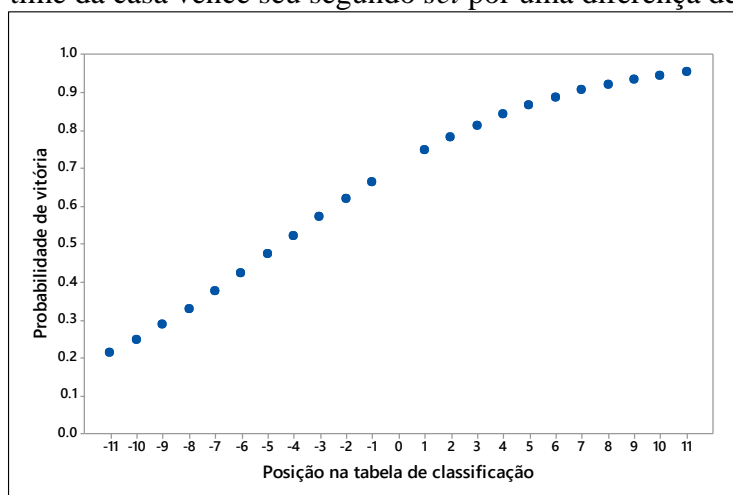
X<sub>1</sub> : Posição do time da casa em relação ao oponente

X<sub>3</sub> : Diferença de 2 pontos no segundo sets ganho

IC 95%: Intervalo de confiança de 95%

FONTE: Dados de pesquisa,2016

**Figura 10:** Probabilidade de vitória do time da casa em função das diferenças de posições na tabela quando o time da casa vence seu segundo set por uma diferença de 2 pontos



A Tabela 11 apresenta as estimativas para o time da casa segundo a tabela de classificação quando ele vence seu segundo set por 6 pontos de diferença ( $X_3 = 6$ ). Por exemplo: quando o time da casa está 5 posições acima e vence seu segundo set por 6 pontos

de diferença, o modelo indica que a probabilidade de vitória esta entre 89% e 98.8%, com 95% de confiança. Mesmo que a diferença de posições entre o time da casa e o oponente seja pequena, por exemplo,  $X_1 = 1$ , a probabilidade de vitória apresentada pelo modelo nunca fica abaixo de 80%, indicando que 6 pontos no segundo set ganho é uma diferença considerável para o aumento da probabilidade de vitória do time da casa.

De acordo com o modelo, se o time da casa está apenas uma posição abaixo do adversário e ganha seu segundo *set* por 6 pontos, sua probabilidade de vitória estará entre 75.5% e 95.2%. Isto mostra que vencer seu segundo *set* por uma diferença maior aumenta sua probabilidade de vencer o jogo, mesmo estando abaixo na tabela em relação ao adversário. Conforme a diferença de posições aumenta, as probabilidades de vitória diminuem (Figura 11).

**Tabela 11-** Previsões das probabilidades de vitória para o time da casa segundo a posição na tabela de classificação considerando uma diferença de 6 pontos no segundo set ganho

X1	X3	Estimativa Pontual	Erro Padrão	IC 95%
11	6	0.988	0.010	(0.939 , 0.997)
10	6	0.986	0.011	(0.934 , 0.997)
9	6	0.983	0.013	(0.927 , 0.996)
8	6	0.979	0.015	(0.920 , 0.994)
7	6	0.975	0.017	(0.911 , 0.993)
6	6	0.969	0.019	(0.901 , 0.990)
5	6	0.963	0.021	(0.890 , 0.988)
4	6	0.955	0.024	(0.877 , 0.984)
3	6	0.946	0.027	(0.860 , 0.979)
2	6	0.934	0.031	(0.841 , 0.974)
1	6	0.921	0.035	(0.818 , 0.967)
-1	6	0.887	0.048	(0.755 , 0.952)
-2	6	0.865	0.056	(0.713 , 0.942)
-3	6	0.840	0.067	(0.664 , 0.932)
-4	6	0.811	0.079	(0.608 , 0.922)
-5	6	0.779	0.094	(0.547 , 0.911)
-6	6	0.743	0.111	(0.480 , 0.900)
-7	6	0.703	0.129	(0.413 , 0.888)
-8	6	0.660	0.148	(0.347 , 0.876)
-9	6	0.614	0.167	(0.285 , 0.863)
-10	6	0.565	0.185	(0.229 , 0.850)
-11	6	0.516	0.200	(0.181 , 0.836)

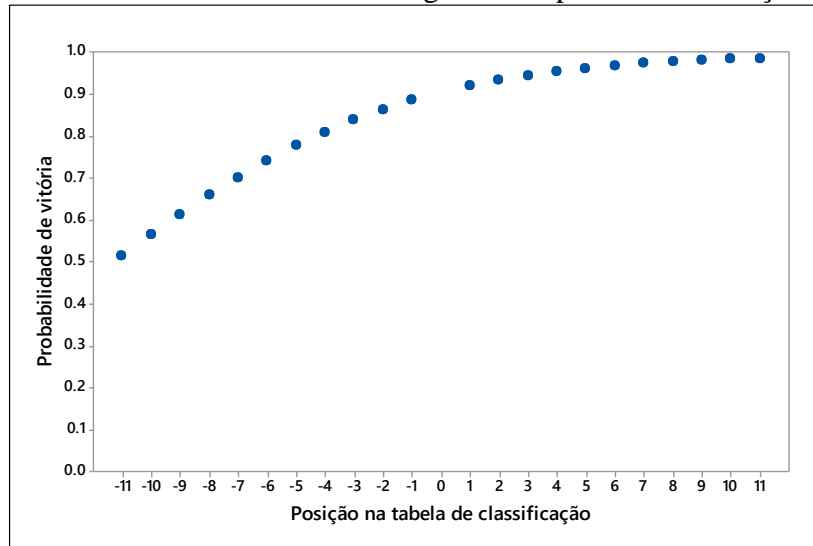
X<sub>1</sub> : Posição do time da casa em relação ao oponente

X<sub>3</sub> : Diferença de 6 pontos no segundo *sets* ganho

IC 95%: Intervalo de confiança de 95%

FONTE: Dados de pesquisa,2016

**Figura 11:** Probabilidade de vitória do time da casa em função das diferenças de posições na tabela quando o time da casa vence seu segundo *set* por uma diferença de 2 pontos



### 6.3. Terceiro Caso

Diferentemente do primeiro e do segundo caso, os jogos analisados são todos de seleções nacionais. Além disso, foram considerados apenas os 82 jogos realizados em campo neutro, ou seja, não temos o fator jogar em casa.

A variável resposta  $Y$  indica se o *time que ganhou dois sets primeiro* venceu o jogo ( $Y=1$ ) ou não ( $Y=0$ ). A base de dados apresenta três potenciais variáveis explicativas:

- $X_1$ : diferença de posição, na tabela de classificação, do time que ganhou os dois *sets* primeiro (valores inteiros menores ou iguais a -1 ou maiores ou iguais a 1);
- $X_2$ : diferença de pontos no primeiro *set* ganho pelo time que ganhou dois *sets* primeiro (valores inteiros maiores ou iguais a 2);
- $X_3$ : diferença de pontos no segundo *set* ganho pelo time que ganhou *sets sets* primeiro (valores inteiros maiores ou iguais a 2).

#### 6.3.1. Análises Descritivas

Dentre os 82 jogos realizados em campo neutro, em 72 (88%) o time que ganhou dois *sets* primeiro ganhou o jogo (tabela 12). Esta porcentagem é um pouco maior que o percentual de 84% encontrada na Seção 6.1, quando não havia jogos em campo neutro.

Em relação à posição dos times na tabela de classificação ( $X_1$ ), na Figura 12 observamos que a proporção de vitórias dos times são muito parecidas entre as faixas de diferenças de posições. A Tabela 12 mostra o resultado final dos times que ganharam dois *sets* primeiro olhando apenas se estes times estavam abaixo ou acima na tabela de classificação. Nota-se que existe uma diferença muito pequena entre o percentual de vitórias dos times que estão abaixo ou acima na tabela de classificação. O teste do Qui-quadrado das proporções de vitória apresentou valor-p igual a 0.726 (Apendice: Quadro 13), o que leva a concluir que as proporções não são significativamente diferentes.

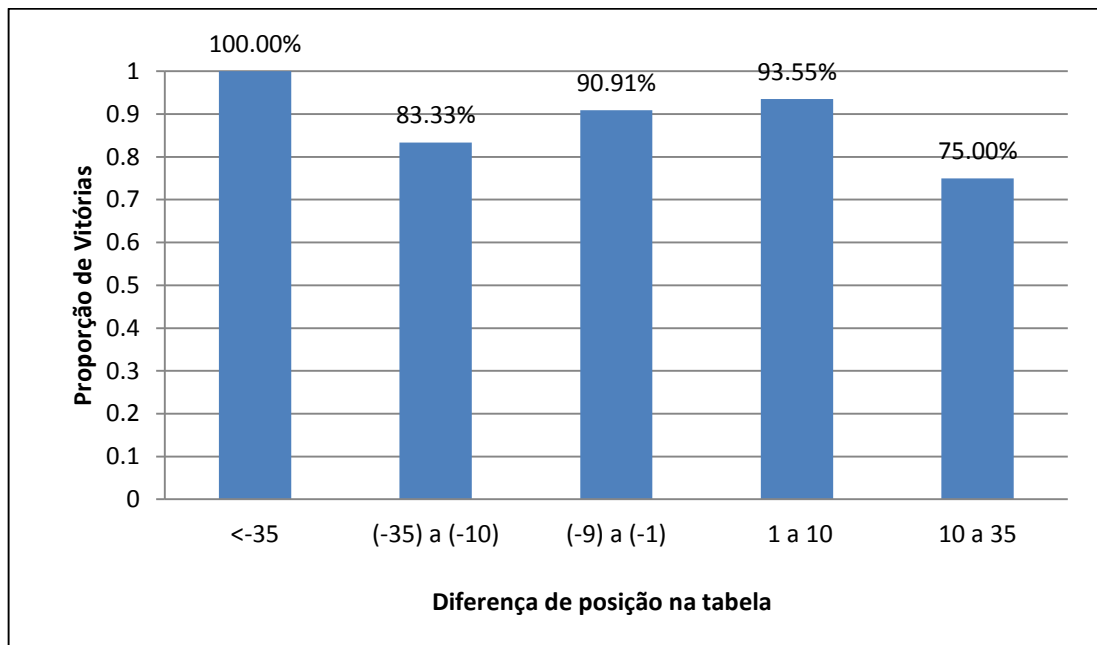


**Tabela 12.** Número de jogos segundo resultado final do time que ganhou dois *sets* primeiro, estratificados por posição na tabela de classificação em relação ao adversário

Posição na tabela de classificação*	Resultado do final do time		Total
	Perdeu (Y=0)	Venceu (Y=1)	
Abaixo ( $X_1 < 0$ )	4 (11%)	33 (89%)	37 (100%)
Acima ( $X_1 > 0$ )	6 (13%)	39 (87%)	45 (100%)
<b>Total</b>	10 (12%)	72 (88%)	82 (100%)

\*Do time que venceu dois *sets* primeiro. FONTE: Dados de pesquisa, 2016.

**Figura 12.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo a diferença de posições na tabela de classificação



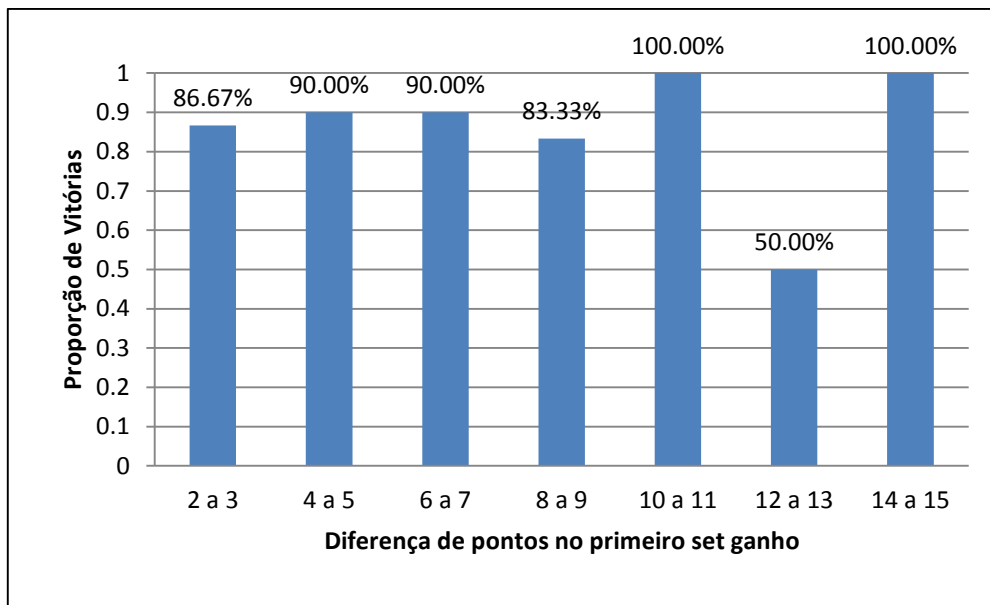
As figuras 13 e 14 mostram a proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores diferença de pontos, no primeiro *set* ganho ( $X_2$ ) e no segundo *set* ganho ( $X_3$ ), respectivamente.

Na Figura 13 é possível perceber uma pequena tendência de aumento da proporção com o aumento nas diferenças de pontos. Entretanto, os valores das proporções são muito parecidos. Além disso, essa tendência é inesperadamente interrompida para valores entre 12 e 13 pontos. Ambas observações indicam que esta variável não é importante para explicar a probabilidade de vitória no jogo do time que ganhou dois *sets* primeiro.

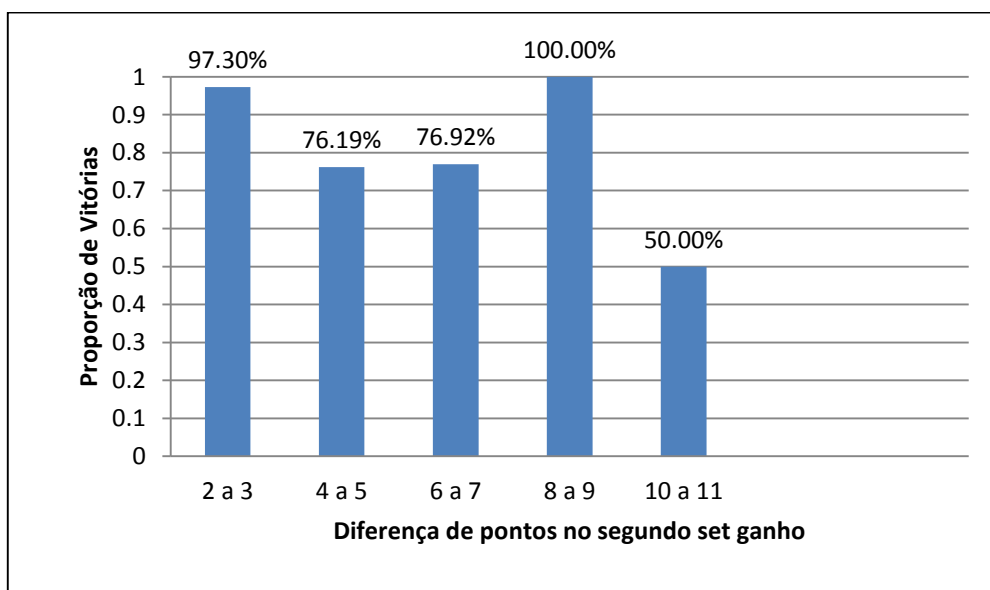
Na Figura 14, observa-se uma oscilação da proporção com o aumento nas diferenças de pontos, indicando que ganhar o segundo *set*, após já ter vencido um *set*, por uma

diferença grande ou pequena não influencia no resultado final do jogo. Isso fica claro pois observou-se que a proporção de vitórias dos times que ganharam o segundo *set* por uma diferença entre 2 e 3 pontos é similar a proporção dos times que ganharam por uma diferença entre 8 e 9 pontos e maior do que a dos times que ganharam por uma diferença entre 10 e 11 pontos.

**Figura 13.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores diferença de pontos no primeiro *set* ganho



**Figura 14.** Proporção de vitórias do time que ganhou dois *sets* primeiro segundo os valores diferença de pontos no segundo *set* ganho



### 6.3.2. Modelos de Regressão Logística

A Tabela 12 mostra os valores dos coeficientes estimados nos modelos de regressão logística para as três variáveis explicativas separadamente (modelos individuais). Nota-se que nenhuma das variáveis foi significativa a 5%.

Para verificar se há efeito conjunto das três variáveis explicativas, foi ajustado um modelo de regressão logística múltipla com todas variáveis da base de dados (Tabela 12; Apendice: Quadro 14). Entretanto, nenhuma das variáveis foi significativa a 5%.

**Tabela 13.** Valores dos coeficientes estimados nos modelos de regressão logística

Variável Explicativa	Modelos Individuais		Modelo Múltiplo (X1, X2, X3)		
	Coef.	Valor-p	Coef.	Valor-p	VIF
<b>X1</b>	-0.014	0.602	-0.0022	0.937	1.19
<b>X2</b>	-0.030	0.805	-0.051	0.704	1.13
<b>X3</b>	-0.153	0.245	-0.158	0.265	1.18

## 7. CONCLUSÕES

### 7.1 Conclusões do Primeiro Caso

Encontrou-se que o time que ganha dois *sets* primeiro vence o jogo 84% das vezes. Há uma diferença de percentual para o time da casa (86%) e para o time de fora (82%) mas essa diferença não é significativa. A diferença de pontos nos dois *sets* ganhos primeiro também é significativa para explicar a variabilidade da variável resposta. Entretanto, quando o time que vence dois *sets* primeiro está acima do adversário na tabela de classificação, a probabilidade de vitória sobe para 90%, contra 73% de quem está abaixo na tabela.

O modelo de regressão logística ajustado indica que, conforme aumenta a diferença de posições na tabela entre o time que ganhou dois *sets* primeiro e o oponente, aumenta a probabilidade de vitória do primeiro. Se o penúltimo colocado joga contra o líder e ganha dois *sets* primeiro, a chance de vencer o jogo é 48%. Quando o time que está em segundo na tabela e joga contra o último e vence dois *sets* primeiro, o modelo estima com uma confiança de 95% que a chance desse time ganhar o jogo está entre 87% e 98.7%. Ou seja, times mais fortes dificilmente deixam a vitória escapar depois de vencer dois *sets* no jogo. Para times mais fracos, mesmo que ganhem dois *sets* primeiro, é muito mais difícil manter o ritmo para vencer o jogo.

Apesar deste modelo ter sido considerado bem ajustado nos três testes realizados, ele apresentou um valor de  $R^2$  ajustado muito baixo (7.79%) e uma Curva ROC insatisfatória, indicando que não se pode prever de maneira confiável a probabilidade do time que ganhou dois *sets* primeiro ganhar o jogo somente em função da diferença de posições na tabela de classificação.

### 7.2 Conclusões do Segundo Caso

Apesar de ser esperado que jogar em casa fizesse grande diferença no resultado final de uma partida, isto não foi verificado no caso da SLMV. O time da casa ganhou seus jogos em apenas 54% das vezes.

O time da casa e o time de fora têm um percentual de vitórias muito semelhante quando vencem dois *sets* primeiro, 84% para o time da casa contra 83% do time de fora. Isso demonstra que ganhar dois *sets* primeiro é extremamente importante, entretanto, não faz

diferença se está jogando em casa ou não. Ganhar dois *sets* primeiro não aumenta significativamente a probabilidade do time da casa ganhar o jogo.

Quando se joga em casa, a diferença de posições é significativa no resultado final do jogo do time da casa. Mesmo quando o time da casa está entre 3 e 5 posições abaixo de seu oponente, seu percentual de vitórias é de 41%. Quando está entre uma e duas posições abaixo, o time da casa tem um percentual de 52%.

Ganhar o segundo *set* por uma boa diferença de pontos aumenta as chances do time da casa vencer o jogo. A diferença de pontuação na vitória do segundo *set* ganho, caso já houvesse ganho um *set* antes, aumenta a chance de vitória em 40% para cada ponto a mais.

Quando analisamos as diferenças de posições na tabela de classificação, conjuntamente com a diferença de pontos na vitória do segundo *set* ganho, nota-se que, mesmo quando existe uma diferença positiva pequena na tabela de classificação entre o time da casa e seu oponente, o time da casa tem uma probabilidade alta de ganhar o jogo se tiver obtido uma boa diferença de pontuação em seu segundo *set* ganho. Quando o time da casa está muito abaixo de seu adversário na tabela, mesmo que ganhe o seu segundo *set*, é necessário que o faça com uma boa diferença de pontos.

### **7.3 Conclusões do Terceiro Caso**

No caso de jogos entre seleções internacionais, as variáveis escolhidas para o estudo não são suficientes para determinar o resultado final de um jogo do time que ganhou dois *sets* primeiro.

Ganhar dois *sets* primeiro com uma diferença grande ou pequena de pontos não interfere no resultado final do time que ganhou dois *sets* primeiro.

Estar acima ou abaixo do oponente na tabela da classificação não influencia na probabilidade de vitória, diferentemente dos jogos entre clubes nacionais, indicando que tal variável interfere no resultado somente quando temos o fator “casa”.

## 8. CONSIDERAÇÕES FINAIS

Em geral, os resultados encontrados indicam que as variáveis estudadas não são suficientes para tentar prever a probabilidade de vitória de um time. Provavelmente, fatores ligados ao lado psicológico dos atletas e em que momento do campeonato ocorre o jogo também devem interferir nessa probabilidade. Além disso, fatores como tempo de viagem para um jogo e tempo de descanso entre um jogo e outro também sejam fatores a ser analisados em um novo estudo.

É recomendado, também, um estudo mais profundo no que diz respeito ao fator “casa”. Fatores como as dimensões do ginásio, quantidade de pessoas assistindo ao jogo e a idade média do time de fora poderiam contribuir para um modelo de previsão melhor.

Uma vez que eu sei que vencer dois *sets* torna a recuperação do adversário mais difícil, mesmo que o time adversário esteja muito acima na tabela, uma estratégia mais arrojada para buscar a vitória nos *sets* iniciais contra esses times pode trazer uma vitória inesperada, levando em consideração que o oponente acima da tabela tem um time mais forte.

Os fatores que interferem no resultado final do jogo ainda não estão claramente explicados. Este estudo consegue identificar apenas alguns fatores que podem influenciar no aumento da probabilidade de vitória de um time.

## 9. REFERÊNCIAS

EOM, H. J., & SCHUTZ, R. W. Statistical analyses of volleyball team performance. *Research Quarterly for Exercise and Sport*, v.63, n.1, p. 11–18. Mar.1992.

HAIR, J. F. et al. *Análise multivariada de dados*. 5 Ed. Porto Alegre: Bookman, 2005. 593p.

HOSMER, D.W.Jr.; LEMESHOW, S. *Applied logistic regression*. 2ª Ed. New York: Wiley-Interscience 2000, p.375

HUGHES, M., & DANIEL, R. Playing patterns of elite and non elite volleyball. *International Journal of Performance Analysis in Sport*, v.3, n.1, p.50–56. 2003

MARCELINO R, MESQUITA I, SAMPAIO J. Estudo dos indicadores de rendimento em Voleibol masculino em função do número do *set*. *Rev. bras. Educ. Fís. Esporte*, São Paulo, v.24, n.1, p.69-78, Jan./Mar. 2010

MARCELINO, R., MESQUITA, I., PALAO, J. M., & SAMPAIO, J. Home advantage in high-level volleyball varies according to *set* number. *Journal of Sports Science and Medicine*, 8, p.352-356. *Set*.2009

MARCELINO, R., MESQUITA, I., SAMPAIO, J., & ANGUERA, M. Home advantage in high-level volleyball [Ventaja de jugar en casa en voleibol de alto rendimiento]. *Revista de Psicología del Deporte*, v.18, n.2, p.181-196. Jul.2009

MEDEIROS, A. I. A. *Performance desportiva no voleibol de praia masculino de alto rendimento: Estudo aplicado nos escalões sub-19, sub-21 e senior*. 2014. 117 f. Tese (Doutorado em Ciências do Desporto) - Faculdade de Desporto da Universidade do Porto, Porto – Portugal, 2014.

PALAO, J. M., SANTOS, J. A., & URENA, A. Effect of team level on skill performance in volleyball. *International Journal of Performance Analysis in Sport*, v.4, n.2, p.50–60. Nov.2004.

ZADRAŽNIK, M., MARELIĆ, N., & REŠETAR, T. Differences in rotations between the winning and losing teams at the youth European volleyball championships for girls. *Acta Universitatis Palackiane Olomucensis Gymnica*, v.38, n.4, p.33-40, Abril. 2009.

Estatísti.co, Estatística no Esporte. Disponível em: <<http://www.estatisti.co/2013/04/estatistica-no-esporte.html>>. Acesso em 6 de agosto de 2016.

Estatísti.co, Estatística na Psicologia. Disponível em: <<http://www.estatisti.co/2013/04/a-estatistica-na-psicologia.html>>. Acesso em 6 de agosto de 2016.

IDRE - Institute for Digital Research and Education, Odds Ratio. Disponível em: <[http://www.ats.ucla.edu/stat/mult\\_pkg/faq/general/odds\\_ratio.htm](http://www.ats.ucla.edu/stat/mult_pkg/faq/general/odds_ratio.htm)>. Acesso em 3 de julho de 2016.

Ministério Do Esporte. Disponível em: < <http://www.esporte.gov.br/diesporte/2.html>>. Acesso em 5 de outubro de 2016.

R7 Esportes, No vôlei, sucesso de Bernardinho está intimamente ligado à estatística. Disponível em: <<http://esportes.r7.com/futebol/noticias/no-volei-sucesso-de-bernardinho-esta-intimamente-ligado-a-estatistica-20110415.html>>. Acesso em 20 de Julho de 2016.



## APÊNDICES

### A - Primeiro Caso

**Quadro 1:** Resultados do ajuste do Modelo 1 (X1, X2, X3 e X4)

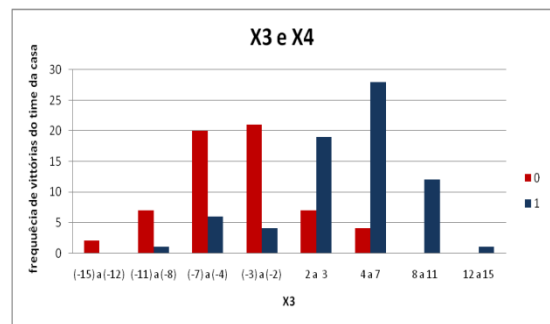
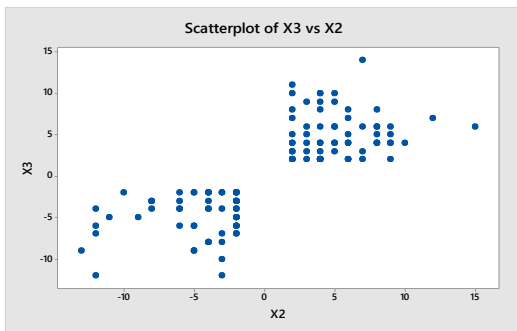
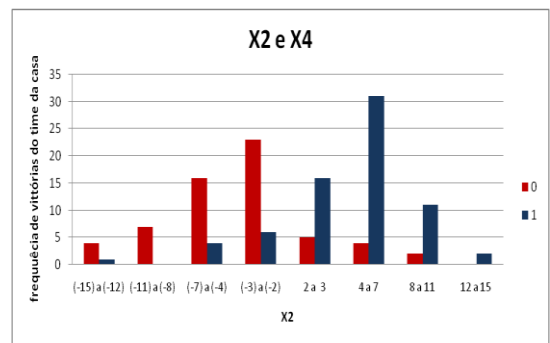
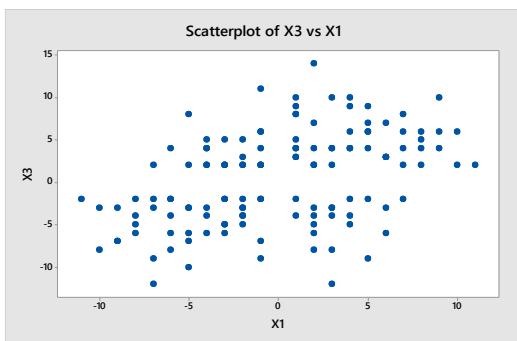
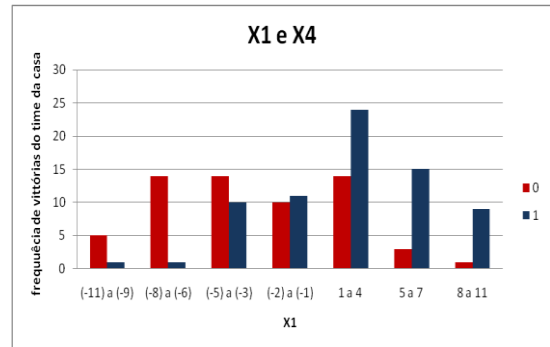
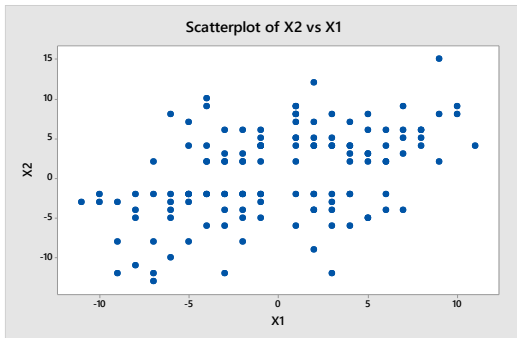
Method						
Link function	Logit					
Rows used	132					
Response Information						
Variable	Value	Count				
Y	1	111	(Event)			
	0	21				
	Total	132				
Deviance Table						
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value	
Regression	4	10.898	2.72447	10.90	0.028	
X1	1	0.063	0.06296	0.06	0.802	
X2	1	1.507	1.50728	1.51	0.220	
X3	1	0.277	0.27739	0.28	0.598	
X4	1	6.949	6.94908	6.95	0.008	
Error	127	104.776	0.82501			
Total	131	115.674				
Model Summary						
Deviance	Deviance					
R-Sq	R-Sq(adj)	AIC				
9.42%	5.96%	114.78				
Coefficients						
Term	Coef	SE Coef	VIF			
Constant	0.717	0.684				
X1	0.0241	0.0966	1.04			
X2	0.128	0.110	1.04			
X3	0.264	0.501	1.01			
X4	0.1433	0.0571	1.04			
Odds Ratios for Continuous Predictors						
	Odds Ratio	95% CI				
X1	1.0243	(0.8477, 1.2378)				
X2	1.1371	(0.9159, 1.4117)				
X3	1.3016	(0.4878, 3.4729)				
X4	1.1540	(1.0318, 1.290)				
Goodness-of-Fit Tests						
Test	DF	Chi-Square	P-Value			
Deviance	127	104.78	0.925			
Pearson	127	121.35	0.625			
Hosmer-Lemeshow	8	12.08	0.148			

**Quadro 2: Resultados do ajuste Modelo Múltiplo 2 (X2 e X4)**

Method						
Link function	Logit					
Rows used	132					
Response Information						
Variable	Value	Count				
Y	1	111	(Event)			
	0	21				
	Total	132				
Deviance Table						
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value	
Regression	2	10.533	5.2663	10.53	0.005	
X2	1	1.522	1.5225	1.52	0.217	
X4	1	7.410	7.4101	7.41	0.006	
Error	129	105.142	0.8151			
Total	131	115.674				
Model Summary						
Deviance	Deviance					
R-Sq	R-Sq(adj)	AIC				
9.11%	7.38%	111.14				
Coefficients						
Term	Coef	SE Coef	VIF			
Constant	0.961	0.512				
X2	0.128	0.110	1.03			
X4	0.1464	0.0567	1.03			
Odds Ratios for Continuous Predictors						
	Odds Ratio	95% CI				
X2	1.1370	(0.9171, 1.4098)				
X4	1.1577	(1.0359, 1.2937)				
Regression Equation						
P(1) = exp(Y') / (1 + exp(Y'))						
Y' = 0.961 + 0.128 X2 + 0.1464 X4						
Goodness-of-Fit Tests						
Test	DF	Chi-Square	P-Value			
Deviance	129	105.14	0.939			
Pearson	129	124.61	0.593			
Hosmer-Lemeshow	8	5.96	0.652			

**B - Segundo Caso**

**Figura 1:** Gráficos de correlação das variáveis do segundo caso ( X1, X2, X3 e X4)



**Quadro 3: Resultados do ajuste do Modelo 1 com todas as variáveis (X1, X2, X3 e X4)**

Method						
Link function	Logit					
Rows used	132					
Response Information						
Variable	Value	Count				
Y	1	71	(Event)			
	0	61				
	Total	132				
Deviance Table						
Source		DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression		4	84.442	21.1104	84.44	0.000
X1		1	11.165	11.1650	11.16	0.001
X2		1	0.064	0.0637	0.06	0.801
X3		1	3.898	3.8983	3.90	0.048
X4		1	0.431	0.4306	0.43	0.512
Error		127	97.457	0.7674		
Total		131	181.898			
Model Summary						
Deviance	Deviance					
R-Sq	R-Sq(adj)	AIC				
46.42%	44.22%	107.46				
Coefficients						
Term	Coef	SE Coef	VIF			
Constant	-0.229	0.717				
X1	0.1928	0.0632	1.02			
X2	0.025	0.101	4.02			
X3	0.227	0.127	4.50			
X4	0.88	1.34	6.78			
Odds Ratios for Continuous Predictors						
	Odds Ratio	95% CI				
X1	1.2127	(1.0715, 1.3725)				
X2	1.0257	(0.8411, 1.2507)				
X3	1.2543	(0.9779, 1.6088)				
X4	2.4121	(0.1745, 33.3397)				
Regression Equation						
P(1) = $\exp(Y') / (1 + \exp(Y'))$						
Y' = -0.229 + 0.1928 X1 + 0.025 X2 + 0.227 X3 + 0.88 X4						
Goodness-of-Fit Tests						
Test	DF	Chi-Square	P-Value			
Deviance	127	97.46	0.976			
Pearson	127	113.62	0.796			
Hosmer-Lemeshow	8	6.59	0.581			

**Quadro 4: Resultados do ajuste do Modelo com X1 e X2**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	75.47	37.7327	75.47	0.000
X1	1	14.74	14.7428	14.74	0.000
X2	1	36.65	36.6548	36.65	0.000
Error	129	106.77	0.8277		
Total	131	182.23			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
41.41%	40.31%	112.77			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	0.063	0.243			
X1	0.2102	0.0603	1.00		
X2	0.2889	0.0575	1.00		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2339	(1.0964, 1.3887)			
X2	1.3349	(1.1927, 1.4941)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = 0.063 + 0.2102 X1 + 0.2889 X2					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	106.77	0.924		
Pearson	129	137.13	0.296		
Hosmer-Lemeshow	8	8.84	0.356		

**Quadro 5: Resultados do ajuste do Modelo com X1 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	81.37	40.6865	81.37	0.000
X1	1	14.52	14.5215	14.52	0.000
X4	1	42.56	42.5625	42.56	0.000
Error	129	100.86	0.7819		
Total	131	182.23			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
44.65%	43.56%	106.86			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-1.332	0.358			
X1	0.2176	0.0632	1.01		
X4	2.976	0.510	1.01		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2430	(1.0983, 1.4069)			
X4	19.6015	(7.2149, 53.2537)			
Regression Equation					
P(1) = exp(Y') / (1 + exp(Y'))					
Y' = -1.332 + 0.2176 X1 + 2.976 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	100.86	0.968		
Pearson	129	140.67	0.227		
Hosmer-Lemeshow	8	6.26	0.618		

**Quadro 6:** Resultados do ajuste do Modelo com X2 e X3

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	73.288	36.6442	73.29	0.000
X2	1	1.125	1.1253	1.13	0.289
X3	1	12.566	12.5658	12.57	0.000
Error	129	108.944	0.8445		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
40.22%	39.12%	114.94			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	0.100	0.244			
X2	0.0850	0.0814	2.74		
X3	0.2979	0.0983	2.74		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X2	1.0887	(0.9281, 1.2770)			
X3	1.3470	(1.1109, 1.6333)			
Regression Equation					
P(1) = exp(Y') / (1 + exp(Y'))					
Y' = 0.100 + 0.0850 X2 + 0.2979 X3					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	108.94	0.899		
Pearson	129	128.76	0.489		
Hosmer-Lemeshow	8	4.58	0.801		

**Quadro 7: Resultados do ajuste do Modelo com X2 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	67.879	33.9393	67.88	0.000
X2	1	1.027	1.0271	1.03	0.311
X4	1	7.156	7.1560	7.16	0.007
Error	129	114.354	0.8865		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
37.25%	36.15%	120.35			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-1.148	0.502			
X2	0.0890	0.0917	3.83		
X4	2.503	0.936	3.83		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X2	1.0931	(0.9132, 1.3084)			
X4	12.2175	(1.9493, 76.5752)			
Regression Equation					
P(1) = exp(Y') / (1 + exp(Y'))					
Y' = -1.148 + 0.0890 X2 + 2.503 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	114.35	0.818		
Pearson	129	132.38	0.401		
Hosmer-Lemeshow	6	4.81	0.568		



**Quadro 8: Resultados do ajuste do Modelo com X3 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	2	73.096	36.5482	73.10	0.000
X3	1	6.245	6.2449	6.24	0.012
X4	1	0.933	0.9333	0.93	0.334
Error	129	109.136	0.8460		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
40.11%	39.01%	115.14			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-0.381	0.577			
X3	0.274	0.124	4.60		
X4	1.02	1.05	4.60		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X3	1.3158	(1.0311, 1.6791)			
X4	2.7635	(0.3562, 21.4383)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = -0.381 + 0.274 X3 + 1.02 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	129	109.14	0.897		
Pearson	129	125.49	0.571		
Hosmer-Lemeshow	7	6.43	0.491		

**Quadro 9: Resultados do ajuste do Modelo com X1, X2 e X3**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	85.228	28.4094	85.23	0.000
X1	1	11.940	11.9397	11.94	0.001
X2	1	0.659	0.6594	0.66	0.417
X3	1	9.763	9.7628	9.76	0.002
Error	128	97.004	0.7578		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
46.77%	45.12%	105.00			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	0.172	0.261			
X1	0.1991	0.0632	1.00		
X2	0.0684	0.0852	2.67		
X3	0.275	0.101	2.67		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2202	(1.0782, 1.3810)			
X2	1.0708	(0.9061, 1.2654)			
X3	1.3159	(1.0797, 1.6038)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = 0.172 + 0.1991 X1 + 0.0684 X2 + 0.275 X3					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	128	97.00	0.981		
Pearson	128	110.51	0.865		
Hosmer-Lemeshow	8	5.07	0.750		

**Quadro 10: Resultados do ajuste do Modelo com X1, X2 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	81.644	27.2148	81.64	0.000
X1	1	13.766	13.7657	13.77	0.000
X2	1	0.271	0.2713	0.27	0.602
X4	1	6.179	6.1789	6.18	0.013
Error	128	100.588	0.7858		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
44.80%	43.16%	108.59			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-1.116	0.546			
X1	0.2145	0.0636	1.02		
X2	0.0511	0.0999	4.00		
X4	2.51	1.02	4.04		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2393	(1.0940, 1.4039)			
X2	1.0525	(0.8653, 1.2802)			
X4	12.3434	(1.6744, 90.9933)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = -1.116 + 0.2145 X1 + 0.0511 X2 + 2.51 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	128	100.59	0.965		
Pearson	128	135.98	0.298		
Hosmer-Lemeshow	8	8.98	0.344		

**Quadro 11: Resultados do ajuste do Modelo com X1, X3 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	85.462	28.4874	85.46	0.000
X1	1	12.366	12.3659	12.37	0.000
X3	1	4.089	4.0892	4.09	0.043
X4	1	0.894	0.8936	0.89	0.345
Error	128	96.770	0.7560		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
46.90%	45.25%	104.77			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-0.333	0.618			
X1	0.2032	0.0637	1.01		
X3	0.231	0.127	4.44		
X4	1.04	1.09	4.47		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X1	1.2254	(1.0815, 1.3884)			
X3	1.2603	(0.9834, 1.6152)			
X4	2.8300	(0.3327, 24.0719)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = -0.333 + 0.2032 X1 + 0.231 X3 + 1.04 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	128	96.77	0.982		
Pearson	128	116.22	0.764		
Hosmer-Lemeshow	8	3.89	0.867		

**Quadro 12: Resultados do ajuste do Modelo com X2, X3 e X4**

Method					
Link function	Logit				
Rows used	132				
Response Information					
Variable	Value	Count			
Y	1	71	(Event)		
	0	61			
	Total	132			
Deviance Table					
Source	DF	Adj Dev	Adj Mean	Chi-Square	P-Value
Regression	3	73.497	24.4989	73.50	0.000
X2	1	0.400	0.4003	0.40	0.527
X3	1	5.618	5.6180	5.62	0.018
X4	1	0.208	0.2082	0.21	0.648
Error	128	108.736	0.8495		
Total	131	182.233			
Model Summary					
Deviance	Deviance				
R-Sq	R-Sq(adj)	AIC			
40.33%	38.68%	116.74			
Coefficients					
Term	Coef	SE Coef	VIF		
Constant	-0.180	0.659			
X2	0.0596	0.0964	4.03		
X3	0.261	0.124	4.61		
X4	0.58	1.26	6.65		
Odds Ratios for Continuous Predictors					
	Odds Ratio	95% CI			
X2	1.0614	(0.8788, 1.2820)			
X3	1.2982	(1.0176, 1.6562)			
X4	1.7786	(0.1510, 20.9539)			
Regression Equation					
P(1) = $\exp(Y') / (1 + \exp(Y'))$					
Y' = -0.180 + 0.0596 X2 + 0.261 X3 + 0.58 X4					
Goodness-of-Fit Tests					
Test	DF	Chi-Square	P-Value		
Deviance	128	108.74	0.890		
Pearson	128	125.97	0.534		
Hosmer-Lemeshow	8	8.76	0.363		

## C - Terceiro Caso

**Quadro 13:** Teste de proporções e IC: Y, X<sub>1</sub>

```

Event = 1

acima/baixo   X   N   Sample p
0              33  37  0.891892
1              39  45  0.866667

Difference = p (0) - p (1)
Estimate for difference:  0.0252252
95% CI for difference:  (-0.115754, 0.166205)
Test for difference = 0 (vs ≠ 0):  Z = 0.35  P-Value = 0.726

* NOTE * The normal approximation may be inaccurate for small samples.

Fisher's exact test: P-Value = 1.000

```

**Quadro 14:** Resultados do ajuste do Modelo Múltiplo

```

Link function  Logit
Rows used     82

Response Information

Variable              Value  Count
Y                    1       72  (Event)
                    0       10
                    Total    82

Deviance Table

Source              DF  Adj Dev  Adj Mean  Chi-Square  P-Value
Regression          3  1.5330  0.51101   1.53        0.675
  X1                 1  0.0063  0.00634   0.01        0.937
  X2                 1  0.1442  0.14415   0.14        0.704
  X3                 1  1.2449  1.24487   1.24        0.265
Error               78  59.2773  0.75997
Total               81  60.8103

Model Summary

Deviance    Deviance
  R-Sq  R-Sq(adj)    AIC
  2.52%    0.00%  67.28

Coefficients

Term              Coef  SE Coef  VIF
Constant          3.00   1.16
X1                -0.0022  0.0280  1.19
X2                -0.051   0.133  1.13
X3                -0.158   0.140  1.18

Odds Ratios for Continuous Predictors

              Odds Ratio      95% CI
X1            0.9978  (0.9445, 1.0541)
X2            0.9504  (0.7325, 1.2332)
X3            0.8538  (0.6485, 1.1242)

Goodness-of-Fit Tests

Test              DF  Chi-Square  P-Value
Deviance          78    59.28    0.943
Pearson           78    79.18    0.441
Hosmer-Lemeshow   8     10.27    0.247

```