

DISTRIBUIÇÃO DE POISSON BIVARIADA APLICADA
À PREVISÃO DE RESULTADOS ESPORTIVOS

WESLEY BERTOLI DA SILVA

UFSCar - São Carlos/SP
Abril de 2014

Universidade Federal de São Carlos
Centro de Ciências Exatas e de Tecnologia
Departamento de Estatística

DISTRIBUIÇÃO DE POISSON BIVARIADA APLICADA
À PREVISÃO DE RESULTADOS ESPORTIVOS

WESLEY BERTOLI DA SILVA

ORIENTADOR

PROF. DR. LUIS ERNESTO BUENO SALASAR

COORIENTADOR

PROF. DR. ADRIANO KAMIMURA SUZUKI

Dissertação apresentada ao Departamento de Estatística da Universidade Federal de São Carlos - DEs/UFSCar - como parte dos requisitos mínimos para a obtenção do título de Mestre em Estatística.

UFSCar - São Carlos/SP
Abril de 2014

**Ficha catalográfica elaborada pelo DePT da
Biblioteca Comunitária da UFSCar**

S586dp Silva, Wesley Bertoli da.
Distribuição de Poisson bivariada aplicada à previsão de resultados esportivos / Wesley Bertoli da Silva. -- São Carlos : UFSCar, 2014.
57 f.

Dissertação (Mestrado) -- Universidade Federal de São Carlos, 2014.

1. Distribuição de Poisson. 2. Simulação. 3. Modelo de efeitos. 4. Previsão de resultados. 5. Verossimilhança ponderada. I. Título.

CDD: 519.24 (22ª)



UNIVERSIDADE FEDERAL DE SÃO CARLOS
Centro de Ciências Exatas e de Tecnologia
Programa de Pós-Graduação em Estatística
Via Washington Luís, Km 235 - C.P.676 - CGC 45358058/0001-40
FONE: (016) 3351-8292 – Email: ppgest@ufscar.br
13565-905 - SÃO CARLOS-SP - BRASIL

FOLHA DE APROVAÇÃO

Aluno(a) : Wesley Bertoli da Silva

DISSERTAÇÃO DE MESTRADO DEFENDIDA E APROVADA EM 23/04/2014
PELA COMISSÃO JULGADORA:

Presidente

Prof. Dr. Luis Ernesto Bueno Salasar (DEs-UFSCar/Orientador)

1º Examinador

Prof. Dr. Adriano Kamimura Suzuki (ICMC-USP/ Co-orientador)

2º Examinador

Prof. Dr. Erlandson Ferreira Saraiva (UFMS)

3º Examinador

Prof. Dr. Francisco Louzada Neto (ICMC-USP)

“Se você tem uma maçã e eu tenho uma maçã, e nós trocamos as maçãs, então você e eu ainda teremos uma maçã. Mas se você tem uma ideia e eu tenho uma ideia, e nós trocamos essas ideias, então cada um de nós terá duas ideias.”

Agradecimentos

Gostaria de expressar meus sinceros agradecimentos

aos meus pais, Edmilson e Sirlene, pela vida, pelo amor e apoio incondicionais, por compreenderem a ausência do primeiro filho ao longo dos últimos anos e por serem os maiores exemplos de luta, trabalho e honestidade que eu levo para a vida,

aos meus familiares que me acompanham e apoiam desde o início e que, apesar da ausência, da distância, acreditam na minha capacidade de ultrapassar barreiras, vencer desafios e de fazer sempre o melhor,

à minha namorada Deisy, também pelo amor e apoio incondicionais, pelo carinho e paciência, pelo companherismo, por ser minha inspiração, por sempre se fazer presente e, principalmente, por me fazer uma pessoa melhor e um homem feliz,

ao Professor Luis Ernesto pela orientação, pelas idéias perspicazes, pelas críticas, pela paciência, pelo exemplo de dedicação/disciplina e, principalmente, por acreditar em minha capacidade para desenvolver este trabalho; ao Professor Adriano pela coorientação e pelas ideias sempre muito construtivas,

aos Professores Galvão, Carlos Diniz, Márcio Diniz, Luis Milan, Vera Lucia, Marinho e Adriano Polpo que, de alguma forma, contribuíram para a minha formação e para crescimento de meu conhecimento,

à Isabel, secretária do Programa de Pós-Graduação em Estatística (PPGEst), responsável por me manter informado e por me salvar em diversas questões burocráticas,

à Universidade Federal de São Carlos (UFSCar) que, por meio de sua estrutura física e humana, forneceu condições para a realização deste trabalho,

aos amigos do Mestrado/Doutorado que, durante esses dois anos, trouxeram momentos de descontração ao longo dos intensos dias de estudo; aos amigos da tradicional *pelada* nas quadras da UFSCar, nossa indispensável diversão semanal,

aos meus amigos de São Paulo, pela amizade, pelos exemplos de luta e pelos divertidos momentos de confraternização,

à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro durante todo o período de desenvolvimento deste trabalho,

e, finalmente, a todos que de alguma forma contribuíram para a realização deste trabalho.

A modelagem de dados provenientes de contagens pareadas é um tópico que vem sendo frequentemente abordado em diversos segmentos de pesquisa. Em particular, podemos citar os casos em que as contagens de interesse são bivariadas, como por exemplo na análise de placares esportivos. Em virtude disso, neste trabalho apresentamos a distribuição Poisson bivariada para os casos em que as contagens de interesse são positivamente correlacionadas. A possível independência entre as contagens também é abordada por meio do modelo Poisson duplo, que surge como caso particular do modelo Poisson bivariado. As principais características e propriedades desses modelos são apresentadas e um estudo de simulação é realizado, visando avaliar o comportamento das estimativas para diferentes tamanhos amostrais. Considerando a possibilidade de se modelar os parâmetros por meio da inserção de variáveis preditoras, apresentamos a estrutura do modelo de regressão Poisson bivariado como caso geral, bem como a estrutura de um modelo de efeitos para aplicação à dados esportivos. Particularmente, neste trabalho vamos considerar aplicações aos dados da Série A do Campeonato Brasileiro de 2012, na qual os efeitos serão estimados por meio dos modelos Poisson duplo e Poisson bivariado. Uma vez obtidos os ajustes, estimam-se as probabilidades de ocorrência dos placares e, a partir destas, obtemos previsões para as partidas de interesse. Com o intuito de se obter previsões mais acuradas para as partidas, apresentamos o método da verossimilhança ponderada, a partir do qual será possível quantificar a relevância dos dados em função do tempo em que estes foram observados.

Palavras-chave: Contagens pareadas, distribuição Poisson bivariada, estudo de simulação, modelo de efeitos, Campeonato Brasileiro, previsão de resultados, verossimilhança ponderada.

Abstract

The modelling of paired counts data is a topic that has been frequently discussed in several threads of research. In particular, we can cite bivariate counts, such as the analysis of sports scores. As a result, in this work we present the bivariate Poisson distribution to modelling positively correlated scores. The possible independence between counts is also addressed through the double Poisson model, which arises as a special case of the bivariate Poisson model. The main characteristics and properties of these models are presented and a simulation study is conducted to evaluate the behavior of the estimates for different sample sizes. Considering the possibility of modeling parameters by insertion of predictor variables, we present the structure of the bivariate Poisson regression model as a general case as well as the structure of an effects model for application in sports data. Particularly, in this work we will consider applications to Brazilian Championship Serie A 2012 data, in which the effects will be estimated by double Poisson and bivariate Poisson models. Once obtained the fits, the probabilities of scores occurrence are estimated and then we obtain forecasts for the outcomes. In order to obtain more accurate forecasts, we present the weighted likelihood method from which it will be possible to quantify the relevance of the data according to the time they were observed.

Keywords: Paired counts, bivariate Poisson distribution, simulation study, effects model, Brazilian Championship, outcomes forecast, weighted likelihood.

Lista de Figuras	xi
Lista de Tabelas	xii
1 Introdução	1
1.1 Objetivos do Trabalho	2
1.2 Organização do Trabalho	3
2 Distribuição Poisson Bivariada	4
2.1 Modelo Probabilístico Poisson Bivariado	5
2.2 Testes de Aderência	7
2.2.1 Teste Chi-Quadrado de Pearson	8
2.2.2 Teste do Índice de Dispersão	8
2.3 Função de Verossimilhança	8
2.4 Estudo de Simulação	9
2.4.1 Estimação e Seleção de Modelos	9
3 Modelo de Regressão Poisson Bivariado	14
3.1 Estrutura Geral do Modelo	14
3.2 Estrutura do Modelo de Efeitos	16
3.3 Medidas de Qualidade Preditiva	17
3.3.1 Medida de DeFinetti	17
3.3.2 Taxa de Sucesso	18
3.4 Aplicação à Dados Reais	19
3.4.1 Campeonato Brasileiro de 2012 - Série A	19
3.4.2 Análise Descritiva	21
3.4.3 Metodologia	22
3.4.4 Resultados Obtidos	24
3.4.5 Previsão de Resultados	31

4	Modelo de Regressão Poisson Bivariado com Ponderação	36
4.1	Método da Verossimilhança Ponderada	37
4.1.1	Escolha da Função Peso	38
4.2	Aplicação à Dados Reais	39
4.2.1	Metodologia	39
4.2.2	Resultados Obtidos	40
4.2.3	Previsão de Resultados	45
5	Considerações Finais	51
A	Identificação e Seleção de Modelos	53
A.1	Critérios AIC, AICc e BIC	53
A.2	Teste da Razão de Verossimilhanças	54
	Referências Bibliográficas	55

Lista de Figuras

2.1	Erros quadráticos médios das estimativas.	11
2.2	Porcentagem de aceitação de \mathcal{H}_0 e poder do teste (TRVG).	12
2.3	Porcentagem de identificação observada por meio do critério AIC.	13
3.1	Frequência do número de gols marcados.	22
3.2	Ilustração dos p-valores associados aos TRV's.	24
3.3	Efeitos ofensivos <i>vs.</i> defensivos das equipes ao término do primeiro turno.	25
3.4	Efeitos ofensivos <i>vs.</i> defensivos das equipes ao término da competição.	26
3.5	Evolução dos efeitos ofensivos e defensivos ao longo do segundo turno.	26
3.6	Evolução do efeito casa ao longo do segundo turno.	27
3.7	Probabilidades estimadas para título, libertadores e despromoção.	30
3.8	Classificação final esperada da Série A do Campeonato Brasileiro de 2012.	31
3.9	Medidas de Definetti para as rodadas previstas.	35
4.1	Comportamento das funções propostas para modelagem dos pesos.	38
4.2	Ilustração dos p-valores associados aos TRV's (ajustes via FP1).	41
4.3	Ilustração dos p-valores associados aos TRV's (ajustes via FP2).	41
4.4	Funções de log-verossimilhança preditiva (ajustes via FP1).	42
4.5	Funções de log-verossimilhança preditiva (ajustes via FP2).	42
4.6	Evolução dos efeitos ofensivos ao longo das rodadas (ajuste via FP1).	43
4.7	Evolução dos efeitos defensivos ao longo das rodadas (ajuste via FP1).	43
4.8	Evolução dos efeitos ofensivos ao longo das rodadas (ajuste via FP2).	44
4.9	Evolução dos efeitos defensivos ao longo das rodadas (ajuste via FP2).	44
4.10	Evolução do efeito casa ao longo das rodadas (ajuste via FP1).	44
4.11	Evolução do efeito casa ao longo das rodadas (ajuste via FP2).	45
4.12	Medidas de Definetti para as rodadas previstas (ajuste via FP1).	48
4.13	Medidas de Definetti para as rodadas previstas (ajuste via FP2).	48

Lista de Tabelas

2.1	Valores atribuídos ao parâmetro de covariância.	9
2.2	Resumo do processo de estimação pontual.	10
3.1	Classificação final do Campeonato Brasileiro de 2012.	20
3.2	Média de gols marcados por turno pelas equipes.	21
3.3	Resumo descritivo do número de gols marcados.	21
3.4	Frequências do número de gols marcados.	22
3.5	Rodadas utilizadas para ajuste e previsão em cada etapa.	23
3.6	Estrutura dos modelos propostos.	23
3.7	Critérios de seleção para comparação dos modelos.	24
3.8	Pontuação final estimada por meio do modelo Poisson duplo.	28
3.9	Pontuação final estimada por meio do modelo Poisson bivariado.	29
3.10	Previsões para a 24 ^a rodada.	32
3.11	Previsões para a 28 ^a rodada.	32
3.12	Previsões para a 34 ^a rodada.	33
3.13	Previsões para a 38 ^a rodada.	33
3.14	Percentual de previsões corretas dos modelos a cada rodada.	34
3.15	Taxa de sucesso dos modelos ajustados.	35
4.1	Rodadas utilizadas para ajuste, teste e previsão em cada etapa.	40
4.2	Critérios de seleção para comparação dos modelos ponderados.	40
4.3	Previsões para a 28 ^a rodada (ajuste via FP1).	46
4.4	Previsões para a 28 ^a rodada (ajuste via FP2).	46
4.5	Previsões para a 38 ^a rodada (ajuste via FP1).	47
4.6	Previsões para a 38 ^a rodada (ajuste via FP2).	47
4.7	Percentual de previsões corretas dos modelos ponderados a cada rodada.	49
4.8	Taxa de sucesso dos modelos ponderados ajustados.	49

A modelagem de dados pareados que exibam correlação não nula tem sido um tópico de bastante interesse para pesquisadores de diversas áreas. Por exemplo, em estudos clínicos como os de caso-controle (ou pré/pós tratamento) pode se estar interessado em avaliar a incidência de duas doenças em um mesmo indivíduo (Shenson e Joshi, 2012), bem como em verificar o número de vezes que um indivíduo reagiu a um certo medicamento, pelo número de vezes que o ingeriu. Em outros segmentos, como na área de *marketing*, pode haver o interesse na modelagem do *cross merchandising* (Drèze e Hoch, 1998), que consiste na realização de estudos baseados na compra conjunta de dois produtos. No âmbito esportivo, as principais abordagens consistem na modelagem do número de gols marcados por duas equipes que se enfrentam, tais como em partidas de futebol (Lee, 1997; Karlis e Ntzoufras, 2003), de futsal, de polo aquático, de hóquei no gelo e assim por diante. Além disso, o interesse na análise de medidas pareadas também pode surgir na econometria, na sociologia (Riphahn *et al.*, 2003), na área industrial (Ho e Singer, 2001), e em diversos outros segmentos. Nesse sentido, em meio as alternativas existentes para a modelagem de tais situações, podemos destacar a distribuição Poisson bivariada para os casos em que os dados disponíveis são representados por contagens pareadas positivamente correlacionadas.

O modelo probabilístico Poisson bivariado pertence à classe Poisson bivariada de distribuições, que inclui o modelo Poisson duplo (Poissons independentes) como caso particular. Essa distribuição foi proposta por Holgate (1964), que determinou a função de probabilidade conjunta utilizando o método da redução trivariada. Além disso, essa distribuição foi amplamente estudada por Kawamura (1973a), que propôs uma derivação alternativa à de Holgate. Desde o surgimento, observou-se que o principal fator que impedia uma maior utilização da distribuição Poisson bivariada para a modelagem de contagens pareadas era o alto custo computacional associado à estimação dos parâmetros, principalmente entre os anos de 1970 e 1980 (Karlis e Ntzoufras, 2005).

Do ponto de vista frequentista, a estimação dos parâmetros da distribuição Poisson bivariada pode ser realizada pelo método dos momentos, pelo método de máxima verossimilhança, pelo método do duplo zero e pelo método dos pontos pares (Kocherlakota e Kocherlakota, 1992). A obtenção de estimativas de máxima verossimilhança não se constitui uma tarefa trivial, uma vez que a função de verossimilhança associada apresenta uma estrutura complicada que depende do

produtório de somatórios. Dentre as principais abordagens que buscam contornar esse problema, podemos citar [Kawamura \(1985\)](#), que propõe um método de estimação baseado na utilização de duas relações de recorrência deduzidas por [Teicher \(1954\)](#). Contudo, esse método pode, eventualmente, ser considerado complicado, uma vez que exige a resolução de um sistema de derivadas indiretas. Nesse sentido, outra alternativa para a estimação via máxima verossimilhança dos parâmetros dessa distribuição é a utilização de métodos numéricos de otimização, tais como o método de Newton-Raphson e do algoritmo EM ([Karlis e Ntzoufras, 2005](#)).

Naturalmente, a distribuição Poisson bivariada permite que a modelagem de seus parâmetros seja realizada a partir da utilização de um conjunto de variáveis preditoras inseridas em uma estrutura log-linear de regressão, por exemplo. O modelo de regressão Poisson bivariado foi introduzido por [Jung e Winkelmann \(1993\)](#) e posteriormente foi discutido em detalhes por [Kocherlakota e Kocherlakota \(2001\)](#). No que tange a estimação, [Ho e Singer \(2001\)](#) propõem estimadores de mínimos quadrados generalizados, ao passo que [Karlis e Ntzoufras \(2003\)](#) desenvolvem um método baseado no algoritmo EM para estimar os coeficientes desse modelo. Como contribuição, [Karlis e Ntzoufras \(2005\)](#) observam que a estimação via EM pode representar uma alternativa interessante para se ultrapassar eventuais problemas que o método de Newton-Raphson ([Jung e Winkelmann, 1993](#)) possa apresentar, fundamentalmente em termos de convergência.

O interesse na utilização do modelo de regressão Poisson bivariado pode ir além da estimação e dos procedimentos de inferência. Em algumas situações, também pode haver o interesse na predição de valores e das respectivas probabilidades de ocorrência. Por exemplo, em estudos clínicos, os pesquisadores podem estar interessados em prever a quantidade de vezes que um novo indivíduo na amostra irá reagir a um certo medicamento, para uma quantidade fixa de pílulas ingeridas. Já no âmbito esportivo, os apostadores estão diretamente interessados em saber quais equipes têm maiores chances de vencer suas partidas na próxima rodada de uma determinada competição. Assim surge o interesse dos pesquisadores em otimizar a capacidade preditiva dos modelos adotados ([Dixon e Coles, 1997](#)). Nesse contexto, o método da verossimilhança ponderada ([Hu e Zidek, 2002](#)) pode, dentre todas as suas utilidades, ser considerado uma alternativa bastante interessante para obtenção de previsões mais acuradas para os fenômenos de interesse, desde que se opte por estimar o modelo adotado via máxima verossimilhança.

1.1 Objetivos do Trabalho

A nossa principal motivação para abordar o modelo Poisson bivariado está em sua possível aplicação a dados de futebol, que é o esporte mais popular no Brasil. Uma partida de futebol é disputada por duas equipes, cada uma com 11 jogadores titulares e até 12 suplentes, durante dois tempos de 45 minutos, além dos acréscimos. O objetivo de cada equipe é marcar mais gols que o adversário. As propostas mais usuais assumem que o resultado de uma partida é um vetor aleatório bivariado cujas componentes representam a contagem do número de gols marcados por cada uma das equipes. Em virtude disso, a partir da estrutura de regressão do modelo Poisson bivariado, podemos estimar as capacidades ofensivas e defensivas de cada equipe, além do número médio de gols marcados em uma partida, levando em consideração as capacidades das equipes que se enfrentam. O modelo ajustado fornece meios pelos quais se torna possível prever qual das equipes tem maior probabilidade de vencer determinada partida futura e, a principal contribuição deste trabalho está na aplicação do método da verossimilhança ponderada, visando melhorar a

capacidade preditiva dos modelos de interesse. Cabe ressaltar que as metodologias apresentadas serão aplicadas aos dados da Série A do Campeonato Brasileiro de Futebol de 2012.

1.2 Organização do Trabalho

Este trabalho foi desenvolvido em cinco capítulos, sendo esta seção parte do Capítulo 1. No Capítulo 2 introduzimos o modelo Poisson bivariado, apresentando suas principais características, propriedades e momentos. Para ilustrar a estimação dos parâmetros desse modelo, um estudo de simulação é realizado, sendo que o método numérico de otimização L-BFGS-B é utilizado para a obtenção das estimativas de máxima verossimilhança. No Capítulo 3 apresentamos as principais características do modelo de regressão Poisson bivariado, inclusive quando escrito sob uma estrutura de efeitos. Além disso, discutimos métodos de verificação da qualidade preditiva e consideramos uma aplicação desse modelo aos dados da Série A do Campeonato Brasileiro de 2012, sendo os coeficientes de regressão estimados via máxima verossimilhança com auxílio do método numérico de otimização BFGS. No Capítulo 4 apresentamos uma breve discussão acerca do método da verossimilhança ponderada e consideramos sua aplicação ao modelo de regressão Poisson bivariado, visando melhorar sua capacidade preditiva. Para efeitos de comparação, os dados utilizados no Capítulo 3 são novamente considerados para aplicação. O Capítulo 5 é dedicado aos comentários e considerações finais.

Distribuição Poisson Bivariada

Em muitas situações, quando existe o interesse nas versões multivariadas dos principais modelos probabilísticos univariados conhecidos, nos deparamos com a situação em que apenas uma versão paramétrica pode ser construída para uma determinada distribuição, como é o caso da distribuição Normal. Em geral, modelos multivariados não normais são mais complexos, uma vez que as funções de probabilidade (ou de densidade) conjunta tendem a apresentar estruturas complicadas. Entretanto, no caso da distribuição de Poisson, existem várias formas eficientes de se construir uma versão multivariada paramétrica generalizada e, atualmente, existem ferramentas que permitem maior tratabilidade da função de probabilidade conjunta para os principais processos inferenciais.

De forma geral, muitas são as maneiras de se construir um modelo probabilístico com marginais Poisson. Qualquer distribuição conjunta cujas marginais são Poisson pode ser referida distribuição Poisson multivariada. O método da convolução é bastante utilizado para se derivar distribuições multivariadas de modelos de contagens. Em particular, temos que uma das formas mais simples de se derivar a distribuição Poisson bivariada é por meio do método da redução trivariada. Em contrapartida, [Hamdan e Al-Bayyati \(1971\)](#) apresentam uma derivação diferente para esse modelo.

Nesse contexto, considere (X, Y) um vetor bivariado de variáveis aleatórias que caracterizam um determinado experimento cujas realizações tem suporte em \mathbb{N}_0^2 , com $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ representando o conjunto dos números naturais, incluindo o zero. Segundo [Kocherlakota e Kocherlakota \(1992\)](#), um vetor aleatório assim determinado segue uma distribuição Poisson bivariada se

1. $f(x, y; \boldsymbol{\theta}) \geq 0$ para todo $(x, y) \in \mathbb{N}_0^2$;
2. $\sum_x \sum_y f(x, y; \boldsymbol{\theta}) = 1$;
3. $f(x; \lambda_1) = \sum_y f(x, y; \boldsymbol{\theta}) = \frac{e^{-\lambda_1} \lambda_1^x}{x!}$, $x \in \mathbb{N}_0$, $\lambda_1 > 0$;
4. $f(y; \lambda_2) = \sum_x f(x, y; \boldsymbol{\theta}) = \frac{e^{-\lambda_2} \lambda_2^y}{y!}$, $y \in \mathbb{N}_0$, $\lambda_2 > 0$.

Neste trabalho vamos considerar $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}^k$, $k \geq 1$, como sendo o vetor paramétrico que indexa a distribuição de interesse. Por conveniência, denotaremos a função de probabilidade conjunta de

X e Y , $P(X = x, Y = y|\boldsymbol{\theta})$, por $f(x, y; \boldsymbol{\theta})$ tal que, para todo (x, y) pertencentes ao espaço dos possíveis resultados de (X, Y) , $f : \mathbb{N}_0 \times \mathbb{N}_0 \rightarrow (0, 1)$.

Após uma breve introdução acerca dos principais aspectos teóricos relacionados, neste capítulo vamos abordar a construção do modelo probabilístico Poisson bivariado utilizando o método da redução trivariada, e vamos apresentar as principais características e propriedades desse modelo. A estimação via máxima verossimilhança dos parâmetros é considerada por meio do método numérico de otimização L-BFGS-B. Visando avaliar o comportamento das estimativas para diferentes tamanhos amostrais, no complemento do capítulo será realizado um estudo de simulação.

2.1 Modelo Probabilístico Poisson Bivariado

A distribuição Poisson bivariada foi formalmente introduzida por [Holgate \(1964\)](#), que utilizou o método da redução trivariada para derivar a função de probabilidade conjunta desse modelo. Em uma abordagem bastante completa, [Kawamura \(1973a\)](#) apresenta importantes aspectos teóricos relativos a essa distribuição. Além disso, [Kawamura \(1976\)](#) utiliza o método da redução trivariada para mostrar que, no caso da distribuição Poisson bivariada, as variáveis aleatórias univariadas independentes seguem distribuição de Poisson. Em virtude disso, é bastante comum nos depararmos com a denominação ‘distribuição Poisson bivariada de Holgate’ para esse modelo probabilístico. Maiores detalhes sobre o método da redução trivariada podem ser encontrados em [Lai \(1995\)](#).

Nesse contexto, considere X_1 , X_2 e X_3 variáveis aleatórias independentes que seguem distribuição de Poisson com parâmetros λ_1 , λ_2 e λ_3 , respectivamente. As variáveis de interesse, digamos X e Y , podem ser definidas pela soma de variáveis aleatórias com uma componente comum nas somas (método da redução trivariada). Portanto, as variáveis aleatórias $X = X_1 + X_3$ e $Y = X_2 + X_3$ seguem uma distribuição Poisson bivariada se a função de probabilidade do vetor aleatório (X, Y) pode ser escrita como

$$f(x, y; \boldsymbol{\theta}) = e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{k=0}^{\min(x, y)} \frac{\lambda_1^{(x-k)}}{(x-k)!} \frac{\lambda_2^{(y-k)}}{(y-k)!} \frac{\lambda_3^k}{k!}, \quad x, y = 0, 1, \dots, \quad (2.1)$$

em que $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \lambda_3)$, com $\lambda_1, \lambda_2 > 0$ e $\lambda_3 \geq 0$. Na prática, as variáveis X_1 , X_2 e X_3 são não observáveis. A seguir, apresentamos a construção via condicionamento da função de probabilidade da distribuição Poisson bivariada (2.1), indexada pelo vetor de parâmetros $\boldsymbol{\theta}$.

Construção via condicionamento

$$\begin{aligned} f(x, y; \boldsymbol{\theta}) &= P(X = x, Y = y|\boldsymbol{\theta}) \\ &= P(X_1 + X_3 = x, X_2 + X_3 = y|\boldsymbol{\theta}) \\ &= \sum_{k=0}^{\infty} P(X_1 + X_3 = x, X_2 + X_3 = y|X_3 = k, \boldsymbol{\theta}) P(X_3 = k|\lambda_3) \\ &= \sum_{k=0}^{\min(x, y)} P(X_1 = x - k, X_2 = y - k|\lambda_1, \lambda_2) P(X_3 = k|\lambda_3) \\ &= \sum_{k=0}^{\min(x, y)} P(X_1 = x - k|\lambda_1) P(X_2 = y - k|\lambda_2) P(X_3 = k|\lambda_3) \end{aligned}$$

$$\begin{aligned}
&= \sum_{k=0}^{\min(x,y)} \frac{e^{-\lambda_1} \lambda_1^{(x-k)}}{(x-k)!} \frac{e^{-\lambda_2} \lambda_2^{(y-k)}}{(y-k)!} \frac{e^{-\lambda_3} \lambda_3^k}{k!} \\
&= e^{-(\lambda_1+\lambda_2+\lambda_3)} \sum_{k=0}^{\min(x,y)} \frac{\lambda_1^{(x-k)}}{(x-k)!} \frac{\lambda_2^{(y-k)}}{(y-k)!} \frac{\lambda_3^k}{k!}.
\end{aligned}$$

Uma forma alternativa de se derivar a distribuição Poisson bivariada é representada pelo processo de composição de uma distribuição Binomial bivariada com uma distribuição de Poisson (Hamdan e Al-Bayyati, 1971). Kawamura (1973b) considera essa alternativa como ferramenta para determinação de uma distribuição diagonal baseada no modelo Poisson bivariado. Além disso, Kocherlakota (1988) propõe um tratamento unificado em modelos probabilísticos conhecidos, como forma de se compor uma distribuição Poisson bivariada.

A distribuição Poisson bivariada é totalmente identificável por seus parâmetros. Ou seja, é impossível os vetores $\mathbf{A} = (X, Y)$ e $\mathbf{B} = (U, V)$ seguirem a mesma distribuição Poisson bivariada com vetor de parâmetros $\boldsymbol{\theta}_A = (\lambda_1, \lambda_2, \lambda_3)$ e $\boldsymbol{\theta}_B = (\mu_1, \mu_2, \mu_3)$ se, para algum $k = 1, 2, 3$ fixo, $\lambda_k \neq \mu_k$. A prova desse fato pode ser encontrada em Arruda (2000).

As principais características associadas à distribuição Poisson bivariada podem ser facilmente derivadas a partir da análise teórica, uma vez que as distribuições marginais das variáveis aleatórias X e Y são Poisson, cujas propriedades são bem conhecidas. Nesse sentido, considere novamente as variáveis X_1 , X_2 e X_3 descritas anteriormente. A esperança e a variância das variáveis aleatórias $X = X_1 + X_3$ e $Y = X_2 + X_3$ são dadas, respectivamente, por

$$E[X] = \text{Var}[X] = \lambda_1 + \lambda_3 \quad \text{e} \quad E[Y] = \text{Var}[Y] = \lambda_2 + \lambda_3.$$

A covariância entre as variáveis X e Y é dada por

$$\begin{aligned}
\text{Cov}[X, Y] &= \text{Cov}[X_1 + X_3, X_2 + X_3] \\
&= \text{Cov}[X_1, X_2] + \text{Cov}[X_1, X_3] + \text{Cov}[X_2, X_3] + \text{Cov}[X_3, X_3] \\
&= \text{Var}[X_3] \\
&= \lambda_3.
\end{aligned}$$

Essa é uma das mais importantes características associadas à distribuição Poisson bivariada, que implica que esse modelo probabilístico apenas se ajusta à dados que apresentam estrutura não negativa de correlação, uma vez que λ_3 representa uma medida de dependência positiva entre as variáveis aleatórias X e Y . A partir da covariância, é possível obter uma medida mais direta de dependência entre X e Y , o coeficiente de correlação ρ . Nesse caso, temos que

$$\begin{aligned}
\rho &= \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \\
&= \frac{\lambda_3}{\sqrt{(\lambda_1 + \lambda_3)(\lambda_2 + \lambda_3)}}.
\end{aligned}$$

Vale ressaltar que o coeficiente de correlação é crescente em λ_3 e acentuadamente decrescente em λ_1 e λ_2 . Sabemos que covariância nula é condição necessária e suficiente para garantir a independência entre X e Y . Em particular, $\lambda_3 = 0$ implica diretamente em $\rho = 0$. Em geral, em

distribuições multivariadas não normais, o coeficiente de correlação ρ não descreve completamente a estrutura de dependência entre as variáveis de interesse. Entretanto, nesse caso, o coeficiente de correlação caracteriza a dependência, sem a necessidade de medidas adicionais.

Outro aspecto relevante refere-se ao fato de que ao se permitir que os parâmetros variem através das observações, como na estrutura de um modelo de regressão, tem-se que a correlação entre as variáveis de interesse também varia através destas, sendo assim possível avaliar o grau de heterogeneidade associado aos dados.

A distribuição Poisson bivariada representa uma generalização do modelo Poisson duplo quando $\lambda_3 > 0$, de forma que, para $\lambda_3 = 0$, a função de probabilidade conjunta (2.1) se resume ao produto de duas Poisson independentes com parâmetros λ_1 e λ_2 . Para verificar essa afirmação, vamos convencionar que $0^0 = 1$ e reescrever (2.1) isolando o primeiro termo da soma ($k = 0$), ou seja

$$\begin{aligned}
 f(x, y; \boldsymbol{\theta}) &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{k=0}^{\min(x, y)} \frac{\lambda_1^{(x-k)} \lambda_2^{(y-k)} \lambda_3^k}{(x-k)! (y-k)! k!} \\
 &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \lambda_1^x \lambda_2^y \sum_{k=0}^{\min(x, y)} \frac{1}{(x-k)! (y-k)! k!} \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k \\
 &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \frac{\lambda_1^x \lambda_2^y}{x! y!} \sum_{k=0}^{\min(x, y)} \frac{x! y! k!}{k! (x-k)! (y-k)!} \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k \\
 &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \frac{\lambda_1^x \lambda_2^y}{x! y!} \sum_{k=0}^{\min(x, y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k \\
 &= e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \frac{\lambda_1^x \lambda_2^y}{x! y!} \left[1 + \sum_{k=1}^{\min(x, y)} \binom{x}{k} \binom{y}{k} k! \left(\frac{\lambda_3}{\lambda_1 \lambda_2} \right)^k \right].
 \end{aligned}$$

Logo, se $\lambda_3 = 0$, em particular para todo $k \geq 1$, teremos

$$\begin{aligned}
 f(x, y; \lambda_1, \lambda_2) &= e^{-(\lambda_1 + \lambda_2)} \frac{\lambda_1^x \lambda_2^y}{x! y!} \\
 &= \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^y}{y!},
 \end{aligned}$$

que é o produto de duas Poisson independentes com parâmetros $\lambda_1, \lambda_2 > 0$.

Dentre as principais ferramentas que podem ser utilizadas para a seleção de modelos, podemos citar os critérios AIC e BIC. A principal motivação para se utilizar tais critérios está no interesse de se obter uma ferramenta que aponte qual dos modelos propostos melhor se adequa aos dados observados. Por outro lado, é possível definir um procedimento para testar se o modelo Poisson duplo melhor se adequa a um certo conjunto de dados em relação ao modelo Poisson bivariado. Nesse caso, podemos definir o teste de hipóteses $\mathcal{H}_0 : \lambda_3 = 0$ vs. $\mathcal{H}_1 : \lambda_3 > 0$. Esse procedimento é baseado no teste da razão de verossimilhanças generalizado (TRVG) assintótico que, na prática, testa a dependência entre as variáveis X e Y .

2.2 Testes de Aderência

Podemos testar a aderência de um vetor aleatório bivariado a uma determinada distribuição de probabilidades através de extensões dos testes univariados. Dentre os mais usuais, podemos citar

o teste chi-quadrado de Pearson, o TRVG, além do teste de Crockett's e do teste do índice de dispersão. Os dois primeiros testes baseiam-se nas frequências observadas e nas frequências esperadas (sob a suposição de que os dados aderem à distribuição de interesse), ao passo que os outros dois testes baseiam-se apenas em quantidades estimadas a partir dos dados. Vale ressaltar que todos esses procedimentos são baseados em resultados assintóticos. Neste trabalho, apresentaremos o teste chi-quadrado de Pearson e o teste do índice de dispersão. Maiores detalhes sobre esses testes de aderência podem ser encontrados em [Kocherlakota e Kocherlakota \(1992\)](#).

2.2.1 Teste Chi-Quadrado de Pearson

A estatística do teste chi-quadrado de Pearson para testar a aderência de um vetor aleatório bivariado à distribuição Poisson bivariada é dada por

$$\chi^2 = \sum_{x=0}^{k_x} \sum_{y=0}^{k_y} \frac{[o_{xy} - nf(x, y; \boldsymbol{\theta})]^2}{nf(x, y; \boldsymbol{\theta})}, \quad (2.2)$$

em que o_{xy} representa a frequência em que um determinado par (x, y) é observado e $f(x, y; \boldsymbol{\theta})$ é a probabilidade de que esse par ocorra sobre o modelo Poisson bivariado. A estatística χ^2 tem distribuição assintótica Chi-quadrado com $(k_x \times k_y) - p - 1$ graus de liberdade, sendo p o número de parâmetros estimados.

2.2.2 Teste do Índice de Dispersão

Proposto por [Loukas e Kemp \(1986\)](#), o teste do índice de dispersão é bastante útil para se avaliar desvios quando a distribuição Poisson bivariada é assumida para a modelagem do vetor bivariado disponível. A estatística de teste é dada por

$$\mathcal{T} = n \frac{\bar{x} \hat{\sigma}_y^2 + \bar{y} \hat{\sigma}_x^2 - 2\hat{\sigma}_{xy}^2}{\bar{x} \bar{y} - \hat{\sigma}_{xy}^2}, \quad (2.3)$$

em que $\hat{\sigma}_x^2$, $\hat{\sigma}_y^2$ e $\hat{\sigma}_{xy}$ são estimadores das variâncias e da covariância entre as variáveis X e Y , respectivamente. Sob a distribuição Poisson bivariada temos que $\hat{\sigma}_x^2 = \hat{\lambda}_1 + \hat{\sigma}_{xy}$, $\hat{\sigma}_y^2 = \hat{\lambda}_2 + \hat{\sigma}_{xy}$ e $\hat{\sigma}_{xy} = \hat{\lambda}_3$. A estatística \mathcal{T} tem distribuição assintótica Chi-Quadrado com $2n - 3$ graus de liberdade, uma vez que três parâmetros devem ser estimados. [Loukas e Kemp \(1986\)](#) mostram que a estatística \mathcal{T} tem maior poder do que a estatística χ^2 e do que a estatística do teste de Crockett's.

2.3 Função de Verossimilhança

Para obtenção da função de verossimilhança, considere $(x_1, y_1), \dots, (x_n, y_n)$ observações provenientes de uma amostra aleatória de tamanho n do modelo Poisson bivariado definido por (2.1). A função de verossimilhança de $\boldsymbol{\theta}$ é dada por

$$\begin{aligned} L(\boldsymbol{\theta} | \mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n f(x_i, y_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \left[e^{-(\lambda_1 + \lambda_2 + \lambda_3)} \sum_{k=0}^{\min(x_i, y_i)} \frac{\lambda_1^{(x_i - k)}}{(x_i - k)!} \frac{\lambda_2^{(y_i - k)}}{(y_i - k)!} \frac{\lambda_3^k}{k!} \right] \end{aligned} \quad (2.4)$$

$$= e^{-n(\lambda_1+\lambda_2+\lambda_3)} \prod_{i=1}^n \left[\sum_{k=0}^{\min(x_i, y_i)} \frac{\lambda_1^{(x_i-k)}}{(x_i-k)!} \frac{\lambda_2^{(y_i-k)}}{(y_i-k)!} \frac{\lambda_3^k}{k!} \right],$$

para $\lambda_1, \lambda_2 > 0$ e $\lambda_3 \geq 0$.

Note que a função de verossimilhança (2.4) é proporcional ao produtório de somatórios, o que a torna bastante dificultosa para cálculos eficientes na estatística frequentista e dificulta a obtenção de estimadores de máxima verossimilhança de forma explícita. Para ultrapassar essa dificuldade, Kawamura (1985) propõe um método de estimação baseado na utilização de duas relações de recorrência deduzidas por Teicher (1954). Entretanto, neste trabalho utilizaremos o método L-BFGS-B para obtenção de estimativas de máxima verossimilhança para os parâmetros da distribuição Poisson bivariada. O método de otimização L-BFGS-B (Limited Memory Boxed BFGS) foi proposto por Byrd *et al.* (1995) como extensão aos métodos BFGS e L-BFGS, pois permite que sejam empregadas quantas restrições de fronteira forem necessárias para otimização de funções cujo domínio dos parâmetros de interesse é restrito.

2.4 Estudo de Simulação

Nesta seção vamos considerar a simulação de variáveis pseudoaleatórias provenientes de uma distribuição Poisson bivariada indexada por $\theta = (\lambda_1, \lambda_2, \lambda_3)$ em diferentes cenários, especificados pelo tamanho amostral. O principal intuito deste estudo de simulação está no interesse em se avaliar o comportamento das estimativas de máxima verossimilhança para diferentes tamanhos amostrais e níveis de dependência entre as variáveis simuladas. O método de simulação a ser utilizado é baseado na redução trivariada. Para o processo de simulação, os parâmetros λ_1 e λ_2 são fixados em 1,5 e 1,0, respectivamente. Esses valores foram escolhidos por se aproximarem daqueles que são comuns na análise de dados de futebol, como na aplicação que será apresentada neste trabalho. O valor do parâmetro λ_3 varia de acordo com o coeficiente de correlação ρ (estabelecido para as variáveis X e Y), com λ_1 e λ_2 fixos. Na Tabela 2.1 são apresentados os valores de λ_3 para cada ρ fixado.

Tabela 2.1: Valores atribuídos ao parâmetro de covariância.

ρ	λ_3	ρ	λ_3	ρ	λ_3	ρ	λ_3	ρ	λ_3	ρ	λ_3
0,00	0,00	0,20	0,31	0,40	0,82	0,60	1,86	0,80	4,98	0,95	23,73

As simulações foram realizadas com $n = 15, 30, 50, 100, 200$ e 500 . Para avaliar o viés e o erro quadrático médio das estimativas, cada amostra de tamanho n_i , $i = 1, \dots, 6$, foi simulada 10000 vezes. Vale ressaltar que, através desse procedimento, é possível avaliar o comportamento das estimativas de máxima verossimilhança a medida que o tamanho amostral cresce. Os procedimentos de simulação e de estimação dos parâmetros foram realizados com auxílio do software R (R Development Core Team, 2009).

2.4.1 Estimação e Seleção de Modelos

A Tabela 2.2 apresenta os vieses obtidos a partir das réplicas das amostras simuladas da distribuição Poisson bivariada com $\lambda_1 = 1,5$, $\lambda_2 = 1,0$ fixos. Os vieses resultam da diferença entre os valores estimados e os verdadeiros valores dos parâmetros e é possível notar que estes decaem rapidamente para zero a medida que n cresce. Observe também que, a medida que o coeficiente

de correlação cresce, menos viesadas são as estimativas obtidas, independentemente do tamanho amostral. Uma possível explicação para esse fato é que quanto maior a correlação entre X e Y , maior a proximidade entre o modelo gerador e o modelo estimado, o que invariavelmente ocasiona a redução dos vieses. Entretanto, é possível notar um ligeiro aumento nos vieses quando $\rho > 0,60$, o que se mostra compreensível uma vez que, quanto maior o valor do parâmetro λ_3 , maior a variância das variáveis X e Y e isso tende viesar as estimativas de máxima verossimilhança dos parâmetros de interesse. De forma geral, podemos considerar que, para $n \geq 100$, as estimativas obtidas para os parâmetros da distribuição Poisson bivariada são bastante satisfatórias.

Tabela 2.2: Resumo do processo de estimação pontual.

Correlação	Parâmetro (Valor)	Viés					
		$n = 15$	$n = 30$	$n = 50$	$n = 100$	$n = 200$	$n = 500$
$\rho = 0,00$	$\lambda_1 (1, 50)$	-0,1501	-0,1058	-0,0775	-0,0561	-0,0403	-0,0259
	$\lambda_2 (1, 00)$	-0,1409	-0,1021	-0,0761	-0,0567	-0,0396	-0,0270
	$\lambda_3 (0, 00)$	0,1464	0,1035	0,0781	0,0563	0,0405	0,0271
$\rho = 0,20$	$\lambda_1 (1, 50)$	-0,0619	-0,0276	-0,0126	-0,0030	-0,0001	-0,0007
	$\lambda_2 (1, 00)$	-0,0656	-0,0295	-0,0108	-0,0023	-0,0010	-0,0014
	$\lambda_3 (0, 31)$	0,0653	0,0278	0,0149	0,0024	0,0015	0,0014
$\rho = 0,40$	$\lambda_1 (1, 50)$	-0,0138	-0,0084	-0,0034	-0,0036	-0,0010	-0,0007
	$\lambda_2 (1, 00)$	-0,0176	-0,0103	-0,0016	-0,0030	-0,0019	-0,0013
	$\lambda_3 (0, 82)$	0,0163	0,0077	0,0052	0,0032	0,0023	0,0014
$\rho = 0,60$	$\lambda_1 (1, 50)$	-0,0105	-0,0144	-0,0072	-0,0051	-0,0017	-0,0010
	$\lambda_2 (1, 00)$	-0,0142	-0,0163	-0,0054	-0,0044	-0,0026	-0,0016
	$\lambda_3 (1, 86)$	0,0105	0,0134	0,0097	0,0047	0,0029	-0,0020
$\rho = 0,80$	$\lambda_1 (1, 50)$	-0,0274	-0,0202	-0,0103	-0,0061	-0,0022	-0,0012
	$\lambda_2 (1, 00)$	-0,0311	-0,0221	-0,0084	-0,0054	-0,0031	-0,0018
	$\lambda_3 (4, 98)$	0,0248	0,0174	0,0119	0,0050	0,0028	0,0023
$\rho = 0,95$	$\lambda_1 (1, 50)$	-0,0492	-0,0252	-0,0144	-0,0081	-0,0043	-0,0012
	$\lambda_2 (1, 00)$	-0,0458	-0,0264	-0,0145	-0,0067	-0,0047	-0,0014
	$\lambda_3 (23, 73)$	0,0374	0,0211	0,0173	0,0098	0,0050	0,0005

A Figura 2.1 ilustra o comportamento dos erros quadráticos médios das estimativas em função do tamanho amostral. Observe que para valores pequenos e moderados de ρ , as estimativas para o parâmetro λ_3 apresentam os menores erros quadráticos médios. Por outro lado, quanto maior o coeficiente de correlação, mais lento é o decaimento dos erros associados às estimativas. Além disso, observe que os erros quadráticos médios das estimativas de λ_3 se afastam dos erros quadráticos médios das estimativas de λ_1 e λ_2 a medida que ρ cresce.

Podemos utilizar o critério AIC, BIC e o TRVG assintótico para identificação e seleção de modelos. Em nosso contexto, o modelo Poisson duplo representa um caso particular do modelo Poisson bivariado quando a covariância entre as variáveis X e Y é assumida nula ($\lambda_3 = 0$). Nesse sentido, podemos utilizar o TRVG assintótico para testar as hipóteses $\mathcal{H}_0 : \lambda_3 = 0$ vs. $\mathcal{H}_1 : \lambda_3 > 0$. Os níveis de significância foram fixados em $\alpha = 0,01; 0,05; 0,10$. A Figura 2.2 ilustra a porcentagem

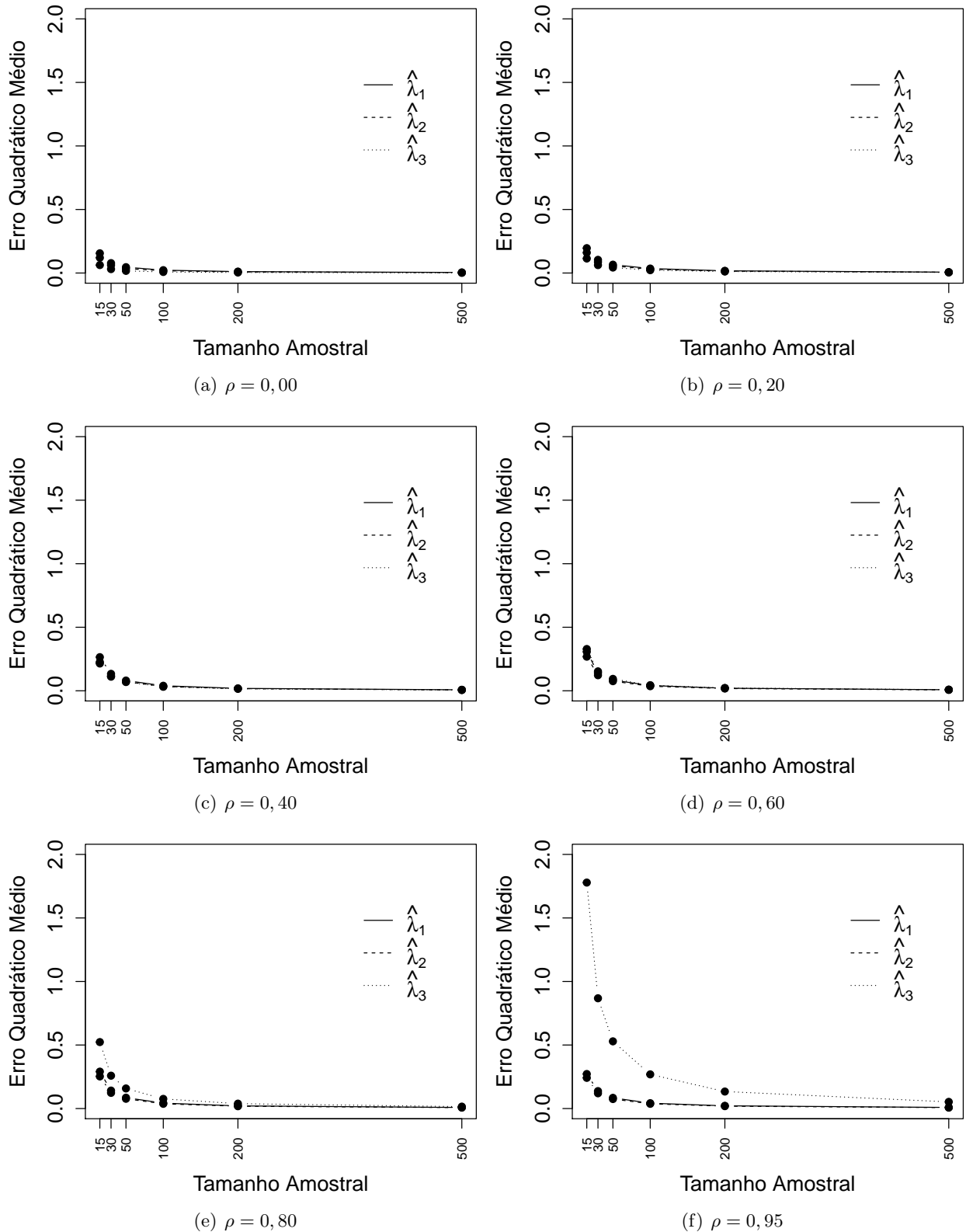


Figura 2.1: Erros quadráticos médios das estimativas.

de aceitação de \mathcal{H}_0 quando esta é verdadeira, além do poder do teste, que corresponde a quantidade relativa de vezes em que a hipótese nula é rejeitada corretamente, em favor do modelo Poisson bivariado. Cada porcentagem é calculada com base na frequência em que o modelo de interesse é selecionado em cada uma das 10000 amostras simuladas para os diferentes tamanhos amostrais, considerando coeficientes de correlação fixos em cada processo de simulação. Observe que, quanto

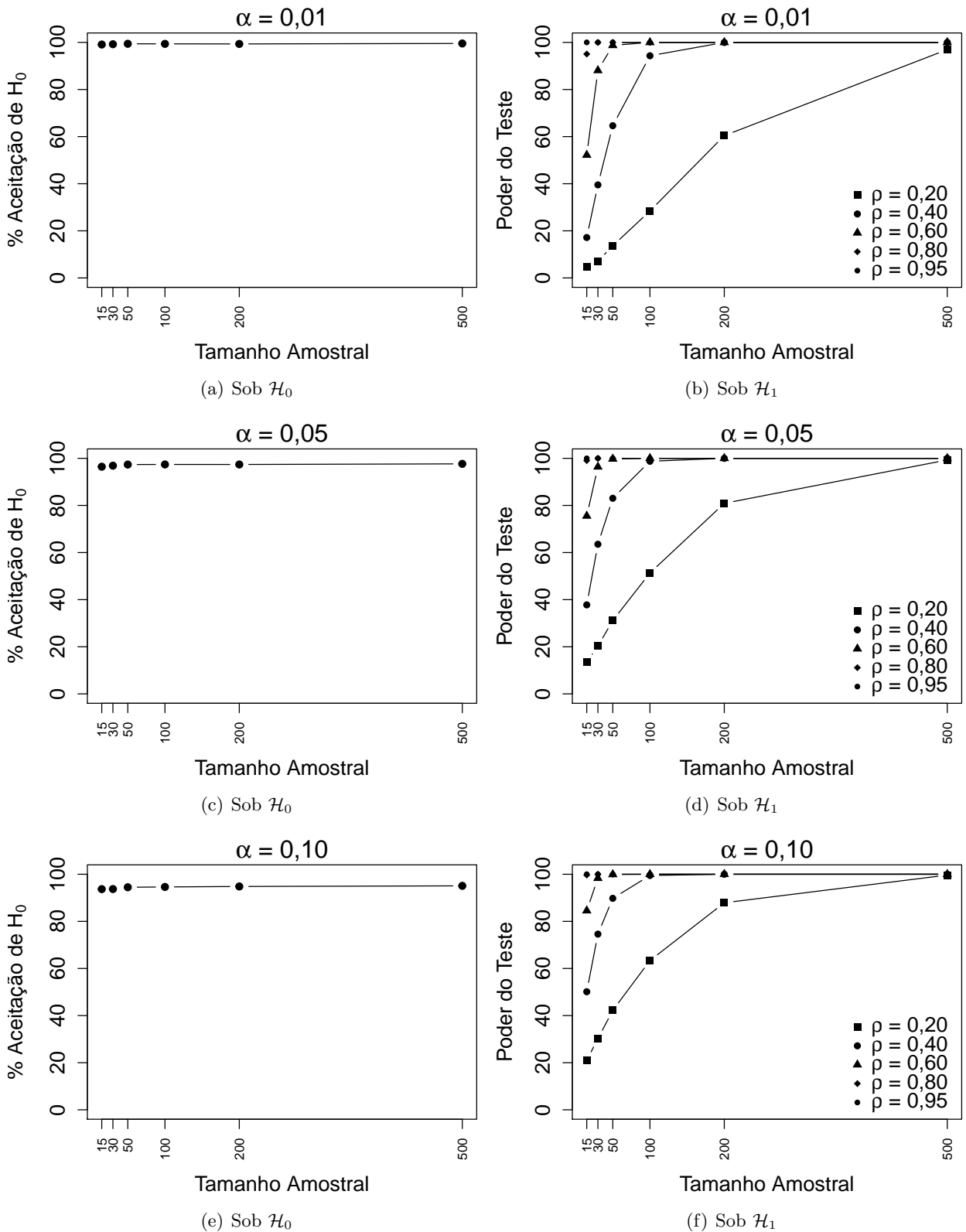


Figura 2.2: Porcentagem de aceitação de \mathcal{H}_0 e poder do teste (TRVG).

maior o nível de significância fixado, menores são as porcentagens de identificação do modelo Poisson duplo, muito embora essa variação se mantenha sempre superior a 90%, o que era esperado. Em contrapartida, a frequência de identificação do modelo Poisson bivariado (poder do teste) aumenta a medida que α cresce. Um aspecto relevante pontua que, quanto maior o coeficiente de

correlação ρ , mais facilmente o modelo gerador dos dados é identificado, uma vez que isso influencia diretamente na magnitude do parâmetro de covariância λ_3 .

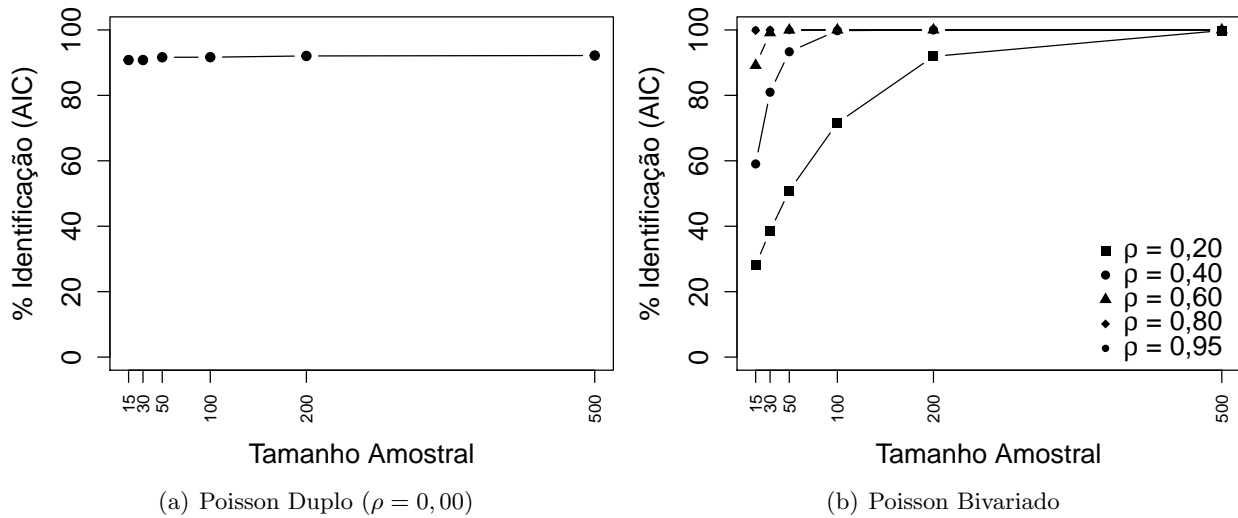


Figura 2.3: Porcentagem de identificação observada por meio do critério AIC.

A Figura 2.3 ilustra a porcentagem de identificação dos modelos Poisson duplo e Poisson bivariado por meio do critério AIC. Os resultados obtidos por meio do critério BIC são análogos. Note que, para tamanhos amostrais moderados e pequenos, o critério AIC apresenta baixa porcentagem de identificação do modelo Poisson bivariado, principalmente quando a correlação entre X e Y é baixa. Entretanto, quando o coeficiente de correlação é nulo, a porcentagem de identificação do modelo Poisson duplo supera os 90%, independentemente do tamanho amostral fixado.

Modelo de Regressão Poisson Bivariado

Diversas são as situações na qual o modelo de regressão Poisson bivariado encontra aplicações. De forma geral, podemos citar a modelagem conjunta da compra de dois produtos (por exemplo, será que a compra de cerveja e de amendoim estão associadas?), o *cross merchandising*, discutido por [Drèze e Hoch \(1998\)](#), além de estudos relacionados à combinação de índices econômicos. Em outros segmentos, podemos destacar a modelagem de dados epidemiológicos, os estudos de caso-controle com observações pareadas, bem como a análise de incidência de doenças em um mesmo indivíduo ([Shenson e Joshi, 2012](#)). No contexto esportivo, o modelo de regressão Poisson bivariado pode ser utilizado para modelar o número de gols marcados por duas equipes oponentes, em esportes como o futebol, o futsal, o polo aquático, o hóquei no gelo e assim por diante. Em particular, a modelagem de dados de futebol via regressão é abordada por [Maher \(1982\)](#) e por [Lee \(1997\)](#), que consideram o ajuste de um modelo Poisson duplo; por [Dixon e Coles \(1997\)](#) que consideram uma abordagem voltada para o mercado de apostas, baseada em um modelo Poisson duplo modificado; por [Rue e Salvesen \(2000\)](#) que propõem a utilização de modelos dinâmicos em um contexto Bayesiano; por [Karlis e Ntzoufras \(2003\)](#) e por [Goddard \(2005\)](#), que propõem abordagens baseadas no modelo Poisson bivariado; dentre outros.

Em vista das muitas oportunidades de aplicação do modelo de regressão Poisson bivariado, neste capítulo apresentaremos sua estrutura, bem como suas principais características. A estimação via máxima verossimilhança dos coeficientes de regressão é considerada por meio do método numérico de otimização BFGS. Além disso, apresentaremos duas medidas úteis para avaliação da qualidade preditiva dos modelos ajustados: a medida de DeFinetti e a taxa de sucesso. No complemento do capítulo apresentamos uma aplicação do modelo Poisson bivariado aos dados da Série A do Campeonato Brasileiro de 2012.

3.1 Estrutura Geral do Modelo

O modelo mais usual para análise de dados de contagens univariadas na presença de variáveis preditoras é o modelo de regressão Poisson, caracterizado por uma estrutura de regressão log-linear. Sobre o modelo Poisson bivariado temos, por suposição, que as distribuições marginais das variáveis X e Y são Poisson. Nesse sentido, considerando funções de ligação apropriadas para

modelar a média dessas variáveis, podemos possível definir uma estrutura de regressão sobre os parâmetros λ_1 , λ_2 e λ_3 da distribuição Poisson bivariada.

Introduzido por Jung e Winkelmann (1993) e amplamente discutido por Ho e Singer (2001), Kocherlakota e Kocherlakota (2001) e Karlis e Ntzoufras (2003), o modelo de regressão Poisson bivariado possui estrutura bastante similar a do modelo de regressão Poisson univariado com função de ligação logarítmica. Seja $(x_1, y_1), \dots, (x_n, y_n)$ uma amostra aleatória de tamanho n do modelo Poisson bivariado. Nossa ferramenta básica para os procedimentos de inferência é a função de verossimilhança que, nesse caso, pode ser escrita da seguinte forma

$$\begin{aligned} L(\boldsymbol{\beta}|\mathbf{x}, \mathbf{y}) &= \prod_{i=1}^n f(x_i, y_i; \boldsymbol{\beta}) \\ &= \prod_{i=1}^n e^{-(\lambda_{1i} + \lambda_{2i} + \lambda_{3i})} \left[\sum_{k=0}^{\min(x_i, y_i)} \frac{\lambda_{1i}^{(x_i-k)}}{(x_i-k)!} \frac{\lambda_{2i}^{(y_i-k)}}{(y_i-k)!} \frac{\lambda_{3i}^k}{k!} \right], \end{aligned} \quad (3.1)$$

em que $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3)$ denota o vetor geral de coeficientes de regressão. Nessa estrutura, os parâmetros da distribuição Poisson bivariada devem ser estimados para cada par observado, sendo então necessário denotá-los por λ_{1i} , λ_{2i} e λ_{3i} , $i = 1, \dots, n$. Note que $3n$ parâmetros devem ser conjuntamente estimados a partir de (3.1), mas, por meio de uma estrutura de regressão, podemos reduzir a dimensionalidade do problema modelando os parâmetros da seguinte forma

$$\begin{aligned} \lambda_{1i} &= \exp\{\mathbf{v}_{1i}^\top \boldsymbol{\beta}_1\}, \\ \lambda_{2i} &= \exp\{\mathbf{v}_{2i}^\top \boldsymbol{\beta}_2\}, \\ \lambda_{3i} &= \exp\{\mathbf{v}_{3i}^\top \boldsymbol{\beta}_3\}, \end{aligned}$$

em que \mathbf{v}_{ki}^\top representa o vetor linha de variáveis preditoras utilizadas para modelar λ_{ki} e $\boldsymbol{\beta}_k$ representa o vetor de coeficientes de regressão associados a λ_k , $k = 1, 2, 3$. De forma geral, esse modelo de regressão deve ser capaz de caracterizar as esperanças marginais por meio de um conjunto de variáveis preditoras, expressando-as por meio de uma estrutura log aditiva.

De fato, o modelo de regressão Poisson bivariado pertence à classe de modelos log-lineares, uma vez que, nesse caso, a escolha da função de ligação logarítmica é natural. Uma restrição desse modelo é que a esperança condicional (aos parâmetros) das variáveis de interesse é igual a variância condicional, e isso pode, em diversas situações, levar o usuário a uma má especificação dos modelos marginais. Por outro lado, no que tange a estrutura funcional desse modelo, podemos ressaltar que os interceptos podem ser introduzidos na modelagem dos parâmetros λ_1 , λ_2 e λ_3 desde que os mesmos tenham interpretação a partir dos dados observados. Outras considerações relevantes sobre a estrutura do modelo de regressão Poisson bivariado são

- As variáveis preditoras utilizadas para a modelagem de cada um dos parâmetros do modelo não necessariamente precisam ser as mesmas;
- Se uma mesma variável preditora modela λ_1 e λ_3 ou λ_2 e λ_3 , então boa parte da influência desta pode ser avaliada por meio da covariância;
- Pode-se não se utilizar nenhuma variável preditora na modelagem de λ_3 , ou seja, assume-se que este é constante entre as observações.

O parâmetro λ_3 exerce um importante papel no modelo de regressão Poisson bivariado. Como citado acima, muitas vezes este pode ser assumido constante, o que proporciona maior facilidade de interpretação do modelo. Contudo, a utilização de variáveis preditoras na modelagem deste parâmetro pode nos fornecer informações mais acuradas sobre o tipo de influência que estas variáveis exercem no par (X, Y) . De forma geral, a escolha das variáveis a serem utilizadas no processo de modelagem é arbitrária e fica a cargo do usuário.

A estimação dos coeficientes do modelo de regressão Poisson bivariado pode ser realizada por meio do procedimento numérico de Newton-Raphson para maximizar a função de log-verossimilhança, como proposto por Jung e Winkelmann (1993). Além disso, Ho e Singer (2001) propõem o método de mínimos quadrados generalizados, ao passo que Karlis e Ntzoufras (2005) consideram um método baseado no algoritmo EM para estimação desse modelo. Neste trabalho vamos considerar o método numérico de otimização BFGS para obtenção de estimativas de máxima verossimilhança para os coeficientes do modelo de regressão Poisson bivariado. Amplamente utilizado em processos de otimização não linear, o método BFGS aproxima o método de Newton para maximização (ou minimização) de funções cujo domínio dos parâmetros de interesse é irrestrito.

3.2 Estrutura do Modelo de Efeitos

Podemos utilizar uma estrutura de regressão para modelagem de dados de futebol desde que, por meio desta, seja possível determinar certas características peculiares à dados esportivos. De forma geral, temos que

- O modelo adotado deve levar em consideração as capacidades de ambas as equipes que disputam uma determinada partida;
- A capacidade de uma determinada equipe deve ser baseada no desempenho recente da mesma;
- Pela natureza do esporte, a qualidade de uma determinada equipe deve ser avaliada por meio de uma medida para a capacidade de marcar gols e de uma medida para a capacidade de não sofrer gols;
- O modelo adotado deve ser flexível ao fato de que as equipes mandantes, em geral, tendem a ter alguma vantagem pelo fato de estarem jogando em casa.

Em geral, é bastante complicado obter estimativas empíricas para as probabilidades associadas ao resultado de uma partida que leve em consideração todas essas condições e, porque não, restrições. Entretanto, podemos utilizar um modelo que consiga incorporar tais características, tal como o modelo de regressão Poisson bivariado com estrutura de efeitos.

Para introduzir esse modelo, denote λ_{kij} , $k = 1, 2, 3$, como sendo o parâmetro associado a partida realizada entre a equipe mandante i , $i = 1, \dots, n$, e a equipe visitante j , $j = 1, \dots, n$. Nesse contexto, podemos expressar λ_{1ij} , λ_{2ij} e λ_{3ij} em termos de um conjunto de fatores explicativos para o número de gols marcados por ambas as equipes que disputam uma determinada partida. Em particular, desejamos obter λ_{1ij} que reflita o potencial ofensivo da equipe mandante, a capacidade defensiva da equipe visitante e o possível efeito do fator ‘casa’; obter λ_{2ij} que reflita o potencial ofensivo da equipe visitante, além da capacidade defensiva da equipe mandante; e obter λ_{3ij} como uma possível combinação desses fatores.

Na prática, esses fatores podem ser traduzidos como efeitos sobre os parâmetros do modelo teórico. Assim, para modelagem de λ_{1ij} e λ_{2ij} , considere os efeitos α_i (ofensivo) e β_i (defensivo) associados a i – *ésima* equipe mandante e os efeitos α_j (ofensivo) e β_j (defensivo) associados a j – *ésima* equipe visitante. Nesse caso, podemos considerar a seguinte estrutura

$$\begin{aligned}\log(\lambda_{1ij}) &= \mu + \omega + \alpha_i + \beta_j, \\ \log(\lambda_{2ij}) &= \mu + \alpha_j + \beta_i,\end{aligned}$$

em que ω representa o efeito casa e μ é um parâmetro constante para ambas as equipes. No que segue, a modelagem do parâmetro λ_3 pode ser considerada a partir da seguinte estrutura

$$\log(\lambda_{3ij}) = \gamma + \delta_1\phi_i + \delta_2\phi_j,$$

em que γ é um parâmetro constante, ϕ_i e ϕ_j são coeficientes que dependem, respectivamente, da equipe mandante e da equipe visitante. As quantidades δ_1 e δ_2 são indicadores que assumem os valores 0 ou 1, dependendo do modelo que estiver sendo considerado. Se $(\delta_1, \delta_2) = (0, 0)$, então a covariância é assumida constante, se $(\delta_1, \delta_2) = (1, 0)$, então a covariância depende somente da equipe mandante e assim por diante. Nesse contexto, o parâmetro λ_3 pode ser interpretado como um efeito aleatório que age aditivamente na média marginal do número de gols marcados por cada uma das equipes, além de refletir as condições de cada partida.

Para evitar a superparametrização do modelo e obter identificabilidade dos parâmetros, podemos utilizar o conjunto de restrições

$$\sum_{i=1}^m \alpha_i = 0 \quad \text{e} \quad \sum_{i=1}^m \beta_i = 0,$$

tal que m denota a quantidade de equipes que disputam a competição. Esse tipo de restrição é denominada restrição de soma zero e, neste caso, a interpretação dos efeitos é realizada comparativamente ao valor zero. Ou seja, um efeito ofensivo positivo indica que a referida equipe tem capacidade ofensiva melhor, ao passo que um efeito defensivo positivo indica que a referida equipe tem capacidade defensiva pior do que uma equipe mediana, de capacidade neutra. Outro possível conjunto de restrições que podem ser impostas em modelos dessa natureza é o conjunto de restrições de ‘canto’, na qual define-se uma equipe como referência (*baseline*), fixando os efeitos ofensivos e defensivos desta em zero. Nesse caso a interpretação dos efeitos também é realizada comparativamente ao valor zero, cabendo ressaltar que, com isso, as demais equipes são comparadas com a equipe de referência.

3.3 Medidas de Qualidade Preditiva

3.3.1 Medida de DeFinetti

A medida de DeFinetti (DeFinetti, 1972) consiste de uma representação geométrica das possíveis previsões probabilísticas obtidas a partir do ajuste de um determinado modelo. Essa medida é baseada em um *2-simplex* contido em \mathbb{R}^3 , cujos vértices correspondem às reais ocorrências, e os demais pontos correspondem a todas as outras possíveis previsões.

Definição 3.3.1. Considere p_1, p_2 e p_3 quantidades, tais que

$$\Delta^2 = \left\{ (p_1, p_2, p_3) \in \mathbb{R}_+^3 ; \sum_{i=1}^3 p_i = 1 \right\}.$$

A partir desse 2-*simplex*, a distância de DeFinetti é definida por

$$DF = (p_1 - a)^2 + (p_2 - b)^2 + (p_3 - c)^2, \quad (3.2)$$

com a, b e c fixos.

A distância de DeFinetti corresponde à distância euclidiana quadrática entre o ponto correspondente à probabilidade estimada e o vértice correspondente ao ponto realmente observado. Quando houver interesse em mais de uma previsão, pode-se definir um índice dado pela média aritmética das distâncias de DeFinetti, denominado Medida de DeFinetti.

Na modelagem de dados de futebol podem ser associados os vértices $(1, 0, 0)$, $(0, 1, 0)$ e $(0, 0, 1)$, que correspondem, respectivamente, à vitória da equipe mandante, ao empate e à vitória da equipe visitante. Dessa forma, podemos estabelecer que p_1 representa a probabilidade de vitória da equipe mandante, p_2 a probabilidade de empate e p_3 a probabilidade de vitória da equipe visitante. No que diz respeito à previsão de placares, um critério simples pode ser estabelecido para avaliar a qualidade das probabilidades estimadas. Para utilizar esse critério, devemos assumir que os três possíveis resultados são equiprováveis, ou seja $p_1 = p_2 = p_3 = 1/3$. Nesse caso, observa-se que $DF = 2/3$, de forma que, podem ser consideradas aceitáveis as previsões cujas distâncias de DeFinetti sejam inferiores a $2/3$. Note que a média aritmética entre as distâncias observadas pode ser utilizada como critério para avaliar a qualidade preditiva do modelo ajustado.

3.3.2 Taxa de Sucesso

A taxa de sucesso é uma medida que nos permite avaliar a qualidade preditiva de um determinado modelo cujas respostas podem ser expressas em termos de k categorias. Essa medida é baseada nos resultados observados e nas probabilidades estimadas para o evento de interesse. Essas probabilidades devem, necessariamente, corresponder aos resultados que não foram considerados para ajustar o modelo proposto (resultados fora da amostra).

Nesse contexto, considere o vetor $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2)$ como sendo aquele que contém as categorias na qual as observações de uma determinada variável aleatória estão inseridas. Considere também que $\dim(\mathbf{z}_1) = n_1$, $\dim(\mathbf{z}_2) = n_2$ e $\dim(\mathbf{z}) = n_1 + n_2 = n$. Agora, suponha que certo modelo é ajustado a partir das informações que determinam \mathbf{z}_1 e que, a partir disso, as probabilidades de ocorrência de \mathbf{z}_2 são estimadas. Se $z_{21}, \dots, z_{2n_2} \in [1, \dots, k]$, a taxa de sucesso é uma medida de qualidade de preditiva definida por

$$TS(\hat{\boldsymbol{\pi}}) = \sum_{i=1}^{n_2} \sum_{j=1}^k \hat{\pi}_{ij} \mathbb{I}(z_{2i} = j), \quad (3.3)$$

tal que para i fixo, $\hat{\pi}_{i1}, \dots, \hat{\pi}_{ik}$ representam as probabilidades de que z_{2i} corresponda a cada uma das k categorias.

Quando houver o interesse na análise de placares esportivos, em esporte como o futebol por exemplo, podemos claramente definir $k = 3$ categorias associadas aos possíveis resultados de uma

determinada partida (1: vitória da equipe mandante, 2: empate e 3: vitória da equipe visitante). Nesse sentido, cada uma das componentes de \mathbf{z} pode ser determinada da seguinte maneira

$$z_q = 1 \times \mathbb{I}(d_q > 0) + 2 \times \mathbb{I}(d_q = 0) + 3 \times \mathbb{I}(d_q < 0),$$

em que d_q , $q = 1, \dots, n$, denota a diferença entre o número de gols marcados pela equipe mandante e pela equipe visitante na q –ésima partida da competição de interesse. Note que a taxa de sucesso corresponde a soma das probabilidades associadas aos placares observados. Além disso, se um determinado modelo prevê todos os placares que ocorreram com probabilidade 1, então a taxa de sucesso será igual ao número de partidas previstas, de forma que, quanto mais próximo deste valor estiver esta medida, maior será a qualidade preditiva do modelo considerado.

3.4 Aplicação à Dados Reais

Esta seção é dedicada a uma aplicação. Inicialmente, nosso interesse reside na obtenção de um ajuste do modelo de regressão Poisson bivariado aos dados da Série A do Campeonato Brasileiro de 2012, e de um ajuste do modelo de regressão Poisson duplo para efeitos de comparação. A estrutura dos modelos que serão ajustados corresponde àquela descrita na seção 3.2, sendo que o parâmetro λ_3 do modelo Poisson bivariado é assumido constante entre as observações. Além da estimação, estamos diretamente interessados em obter previsões para as rodadas do campeonato que não foram consideradas para ajuste e verificar qual dos modelos se mostra mais adequado do ponto de vista preditivo. Os procedimentos de estimação dos coeficientes e de previsão dos resultados foram realizados com auxílio do software R ([R Development Core Team, 2009](#)).

O Campeonato Brasileiro de futebol é a principal competição interclubes do Brasil. Trata-se de um torneio realizado anualmente desde 1959, sendo que, em dezembro de 2010, a Taça Brasil, realizada entre os anos de 1959 a 1968, e as edições de 1967 a 1970 do Torneio Roberto Gomes Pedrosa foram unificadas com o Campeonato Brasileiro propriamente dito, que teve início em 1971.

Atualmente, o Campeonato Brasileiro é disputado em quatro divisões (Séries A, B, C e D), que são organizadas pela CBF (Confederação Brasileira de Futebol). As duas principais divisões (Séries A e B) são disputadas no sistema de turno e retorno em pontos corridos desde 2003 e 2006, respectivamente. Na Série A, além do campeão, o segundo, o terceiro e o quarto colocados classificam-se para a Taça Libertadores da América; as equipes classificadas da quinta à décima segunda posição classificam-se para a fase nacional da Copa Sul-Americana e as quatro últimas são despromovidos à Série B. Entre os anos de 2004 a 2006, o número de equipes que disputavam a Série A reduziu de 24 para 20, o que tornou a competição mais seletiva e menos desgastante para os atletas. Os maiores vencedores da principal divisão do Campeonato Brasileiro são: o Santos e o Palmeiras, ambos com 8 títulos; o São Paulo com 6 títulos; o Corinthians e o Flamengo, com 5 títulos. O atual *ranking* da IFFHS (*International Federation of Football History & Statistics*) aponta que a Série A do Campeonato Brasileiro é a segunda melhor competição de futebol do mundo, superado apenas pela Liga BBVA (Primeira Divisão Espanhola).

3.4.1 Campeonato Brasileiro de 2012 - Série A

A Série A do Campeonato Brasileiro de Futebol de 2012 foi a 56^a edição da principal divisão do futebol brasileiro, disputada entre 02 de maio e 02 de dezembro de acordo com o regulamento

estabelecido em 2003. O campeão da competição foi o Fluminense, que garantiu o título com 3 rodadas de antecedência. Além do Fluminense, Atlético Mineiro, Grêmio e São Paulo classificaram-se para a Taça Libertadores da América. As equipes do Sport, do Palmeiras, do Figueirense e do Atlético Goianiense foram despromovidas à Série B. Nessa edição foram marcados 940 gols, sendo 467 no primeiro turno e 473 no segundo turno. O melhor ataque da competição foi o do Atlético Mineiro, que marcou 64 gols, e a melhor defesa foi a do Fluminense, que sofreu apenas 33 gols em 38 partidas. O maior público registrado na competição foi de 62.207 pessoas no estádio do Morumbi, durante a partida entre São Paulo e Náutico válida pela 36ª rodada. O campeonato teve uma média de público de 12.971 pessoas. Entretanto, o Corinthians apresentou uma média de 25.222 pessoas por partida, que foi a maior registrada entre as equipes.

Tabela 3.1: Classificação final do Campeonato Brasileiro de 2012.

Pos.	Equipe	J	V	E	D	GM	GS	SG	P
1	Fluminense	38	22	11	5	61	33	28	77
2	Atlético-MG	38	20	12	6	64	37	27	72
3	Grêmio	38	20	11	7	56	33	23	71
4	São Paulo	38	20	6	12	59	37	22	66
5	Vasco da Gama	38	16	10	12	45	44	1	58
6	Corinthians	38	15	12	11	51	39	12	57
7	Botafogo	38	15	10	13	60	50	10	55
8	Santos	38	13	14	11	50	44	6	53
9	Cruzeiro	38	15	7	16	47	51	-4	52
10	Internacional	38	13	13	12	44	40	4	52
11	Flamengo	38	12	14	12	39	46	-7	50
12	Náutico	38	14	7	17	44	51	-7	49
13	Coritiba	38	14	6	18	53	60	-7	48
14	Ponte Preta	38	12	12	14	37	44	-7	48
15	Bahia	38	11	14	13	37	41	-4	47
16	Portuguesa	38	10	15	13	39	41	-2	45
17	Sport Recife	38	10	11	17	39	56	-17	41
18	Palmeiras	38	9	7	22	39	54	-15	34
19	Atlético-GO	38	7	9	22	37	67	-30	30
20	Figueirense	38	7	9	22	39	72	-33	30

A Tabela 3.1 apresenta a classificação final da Série A do Campeonato Brasileiro de 2012. Vale ressaltar que durante as 38 rodadas da competição, houveram 183 vitórias das equipes mandantes (48% do total), 92 vitórias das equipes visitantes (24% do total) e 105 empates (28% do total). Considerando a condição de mandante e de visitante em cada um dos dois turnos que foram disputados, a Tabela 3.2 apresenta o número médio de gols que foram marcados por cada uma das 20 equipes que disputaram a Série A Campeonato Brasileiro de 2012. Os valores entre parênteses indicam a quantidade de jogos disputados pela equipe no referido turno. Note que a equipe do Atlético Mineiro apresenta a maior média geral de gols marcados (1,67), seguido pelas equipes do Fluminense (1,61) e do Botafogo (1,60). Além disso, o Atlético Mineiro foi responsável pela maior média de gols no segundo turno, na condição de equipe mandante (média de 2,30 gols marcados por jogo). Por outro lado, a menor média de gols pertence à equipe do Bahia que, juntamente com Ponte Preta e Atlético Goianiense, teve o pior ataque da competição com apenas 37 gols marcados.

Tabela 3.2: Média de gols marcados por turno pelas equipes.

Equipe	Mandante		Visitante		Geral		Final
	1º Turno	2º Turno	1º Turno	2º Turno	1º Turno	2º Turno	
Fluminense	1,67 (9)	1,50 (10)	1,60 (10)	1,67 (9)	1,64 (19)	1,58 (19)	1,61 (38)
Atlético-MG	2,11 (9)	2,30 (10)	1,50 (10)	0,78 (9)	1,80 (19)	1,54 (19)	1,67 (38)
Grêmio	1,89 (9)	1,30 (10)	1,20 (10)	1,56 (9)	1,54 (19)	1,43 (19)	1,48 (38)
São Paulo	1,67 (9)	2,20 (10)	1,40 (10)	0,89 (9)	1,53 (19)	1,55 (19)	1,54 (38)
Vasco da Gama	1,70 (10)	1,22 (9)	0,89 (9)	0,90 (10)	1,29 (19)	1,06 (19)	1,18 (38)
Corinthians	1,20 (10)	2,11 (9)	0,89 (9)	1,20 (10)	1,04 (19)	1,65 (19)	1,34 (38)
Botafogo	1,50 (10)	2,22 (9)	1,67 (9)	1,00 (10)	1,58 (19)	1,61 (19)	1,60 (38)
Santos	2,00 (9)	1,40 (10)	0,50 (10)	1,44 (9)	1,25 (19)	1,42 (19)	1,34 (38)
Cruzeiro	1,20 (10)	1,56 (9)	1,33 (9)	0,90 (10)	1,27 (19)	1,23 (19)	1,25 (38)
Internacional	1,20 (10)	1,78 (9)	1,11 (9)	0,60 (10)	1,16 (19)	1,19 (19)	1,17 (38)
Flamengo	1,67 (9)	0,80 (10)	0,80 (10)	0,89 (9)	1,23 (19)	0,84 (19)	1,03 (38)
Náutico	1,89 (9)	1,40 (10)	0,70 (10)	0,67 (9)	1,29 (19)	1,03 (19)	1,16 (38)
Coritiba	1,89 (9)	1,60 (10)	1,40 (10)	0,67 (9)	1,64 (19)	1,14 (19)	1,39 (38)
Ponte Preta	1,67 (10)	1,11 (9)	1,00 (9)	0,60 (10)	1,34 (19)	0,86 (19)	1,10 (38)
Bahia	0,80 (10)	0,89 (9)	0,78 (9)	1,40 (10)	0,79 (19)	1,15 (19)	0,97 (38)
Portuguesa	1,11 (9)	1,50 (10)	0,70 (10)	0,78 (9)	0,90 (19)	1,14 (19)	1,02 (38)
Sport Recife	0,70 (10)	1,67 (9)	0,67 (9)	1,10 (10)	0,68 (19)	1,38 (19)	1,03 (38)
Palmeiras	1,10 (10)	1,44 (9)	0,78 (9)	0,80 (10)	0,94 (19)	1,12 (19)	1,03 (38)
Atlético-GO	1,44 (9)	0,80 (10)	0,70 (10)	1,00 (9)	1,07 (19)	0,90 (19)	0,99 (38)
Figueirense	1,10 (10)	1,33 (9)	0,78 (9)	0,90 (10)	0,94 (19)	1,12 (19)	1,03 (38)

3.4.2 Análise Descritiva

Para introduzir maior simplicidade na apresentação dos resultados obtidos, considere as variáveis aleatórias X : contagem do número de gols marcados pelas equipes mandantes e Y : contagem do número de gols marcados pelas equipes visitantes. Utilizaremos tal notação para essas variáveis ao decorrer deste trabalho. Na Tabela 3.3 é apresentado um breve resumo descritivo das variáveis X e Y a partir dos dados da Série A do Campeonato Brasileiro de 2012.

Observe que a média do número de gols dos mandantes é superior à dos visitantes. Além disso, note que em 75% das partidas disputadas, ambas as equipes mandantes e visitantes marcaram no máximo 2 gols. Ao término da competição, a covariância amostral observada entre as variáveis X e Y foi de aproximadamente 0,04. Note que as médias e as variâncias são relativamente próximas,

Tabela 3.3: Resumo descritivo do número de gols marcados.

Variáveis	Medidas					
	Média	Variância	CV(%)	1º Quartil	Mediana	3º Quartil
X	1,47	1,37	79,68	1,00	1,00	2,00
Y	1,00	0,96	97,88	0,00	1,00	2,00

o que nos permite especificar o modelo Poisson para a modelagem univariada do número de gols marcados tanto pelas equipes mandantes quanto pelas equipes visitantes. A Tabela 3.4 apresenta

as frequências cruzadas das variáveis X e Y .

Tabela 3.4: Frequências do número de gols marcados.

X	Y				
	0	1	2	3	4
0	36	24	20	2	2
1	49	50	22	9	2
2	28	34	18	7	2
3	19	23	11	1	2
4	5	5	4	1	0
5	0	3	0	0	0
6	1	0	0	0	0

Note que o placar mais frequente durante a disputa da competição foi o empate em 1–1 (50 vezes), seguido pela vitória da equipe mandante por 1–0 (49 vezes). Por meio da tabela também é possível observar que a maior goleada do campeonato foi de 6–0, que ocorreu na partida entre Atlético Mineiro (mandante) e Figueirense (visitante) válida pela 28ª rodada. Em particular, observe que a equipe do Atlético Mineiro foi a única a marcar 6 gols em uma mesma partida ao longo da competição.

A Figura 3.1 ilustra as frequências marginais das variáveis X e Y . Observe que as maiores quantidades de gols apresentam baixas frequências, uma vez que trata-se de uma competição historicamente equilibrada. Além disso, note que em uma quantidade maior de partidas, as equipes

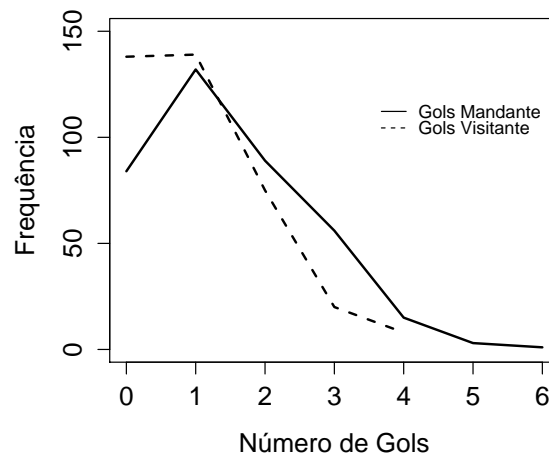


Figura 3.1: Frequência do número de gols marcados.

visitantes marcaram exatamente 1 gol. Observe também que as equipes mandantes marcaram no máximo 6 gols (em apenas uma partida), ao passo que as equipes visitantes marcaram no máximo 4 gols (em exatamente oito partidas).

3.4.3 Metodologia

Para obtenção de ajustes do modelo de regressão Poisson bivariado aos dados da Série A do Campeonato Brasileiro de 2012 vamos, no mínimo, fazer uso das 19 primeiras rodadas (primeiro

turno). Essa restrição se justifica pelo simples fato de que, dessa forma, estaremos permitindo que se observe, pelo menos uma vez, o confronto direto entre todas as equipes. Nesse sentido, vamos considerar uma estrutura baseada na ideia de ajuste-previsão para os dados do segundo turno da competição. Essa estrutura define um algoritmo na qual se obtém um ajuste do modelo de interesse aos dados das r primeiras rodadas e, conseqüentemente, realizam-se previsões para as partidas da rodada $r + 1$, $r < 38$. A Tabela 3.5 ilustra a estrutura ajuste-previsão que será utilizada. Note que RC refere-se as rodadas que serão consideradas para ajuste em cada etapa e RP indica a rodada que será prevista. TCQ refere-se aos testes chi-quadrado de aderência à distribuição Poisson bivariada que foram realizados para cada particular sub-amostra observada do número de gols marcados pelas equipes.

Tabela 3.5: Rodadas utilizadas para ajuste e previsão em cada etapa.

RC	TCQ (p-valor)	RP	RC	TCQ (p-valor)	RP
1 ^a – 19 ^a	16,35 (0,43)	20 ^a	1 ^a – 29 ^a	22,05 (0,34)	30 ^a
1 ^a – 20 ^a	17,23 (0,37)	21 ^a	1 ^a – 30 ^a	21,23 (0,38)	31 ^a
1 ^a – 21 ^a	16,69 (0,41)	22 ^a	1 ^a – 31 ^a	19,82 (0,47)	32 ^a
1 ^a – 22 ^a	16,52 (0,42)	23 ^a	1 ^a – 32 ^a	20,33 (0,44)	33 ^a
1 ^a – 23 ^a	16,23 (0,43)	24 ^a	1 ^a – 33 ^a	18,72 (0,54)	34 ^a
1 ^a – 24 ^a	15,63 (0,48)	25 ^a	1 ^a – 34 ^a	18,59 (0,55)	35 ^a
1 ^a – 25 ^a	16,08 (0,45)	26 ^a	1 ^a – 35 ^a	19,37 (0,50)	36 ^a
1 ^a – 26 ^a	17,93 (0,33)	27 ^a	1 ^a – 36 ^a	19,01 (0,52)	37 ^a
1 ^a – 27 ^a	18,27 (0,31)	28 ^a	1 ^a – 37 ^a	19,54 (0,49)	38 ^a
1 ^a – 28 ^a	20,34 (0,44)	29 ^a	-	-	-

Para um nível de significância fixado em 5%, podemos afirmar que, em cada particular sub-amostra considerada, existe uma dependência entre o número de gols marcados pelas equipes mandantes e visitantes que pode ser modelada por meio do modelo Poisson bivariado. Esse modelo será ajustado e utilizado para os procedimentos de previsão, mas, para efeitos comparativos, vamos forçar uma má especificação, também sugerindo que o modelo Poisson duplo seja utilizado para os mesmos fins. O teste do índice de dispersão também fora utilizado, sendo aqui omitido por apresentar resultados idênticos ao do teste chi-quadrado. Na Tabela 3.6 apresentamos os modelos de regressão Poisson duplo (PD) e Poisson bivariado (PB) que serão utilizados. Esses modelos são formulados de acordo com a estrutura descrita na seção 3.2. Note que o parâmetro λ_3 do modelo Poisson bivariado é assumido constante para todas as observações.

Tabela 3.6: Estrutura dos modelos propostos.

Modelo	Efeitos		
	$\log(\lambda_{1ij})$	$\log(\lambda_{2ij})$	$\log(\lambda_{3ij})$
PD	$\mu + \omega + \alpha_i + \beta_j$	$\mu + \alpha_j + \beta_i$	$-\infty$
PB	$\mu + \omega + \alpha_i + \beta_j$	$\mu + \alpha_j + \beta_i$	γ

Para evitar a superparametrização dos modelos e garantir a identificabilidade dos parâmetros, vamos utilizar o conjunto de restrições de soma zero sobre os efeitos ofensivos e defensivos das equipes.

3.4.4 Resultados Obtidos

Os modelos de regressão PD e PB especificados anteriormente foram ajustados a cada uma das sub-amostras do número de gols marcados. Na Tabela 3.7 encontram-se os valores de AICc (AIC corrigido) e BIC obtidos para cada particular ajuste dos modelos. Vale ressaltar que o número de coeficientes estimados é de 40 (modelo PD) e 41 (modelo PB), respectivamente.

Tabela 3.7: Critérios de seleção para comparação dos modelos.

RC	Modelo PD		Modelo PB		RC	Modelo PD		Modelo PB	
	AICc	BIC	AICc	BIC		AICc	BIC	AICc	BIC
1 ^a – 19 ^a	1110,69	1218,55	1099,42	1209,28	1 ^a – 29 ^a	1653,29	1786,91	1649,70	1786,28
1 ^a – 20 ^a	1169,98	1281,28	1161,70	1275,13	1 ^a – 30 ^a	1702,22	1837,71	1699,66	1838,16
1 ^a – 21 ^a	1223,67	1338,14	1216,60	1333,33	1 ^a – 31 ^a	1762,66	1899,93	1760,37	1900,71
1 ^a – 22 ^a	1274,58	1392,00	1266,90	1386,69	1 ^a – 32 ^a	1815,74	1954,72	1812,73	1954,84
1 ^a – 23 ^a	1336,08	1456,25	1331,18	1453,82	1 ^a – 33 ^a	1867,18	2007,79	1862,83	2006,63
1 ^a – 24 ^a	1381,14	1503,88	1376,03	1501,35	1 ^a – 34 ^a	1926,93	2069,12	1925,54	2070,97
1 ^a – 25 ^a	1438,15	1563,31	1433,47	1561,30	1 ^a – 35 ^a	1983,62	2127,32	1981,72	2128,71
1 ^a – 26 ^a	1494,17	1621,62	1488,13	1618,32	1 ^a – 36 ^a	2040,79	2185,95	2040,47	2188,97
1 ^a – 27 ^a	1541,41	1671,02	1535,76	1668,19	1 ^a – 37 ^a	2095,02	2241,59	2094,74	2244,70
1 ^a – 28 ^a	1593,69	1725,36	1587,00	1721,56	-	-	-	-	-

Note que o critério BIC tende a levemente penalizar o modelo PB a medida que são incrementadas rodadas ao ajuste. Além disso, observe que quanto maior o número de rodadas utilizadas, menor é melhoria na verossimilhança promovida pela inserção do parâmetro de covariância. De forma geral, podemos nos basear no critério AICc para indicar o modelo PB como sendo aquele que apresenta melhor ajuste aos dados provenientes de cada uma das sub-amostras consideradas. Ainda que esses critérios (AICc e BIC) sejam úteis para o procedimento de seleção, podemos recorrer ao TRV assintótico para comparar os modelos PD e PB, uma vez que estes são encaixados. Os p-valores obtidos para cada um dos testes que foram realizados são apresentados graficamente na Figura 3.2.

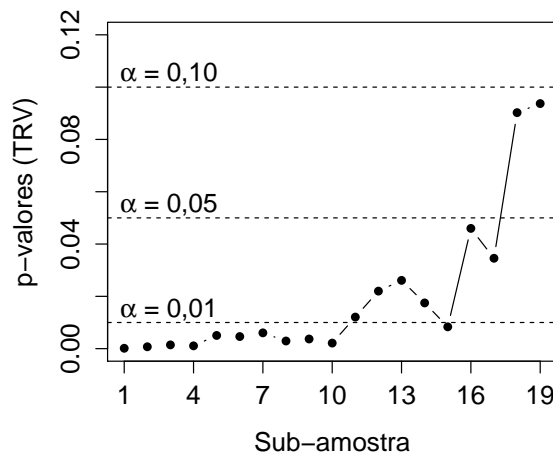


Figura 3.2: Ilustração dos p-valores associados aos TRV's.

Para um nível de significância fixado em 5% podemos afirmar que, em aproximadamente 90% dos ajustes obtidos, o modelo PB melhor se adequou aos dados considerados. Esse fato evidencia que, para essas particulares sub-amostras, existe uma dependência entre as variáveis X e Y que

pôde ser captada por meio da inserção de um parâmetro de covariância ao modelo PD. Note que praticamente não existe diferença entre os modelos PD e PB nos dois últimos ajustes. De forma geral, o TRV assintótico concorda com os critérios AICc e BIC na seleção do modelo que apresenta melhor ajuste.

A partir do ajuste sequencial dos modelos PD e PB foram obtidas estimativas para os efeitos ofensivos e defensivos das equipes ao longo do segundo turno da competição. Esses efeitos indicam a capacidade relativa de cada uma das equipes, sendo que a média marginal de gols é diretamente afetada pela magnitude dos mesmos. Como ilustração, as Figuras 3.3 e 3.4 apresentam o comportamento dos efeitos ofensivos *vs.* efeitos defensivos ao término do primeiro turno e ao término da competição. Note que os efeitos defensivos foram multiplicados por -1, de tal forma que maiores valores indicam equipes com melhores capacidades defensivas.

Note, na primeira ilustração, uma alta proximidade entre as equipes, principalmente quando os efeitos são estimados por meio do modelo Poisson duplo. Além disso, observe que a equipe do Sport Recife realizou uma campanha bastante fraca durante o primeiro turno, o que acabou por sendo determinante para a sua despromoção ao término da competição. Curiosamente, a equipe do Palmeiras, que também acabou despromovida, realizou uma campanha razoável durante o primeiro turno. De forma geral, o modelo Poisson bivariado indicou, de forma mais clara, a superioridade da equipe do Fluminense ao longo do primeiro turno da competição.

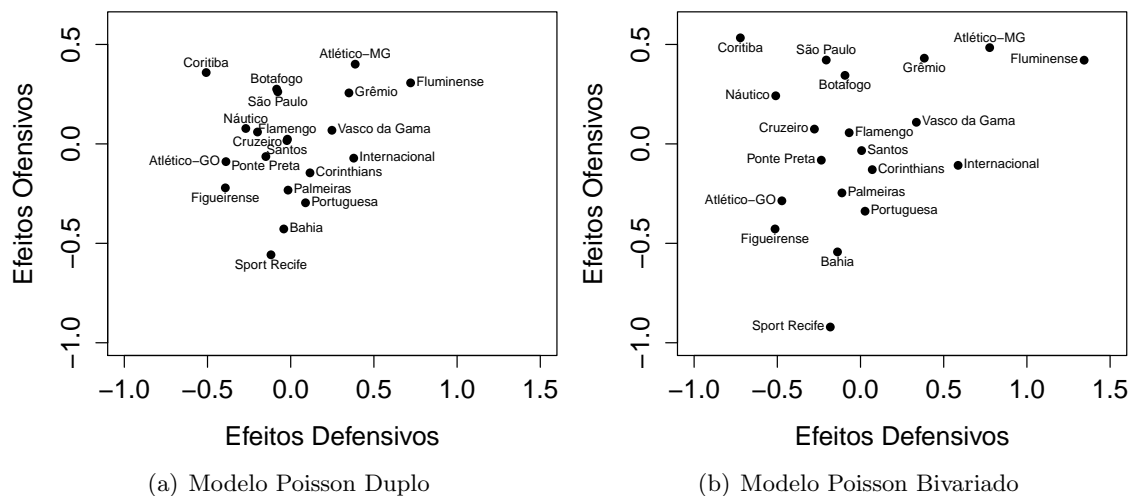


Figura 3.3: Efeitos ofensivos *vs.* defensivos das equipes ao término do primeiro turno.

Pela segunda ilustração observe que as melhores equipes, ou seja, aquelas que apresentaram melhores performances ofensivas e defensivas durante a competição, estão concentradas na parte superior direita dos gráficos, tais como as equipes do Fluminense e do Atlético Mineiro. Note que os efeitos estimados por ambos os modelos são muito semelhantes, de forma que as médias marginais finais são aproximadamente iguais, a menos da covariância introduzida pelo modelo PB. Para avaliar o quão diferentes foram as quatro primeiras equipes e o quanto suas capacidades variaram durante o segundo turno da competição, apresentamos na Figura 3.5 a sequência de estimativas dos efeitos ofensivos e defensivos. Note que os efeitos não são uniformes, indicando que a performance das equipes é altamente dinâmica.

Observe que as três primeiras equipes classificadas apresentaram performances ofensivas bastante similares. Em termos defensivos, o Fluminense e o Atlético Mineiro se destacaram, e a equi-

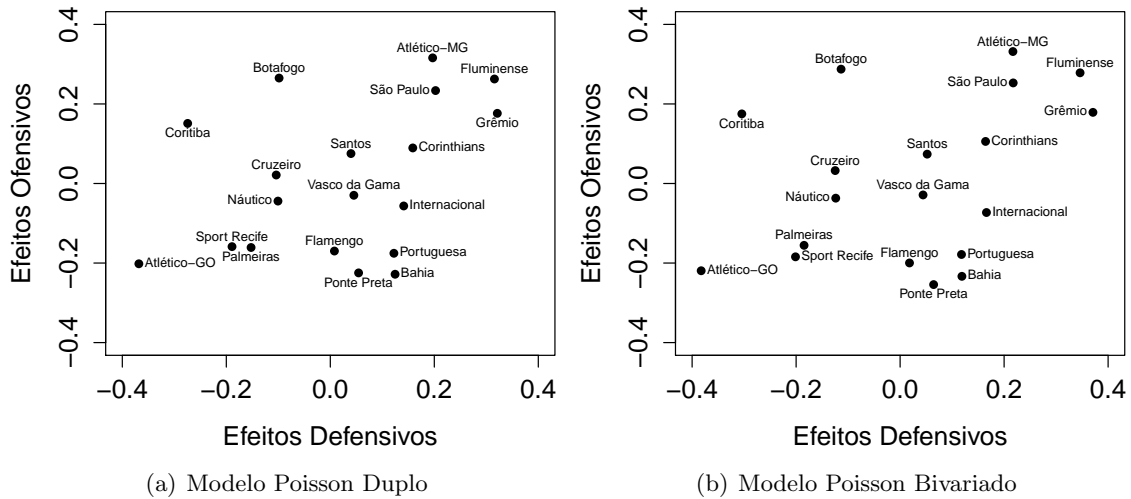


Figura 3.4: Efeitos ofensivos *vs.* defensivos das equipes ao término da competição.

pe do São Paulo melhorou seu desempenho ao longo do segundo turno. Houve um grande equilíbrio entre o Atlético Mineiro e o Grêmio, sendo que, devido ao melhor desempenho ofensivo, o Atlético acabou por terminar à frente. Em contrapartida, note que um melhor desempenho defensivo propi-

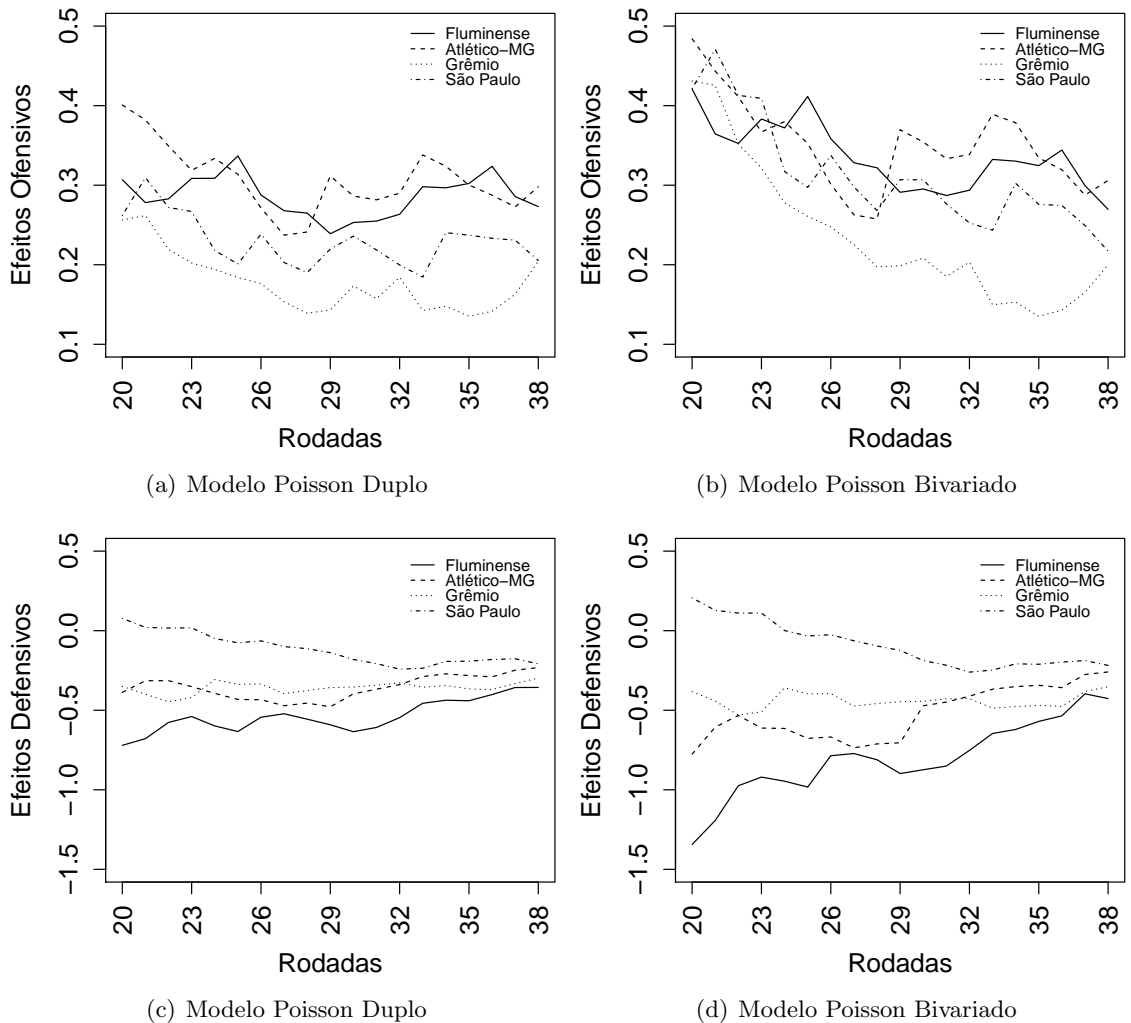


Figura 3.5: Evolução dos efeitos ofensivos e defensivos ao longo do segundo turno.

ciou ao Grêmio terminar à frente do São Paulo, ainda que durante o período avaliado esta equipe tenha apresentado performance ofensiva levemente superior. De forma geral, a equipe do Fluminense se mostrou a mais forte durante o segundo turno da competição. Adicionalmente, na Figura 3.6 apresentamos a sequência de estimativas para o efeito casa. Como era esperado, esse efeito se mantém aproximadamente constante ao longo do tempo.

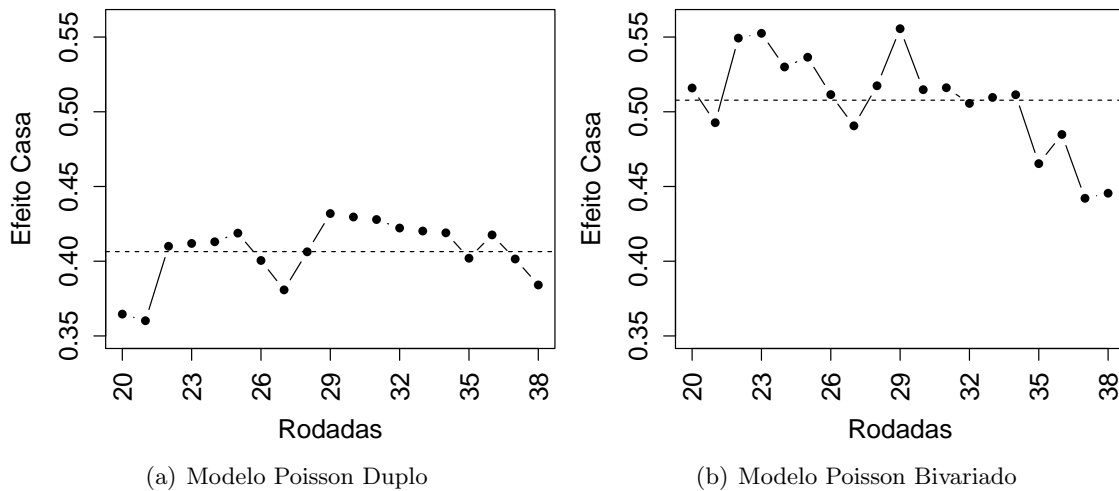


Figura 3.6: Evolução do efeito casa ao longo do segundo turno.

As estimativas de máxima verossimilhança obtidas para os efeitos ao longo do tempo podem ser utilizadas para diversos fins, tais como a realização de uma simulação da competição. Nas Tabelas 3.8 e 3.9 apresentamos a pontuação final estimada (por rodada de referência) a partir de 10000 simulações dos modelos PD e PB, respectivamente. A rodada de referência indica que os efeitos atuais refletem as capacidades ofensivas e defensivas de cada uma das equipes até a rodada imediatamente anterior. Para fins de comparação, na última linha de cada uma das tabelas encontra-se a pontuação final observada de cada uma das equipes. Note que as pontuações finais estimadas por ambos os modelos são muito próximas entre si e daquilo que de fato fora observado. Uma evidência da superioridade da equipe do Fluminense é que sua pontuação final fora estimada em 79,2 por ambos os modelos, sendo que, na verdade, a equipe obteve 77 pontos. Além disso, observe que a pontuação de equipes como Botafogo, Internacional e Ponte Preta foram estimadas com bastante acurácia.

Note que a partir da 28^a rodada, uma leve queda de rendimento da equipe do Atlético Mineiro fez com que a probabilidade do Fluminense tornar-se o campeão da competição ultrapassasse os 80%. Além disso, um bom início de segundo turno fez com que a equipe do Grêmio fosse postulada como concorrente ao título, mas a superioridade do Fluminense fez com que as chances do Grêmio decaíssem rapidamente para zero. Por outro lado, podemos observar que Fluminense, Atlético Mineiro e Grêmio tiveram bastante facilidade para se manter entre os quatro primeiros e se classificarem para a Taça Libertadores da América. Durante a competição, houve uma disputa bastante acirrada entre as equipes do São Paulo e do Vasco da Gama pela quarta vaga na competição internacional. Como se nota, essa disputa perdurou até a 29^a rodada, quando o São Paulo disparou e acabou por conseguir a quarta vaga. Na parte inferior da tabela, a luta contra a despromoção envolvia seis equipes no início do segundo turno. Porém, as equipes do Bahia e da Portuguesa mantiveram-se fora da zona de rebaixamento e as equipes do Sport Recife e do Palmeiras, que lu-

Tabela 3.8: Pontuação final estimada por meio do modelo Poisson duplo.

Rodada	Classificação Final																			
	Referência	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
20 ^a	82,6	81,6	72,9	60,9	65,3	49,2	57,5	52,2	51,8	60,5	55,4	47,3	43,5	45,8	35,9	44,9	31,3	38,0	34,9	30,2
21 ^a	80,6	79,2	74,8	63,7	63,6	50,5	54,8	49,0	54,5	58,4	53,7	48,1	45,6	47,4	39,9	47,9	32,1	35,5	32,3	29,7
22 ^a	78,5	77,1	73,9	61,6	65,4	52,8	55,9	47,0	56,5	60,6	51,9	46,4	44,5	48,7	42,4	46,3	34,5	35,7	31,6	30,7
23 ^a	79,2	76,3	73,8	61,1	64,8	50,6	58,6	46,5	53,8	60,5	49,4	46,1	42,7	51,3	42,4	48,8	33,1	37,8	31,7	33,0
24 ^a	81,4	77,7	71,0	60,9	60,6	53,7	59,7	46,2	52,0	58,9	47,4	45,1	44,6	49,7	47,0	48,1	35,2	36,7	31,3	34,0
25 ^a	83,3	78,6	71,6	60,3	61,8	52,7	59,0	48,1	50,1	59,0	46,0	44,6	46,5	50,0	46,7	46,6	34,8	35,6	29,5	36,1
26 ^a	79,4	76,3	71,0	61,7	61,7	54,8	58,8	50,6	49,8	57,3	46,1	47,2	44,2	49,4	47,8	45,3	35,7	33,6	33,1	35,2
27 ^a	79,5	75,5	71,3	62,5	61,6	55,0	57,9	48,1	49,1	58,9	48,2	47,0	42,4	48,8	46,5	48,1	37,4	36,4	31,3	33,0
28 ^a	80,8	74,7	70,0	62,0	62,5	56,4	56,2	48,9	48,7	58,4	47,0	48,0	42,1	46,8	48,4	48,3	36,5	38,8	30,2	32,5
29 ^a	81,4	75,9	70,6	63,4	64,0	54,4	55,8	48,6	48,1	58,1	46,4	50,0	43,5	45,6	48,3	50,1	35,2	37,7	29,1	31,7
30 ^a	83,1	73,9	71,9	65,6	62,0	55,3	53,5	51,5	49,9	60,6	45,8	48,8	45,7	46,7	46,9	48,6	34,3	35,5	28,0	33,0
31 ^a	83,6	74,1	70,7	65,9	60,6	55,1	54,1	53,2	49,8	58,2	45,2	50,0	47,2	46,5	45,7	48,1	34,2	34,6	30,5	32,6
32 ^a	82,5	73,5	71,1	66,6	59,7	53,7	55,3	53,0	51,7	55,5	45,4	49,1	48,3	44,8	43,6	47,2	36,1	36,9	30,1	35,4
33 ^a	81,2	75,0	69,6	64,9	57,9	52,9	56,9	51,8	50,1	57,5	47,3	48,4	49,2	46,5	43,9	47,0	38,2	38,7	28,1	33,9
34 ^a	81,7	73,5	69,0	66,2	57,0	54,2	57,6	50,7	49,0	58,3	48,0	48,6	48,8	48,0	43,9	46,5	37,2	38,1	27,5	33,6
35 ^a	81,6	72,0	70,0	65,6	54,4	55,6	57,2	53,0	47,1	56,7	49,0	50,4	51,0	47,4	46,1	44,7	39,9	37,6	26,2	32,7
36 ^a	82,7	71,3	71,3	64,6	54,3	56,8	58,4	51,2	48,6	55,3	51,1	48,5	50,0	49,2	44,8	43,7	39,8	36,8	28,3	32,0
37 ^a	80,1	69,5	70,9	65,3	56,4	58,8	56,8	51,8	51,4	53,6	50,3	47,9	48,2	48,2	46,1	43,6	41,7	36,8	29,1	31,5
38 ^a	79,2	71,2	72,0	64,8	55,6	58,0	55,3	52,0	52,6	51,7	50,4	47,9	47,2	48,0	45,4	45,7	41,9	34,7	31,3	30,6
Observado	77	72	71	66	58	57	55	53	52	52	50	49	48	48	47	45	41	34	30	30

Tabela 3.9: Pontuação final estimada por meio do modelo Poisson bivariado.

Rodada	Classificação Final																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
20 ^a	83,8	82,7	73,7	60,7	64,9	48,4	57,4	51,5	51,4	59,9	55,0	46,6	43,6	45,0	34,5	43,7	29,3	37,1	33,1	28,3
21 ^a	81,7	80,0	75,5	63,5	63,3	49,7	54,6	48,1	54,1	57,8	53,3	47,2	45,7	46,7	39,2	47,1	30,3	34,7	30,7	28,5
22 ^a	79,5	77,5	74,5	61,5	65,1	52,2	55,7	46,3	56,1	60,3	51,3	45,8	45,0	48,2	41,4	45,6	33,0	34,8	30,3	29,5
23 ^a	80,2	76,9	74,3	61,1	64,5	50,0	58,5	45,7	53,3	60,3	48,6	45,5	43,2	51,0	41,7	47,9	31,6	37,1	30,4	31,9
24 ^a	82,4	78,2	71,3	60,9	60,2	53,5	59,3	45,6	51,5	58,7	46,9	44,7	44,9	49,2	46,9	47,5	34,1	36,2	30,2	33,1
25 ^a	84,3	79,0	71,8	60,6	61,4	52,5	58,7	47,5	49,5	58,8	45,6	44,3	46,6	49,7	46,4	46,1	33,8	35,2	28,4	35,3
26 ^a	80,1	76,6	71,1	61,9	61,3	54,7	58,6	50,2	49,3	57,1	45,7	47,1	44,1	49,0	47,5	45,0	34,7	33,0	32,1	34,8
27 ^a	80,0	75,8	71,6	62,7	61,3	55,0	57,8	47,6	48,9	58,9	47,7	47,0	42,3	48,2	46,1	47,8	36,6	36,0	30,4	32,3
28 ^a	81,3	74,9	70,1	62,1	62,2	56,3	56,1	48,5	48,4	58,4	46,6	47,9	42,1	46,3	48,2	48,2	35,7	38,4	29,4	32,0
29 ^a	81,8	76,0	70,7	63,4	63,7	54,4	55,7	48,4	47,9	58,0	46,0	49,9	43,2	45,2	48,0	50,0	34,5	37,5	28,3	31,3
30 ^a	83,4	73,9	71,9	65,7	61,7	55,3	53,6	51,4	49,8	60,7	45,5	48,8	45,6	46,4	46,8	48,4	33,7	35,3	27,4	32,6
31 ^a	83,9	74,1	70,7	66,0	60,3	55,2	54,1	53,1	49,7	58,3	44,9	49,8	47,1	46,2	45,7	48,0	33,6	34,3	30,2	32,3
32 ^a	82,7	73,5	71,2	66,5	59,5	53,8	55,3	53,0	51,5	55,5	45,1	49,0	48,1	44,7	43,5	47,0	35,6	36,9	29,8	35,2
33 ^a	81,4	75,2	69,6	64,8	57,8	52,8	56,9	51,7	50,1	57,4	47,2	48,3	49,0	46,3	43,8	46,8	37,9	38,5	27,8	33,8
34 ^a	81,9	73,6	69,1	66,1	56,8	54,1	57,6	50,7	49,0	58,3	47,8	48,5	48,8	47,7	43,8	46,3	36,8	38,0	27,2	33,4
35 ^a	81,7	72,0	70,0	65,6	54,3	55,5	57,2	53,0	47,0	56,6	48,9	50,3	51,0	47,3	46,0	44,6	39,7	37,6	26,0	32,5
36 ^a	82,8	71,4	71,3	64,6	54,2	56,7	58,3	51,2	48,6	55,3	51,1	48,5	50,0	49,1	44,9	43,6	39,6	36,7	28,1	31,8
37 ^a	80,1	69,6	70,9	65,3	56,3	58,8	56,8	51,9	51,4	53,6	50,3	47,9	48,2	48,2	46,1	43,6	41,7	36,8	29,0	31,4
38 ^a	79,2	71,2	72,0	64,8	55,6	57,9	55,3	52,1	52,5	51,7	50,4	47,9	47,3	47,9	45,4	45,8	41,8	34,7	31,3	30,6
Observado	77	72	71	66	58	57	55	53	52	52	50	49	48	48	47	45	41	34	30	30

taram até a última rodada contra a queda, acabaram sendo despromovidas juntamente com Atlético Goianiense e Figueirense que, por realizarem campanhas muito fracas durante toda a competição, sempre mantiverem bastante elevadas as probabilidades de despromoção.

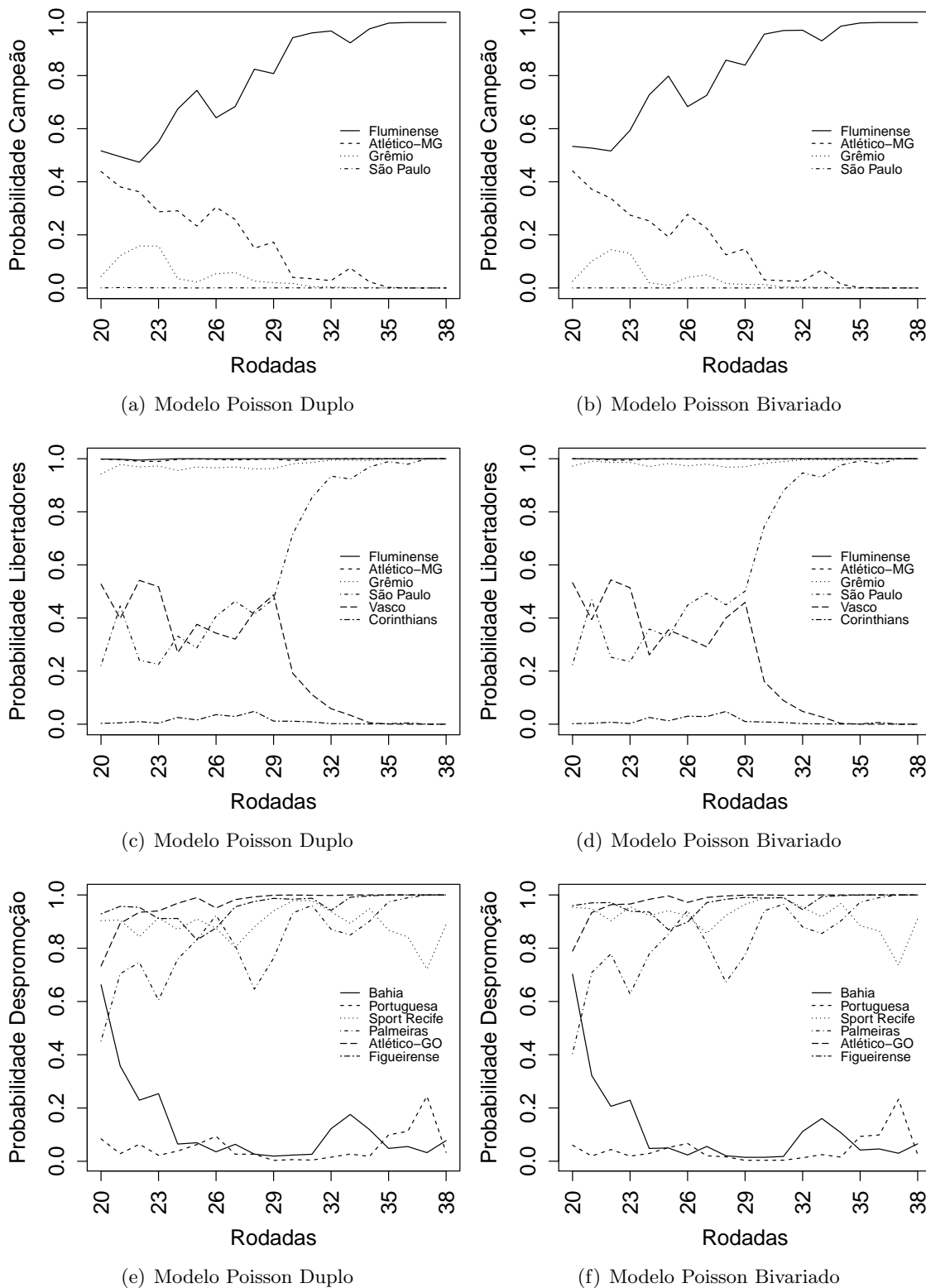


Figura 3.7: Probabilidades estimadas para título, libertadores e despromoção.

A partir do procedimento de simulação da competição também podemos estimar as probabilidades de que uma determinada equipe vença a competição, se classifique para a Taça Libertadores

da América ou seja despromovida à Série B. A Figura 3.7 ilustra a evolução dessas probabilidades ao longo do segundo turno, considerando as equipes que participaram diretamente de tais disputas.

A Figura 3.8 ilustra de forma mais completa a classificação final observada e a classificação final esperada segundo os modelos PD e PB. A classificação final esperada é obtida de forma similar à pontuação final estimada, por meio das 1000 simulações dos modelos PD e PB. Nesse caso, a posição final, estimada com base no desempenho recente das equipes, resulta da média entre as posições finais estimadas ao término de cada uma das rodadas do segundo turno. Além disso, em ambas as representações, as linhas tracejadas indicam os limites empíricos (baseados nos valores mínimos e máximos observados no processo de simulação) para a classificação final de cada uma das 20 equipes que disputaram a competição.

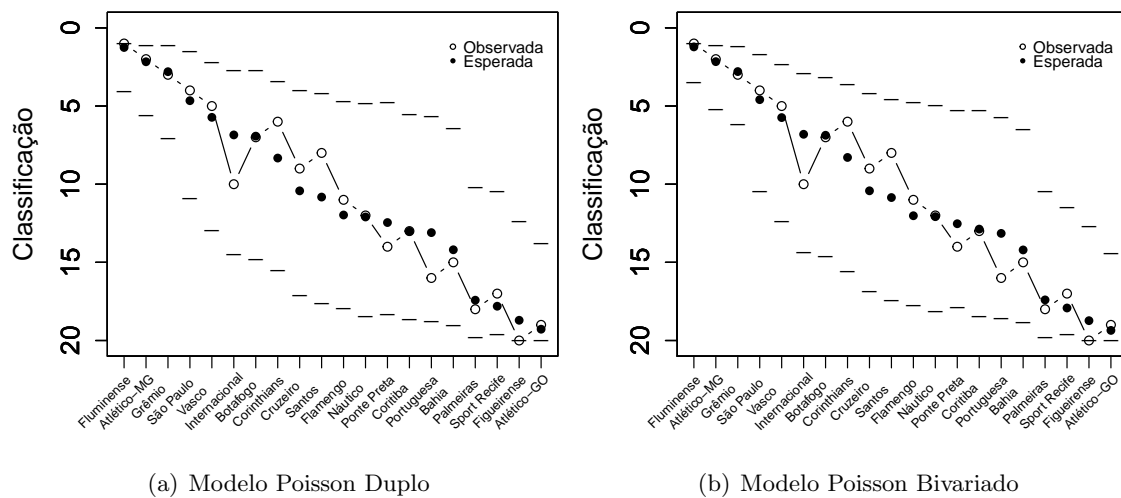


Figura 3.8: Classificação final esperada da Série A do Campeonato Brasileiro de 2012.

Por meio dos gráficos podemos observar que as equipes do Corinthians e do Santos realizaram melhores campanhas do que se esperava, ao passo que Internacional e Portuguesa estiveram aquém das expectativas. Uma possível justificativa para que o desempenho esperado do Corinthians tenha sido abaixo do observado está no fato de que esta equipe esteve, durante boa parte do primeiro turno, dedicada à Taça Libertadores, realizando uma fraca campanha durante esse período da competição. Além disso, observe que a incerteza acerca da classificação final é maior para as equipes que terminaram a competição na zona intermediária, ainda que os limites obtidos por meio do modelo PB sejam inferiores àqueles obtidos por meio do modelo PD. Por exemplo, o Flamengo, que terminou na 11^a colocação, poderia tanto ter terminado em 5^o, como em 18^o (o que acarretaria uma despromoção à Série B). De fato, a instabilidade das equipes intermediárias tende a causar grandes incertezas acerca da classificação final. É interessante notar que as equipes naturalmente se dividiram em quatro grupos: o grupo dos três primeiros, o grupo dos seis intermediários da parte de cima, o grupo dos oito intermediários da parte de baixo e o grupo dos quatro últimos.

3.4.5 Previsão de Resultados

Uma vez obtidos ajustes sequenciais dos modelos PD e PB, podemos realizar procedimentos de previsão. Esses procedimentos baseiam-se no cálculo das probabilidades de vitória das equipes mandantes (PVM) e das probabilidades de vitória das equipes visitantes (PVV), sendo que as probabilidades de empate (PE) são dadas pelos complementares das somas de PVM e PVV. Essas

Tabela 3.10: Previsões para a 24^a rodada.

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti
Modelo Poisson Duplo						
Atlético-GO	0,3963	0,2173	0,3864	Coritiba	1-2	0,5808
Atlético-MG	0,6333	0,2168	0,1499	São Paulo	1-0	0,2040
Botafogo	0,4298	0,2890	0,2812	Internacional	1-1	0,7694
Corinthians	0,5314	0,2694	0,1991	Ponte Preta	1-1	0,8559
Figueirense	0,4017	0,2431	0,3552	Cruzeiro	2-0	0,5433
Grêmio	0,7421	0,1634	0,0945	Náutico	2-0	0,1021
Portuguesa	0,2110	0,3231	0,4659	Fluminense	0-2	0,4342
Santos	0,5173	0,2560	0,2267	Flamengo	2-0	0,3499
Sport Recife	0,3117	0,3387	0,3495	Bahia	1-1	0,6566
Vasco da Gama	0,5820	0,2602	0,1577	Palmeiras	3-1	0,2673
Modelo Poisson Bivariado						
Atlético-GO	0,3550	0,2463	0,3987	Coritiba	1-2	0,5482
Atlético-MG	0,6573	0,2359	0,1067	São Paulo	1-0	0,1845
Botafogo	0,3809	0,3638	0,2554	Internacional	1-1	0,6150
Corinthians	0,5224	0,3136	0,1641	Ponte Preta	1-1	0,7709
Figueirense	0,3789	0,2836	0,3375	Cruzeiro	2-0	0,5802
Grêmio	0,7856	0,1500	0,0644	Náutico	2-0	0,0726
Portuguesa	0,1479	0,3918	0,4603	Fluminense	0-2	0,4666
Santos	0,4845	0,3201	0,1954	Flamengo	2-0	0,4064
Sport Recife	0,2593	0,4172	0,3234	Bahia	1-1	0,5115
Vasco da Gama	0,5946	0,2785	0,1269	Palmeiras	3-1	0,2580

Tabela 3.11: Previsões para a 28^a rodada.

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti
Modelo Poisson Duplo						
Atlético-GO	0,2704	0,2664	0,4631	Vasco da Gama	0-1	0,4323
Atlético-MG	0,8263	0,1207	0,0530	Figueirense	6-0	0,0475
Coritiba	0,5183	0,2256	0,2561	Ponte Preta	1-0	0,3485
Flamengo	0,4004	0,3002	0,2994	Bahia	0-0	0,7397
Fluminense	0,6933	0,1950	0,1117	Botafogo	1-0	0,1446
Grêmio	0,6690	0,2141	0,1168	Cruzeiro	2-1	0,1690
Náutico	0,3523	0,2782	0,3695	Corinthians	2-1	0,6335
Portuguesa	0,5852	0,2632	0,1517	Sport Recife	5-1	0,2643
Santos	0,3434	0,3064	0,3502	Internacional	1-1	0,7216
São Paulo	0,6562	0,2100	0,1338	Palmeiras	3-0	0,1802
Modelo Poisson Bivariado						
Atlético-GO	0,2299	0,3203	0,4498	Vasco da Gama	0-1	0,4582
Atlético-MG	0,8414	0,1267	0,0319	Figueirense	6-0	0,0422
Coritiba	0,5402	0,2404	0,2194	Ponte Preta	1-0	0,3174
Flamengo	0,3843	0,3431	0,2726	Bahia	0-0	0,6535
Fluminense	0,7112	0,2141	0,0747	Botafogo	1-0	0,1348
Grêmio	0,6824	0,2311	0,0865	Cruzeiro	2-1	0,1618
Náutico	0,3456	0,3026	0,3518	Corinthians	2-1	0,6437
Portuguesa	0,5870	0,3009	0,1121	Sport Recife	5-1	0,2737
Santos	0,2768	0,4088	0,3143	Internacional	1-1	0,5250
São Paulo	0,6889	0,2071	0,1041	Palmeiras	3-0	0,1505

Tabela 3.12: Previsões para a 34^a rodada.

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti
Modelo Poisson Duplo						
Atlético-GO	0,2653	0,2712	0,4635	Corinthians	0-2	0,4318
Coritiba	0,2928	0,2395	0,4676	Atlético-MG	1-0	0,7761
Cruzeiro	0,4474	0,2740	0,2787	Santos	0-4	0,7955
Flamengo	0,6003	0,2217	0,1780	Figueirense	1-0	0,2405
Grêmio	0,6580	0,2244	0,1176	Ponte Preta	1-0	0,1812
Náutico	0,3438	0,2845	0,3717	Internacional	3-0	0,6497
Palmeiras	0,3496	0,2715	0,3788	Botafogo	2-2	0,7964
Portuguesa	0,4642	0,3191	0,2168	Bahia	0-1	0,9307
São Paulo	0,3876	0,2905	0,3219	Fluminense	1-1	0,7572
Vasco da Gama	0,6481	0,2152	0,1367	Sport Recife	0-3	1,2117
Modelo Poisson Bivariado						
Atlético-GO	0,2328	0,3135	0,4536	Corinthians	0-2	0,4510
Coritiba	0,2714	0,2622	0,4663	Atlético-MG	1-0	0,8170
Cruzeiro	0,4334	0,3144	0,2522	Santos	0-4	0,8459
Flamengo	0,5973	0,2471	0,1556	Figueirense	1-0	0,2474
Grêmio	0,6536	0,2563	0,0901	Ponte Preta	1-0	0,1938
Náutico	0,3171	0,3282	0,3547	Internacional	3-0	0,7000
Palmeiras	0,3458	0,2921	0,3621	Botafogo	2-2	0,7518
Portuguesa	0,4593	0,3503	0,1903	Bahia	0-1	0,9892
São Paulo	0,3444	0,3445	0,3111	Fluminense	1-1	0,6452
Vasco da Gama	0,6574	0,2337	0,1090	Sport Recife	0-3	1,2807

Tabela 3.13: Previsões para a 38^a rodada.

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti
Modelo Poisson Duplo						
Atlético-GO	0,3453	0,2917	0,3630	Bahia	0-1	0,6101
Atlético-MG	0,6803	0,1945	0,1252	Cruzeiro	3-2	0,1557
Coritiba	0,6716	0,1732	0,1552	Figueirense	3-0	0,1619
Flamengo	0,3799	0,2712	0,3489	Botafogo	2-2	0,7971
Fluminense	0,6565	0,2191	0,1244	Vasco da Gama	1-2	1,2456
Grêmio	0,5858	0,2521	0,1621	Internacional	0-0	0,9287
Náutico	0,5466	0,2415	0,2119	Sport Recife	1-0	0,3088
Portuguesa	0,4753	0,3053	0,2194	Ponte Preta	0-0	0,7567
Santos	0,5915	0,2388	0,1698	Palmeiras	3-1	0,2527
São Paulo	0,5044	0,2707	0,2249	Corinthians	3-1	0,3695
Modelo Poisson Bivariado						
Atlético-GO	0,3218	0,3167	0,3615	Bahia	0-1	0,6115
Atlético-MG	0,6828	0,2032	0,1139	Cruzeiro	3-2	0,1549
Coritiba	0,7030	0,1673	0,1297	Figueirense	3-0	0,1330
Flamengo	0,3779	0,2916	0,3305	Botafogo	2-2	0,7538
Fluminense	0,6537	0,2360	0,1103	Vasco da Gama	1-2	1,2746
Grêmio	0,5763	0,2767	0,1470	Internacional	0-0	0,8769
Náutico	0,5478	0,2559	0,1963	Sport Recife	1-0	0,3086
Portuguesa	0,4703	0,3285	0,2013	Ponte Preta	0-0	0,7126
Santos	0,5947	0,2508	0,1545	Palmeiras	3-1	0,2510
São Paulo	0,5008	0,2882	0,2110	Corinthians	3-1	0,3768

probabilidades são calculadas com base na distribuição Poisson bivariada, cujos parâmetros são aqueles estimados via regressão. Segundo a estrutura ajuste-previsão, foram obtidas previsões para todos os jogos do segundo turno da Série A do Campeonato Brasileiro de 2012. Para abreviar a apresentação dos resultados, apenas as previsões para as rodadas 24, 28, 34 e 38 são apresentadas. Além das previsões, nas Tabelas 3.10, 3.11, 3.12 e 3.13 encontram-se os valores da distância de DeFinetti obtidos para cada particular partida prevista.

Note que a magnitude da distância de DeFinetti é diretamente proporcional à probabilidade prevista e ao placar observado. Ou seja, quanto maior a probabilidade que o modelo atribui para um determinado evento que ocorreu, menor é a distância de DeFinetti. Em contrapartida, distâncias superiores a 1 indicam um alto erro de previsão, tendo em vista que um resultado bastante improvável ocorreu. Pontualmente, tais erros podem ocorrer com alta frequência, dada a imprevisibilidade desse tipo de esporte. Entretanto, ao longo de uma determinada competição, altos erros de previsão tendem a ocorrer menos frequentemente a medida que as capacidades ofensivas e defensivas de cada uma das equipes vão se tornando evidentes. Outro aspecto relevante é que o modelo PB melhor prevê os empates em relação ao modelo PD, sendo que, em alguns casos, as previsões fornecidas por este modelo coincidem com os resultados de empate que de fato ocorreram.

Na Tabela 3.14 apresentamos o percentual de previsões corretas fornecidas pelos modelos PD e PB a cada rodada. Note que, em geral, o modelo PB apresenta melhor desempenho, sendo responsável por prever corretamente pelo menos 70% dos placares em quatro rodadas. Adicionalmente, apresentamos uma estatística da quantidade de vezes em que se observa medidas de DeFinetti obtidas por meio do modelo PB inferiores àquelas obtidas por meio do modelo PD em cada particular partida prevista. Esses valores são denotados por $\# PB < PD$ e nos auxiliam a verificar o quão melhor é o desempenho preditivo do modelo PB. De forma geral, observe que em aproximadamente 50% das rodadas previstas, o modelo PB fornece probabilidades mais acuradas do que o modelo PD para mais de seis partidas por rodada.

Tabela 3.14: Percentual de previsões corretas dos modelos a cada rodada.

RP	Modelo		# PB < PD	RP	Modelo		# PB < PD
	PD	PB			PD	PB	
20 ^a	40	50	5	30 ^a	60	60	7
21 ^a	50	50	4	31 ^a	50	40	5
22 ^a	40	40	5	32 ^a	40	30	2
23 ^a	30	30	6	33 ^a	60	70	7
24 ^a	60	80	7	34 ^a	30	40	2
25 ^a	30	20	6	35 ^a	60	50	4
26 ^a	40	40	5	36 ^a	30	30	4
27 ^a	60	70	6	37 ^a	20	20	6
28 ^a	70	80	7	38 ^a	60	60	7
29 ^a	60	60	2	-	-	-	-

A Figura 3.9 ilustra as médias aritméticas por rodada prevista das distâncias de DeFinetti. Essas distâncias são calculadas com base nas probabilidades estimadas de vitória da equipe mandante, empate e vitória da equipe visitante. A linha tracejada em 2/3 refere-se ao ponto de corte estabelecido para validação da qualidade preditiva dos modelos ajustados.

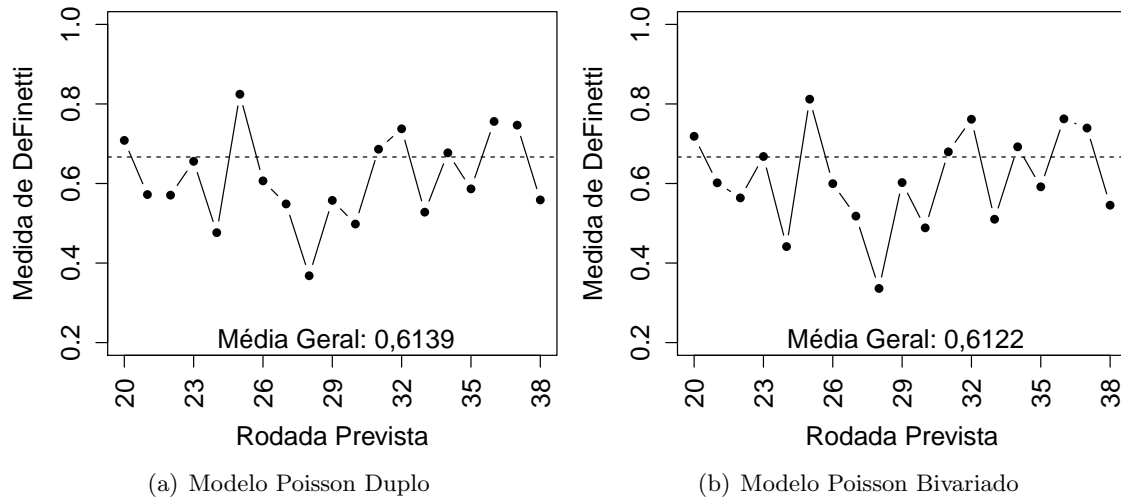


Figura 3.9: Medidas de Definetti para as rodadas previstas.

Observe que as médias mais elevadas (superiores a $2/3$) indicam que a referida rodada foi atípica, com a ocorrência de resultados pouco prováveis. Note que, a partir das previsões obtidas por meio do modelo PB, a média geral das distâncias de DeFinetti é levemente inferior em relação àquela do modelo PD (0,6139 contra 0,6122). De forma geral, ambos os modelos podem ser considerados válidos para previsão de resultados da Série A do Campeonato Brasileiro de 2012, podendo o modelo PB, invariavelmente, ser preferido ao PD. No que segue, apresentamos na Tabela 3.15 as taxas de sucesso obtidas para cada particular rodada prevista. Para efeitos de comparação foram obtidas previsões e as respectivas taxas de sucesso associadas ao modelo PD saturado (ST), no qual os valores estimados para os parâmetros λ_1 e λ_2 correspondem aos próprios valores observados do vetor aleatório bivariado (X, Y) .

Tabela 3.15: Taxa de sucesso dos modelos ajustados.

RP	Modelo			PD/ST (%)	PB/ST (%)	RP	Modelo			PD/ST (%)	PB/ST (%)
	PD	PB	ST				PD	PB	ST		
20 ^a	3,62	3,64	3,38	107,22	107,90	30 ^a	4,84	4,93	3,38	143,18	145,87
21 ^a	4,16	4,09	3,51	118,79	116,78	31 ^a	3,70	3,77	3,38	109,55	111,60
22 ^a	4,25	4,32	3,38	125,78	128,01	32 ^a	3,19	3,08	3,38	94,38	91,06
23 ^a	3,77	3,79	3,38	111,47	112,23	33 ^a	4,59	4,73	3,24	141,48	145,79
24 ^a	4,63	4,85	3,38	136,93	143,71	34 ^a	3,55	3,48	3,51	101,18	99,14
25 ^a	3,05	3,14	3,24	93,87	96,80	35 ^a	3,93	3,91	3,51	111,93	111,46
26 ^a	4,00	4,08	3,38	118,45	120,79	36 ^a	3,48	3,49	3,38	103,03	103,21
27 ^a	4,28	4,48	3,25	131,91	137,91	37 ^a	3,19	3,23	3,11	102,59	103,82
28 ^a	5,37	5,60	3,51	153,07	159,57	38 ^a	4,31	4,40	3,38	127,64	130,23
29 ^a	4,11	3,93	3,78	108,75	103,88	-	-	-	-	-	-

Note que a taxa de sucesso reforça a similaridade entre os modelos, embora, no geral, o modelo PB apresente taxas levemente superiores as do modelo PD. Além disso, podemos assumir que praticamente não existe perda em termos de predição, tendo em vista que, percentualmente, ambos os modelos apresentam taxas de sucesso próximas ou superiores a do modelo PD saturado.

Modelo de Regressão Poisson Bivariado com Ponderação

A ponderação de dados é uma técnica que pode ser bastante útil para muitas áreas, principalmente no que se refere à tomada de decisão. Essa afirmação é válida pelo fato de que, nos mais diversos segmentos de pesquisa, normalmente existe o interesse em se utilizar a maior quantidade de informações possível acerca do fenômeno em estudo, sendo que, em algumas situações, tais informações são históricas. Nesses casos, nem sempre a utilização de uma grande quantidade de observações fornece resultados acurados (a não ser que estas sejam dependentes e formem uma série temporal), pois se o fenômeno estudado for dinâmico, as observações mais antigas tendem a não mais refletir corretamente a natureza do mesmo. Excluindo os casos em que as observações são tomadas simultaneamente, ou em um espaço suficientemente curto de tempo, voltamos nossas atenções para os conjuntos de dados composto por informações históricas na qual as observações são independentes entre si. O método da verossimilhança ponderada surge como alternativa para os casos em que se deseja ponderar essas observações.

Proposto por [Hu e Zidek \(2002\)](#), o método da verossimilhança ponderada tem suas raízes ligadas ao paradigma de Wald (teoria da decisão). A literatura relacionada envolve importantes contribuições tais como a proposta de estimação robusta via minimização de distâncias (estimação por menor disparidade), baseada em equações de verossimilhanças ponderadas ([Lindsay, 1994](#); [Basu e Lindsay, 1994](#); [Markatou et al., 1997](#)). [Wang et al. \(2004\)](#) estudam as propriedades assintóticas dos estimadores de máxima verossimilhança ponderados, considerando pesos fixos e adaptativos. [Ibrahim e Chen \(2000\)](#) propõem o método da Power Prior, que consiste na utilização de dados históricos para a construção de distribuições *a priori*, e consideram aplicações deste método para uma ampla classe de modelos de regressão. [Wang \(2006\)](#) propõe uma abordagem Bayesiana utilizando uma função de verossimilhança ponderada na ausência de prioris informativas.

Muitas são as situações na qual o método da verossimilhança ponderada encontra aplicações. Por exemplo, se estivermos interessados na modelagem do número médio de gols marcados por cada uma das equipes que disputaram um determinado torneio de pontos corridos, podemos considerar que as informações provenientes das rodadas iniciais não necessariamente refletem a capacidade ofensiva das equipes ao término da competição e, nesse caso, as observações contidas em tais rodadas devem receber menos peso na função de verossimilhança. [Hu e Zidek \(2002\)](#) discutem diversos exemplos

de aplicação do método da verossimilhança ponderada e, em cada caso, determinam os estimadores apropriados. Além disso, mostram que o estimador de James-Stein representa uma alternativa interessante para se determinar os pesos em algumas situações. De modo geral, quando não for possível determinar expressões analíticas para os estimadores de interesse, métodos numéricos de otimização podem ser utilizados.

Como podemos observar, a construção de uma função de verossimilhança a partir da qual seja possível ponderar os dados de acordo com sua relevância (cronológica, por exemplo), representa uma alternativa bastante interessante em diversas situações práticas. Tomando isto como motivação, neste capítulo vamos abordar a construção de funções de verossimilhança ponderadas para os modelos de regressão apresentados no Capítulo 3 utilizando, para tal, estruturas adequadas de ponderação. A obtenção de estimativas ponderadas de máxima verossimilhança para os coeficientes de regressão é considerada por meio do método numérico de otimização BFGS. A qualidade preditiva dos modelos ajustados poderá novamente ser acessada por meio da medida de DeFinetti e da taxa de sucesso. Visando comparar os modelos (com e sem ponderação) fundamentalmente do ponto de vista preditivo, reproduziremos a aplicação aos dados da Série A do Campeonato Brasileiro de 2012 considerada anteriormente.

4.1 Método da Verossimilhança Ponderada

O principal intuito em se considerar uma versão ponderada da função de verossimilhança reside no interesse em se obter soluções para as situações em que se deseja reduzir (ou aumentar) a relevância das observações presentes em um determinado conjunto de dados. De forma geral, essa relevância pode ser quantificada em função do tempo, atribuindo-se maiores pesos para as observações mais recentes, ou até mesmo de acordo com o interesse do pesquisador em ponderar as observações segundo critérios arbitrários. Dentre as vantagens que a ponderação de dados pode oferecer, podemos destacar a possível redução do viés das estimativas dos parâmetros, além da melhoria da capacidade preditiva do modelo adotado. A principal desvantagem é que tal procedimento introduz complicações na função de verossimilhança.

Neste capítulo estamos diretamente interessados em obter um método pelo qual seja possível melhorar a capacidade preditiva dos modelos apresentados no Capítulo 3, sob os quais cada particular observação recebe peso máximo na função de verossimilhança. Nesse contexto, parece bastante natural que a relevância dos dados seja quantificada em função do tempo em que estes foram observados. Portanto, para modelos de regressão bivariados, podemos genericamente definir uma função de verossimilhança ponderada da seguinte forma

$$L_t(\boldsymbol{\beta}|\mathbf{x}, \mathbf{y}) = \prod_{i=1}^n [f(x_i, y_i; \boldsymbol{\beta})]^{\psi(w_i)}, \quad (4.1)$$

em que $\boldsymbol{\beta}$ denota o vetor geral de coeficientes de regressão e ψ é uma função decrescente no tempo, aqui denominada função peso. Estabelecida uma unidade de tempo coerente com os dados de interesse, o termo $w > 0$ deve ser capaz de refletir, para cada i , a distância entre a i -ésima observação e tempo atual t , previamente fixado. Naturalmente, podemos definir $w = t - t_i$, tal que t_i denota o tempo em que a i -ésima observação ocorreu. Assim, podemos assumir que os parâmetros são constantes em cada unidade de tempo, sendo então possível obter estimativas de máxima verossimilhança a partir das observações tomadas até o ponto t .

4.1.1 Escolha da Função Peso

Existem diversas escolhas possíveis para a função peso (FP) ψ . De forma geral, essa função deve ser escolhida de tal sorte que sejam atribuídos maiores pesos para as observações mais recentes. Uma possível escolha é a função indicadora, ou seja,

$$\psi(w) = \begin{cases} 1, & \text{se } w \leq t^*, \\ 0, & \text{se } w > t^*, \end{cases}$$

tal que para t^* fixo ($t^* < t$), as inferências via verossimilhança baseiam-se totalmente nas informações que foram observadas dentro das últimas t^* unidades de tempo. Entretanto, tal função pode não ser a escolha adequada em muitas situações, tendo em vista que a atribuição de peso zero acarreta no descarte de informação. Em virtude disso, neste trabalho vamos considerar a utilização de funções na qual a ‘quantidade’ de peso atribuída a cada observação é controlada por meio de um parâmetro $\xi > 0$. Em particular, estamos interessados em modelar os pesos da seguinte forma

$$\psi_1(w) = e^{-\xi w} \quad \text{e} \quad \psi_2(w) = (w + 1)^{-\xi}.$$

Observe que $\psi_1(0) = \psi_2(0) = 1$. Além disso, note que a partir de ψ_1 o peso das observações decai exponencialmente de acordo com a magnitude de ξ , ao passo que em ψ_2 esse decaimento se mostra mais suave. Cabe ressaltar que o modelo sem ponderação surge como caso particular se $\xi = 0$. A Figura 4.1 ilustra o comportamento de ambas as funções para diferentes valores do parâmetro ξ . Observe que, quanto maior o valor desse parâmetro, menor é o peso que se atribui para as observações. De forma geral, variar a escolha de ψ permite que sejam atribuídos maiores ou menores pesos para as observações mais antigas.

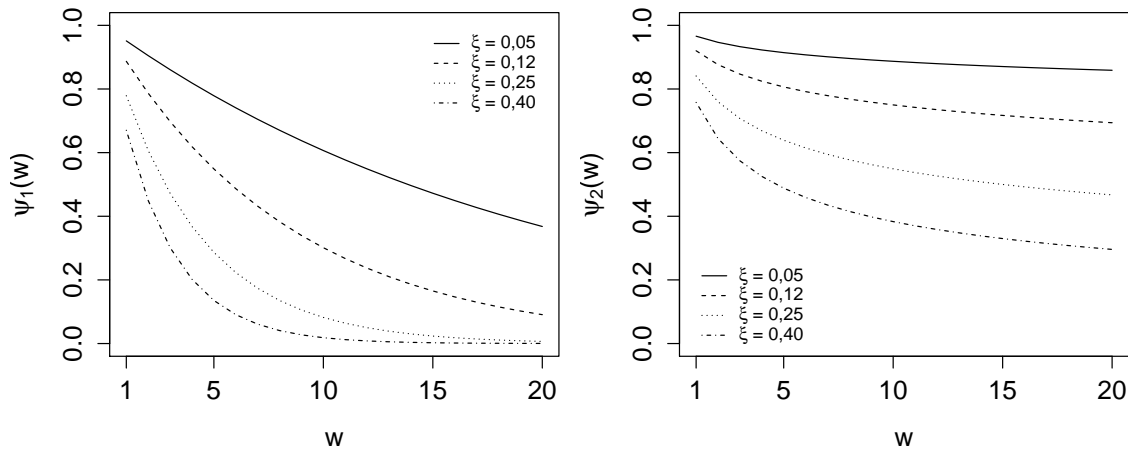


Figura 4.1: Comportamento das funções propostas para modelagem dos pesos.

Otimizar a escolha de ξ é uma tarefa bastante complicada, um vez que as parcelas da função de verossimilhança (4.1) não são independentes (todas dependem de ξ). Porém, tal problemática não representa necessariamente uma perturbação, uma vez que, em nosso contexto, o interesse reside na obtenção do valor de ξ que maximiza a capacidade preditiva do modelo adotado. Em contrapartida, determinar esse parâmetro do ponto de vista preditivo pode ser algo bastante pragmático, tendo em vista que, a parte dos dados e da função de verossimilhança, devemos necessariamente voltar atenções para as categorias nas quais as observações estão inseridas.

Como ferramenta para obtenção de ξ , podemos considerar uma generalização da proposta de [Dixon e Coles \(1997\)](#) para avaliação do perfil preditivo de um determinado modelo. Similar à taxa de sucesso, a função de log-verossimilhança preditiva é definida por

$$S(\hat{\boldsymbol{\pi}}, \xi) = \sum_{i=1}^{n_2} \sum_{j=1}^k \log(\hat{\pi}_{ij}) \mathbb{I}(z_{2i} = j), \quad (4.2)$$

em que k representa o número de categorias e $\hat{\pi}_{ij}$ denota a probabilidade estimada de que a i -ésima observação pertença a j -ésima categoria. A leitura da subseção [3.3.2](#) é suficiente para compreensão da equação (4.2). Observe que para cada possível valor de ξ , o modelo adotado aumenta sua capacidade preditiva a medida que $S(\hat{\boldsymbol{\pi}}, \xi)$ se aproxima de zero. Nesse sentido, podemos estabelecer um *grid* de valores para o parâmetro ξ e, por meio da FP escolhida, aplicá-los na função de verossimilhança (4.1) durante o processo de estimação dos coeficientes de regressão. Assim, uma vez estimados os coeficientes e, conseqüentemente, as probabilidades de interesse, o valor ótimo de ξ será aquele que maximizar a função de log-verossimilhança preditiva, como se deseja.

4.2 Aplicação à Dados Reais

Esta seção é dedicada a uma aplicação. Para efeitos de comparação com os resultados obtidos no Capítulo 3, vamos considerar o ajuste dos modelos de regressão Poisson bivariado e Poisson duplo aos dados da Série A do Campeonato Brasileiro de 2012, utilizando o método da verossimilhança ponderada. Novamente, a estrutura dos modelos que serão ajustados corresponde àquela descrita na seção [3.2](#), sendo que o parâmetro λ_3 do modelo Poisson bivariado é assumido constante entre as observações. Visando otimizar a qualidade preditiva dos modelos, as funções ψ_1 e ψ_2 , especificadas anteriormente, serão utilizadas para modelar os pesos individualmente, sendo que os resultados serão comparados do ponto de vista preditivo. Os procedimentos de estimação dos coeficientes, de seleção do parâmetro ξ e de previsão dos resultados foram realizados com auxílio do software R ([R Development Core Team, 2009](#)).

4.2.1 Metodologia

A metodologia utilizada para obtenção dos ajustes neste capítulo é similar àquela proposta no capítulo anterior no sentido de que também serão utilizadas, no mínimo, as 19 primeiras rodadas da competição. Uma vez que o nosso interesse reside em aplicar o método da verossimilhança ponderada, modelando os pesos de acordo com uma estrutura paramétrica, devemos, como discutido na subseção [4.1.1](#), ser capazes de estimar o parâmetro ξ de forma conveniente. Em virtude disso, vamos considerar a estrutura ajuste-previsão, introduzindo, porém, algumas modificações que irão permitir que o parâmetro ξ seja corretamente incorporado na função de verossimilhança.

Essa nova estrutura define um algoritmo na qual são obtidos ajustes do modelo de interesse aos dados das r primeiras rodadas, para cada particular valor do *grid* de valores de ξ . Contudo, para que seja possível estimar esse parâmetro, o algoritmo exige que seja utilizada pelo menos uma rodada de teste, que será prevista a partir do ajuste do modelo estimado para um dado ξ . Assim, obtidas previsões para a rodada teste, podemos calcular a função de log-verossimilhança preditiva para cada ξ , sendo que sua estimativa será o valor do *grid* que maximizar esta função. Neste trabalho vamos estabelecer, arbitrariamente, a utilização de quatro rodadas teste para a estimação

do parâmetro ξ . Dessa forma estamos definindo um algoritmo na qual o modelo de interesse utiliza os dados das r primeiras rodadas para ajuste e, conseqüentemente, os resultados das partidas da rodada $r + 5$ ($r < 34$) são previstos.

Tabela 4.1: Rodadas utilizadas para ajuste, teste e previsão em cada etapa.

RC	RT	RP	RC	RT	RP	RC	RT	RP
1 ^a – 19 ^a	20 ^a – 23 ^a	24 ^a	1 ^a – 24 ^a	25 ^a – 28 ^a	29 ^a	1 ^a – 29 ^a	30 ^a – 33 ^a	34 ^a
1 ^a – 20 ^a	21 ^a – 24 ^a	25 ^a	1 ^a – 25 ^a	26 ^a – 29 ^a	30 ^a	1 ^a – 30 ^a	31 ^a – 34 ^a	35 ^a
1 ^a – 21 ^a	22 ^a – 25 ^a	26 ^a	1 ^a – 26 ^a	27 ^a – 30 ^a	31 ^a	1 ^a – 31 ^a	32 ^a – 35 ^a	36 ^a
1 ^a – 22 ^a	23 ^a – 26 ^a	27 ^a	1 ^a – 27 ^a	28 ^a – 31 ^a	32 ^a	1 ^a – 32 ^a	33 ^a – 36 ^a	37 ^a
1 ^a – 23 ^a	24 ^a – 27 ^a	28 ^a	1 ^a – 28 ^a	29 ^a – 32 ^a	33 ^a	1 ^a – 33 ^a	34 ^a – 37 ^a	38 ^a

A Tabela 4.1 ilustra a estrutura que será utilizada neste capítulo para ajuste dos modelos. Note que RC refere-se as rodadas que serão consideradas para ajuste em cada etapa e RT indica quais rodadas serão utilizadas para determinar o ξ ótimo. Por fim, RP indica a rodada que será prevista com base nas estimativas de máxima verossimilhança ponderadas. Note que, ao incorporar o parâmetro ξ da forma que estamos propondo, não somos capazes de prever os resultados da 20^a a 23^a rodada, o que se configura uma desvantagem em relação à metodologia aplicada no Capítulo 3, no qual nenhuma estrutura de ponderação de dados foi considerada.

4.2.2 Resultados Obtidos

Os modelos de regressão PD e PB especificados pela estrutura de ponderação definida por me-

Tabela 4.2: Critérios de seleção para comparação dos modelos ponderados.

Função Peso 1									
RC	Modelo PD		Modelo PB		RC	Modelo PD		Modelo PB	
	AICc	BIC	AICc	BIC		AICc	BIC	AICc	BIC
1 ^a – 19 ^a	1099,60	1207,46	959,54	1069,40	1 ^a – 27 ^a	1521,26	1650,87	956,41	1088,84
1 ^a – 20 ^a	581,93	693,23	872,91	986,34	1 ^a – 28 ^a	1569,86	1701,53	976,88	1111,43
1 ^a – 21 ^a	658,85	773,33	836,07	952,80	1 ^a – 29 ^a	1630,04	1763,66	679,31	815,89
1 ^a – 22 ^a	625,54	742,96	844,92	964,71	1 ^a – 30 ^a	727,88	863,37	935,96	1074,46
1 ^a – 23 ^a	1149,50	1296,67	728,90	851,54	1 ^a – 31 ^a	1730,94	1868,21	886,44	1026,79
1 ^a – 24 ^a	1157,73	1280,47	933,98	1059,29	1 ^a – 32 ^a	819,04	958,01	873,65	1015,76
1 ^a – 25 ^a	1420,85	1546,01	912,99	1040,81	1 ^a – 33 ^a	1686,38	1827,00	694,55	838,35
1 ^a – 26 ^a	1398,66	1526,11	941,21	1071,40	-	-	-	-	-

Função Peso 2									
RC	Modelo PD		Modelo PB		RC	Modelo PD		Modelo PB	
	AICc	BIC	AICc	BIC		AICc	BIC	AICc	BIC
1 ^a – 19 ^a	1104,84	1212,70	979,53	1089,39	1 ^a – 27 ^a	1537,76	1667,38	1523,15	1655,58
1 ^a – 20 ^a	1167,55	1278,86	1010,94	1124,38	1 ^a – 28 ^a	1588,34	1720,00	1566,62	1701,17
1 ^a – 21 ^a	1014,33	1128,81	1012,53	1129,26	1 ^a – 29 ^a	1649,26	1782,88	1638,17	1774,75
1 ^a – 22 ^a	1052,06	1169,48	1049,58	1169,37	1 ^a – 30 ^a	1605,45	1740,94	1512,66	1651,17
1 ^a – 23 ^a	1333,14	1453,31	1232,93	1355,58	1 ^a – 31 ^a	1758,27	1895,54	1749,09	1889,44
1 ^a – 24 ^a	1378,03	1500,78	1365,91	1491,22	1 ^a – 32 ^a	1451,52	1590,49	1765,66	1907,77
1 ^a – 25 ^a	1434,86	1560,03	1399,70	1527,52	1 ^a – 33 ^a	1769,90	1910,51	1835,59	1979,39
1 ^a – 26 ^a	1485,52	1612,97	1477,85	1608,04	-	-	-	-	-

io da função de verossimilhança (4.1) foram ajustados a cada uma das sub-amostras do número de gols marcados (ver Tabela 4.1), considerando primeiramente a FP1 e, em seguida, a FP2. Na Tabela 4.2 encontram-se os valores de AICc e BIC obtidos para cada particular ajuste dos modelos. Cabe ressaltar que as medidas apresentadas são calculadas com base nos valores das funções de log-verossimilhança ponderadas, já levando em consideração o valor de ξ , que é estimado com auxílio da função de log-verossimilhança preditiva, cujos valores são obtidos a partir das previsões das rodadas fixadas para teste em cada uma das etapas.

Note que em 2/3 dos ajustes obtidos por meio da FP1, o modelo PB melhor se adequa aos dados provenientes de cada uma das sub-amostras consideradas. Além disso, quando a FP2 é utilizada, a superioridade desse modelo se mostra ainda mais evidente. De forma geral, independentemente da função peso, podemos apontar o modelo PB como sendo a escolha mais apropriada em termos de ajuste. De forma complementar, podemos utilizar o TRV assintótico para comparar os modelos

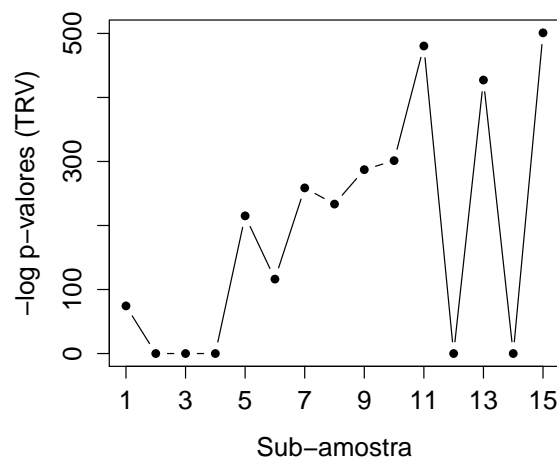


Figura 4.2: Ilustração dos p-valores associados aos TRV's (ajustes via FP1).

PD e PB, considerando as diferentes FP's utilizadas. Em virtude da proximidade de 0 ou 1, os p-valores são apresentados em escala logarítmica para que a interpretação dos mesmos seja facilitada. As Figuras 4.2 e 4.3 apresentam o negativo do logaritmo dos p-valores obtidos para cada um dos testes que foram realizados.

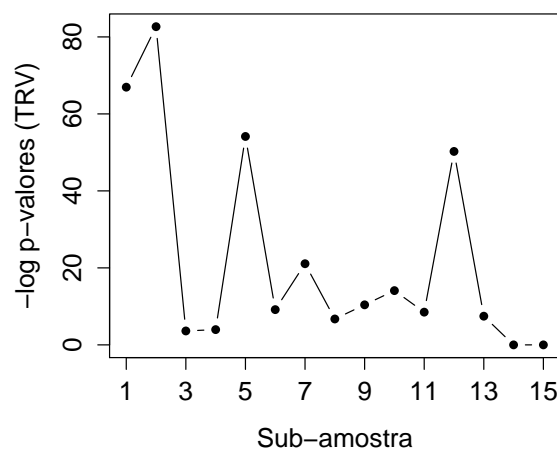


Figura 4.3: Ilustração dos p-valores associados aos TRV's (ajustes via FP2).

Por meio das ilustrações é possível concluir que, quanto mais distante de zero estiver o ponto, maior é a evidência a favor do modelo PB. Em qualquer caso, todos os p-valores favoráveis ao modelo

PB se mostraram bastante significativos a 1%. Nessas situações, verifica-se que a dependência existente entre as variáveis X e Y pôde ser modelada por meio da inserção de um parâmetro de covariância ao modelo PD. Além disso, novamente temos que o TRV assintótico concorda com os critérios AICc e BIC na seleção do modelo que apresenta melhor ajuste.

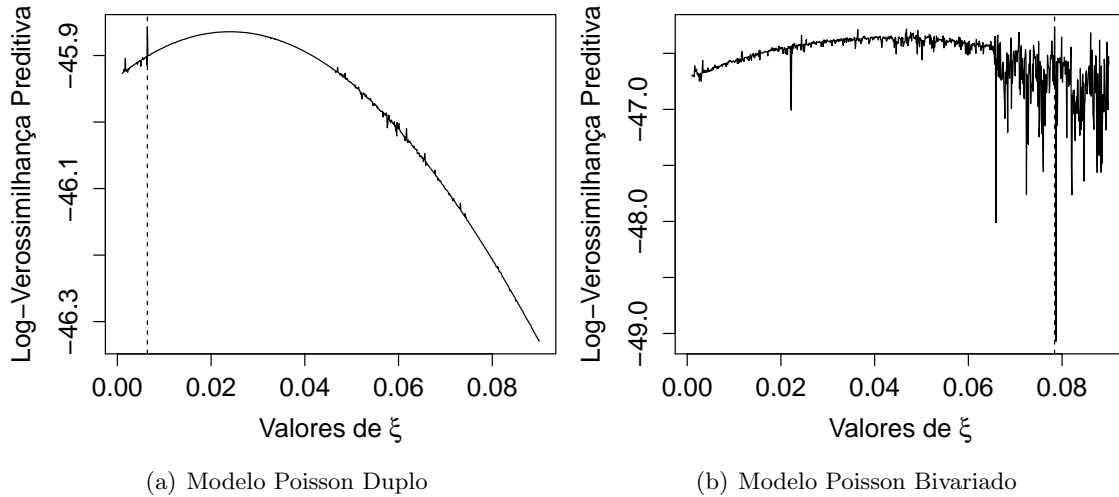


Figura 4.4: Funções de log-verossimilhança preditiva (ajustes via FP1).

O conjunto de efeitos estimados por cada um dos modelos até o ponto t é diretamente afetado pela estimativa atual do parâmetro ξ . Como citado anteriormente, ao se estabelecer um *grid* de valores para ξ , a função de log-verossimilhança preditiva pode ser utilizada como ferramenta para obtenção de estimativas desse parâmetro. Ou seja, para cada sub-amostra considerada, os dados são ajustados para um dado valor de ξ a partir do qual a função de log-verossimilhança preditiva é maximizada. Gráficos de $S(\hat{\pi}, \xi)$ pelos valores de ξ são apresentados nas Figuras 4.4 e 4.5, naturalmente segmentados pela FP utilizada. A linha tracejada na vertical em cada uma das ilustrações refere-se ao valor de ξ que maximiza a função de log-verossimilhança preditiva. Para abreviar a apresentação, ilustramos apenas os resultados obtidos a partir dos ajustes aos dados da última sub-amostra considerada (1^a a 33^a rodada).

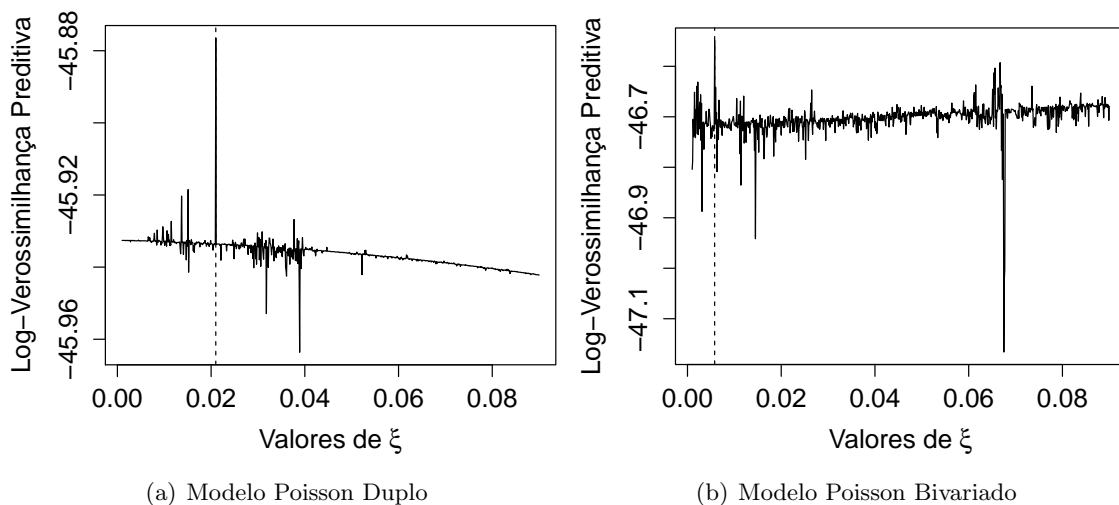


Figura 4.5: Funções de log-verossimilhança preditiva (ajustes via FP2).

Para avaliar o quão diferentes foram as quatro primeiras equipes e o quanto suas capacidades

variaram a partir da 24^a rodada da competição, apresentamos nas Figuras 4.6 e 4.8 a sequência de estimativas dos efeitos ofensivos e nas Figuras 4.7 e 4.9 a sequência de efeitos defensivos obtidos a partir dos ajustes dos modelos, considerando ambas as FP's que foram utilizadas. Note que, como citado no capítulo anterior, os efeitos não são uniformes, indicando que a performance das equipes é altamente dinâmica. Em particular, a ponderação de dados tende a fornecer uma noção muito mais real do dinamismo ofensivo e defensivo das equipes ao longo do período avaliado da competição.

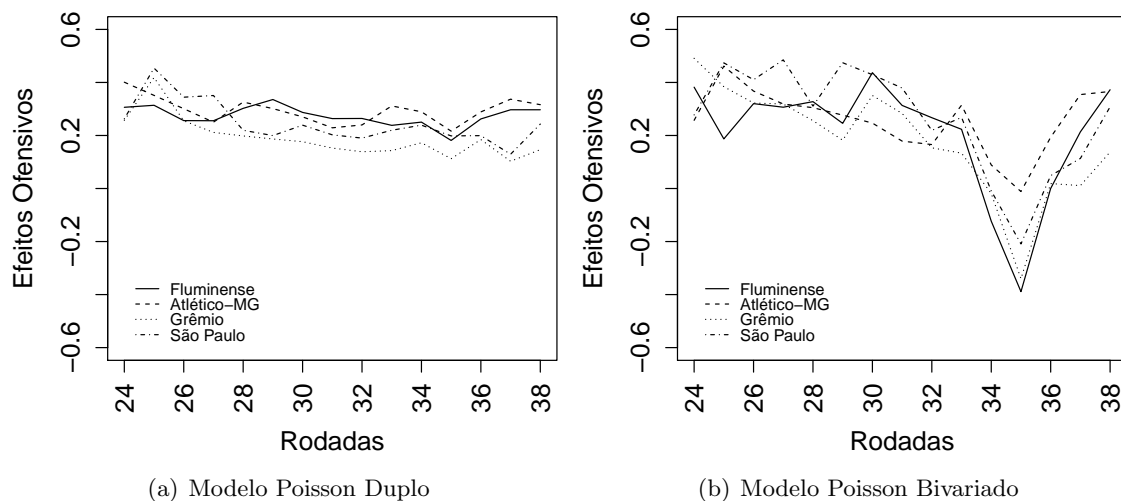


Figura 4.6: Evolução dos efeitos ofensivos ao longo das rodadas (ajuste via FP1).

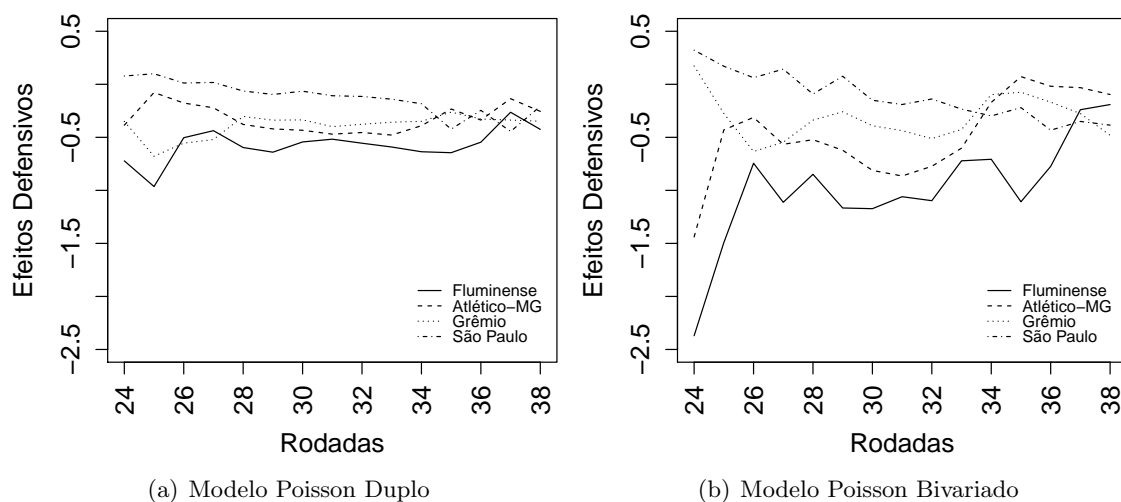


Figura 4.7: Evolução dos efeitos defensivos ao longo das rodadas (ajuste via FP1).

Note que as quatro primeiras equipes classificadas apresentaram performances ofensivas bastante similares a partir da 24^a rodada da competição. Além disso, observe que por meio do ajuste do modelo PB, utilizando a FP1, houve uma queda significativa de rendimento ofensivo das equipes do Fluminense, do Grêmio e do São Paulo entre a 33^a e a 36^a rodadas. Em termos defensivos, a superioridade do Fluminense em relação as demais equipes ficou evidente por meio de ambos os modelos, restando ao Atlético Mineiro e ao Grêmio uma equilibrada disputa pela segunda colocação. Ao término da competição um melhor desempenho ofensivo propiciou ao Atlético Mineiro terminar à frente do Grêmio que, por sua vez, acabou por terminar à frente do São Paulo, muito em função do melhor desempenho defensivo ao longo do período avaliado. De forma geral, a ponderação de

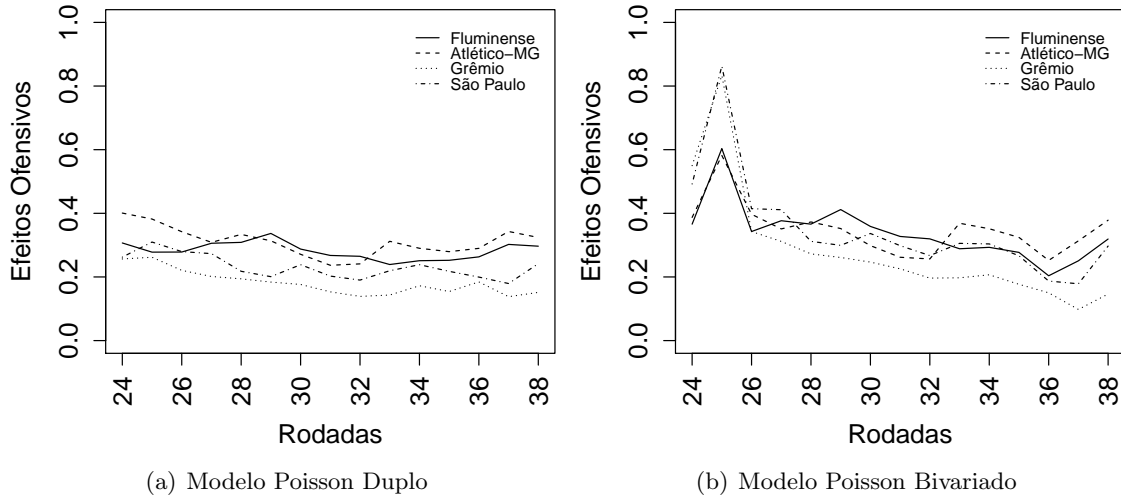


Figura 4.8: Evolução dos efeitos ofensivos ao longo das rodadas (ajuste via FP2).

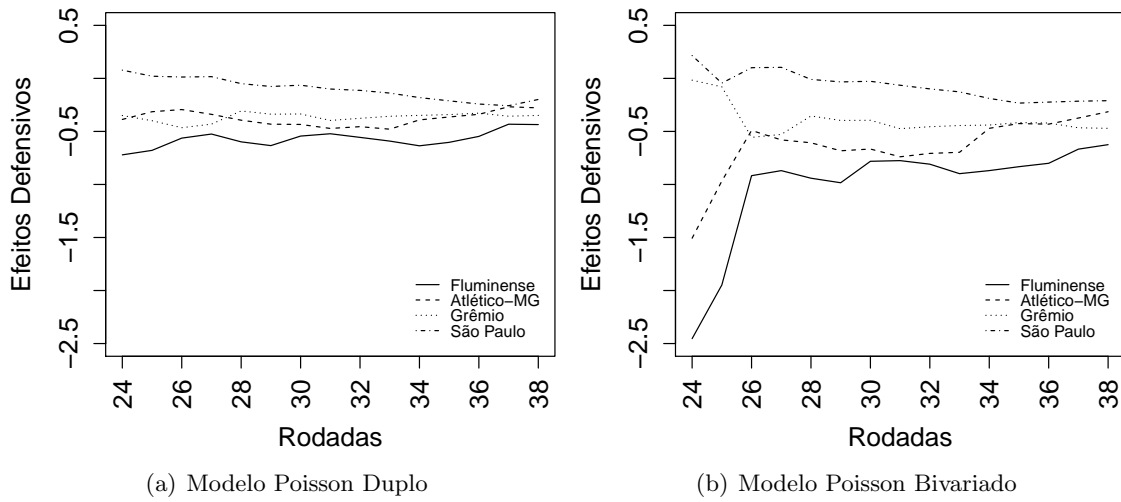


Figura 4.9: Evolução dos efeitos defensivos ao longo das rodadas (ajuste via FP2).

dados acabou por ressaltar a proximidade entre as equipes posicionadas na parte superior da tabela de classificação, principalmente em termos defensivos. De forma complementar, nas Figuras 4.10 e

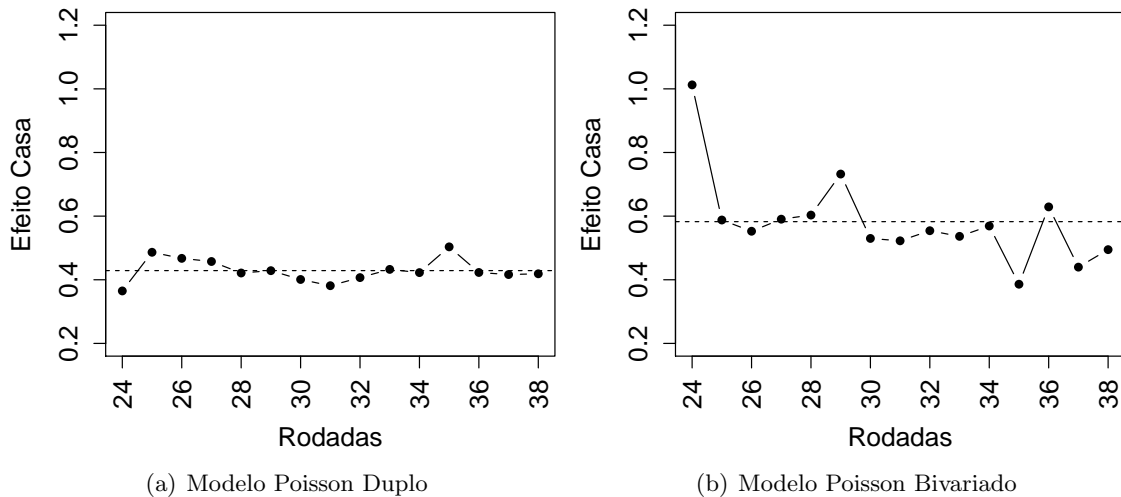


Figura 4.10: Evolução do efeito casa ao longo das rodadas (ajuste via FP1).

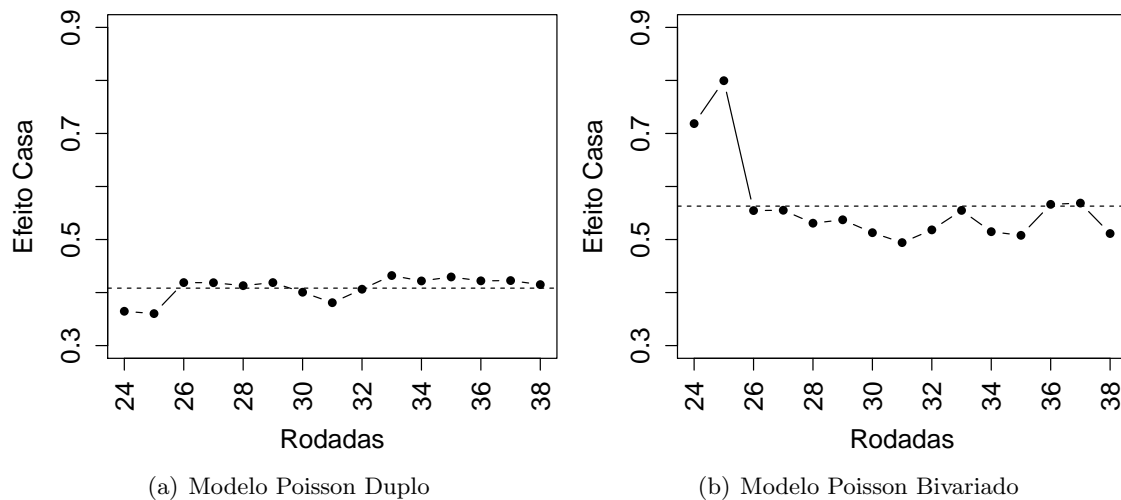


Figura 4.11: Evolução do efeito casa ao longo das rodadas (ajuste via FP2).

4.11 são apresentadas as sequências de estimativas para o efeito casa, de acordo com as diferentes FP's utilizadas para ajuste dos modelos. Observe que segundo o modelo PB, ajustado por meio da FP1, o efeito casa estimado para a 24^a rodada supera o valor 1, o que vem a ser um indicativo da força das equipes mandantes durante aquele período da competição, captada apenas por este modelo. Para além desse fato, o efeito casa se mantém aproximadamente constante ao longo do tempo, principalmente quando o modelo PD é considerado.

4.2.3 Previsão de Resultados

Uma vez obtidos ajustes sequenciais dos modelos PD e PB, podemos novamente realizar procedimentos de previsão, agora no contexto de ponderação. Como citado no capítulo anterior, esses procedimentos baseiam-se no cálculo das probabilidades de vitória das equipes mandantes (PVM) e das probabilidades de vitória das equipes visitantes (PVV), sendo que as probabilidades de empate (PE) são dadas pelos complementares das somas de PVM e PVV. Tais probabilidades são calculadas com base nas distribuições dos modelos adotados, com parâmetros estimados via regressão. Segundo a estrutura ajuste-previsão proposta neste capítulo, foram obtidas previsões para os jogos da 24^a a 38^a rodadas da Série A do Campeonato Brasileiro de 2012. Para abreviar a apresentação dos resultados, apenas as previsões para as rodadas 28 e 38 são apresentadas. Além das previsões, nas Tabelas 4.3, 4.4, 4.5 e 4.6 encontram-se os valores da distância de DeFinetti obtidos para cada particular partida prevista no contexto de ponderação e sem ponderação (SP), sendo este último apresentado para efeitos meramente comparativos. É de se esperar que, por meio do procedimento de ponderação, ocorra uma melhoria geral na capacidade preditiva dos modelos PD e PB, muito embora os ganhos mais significativos devam ser observados apenas nas previsões específicas para cada partida.

Observe que, assim como nos procedimentos de previsão sem ponderação, o modelo PB melhor prevê os empates em relação ao modelo PD, sendo que, em alguns casos, as previsões fornecidas por este modelo coincidem com os resultados de empate que de fato ocorreram. De forma geral, nota-se que a metodologia de ponderação de dados introduziu uma melhoria na capacidade preditiva dos modelos considerados, sendo que, quando a FP2 é utilizada, essa melhora é ainda mais evidente. As Figuras 4.12 e 4.13 ilustram as médias aritméticas por rodada prevista das distâncias de DeFinetti.

Tabela 4.3: Previsões para a 28ª rodada (ajuste via FP1).

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti	DeFinetti (SP)
Modelo Poisson Duplo							
Atlético-GO	0,3128	0,2726	0,4146	Vasco da Gama	0-1	0,5148	0,4323
Atlético-MG	0,8389	0,1110	0,0501	Figueirense	6-0	0,0408	0,0475
Coritiba	0,5330	0,2088	0,2581	Ponte Preta	1-0	0,3283	0,3485
Flamengo	0,3793	0,3147	0,3060	Bahia	0-0	0,7072	0,7397
Fluminense	0,7007	0,1923	0,1071	Botafogo	1-0	0,1380	0,1446
Grêmio	0,6933	0,1890	0,1177	Cruzeiro	2-1	0,1437	0,1690
Náutico	0,3589	0,2799	0,3612	Corinthians	2-1	0,6198	0,6335
Portuguesa	0,5825	0,2782	0,1393	Sport Recife	5-1	0,2711	0,2643
Santos	0,3316	0,3145	0,3539	Internacional	1-1	0,7050	0,7216
São Paulo	0,6704	0,2076	0,1221	Palmeiras	3-0	0,1667	0,1802
Modelo Poisson Bivariado							
Atlético-GO	0,3172	0,3155	0,3673	Vasco da Gama	0-1	0,6004	0,4582
Atlético-MG	0,8333	0,1291	0,0376	Figueirense	6-0	0,0459	0,0422
Coritiba	0,5978	0,2134	0,1889	Ponte Preta	1-0	0,2430	0,3174
Flamengo	0,2644	0,3847	0,3509	Bahia	0-0	0,5716	0,6535
Fluminense	0,6981	0,2325	0,0694	Botafogo	1-0	0,1500	0,1348
Grêmio	0,7477	0,1743	0,0780	Cruzeiro	2-1	0,1001	0,1618
Náutico	0,3073	0,3368	0,3558	Corinthians	2-1	0,7198	0,6437
Portuguesa	0,6273	0,2964	0,0764	Sport Recife	5-1	0,2326	0,2737
Santos	0,2793	0,4173	0,3034	Internacional	1-1	0,5097	0,5250
São Paulo	0,7242	0,1980	0,0778	Palmeiras	3-0	0,1214	0,1505

Tabela 4.4: Previsões para a 28ª rodada (ajuste via FP2).

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti	DeFinetti (SP)
Modelo Poisson Duplo							
Atlético-GO	0,2936	0,2700	0,4364	Vasco da Gama	0-1	0,4767	0,4323
Atlético-MG	0,8398	0,1110	0,0492	Figueirense	6-0	0,0404	0,0475
Coritiba	0,5248	0,2106	0,2646	Ponte Preta	1-0	0,3402	0,3485
Flamengo	0,4044	0,3076	0,2880	Bahia	0-0	0,7259	0,7397
Fluminense	0,7045	0,1887	0,1067	Botafogo	1-0	0,1343	0,1446
Grêmio	0,6825	0,1956	0,1220	Cruzeiro	2-1	0,1539	0,1690
Náutico	0,3661	0,2799	0,3539	Corinthians	2-1	0,6054	0,6335
Portuguesa	0,5663	0,2854	0,1484	Sport Recife	5-1	0,2916	0,2643
Santos	0,3248	0,3191	0,3561	Internacional	1-1	0,6960	0,7216
São Paulo	0,6612	0,2114	0,1274	Palmeiras	3-0	0,1757	0,1802
Modelo Poisson Bivariado							
Atlético-GO	0,2567	0,3220	0,4213	Vasco da Gama	0-1	0,5045	0,4582
Atlético-MG	0,8550	0,1151	0,0300	Figueirense	6-0	0,0352	0,0422
Coritiba	0,5549	0,2194	0,2256	Ponte Preta	1-0	0,2971	0,3174
Flamengo	0,3747	0,3604	0,2649	Bahia	0-0	0,6196	0,6535
Fluminense	0,7289	0,2057	0,0655	Botafogo	1-0	0,1201	0,1348
Grêmio	0,7092	0,2005	0,0903	Cruzeiro	2-1	0,1329	0,1618
Náutico	0,3542	0,3134	0,3324	Corinthians	2-1	0,6257	0,6437
Portuguesa	0,5601	0,3333	0,1066	Sport Recife	5-1	0,3159	0,2737
Santos	0,2623	0,4216	0,3161	Internacional	1-1	0,5032	0,5250
São Paulo	0,6907	0,2127	0,0966	Palmeiras	3-0	0,1503	0,1505

Tabela 4.5: Previsões para a 38ª rodada (ajuste via FP1).

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti	DeFinetti (SP)
Modelo Poisson Duplo							
Atlético-GO	0,3285	0,2950	0,3765	Bahia	0-1	0,5837	0,6101
Atlético-MG	0,7047	0,1867	0,1086	Cruzeiro	3-2	0,1339	0,1557
Coritiba	0,6917	0,1661	0,1422	Figueirense	3-0	0,1429	0,1619
Flamengo	0,3694	0,2732	0,3574	Botafogo	2-2	0,7923	0,7971
Fluminense	0,6731	0,2140	0,1129	Vasco da Gama	1-2	1,2858	1,2456
Grêmio	0,5236	0,2813	0,1952	Internacional	0-0	0,8288	0,9287
Náutico	0,5987	0,2240	0,1773	Sport Recife	1-0	0,2427	0,3088
Portuguesa	0,5095	0,2882	0,2023	Ponte Preta	0-0	0,8071	0,7567
Santos	0,5588	0,2558	0,1854	Palmeiras	3-1	0,2944	0,2527
São Paulo	0,5690	0,2509	0,1802	Corinthians	3-1	0,2812	0,3695
Modelo Poisson Bivariado							
Atlético-GO	0,2733	0,3231	0,4035	Bahia	0-1	0,5349	0,6115
Atlético-MG	0,6943	0,1982	0,1075	Cruzeiro	3-2	0,1443	0,1549
Coritiba	0,7651	0,1501	0,0848	Figueirense	3-0	0,0849	0,1330
Flamengo	0,3346	0,3416	0,3238	Botafogo	2-2	0,6503	0,7538
Fluminense	0,7250	0,1883	0,0868	Vasco da Gama	1-2	1,3950	1,2746
Grêmio	0,5775	0,2795	0,1431	Internacional	0-0	0,8731	0,8769
Náutico	0,5894	0,2482	0,1624	Sport Recife	1-0	0,2566	0,3086
Portuguesa	0,5294	0,3185	0,1521	Ponte Preta	0-0	0,7678	0,7126
Santos	0,5810	0,2551	0,1640	Palmeiras	3-1	0,2675	0,2510
São Paulo	0,6105	0,2513	0,1383	Corinthians	3-1	0,2340	0,3768

Tabela 4.6: Previsões para a 38ª rodada (ajuste via FP2).

Mandante	PVM	PE	PVV	Visitante	Placar	DeFinetti	DeFinetti (SP)
Modelo Poisson Duplo							
Atlético-GO	0,3400	0,2964	0,3636	Bahia	0-1	0,6085	0,6101
Atlético-MG	0,7106	0,1842	0,1051	Cruzeiro	3-2	0,1287	0,1557
Coritiba	0,6872	0,1671	0,1457	Figueirense	3-0	0,1470	0,1619
Flamengo	0,3696	0,2698	0,3606	Botafogo	2-2	0,7998	0,7971
Fluminense	0,6693	0,2160	0,1147	Vasco da Gama	1-2	1,2784	1,2456
Grêmio	0,5205	0,2826	0,1969	Internacional	0-0	0,8243	0,9287
Náutico	0,5992	0,2232	0,1776	Sport Recife	1-0	0,2420	0,3088
Portuguesa	0,5056	0,2887	0,2057	Ponte Preta	0-0	0,8039	0,7567
Santos	0,5563	0,2588	0,1849	Palmeiras	3-1	0,2981	0,2527
São Paulo	0,5661	0,2525	0,1814	Corinthians	3-1	0,2849	0,3695
Modelo Poisson Bivariado							
Atlético-GO	0,3051	0,3356	0,3593	Bahia	0-1	0,6163	0,6115
Atlético-MG	0,7214	0,1932	0,0854	Cruzeiro	3-2	0,1223	0,1549
Coritiba	0,7221	0,1618	0,1161	Figueirense	3-0	0,1169	0,1330
Flamengo	0,3587	0,3048	0,3364	Botafogo	2-2	0,7251	0,7538
Fluminense	0,6732	0,2409	0,0859	Vasco da Gama	1-2	1,3468	1,2746
Grêmio	0,4879	0,3460	0,1660	Internacional	0-0	0,6933	0,8769
Náutico	0,6096	0,2422	0,1483	Sport Recife	1-0	0,2331	0,3086
Portuguesa	0,4957	0,3258	0,1785	Ponte Preta	0-0	0,7321	0,7126
Santos	0,5505	0,2858	0,1637	Palmeiras	3-1	0,3105	0,2510
São Paulo	0,5670	0,2762	0,1568	Corinthians	3-1	0,2884	0,3768

Novamente, essas distâncias são calculadas com base nas probabilidades estimadas de vitória da equipe mandante, empate e vitória da equipe visitante e a linha tracejada em $2/3$ refere-se ao ponto de corte estabelecido para validação da qualidade preditiva dos modelos ajustados. Médias superiores a $2/3$ indicam que a referida rodada foi atípica, tendo sido registrada a ocorrência de resultados pouco prováveis.

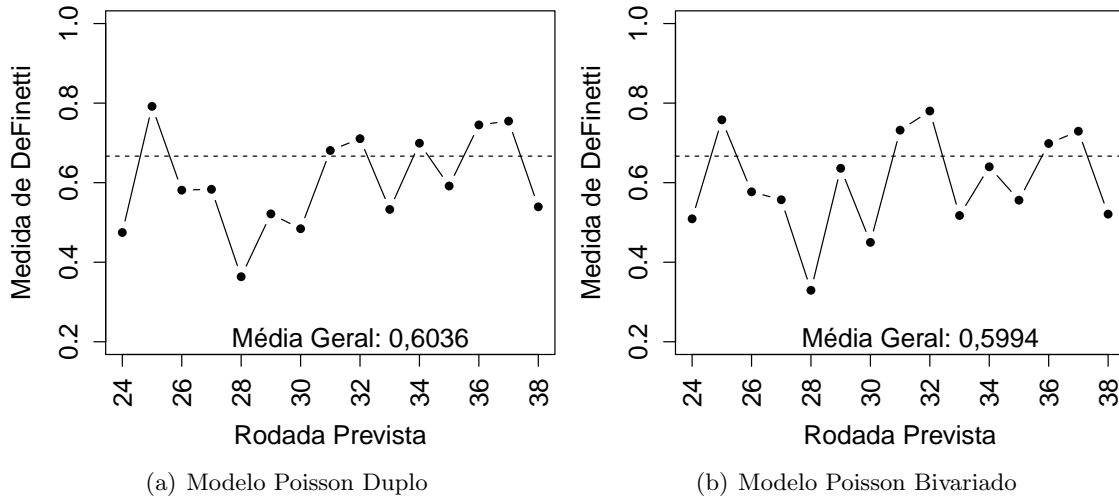


Figura 4.12: Medidas de Definetti para as rodadas previstas (ajuste via FP1).

Na Tabela 4.7 apresentamos o percentual de previsões corretas fornecidas pelos modelos PD e PB a cada rodada, considerando ambas as FP's adotadas. Observe que, por meio da FP1, o modelo PB apresenta melhor desempenho, sendo responsável por prever corretamente pelo menos 70% dos placares em seis rodadas (o que se configura um ganho em relação ao procedimento sem ponderação). Além disso, esses mesmo modelo fornece, em exatamente dez rodadas da competição, probabilidades mais acuradas do que o modelo PD para mais de seis partidas por rodada quando a FP1 é utilizada, sendo esse número reduzido para oito quando da utilização da FP2. De forma geral, as estatísticas apresentadas apontam para o modelo PB como sendo o mais adequado para os procedimentos de previsão.

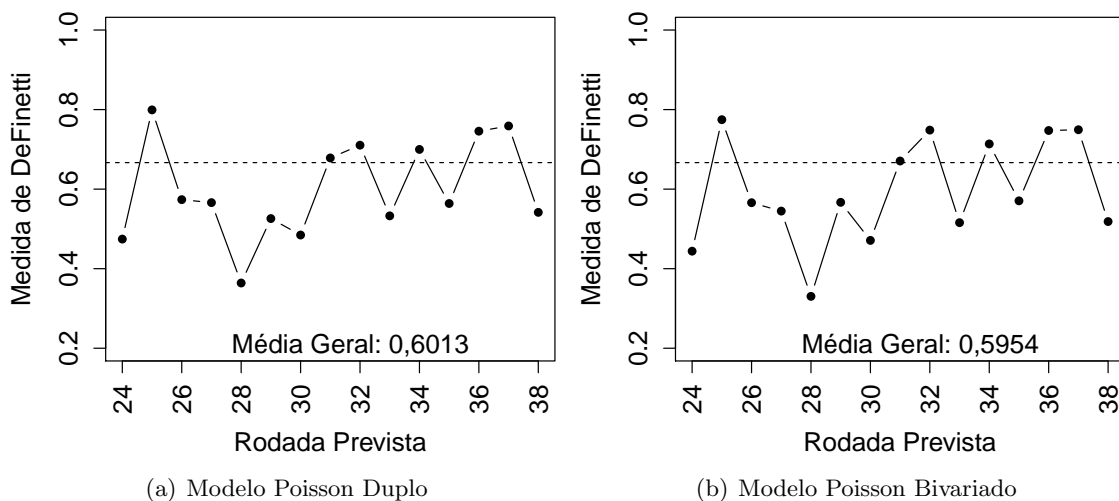


Figura 4.13: Medidas de Definetti para as rodadas previstas (ajuste via FP2).

Note que, a partir das previsões obtidas por meio do modelo PB, utilizando a FP1, a média

geral das distâncias de DeFinetti novamente se mantem inferior em relação àquela obtida por meio

Tabela 4.7: Percentual de previsões corretas dos modelos ponderados a cada rodada.

Função Peso 1											
RP	Modelo		# PB > PD	RP	Modelo		# PB > PD	RP	Modelo		# PB > PD
	PD	PB			PD	PB			PD	PB	
24 ^a	50	60	3	29 ^a	60	40	1	34 ^a	30	50	6
25 ^a	30	20	7	30 ^a	60	70	6	35 ^a	60	50	6
26 ^a	30	30	6	31 ^a	50	30	4	36 ^a	30	50	7
27 ^a	60	70	6	32 ^a	40	70	2	37 ^a	20	30	6
28 ^a	70	90	6	33 ^a	60	70	5	38 ^a	60	70	6

Função Peso 2											
RP	Modelo		# PB < PD	RP	Modelo		# PB < PD	RP	Modelo		# PB < PD
	PD	PB			PD	PB			PD	PB	
24 ^a	60	50	7	29 ^a	60	60	1	34 ^a	30	40	2
25 ^a	30	20	6	30 ^a	60	70	7	35 ^a	50	50	4
26 ^a	40	50	5	31 ^a	50	40	5	36 ^a	30	40	5
27 ^a	60	60	6	32 ^a	40	30	3	37 ^a	30	20	6
28 ^a	80	90	7	33 ^a	60	70	6	38 ^a	60	60	6

do modelo PD (0,6036 contra 0,5994), tal como observado no Capítulo 3. Para os ajustes baseados

Tabela 4.8: Taxa de sucesso dos modelos ponderados ajustados.

Função Peso 1									
RP	Modelo		PD/PD* (%)	PB/PB* (%)	RP	Modelo		PD/PD* (%)	PB/PB* (%)
	PD	PB				PD	PB		
24 ^a	4,65	4,67	100,43	96,29	32 ^a	3,34	3,06	104,70	99,35
25 ^a	3,31	3,30	108,52	105,10	33 ^a	4,61	4,72	100,44	99,79
26 ^a	4,08	4,14	102,00	101,47	34 ^a	3,48	3,70	98,03	106,32
27 ^a	4,18	4,36	97,66	97,32	35 ^a	3,92	4,16	99,75	106,39
28 ^a	5,42	5,70	100,93	101,79	36 ^a	3,51	3,67	100,86	105,16
29 ^a	4,34	3,83	105,60	97,46	37 ^a	3,13	3,18	98,12	98,45
30 ^a	4,77	5,03	98,55	102,03	38 ^a	4,46	4,67	103,48	106,14
31 ^a	3,66	3,59	98,92	95,23	-	-	-	-	-

Função Peso 2									
RP	Modelo		PD/PD* (%)	PB/PB* (%)	RP	Modelo		PD/PD* (%)	PB/PB* (%)
	PD	PB				PD	PB		
24 ^a	4,66	5,02	100,65	103,51	32 ^a	3,34	3,18	104,70	103,25
25 ^a	3,10	3,54	101,64	112,74	33 ^a	4,61	4,76	100,44	100,63
26 ^a	4,13	4,22	103,25	103,43	34 ^a	3,48	3,42	98,03	98,28
27 ^a	4,23	4,39	98,83	97,99	35 ^a	4,08	4,07	103,82	104,09
28 ^a	5,41	5,66	100,74	101,07	36 ^a	3,51	3,48	100,86	99,71
29 ^a	4,31	4,14	104,87	105,34	37 ^a	3,15	3,19	98,75	98,76
30 ^a	4,77	4,89	98,55	99,19	38 ^a	4,44	4,59	103,02	104,32
31 ^a	3,67	3,76	99,19	99,73	-	-	-	-	-

* Modelo sem ponderação.

na FP2, essas distâncias decaem para 0,6013 e 0,5954, respectivamente. Ainda que o ganho médio introduzido pela ponderação de dados aparentemente seja relativamente baixo em relação ao procedimento sem ponderação, individualmente, quando a melhoria ocorre, esta é significativa. No que segue, apresentamos na Tabela 4.8 as taxas de sucesso obtidas para cada rodada prevista. Para efeitos de comparação, as taxas de sucesso obtidas por meio das probabilidades estimadas no contexto de ponderação são comparadas com aquelas obtidas no capítulo anterior, sem ponderação, visando identificar uma eventual perda dos modelos ponderados em termos preditivos.

Em linhas gerais temos que, dentre os modelos que foram ajustados no contexto de ponderação, podemos observar que o PB é aquele que apresenta melhor desempenho do ponto de vista preditivo (levemente superior ao do modelo PD). Além disso, note que praticamente não existe perda dos modelos com ponderação em relação aos modelos sem ponderação em termos de qualidade preditiva, ou seja, de forma bastante geral, a utilização de qualquer uma das FP's na função de verossimilhança (4.1) nos conduz a estimativas e, conseqüentemente, a procedimentos de previsão a partir dos quais as taxas de sucesso dos modelos ponderados são maiores do que a dos modelos sem ponderação.

Considerações Finais

Neste trabalho apresentamos a distribuição Poisson bivariada, bem como suas principais características e propriedades. Além disso, apresentamos o modelo Poisson duplo como caso particular, que surge quando a covariância entre as variáveis de interesse é assumida nula. A construção dessas distribuições pode ser realizada por meio do método da redução trivariada, o que faz com que suas marginais sejam positivamente correlacionadas. Para ilustrar a aplicação desses modelos, foi realizado um estudo de simulação na qual as medidas de viés, erro quadrático e erro quadrático médio foram utilizadas para avaliar o comportamento das estimativas para os diferentes tamanhos amostrais considerados. Por outro lado, utilizamos os critérios AIC, BIC e o teste da razão de verossimilhanças assintótico para comparar os modelos ajustados. Cabe ressaltar que os procedimentos de estimação de parâmetros em todo o trabalho foram realizados por meio de métodos numéricos de otimização.

Uma vez que o objetivo principal deste trabalho esteve na modelagem e previsão de placares esportivos, apresentamos a estrutura do modelo de regressão Poisson bivariado como caso geral, além da estrutura de um modelo de efeitos para aplicação à dados de futebol. Nesse sentido, foi considerada uma aplicação aos dados da Série A do Campeonato Brasileiro de 2012, na qual os efeitos ofensivos e defensivos foram estimados sequencialmente, levando em consideração o desempenho de cada uma das 20 equipes que disputaram a competição. Por meio dos modelos ajustados foram simulados mil torneios e, com isso, foi possível estimar as probabilidades de título, libertadores e despromoção associadas a cada equipe. A partir dos efeitos estimados, foram obtidas previsões para as partidas do segundo turno e, com auxílio da medida de DeFinetti e da taxa de sucesso, verificamos que o modelo Poisson bivariado forneceu melhores previsões nesse contexto.

A parte final deste trabalho foi dedicada a tentativa de se obter previsões mais acuradas, a partir dos modelos Poisson duplo e Poisson bivariado. Nesse sentido, apresentamos e caracterizamos o método da verossimilhança ponderada a partir do qual foi possível quantificar a relevância dos dados em função do tempo em que estes foram observados. Em outras palavras, tal abordagem permitiu que fossem atribuídos maiores pesos para os placares observados em rodadas mais recentes em relação a rodada na qual o ajuste fora realizado. Com auxílio da log-verossimilhança preditiva foi possível estimar os efeitos e, conseqüentemente, obter previsões para as partidas da 24^a rodada

em diante. Esse procedimento se mostrou bastante coerente, uma vez que captou de forma mais eficiente o dinamismo no desempenho das equipes ao longo da competição. De forma geral, também no contexto de estimação ponderada, o modelo Poisson bivariado apresentou-se melhor em termos de capacidade preditiva, quando comparado com o modelo Poisson duplo.

A principal característica dos modelos bivariados aplicados neste trabalho está na suposição de que as distribuições marginais são positivamente correlacionadas. Entretanto, na prática, nem sempre as variáveis de interesse apresentam tal comportamento e essa violação faz com que os modelos Poisson duplo e Poisson bivariado não sejam apropriados para os procedimentos de modelagem. Nesse contexto, surge a necessidade de se obter modelos mais flexíveis que permitam correlação negativa. O problema de se construir modelos bivariados para a modelagem de dados que podem não ser positivamente correlacionados é abordado por diversos autores. Em particular, para o caso em que as marginais são Poisson, [Griffiths et al. \(1979\)](#) apresentam um método na qual utilizam uma função de probabilidades bivariada modificada, dada por

$$f(x, y; \boldsymbol{\theta}) = \frac{e^{-\lambda_1} \lambda_1^x}{x!} \frac{e^{-\lambda_2} \lambda_2^y}{y!} \{1 + \alpha [g_1(x) g_2(y)]\}, \quad x, y = 0, 1, \dots, \quad (5.1)$$

em que $\boldsymbol{\theta} = (\lambda_1, \lambda_2, \alpha)$, com $\lambda_1, \lambda_2 > 0$ e $\alpha \in \mathbb{R}$. Além disso, g_1 e g_2 são funções limitadas em \mathbb{R}^2 , ambas com esperança igual a zero. Nesse caso, os autores propõem que

$$g_1(k) = g_2(k) = \frac{(\gamma^k e^{1-\gamma} - 1)}{1 + e^2},$$

com $-1 \leq \gamma \leq 1$, o que permite correlação negativa sempre que $\alpha > 0$. Uma ideia similar foi utilizada por [Lee \(1996\)](#) e [Lakshminarayana et al. \(1999\)](#), que aplicam em (5.1) as funções

$$g_1(x) = e^{-x} - e^{\lambda_1(1-e^{-1})} \quad \text{e} \quad g_2(y) = e^{-y} - e^{\lambda_2(1-e^{-1})},$$

e, assim, constroem uma distribuição Poisson bivariada com marginais pertencentes à família *Sarmanov* de distribuições. Nesse caso, também sob a condição $\alpha < 0$, será possível obter correlação negativa para as contagens de interesse, muito embora tal valor tenda para zero a medida que λ_1 e λ_2 crescem.

Com base nessa possível modificação para os modelos bivariados abordados neste trabalho, podem ser estabelecidas diversas propostas de trabalhos futuros, muito em virtude das inúmeras possibilidades de aplicações práticas, nas quais se observam variáveis que podem ser negativamente correlacionadas. De forma geral, a obtenção de métodos para a estimação pontual via máxima verossimilhança, bem como a introdução de uma estrutura de regressão e, possivelmente, de ponderação no modelo (5.1) representaria, sem dúvida alguma, uma contribuição bastante significativa para a literatura relacionada.

Identificação e Seleção de Modelos

A.1 Critérios AIC, AICc e BIC

Considere $\mathbf{y} = (y_1, y_2, \dots, y_n)$ um vetor de observações da variável aleatória Y , cujo modelo probabilístico associado é $f(y; \boldsymbol{\theta})$ e $\boldsymbol{\theta} \in \Theta$ é o vetor paramétrico de interesse. Os critérios de seleção de modelos AIC, AICc e BIC são dados por

- **Critério AIC:**

$$\text{AIC} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{y}) + 2p, \quad (\text{A.1})$$

em que p representa o número de parâmetros do modelo ajustado e $L(\hat{\boldsymbol{\theta}}|\mathbf{y})$ representa a função de verossimilhança avaliada nos estimadores de máxima verossimilhança.

- **Critério AIC Corrigido (AICc):**

$$\text{AICc} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{y}) + 2p + 2 \frac{p(p+1)}{n-p-1}, \quad (\text{A.2})$$

em que n representa o tamanho amostral. O critério AICc é amplamente utilizado no contexto de comparação entre modelos de regressão.

- **Critério BIC:**

$$\text{BIC} = -2 \log L(\hat{\boldsymbol{\theta}}|\mathbf{y}) + p \log(n), \quad (\text{A.3})$$

em que $\log(n)$ representa uma correção no tamanho amostral.

Vale ressaltar que, para cada um desses critérios, menores valores indicam melhores modelos. Note que os esses critérios apresentem diferenças conceituais importantes, principalmente na forma como os modelos são avaliados. Entretanto, os três utilizam a estimativa de máxima verossimilhança como ferramenta de ajustamento. De forma geral, ao se utilizar o critério AIC (e AICc), admite-se que dentre os modelos avaliados, nenhum é tido como sendo o verdadeiro e, então, ocorre a tentativa de se escolher o modelo que minimize a divergência Kullback-Leibler. Em contrapartida, ao se utilizar o critério BIC, supõe-se que existe um modelo que descreve completamente a incerteza

sobre os parâmetros de interesse, de forma que o critério busca maximizar a probabilidade de que esse modelo verdadeiro seja escolhido.

A.2 Teste da Razão de Verossimilhanças

O teste da razão de verossimilhanças é um procedimento de teste utilizado para comparar dois modelos que sejam encaixados. Por exemplo, no ajuste de um modelo de regressão linear simples com uma única variável preditora, esse teste pode ser utilizado para testar o modelo nulo (somente o intercepto) contra o modelo com a variável preditora. Nesse sentido, considere $\mathbf{y} = (y_1, y_2, \dots, y_n)$ um vetor de observações da variável aleatória Y , cujo modelo probabilístico associado é $f(y; \boldsymbol{\theta})$ e $\boldsymbol{\theta} \in \Theta$ é o vetor paramétrico de interesse. A estatística do teste da razão de verossimilhanças pode ser escrita da seguinte forma

$$\mathcal{G}^2 = -2 \log \left[\frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta} | \mathbf{y})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta} | \mathbf{y})} \right]. \quad (\text{A.4})$$

Para grandes amostras mostra-se que, sob \mathcal{H}_0 , a estatística \mathcal{G}^2 tem distribuição assintótica Chi-Quadrado com $k - p$ graus de liberdade, sendo k o número de parâmetros do modelo completo (espaço paramétrico irrestrito), e p o número de parâmetros do modelo em teste, com $\Theta_0 \subset \Theta$.

Na prática, o teste da razão de verossimilhanças é geralmente preferido para o teste de hipóteses precisas, uma vez que, se existe um teste φ UMP, o lema de Neyman-Pearson o define. Em outras palavras, dentre todos os procedimentos para o teste de hipóteses precisas, o teste da razão de verossimilhanças é o uniformemente mais poderoso. Por outro lado, para testes de hipóteses compostas, podemos utilizar a extensão do teste denominada razão de verossimilhanças generalizada para um nível α previamente fixado.

Referências Bibliográficas

- Arruda (2000)** M. L. Arruda. Poisson, Bayes, Futebol e DeFinetti. *São Paulo, Dissertação de Mestrado IME/USP*. Citado na pág. [6](#)
- Basu e Lindsay (1994)** A. Basu e B. G. Lindsay. Minimum disparity estimation for continuous models: efficiency, distributions and robustness. *Annals of the Institute of Statistical Mathematics*, 46(4):683–705. Citado na pág. [36](#)
- Byrd et al. (1995)** R. H. Byrd, P. Lu, J. Nocedal e C. Zhu. A limited memory algorithm for bound constrained optimization. *Journal on Scientific Computing*, 16(5):1190–1208. Citado na pág. [9](#)
- DeFinetti (1972)** B. DeFinetti. *Probability, induction and statistics*. John Wiley & Sons. Citado na pág. [17](#)
- Dixon e Coles (1997)** M. J. Dixon e S. G. Coles. Modelling association football scores and inefficiencies in the football betting market. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 46(2):265–280. Citado na pág. [2](#), [14](#), [39](#)
- Drèze e Hoch (1998)** X. Drèze e S. J. Hoch. Exploiting the installed base using cross-merchandising and category destination programs. *International Journal of Research in Marketing*, 15(5):459–471. Citado na pág. [1](#), [14](#)
- Goddard (2005)** J. Goddard. Regression models for forecasting goals and match results in association football. *International Journal of Forecasting*, 21(2):331–340. Citado na pág. [14](#)
- Griffiths et al. (1979)** R. C. Griffiths, R. K. Milne e R. Wood. Aspects of correlation in bivariate poisson distributions and processes. *Australian Journal of Statistics*, 21(3):238–255. Citado na pág. [52](#)
- Hamdan e Al-Bayyati (1971)** M. A. Hamdan e H. A. Al-Bayyati. Canonical expansion of the compound correlated bivariate Poisson distribution. *Journal of the American Statistical Association*, 66(334):390–393. Citado na pág. [4](#), [6](#)

- Ho e Singer (2001)** L. L. Ho e J. M. Singer. Generalized least squares methods for bivariate Poisson regression. *Communications in Statistics-Theory and Methods*, 30(2):263–277. Citado na pág. [1](#), [2](#), [15](#), [16](#)
- Holgate (1964)** P. Holgate. Estimation for the bivariate Poisson distribution. *Biometrika*, 51(1-2):241–287. Citado na pág. [1](#), [5](#)
- Hu e Zidek (2002)** F. Hu e J. V. Zidek. The weighted likelihood. *The Canadian Journal of Statistics*, 30(3):347–371. Citado na pág. [2](#), [36](#)
- Ibrahim e Chen (2000)** J. G. Ibrahim e M-H. Chen. Power Prior distributions for regression models. *Statistical Science*, páginas 46–60. Citado na pág. [36](#)
- Jung e Winkelmann (1993)** R. C. Jung e R. Winkelmann. Two aspects of labor mobility: A bivariate Poisson regression approach. *Empirical Economics*, 18(3):543–556. Citado na pág. [2](#), [15](#), [16](#)
- Karlis e Ntzoufras (2003)** D. Karlis e I. Ntzoufras. Analysis of sports data by using bivariate Poisson models. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 52(3):381–393. Citado na pág. [1](#), [2](#), [14](#), [15](#)
- Karlis e Ntzoufras (2005)** D. Karlis e I. Ntzoufras. Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of Statistical Software*. Citado na pág. [1](#), [2](#), [16](#)
- Kawamura (1973a)** K. Kawamura. The structure of bivariate Poisson distribution. *Kodai Mathematical Journal*, 25(2):246–256. Citado na pág. [1](#), [5](#)
- Kawamura (1973b)** K. Kawamura. The diagonal distribution of the bivariate Poisson distribution. *Kodai Mathematical Journal*, 25(3):379–384. Citado na pág. [6](#)
- Kawamura (1976)** K. Kawamura. The structure of trivariate Poisson distribution. *Kodai Mathematical Journal*, 28(1):1–8. Citado na pág. [5](#)
- Kawamura (1985)** K. Kawamura. A note on the recurrent relations for the bivariate Poisson distribution. *Kodai Mathematical Journal*, 8(1):70–78. Citado na pág. [2](#), [9](#)
- Kocherlakota (1988)** S. Kocherlakota. On the compounded bivariate Poisson distribution: A unified treatment. *Annals of the Institute of Statistical Mathematics*, 40(1):61–76. Citado na pág. [6](#)
- Kocherlakota e Kocherlakota (1992)** S. Kocherlakota e K. Kocherlakota. *Bivariate discrete distributions*, volume 132. CRC Press LLC. Citado na pág. [1](#), [4](#), [8](#)
- Kocherlakota e Kocherlakota (2001)** S. Kocherlakota e K. Kocherlakota. Regression in the bivariate Poisson distribution. Citado na pág. [2](#), [15](#)
- Lai (1995)** C. D. Lai. Construction of bivariate distributions by a generalised trivariate reduction technique. *Statistics & Probability Letters*, 25(3):265–270. Citado na pág. [5](#)
- Lakshminarayana et al. (1999)** J. Lakshminarayana, S. N. N. Pandit e K. Srinivasa Rao. On a bivariate poisson distribution. *Communications in Statistics - Theory and Methods*, 28(2):267–276. Citado na pág. [52](#)

- Lee (1997)** A. J. Lee. Modeling scores in the Premier League: Is Manchester United really the best? *Chance*, 10(1):15–19. Citado na pág. [1](#), [14](#)
- Lee (1996)** M-L. T. Lee. Properties and applications of the sarmanov family of bivariate distributions. *Communications in Statistics - Theory and Methods*, 25(6):1207–1222. Citado na pág. [52](#)
- Lindsay (1994)** B. G. Lindsay. Efficiency versus robustness: the case for minimum Hellinger distance and related methods. *The Annals of Statistics*, 22(2):1081–1114. Citado na pág. [36](#)
- Loukas e Kemp (1986)** S. Loukas e C. D. Kemp. On the chi-square goodness-of-fit statistic for bivariate discrete distributions. *The Statistician*, páginas 525–529. Citado na pág. [8](#)
- Maher (1982)** M. J. Maher. Modelling association football scores. *Statistica Neerlandica*, 36(3): 109–118. Citado na pág. [14](#)
- Markatou et al. (1997)** M. Markatou, A. Basu e B. Lindsay. Weighted likelihood estimating equations: The discrete case with applications to logistic regression. *Journal of Statistical Planning and Inference*, 57(2):215–232. Citado na pág. [36](#)
- R Development Core Team (2009)** R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2009. URL <http://www.R-project.org>. Citado na pág. [9](#), [19](#), [39](#)
- Riphahn et al. (2003)** R. T. Riphahn, A. Wambach e A. Million. Incentive effects in the demand for health care: A bivariate panel count data estimation. *Journal of Applied Econometrics*, 18 (4):387–405. Citado na pág. [1](#)
- Rue e Salvesen (2000)** H. Rue e O. Salvesen. Prediction and retrospective analysis of soccer matches in a league. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 49(3): 399–418. Citado na pág. [14](#)
- Shenson e Joshi (2012)** J. Shenson e A. Joshi. Visualizing disease incidence in the context of socioeconomic factors. Em *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction*, páginas 29–38. ACM. Citado na pág. [1](#), [14](#)
- Teicher (1954)** H. Teicher. On the multivariate Poisson distribution. *Scandinavian Actuarial Journal*, 1954(1):1–9. Citado na pág. [2](#), [9](#)
- Wang (2006)** X. Wang. Approximating Bayesian inference by weighted likelihood. *Canadian Journal of Statistics*, 34(2):279–298. Citado na pág. [36](#)
- Wang et al. (2004)** X. Wang, C. van Eeden e J. V. Zidek. Asymptotic properties of maximum weighted likelihood estimators. *Journal of Statistical Planning and Inference*, 119(1):37–54. Citado na pág. [36](#)