



IBMEC SÃO PAULO
Faculdade de Economia e Administração

Thomas Osborn Mahfuz

**MODELAGEM DO NÚMERO DE GOLS MARCADOS
USANDO A REGRESSÃO DE POISSON**

São Paulo
2009

Thomas Osborn Mahfuz

**MODELAGEM DO NÚMERO DE GOLS MARCADOS
USANDO A REGRESSÃO DE POISSON**

Monografia entregue ao curso de Ciências
Econômicas, como requisito final para a disciplina
Monografia II.

Orientadora:
Adriana Bruscato Bortoluzzo – Ibmecc SP

**São Paulo
2009**

Thomas Osborn Mahfuz

Modelagem do número de gols marcados usando a regressão de Poisson

Monografia apresentada à Faculdade de Economia, do Ibmec como parte dos requisitos para conclusão do curso de graduação em Economia.

Aprovado em Dezembro 2009

EXAMINADORES

Prof. Dra. Adriana Bruscato Bortoluzzo
Orientadora

Prof. Dr. Eduardo Andrade
Examinador

Prof. Ms. Sérgio Martins
Examinador

Resumo

Mahfuz, Thomas Osborn. Modelagem do número de gols marcados usando a regressão de Poisson. São Paulo, 2009. 29p. Monografia – Faculdade de Economia do Ibmec.

Essa monografia utiliza um método de previsão para o resultado de partidas de futebol, baseado num modelo de gols marcados tratados como variáveis com distribuição de Poisson, usando o sistema de pontos do ranking da Fédération Internationale de Football Association (FIFA) e o mando de jogo como variáveis preditoras. O modelo apresentou um coeficiente de explicação elevado, indicando boa previsibilidade do número de gols marcados numa determinada partida da Copa do Mundo de 2002. Além disso, o ranking da FIFA e o mando de jogo foram relevantes e a análise revelou a adequabilidade do modelo.

Palavras-chave: esportes, futebol, previsão, regressão de Poisson.

Abstract

Mahfuz, Thomas Osborn. A ratings based model for the number of goals scored using a Poisson based regression. São Paulo, 2009. 29p. Monograph – Faculdade de Economia do Ibmecc.

This monograph uses a method of prediction for results of football matches based on a model that regresses the number of goals scored as independent Poisson variables dependent on the Fédération Internationale de Football Association (FIFA) rating of each team, and the match venue. The FIFA rating and the match venue were found to be statistically relevant in the final model. To evaluate the final model's adequacy, results of a few matches of the 2002 FIFA World Cup were estimated, and a residual analysis carried out.

Keywords: football, forecasting, Poisson regression, prediction, sports.

Dedicatória

A minha orientadora Adriana, cuja paciência, energia e dedicação foram ingredientes essenciais para a elaboração desta monografia, valorizando as quatrocentas e vinte horas que investi na mesma.

Sumário

1 Introdução	7
2 Metodologia	9
3 Resultados	12
3.1 Análise Descritiva	12
3.2 Análise Inferencial	15
3.3 Análise de Resíduos	16
3.4 Previsão das Partidas	18
4 Conclusões	21
5 Referências	22
6 Apêndices	23
5.1 Apêndice A	23
5.2 Apêndice B	28
5.2 Apêndice C	29

1. Introdução

A Copa do Mundo promovida pela Federal International Football Association (FIFA), disputada a cada quatro anos, é o evento esportivo mais assistido no mundo. A etapa eliminatória é disputada por 192 nações e as seleções que obtêm a classificação participam do torneio num país sede pré-definido. Desde a Copa do Mundo de 1998, sediada na França, o número de seleções disputando o torneio expandiu para 32 times. Os times são divididos em oito grupos e após a disputa de jogos entre todos os quatro times de cada grupo, o primeiro e o segundo colocados de cada grupo classificam-se para as fases seguintes da Copa.

O índice da Coca-Cola FIFA (índice da FIFA), elaborado para classificar as seleções num ranking baseado em um sistema de pontos, premia fatores como o número de gols marcados, a força regional relativa das equipes e a importância do jogo. O interesse do autor por esse tema foi despertado graças à possibilidade de avaliar a qualidade do ranking da FIFA como previsor do número de gols marcados para uma dada partida.

Com isso, espera-se contribuir para a literatura desportiva ao avaliar o poder de previsão preditivo de um modelo onde o número de gols marcados é tratado como variável dependente com distribuição de Poisson, tendo o sistema de pontos do ranking da FIFA como variável preditora. Outra expectativa é a de despertar o interesse matemático e estatístico de leitores, considerando que o tema aborda uma área de amplo interesse populacional.

Stefani (1997) compilou uma análise sintetizada dos índices oficiais utilizado pelas instituições que governam os esportes como o esqui, tênis, futebol e golfe e concluiu que esses índices dependem mais do resultado do evento, seja uma vitória ou uma derrota (e um empate se for aplicável ao esporte em questão), do que o placar observado. A exceção é o futebol, segundo o estudo, devido aos rankings levarem em consideração o placar observado nos jogos em suas compilações. O ranking da FIFA, por exemplo, premia as seleções pelo número de gols marcados e os pune pelo número de gols sofridos. Já os autores Clarke e Dyte (1999) elaboraram um modelo para prever o resultado de sets de tênis baseado somente na diferença entre os pontos dos jogadores no ranking da ATP.

Dixon e Robinson (1998) estudaram aproximadamente 4000 partidas de futebol entre clubes ingleses e modelaram o número de gols marcados como dois processos interativos, usando parâmetros como força ofensiva e defensiva, mando de jogo, placar do jogo atual e o tempo restante para o fim da partida. Os resultados modelaram os dados muito bem e um modelo robusto foi elaborado ao fim da pesquisa. No entanto, tais detalhes estendem-se fora

do escopo desse trabalho, em que se pretende seguir a linha da modelagem feita por Clarke e Dyte (2000), em que se busca apenas medir, de alguma forma, o poder preditivo do ranking oficial da FIFA. No modelo desses autores, as mecânicas mais profundas de uma partida completa de futebol foram deixadas de lado para que se pudesse dar ênfase na avaliação da qualidade de previsão do índice da FIFA ao estimar o número de gols marcados para um dado time utilizando uma regressão de Poisson.

O objetivo dessa monografia é o de elaborar um modelo de previsão de gols marcados, tratando o número de gols de cada seleção como variáveis com distribuição de Poisson independentes, usando como variável explicativa o índice de cada país gerado pelo ranking da FIFA e o mando de jogo. Utilizando os resultados desse modelo de regressão de Poisson, pode-se derivar a distribuição de probabilidades para os gols marcados por cada time em cada jogo, e, portanto, o resultado esperado da partida.

A escolha de 2002 deve-se ao fato de que não há nenhum trabalho na literatura que simulou esta Copa usando como previsor o índice elaborado pela FIFA. Além disso, o autor é brasileiro e esse foi o último ano em que o Brasil venceu a Copa do Mundo. O autor entende que a obtenção de um previsor perfeito para um jogo que envolva centenas de variáveis provavelmente seja um objetivo inalcançável. No entanto, caso esse trabalho consiga avaliar o quanto adequado seja um modelo baseado no índice FIFA, utilizando uma regressão de Poisson, para estimar o número de gols marcados para um dado jogo da Copa do Mundo, o mesmo terá atingido os objetivos do autor.

É de profundo interesse do autor e de fãs de futebol, principalmente daqueles que apostam em resultados de partidas, investigar a capacidade de previsão de um índice deste tipo. A escolha do índice da FIFA está enraizada no fato de que esse é o órgão máximo do esporte em questão e, portanto, espera-se que sua avaliação das seleções seja o mais próximo da realidade possível. Ademais, a disponibilidade dos dados auxiliou na motivação do autor na escolha do índice FIFA para servir como principal variável explicativa do modelo construído.

2. Metodologia

A coleta de dados foi feita a partir do site oficial da FIFA. Dados como a data da partida, os participantes envolvidos, o local de jogo e o placar foram extraídos e compilados com o auxílio de macros elaborados em VBA pelo próprio autor. O período coletado foi de 12 meses antes da Copa, pois se espera que esses resultados sejam os que melhor descrevam o desempenho esperado para a Copa de 2002. Vale ressaltar que a FIFA não atualiza o índice durante o andamento da Copa, portanto a regressão dos dados foi limitada ao pré-Copa, data mais próxima da coleta de dados do trabalho. Ademais, foi extraído o índice da FIFA para o mesmo período através do site oficial da entidade. A amostra final contava com 886 observações, que correspondem aos jogos 12 meses antes da Copa.

Anderson *et al.* (1998) discutem sobre a aplicação da regressão de Poisson, em que ela é uma forma de análise utilizada para modelar dados cujos valores observados devem ser números inteiros e não negativos, além de servir na modelagem de tabelas de contingência. Uma característica de uma variável que segue a distribuição de Poisson é que a média é igual à variância da variável. Em certas circunstâncias, será observado que a variância é maior do que a média. Esse fenômeno é conhecido como excesso de dispersão (superdispersão) na literatura e indica que o modelo não é apropriado. Segundo Boland (2007), causas comuns são a omissão de variáveis explicativas relevantes ou a presença de outliers. Para esses casos, a literatura recomenda outros modelos generalizados, como a distribuição binomial negativa. Anderson *et al.* (1998) sugerem que, como via de regra, a distribuição de Poisson pode ser aproximada por uma distribuição binomial e será bem especificada quando a probabilidade de sucesso do evento for inferior a 5% e o número de tentativas for maior ou igual a 20.

Seguindo a descrita em Paula (2004), dizemos que uma variável aleatória Y tem distribuição de Poisson com parâmetros λ , ou seja $Y \sim P(\lambda)$, quando sua distribuição de probabilidades é dada por:

$$\Pr(Y = y) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad y = 0, 1, 2, \dots, \lambda > 0 \quad (2.1)$$

Este trabalho será desenvolvido com base num modelo de regressão de Poisson, em que o número de gols marcados para uma dada equipe é a variável resposta com distribuição de Poisson e as variáveis explicativas são o índice da FIFA para o time e para o oponente e o mando de campo (casa, fora de casa ou neutro).

Desta forma, o modelo de regressão linear múltipla para cada um dos jogos da Copa de 2002 é dado por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \varepsilon_i \quad (2.2)$$

em que

y_i é o número de gols marcados no i -ésimo jogo,

x_{1i} é o índice FIFA para o time considerado,

x_{2i} é o índice FIFA para o time oponente,

x_{3i} é uma dummy, que vale 1 caso o mando de jogo seja do time considerado,

x_{4i} é uma dummy, que vale 1 caso o mando de jogo seja neutro e

ε_i é o erro aleatório.

O modelo linear geral, em sua forma matricial, pode ser escrito da seguinte forma:

$$\underset{\sim}{y} = \underset{\sim}{X} \underset{\sim}{\beta} + \underset{\sim}{\varepsilon} \quad (2.3)$$

em que

$\underset{\sim}{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ é um vetor de dimensão $n \times 1$, contendo o número de gols marcados para cada um dos jogos da Copa de 2002,

$\underset{\sim}{X} = \begin{bmatrix} 1 & x_1 & \dots & x_4 \end{bmatrix}$ é uma matriz de dimensão $n \times 5$, contendo as variáveis explicativas do modelo,

$\underset{\sim}{\beta}$ é um vetor de dimensão 5×1 , contendo os parâmetros do modelo e

$\underset{\sim}{\varepsilon}$ é um vetor de dimensão $n \times 1$, contendo os erros aleatórios.

A função de ligação utilizada para a média será logaritmo neperiano (log), ou seja, o valor esperado de gols marcados pelo time considerado é dado por:

$$\hat{y} = e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 x_4} \quad (2.4)$$

Para que o modelo seja estimado, algumas suposições precisam ser feitas:

1. o número de gols marcados por um time numa partida de futebol segue uma distribuição de Poisson; e

2. o número de gols marcados por um time num determinado jogo não é correlacionado com o número de gols marcados numa outra partida.

Para que uma variável aleatória seja considerada de Poisson há necessidade que ela tenha resultados não negativos e siga uma distribuição de Poisson, o que será verificado na análise de resíduos. A ausência de correlação entre o número de gols marcados em diferentes partidas é um fenômeno natural.

Maher (1982) sugere que se for encontrado um método apropriado para calcular o valor esperado de gols para cada possível par de times, uma distribuição de Poisson providencia um bom ajuste aos dados observados. Seguindo essa sugestão, Clarke e Dyte (2000) elaboraram uma regressão de Poisson para a Copa de 1998. Este trabalho seguirá um rumo semelhante, utilizando a distribuição de Poisson para modelar o número de gols marcados na Copa de 2002.

McCullagh e Nelder (1989) mostram que, quando $\lambda \rightarrow \infty$, $\frac{(Y - \lambda)}{\sqrt{\lambda}} \xrightarrow{d} N(0,1)$. Este resultado será usado na análise de resíduos. Em outras palavras, para um λ suficientemente grande, temos que Y segue aproximadamente uma distribuição normal de média e variância iguais a λ . Entretanto, não é possível utilizar um modelo normal linear homocedástico devido à inconveniência do desvio padrão depender da média.

Paula (2004) sugere um meio de contornar o problema, através da transformação da variável resposta Y' , de modo a alcançar a normalidade e constância de variância, mesmo que de maneira aproximada. Nesse caso, por exemplo, se Y tem distribuição de Poisson, quando $\lambda \rightarrow \infty$, temos que $\left[\sqrt{Y} - E(\sqrt{Y}) \right] \xrightarrow{d} N(0, 1/4)$. Logo, quando λ é grande, a variável aleatória $2\left[\sqrt{Y} - E(\sqrt{Y}) \right]$ segue aproximadamente uma distribuição $N(0,1)$.

Os resultados das partidas da Copa de 2002 também serão compilados por meio da base disponível na página oficial da FIFA na internet. As análises e regressões foram feitas por meio dos softwares Eviews (Versão 6.0), S-PLUS e Microsoft Excel 2007. Além disso, os trabalhos acadêmicos necessários para a pesquisa e utilizados para a elaboração desta monografia foram obtidos por meio das bases de dados JSTOR e EconLit.

3. Resultados

3.1. Análise Descritiva

A Tabela 1 lista os países participantes da Copa de 2002, além de algumas estatísticas relevantes, como o número de jogos disputados por cada seleção, o número de vitórias, gols marcados, gols sofridos, as médias de gols marcados e sofridos por jogo e o percentual de vitórias.

É possível extrair da Tabela 1 que as três seleções favoritas à conquista da Copa do Mundo de 2002, utilizando apenas o índice FIFA como critério de avaliação, foram a França, o Brasil e a Argentina. A França obteve um índice de 802, o Brasil e a Argentina tiveram um índice de 784. Porém, o desempenho da França foi extremamente desapontante comparado a seu grau de favoritismo. Isso sugere uma possível má distribuição de pontos do índice da FIFA a favor da seleção que havia vencido a edição anterior da Copa, já que a França perdeu todos os jogos que disputou na Copa de 2002.

O mau desempenho Argentino, apesar de ser uma decepção se considerado o alto índice FIFA que esta seleção obteve antes da Copa, pode ser parcialmente atribuído ao fato de que ela foi sorteada para o “Grupo da Morte”, composto das seleções da Inglaterra, Suécia e Nigéria. A Itália foi outra seleção cujo desempenho na Copa de 2002 deixou a desejar, considerando seu índice FIFA e a tradição da “Azurra” no futebol mundial. No entanto, a Itália, a despeito do que ocorreu na Copa de 2006 quando foi beneficiada por erros de arbitragem no jogo contra a Austrália, teve 5 gols anulados erroneamente pela arbitragem na Copa de 2002. Por outro lado, o desempenho extraordinário do Brasil reforça que os critérios de pontuação da FIFA estejam relativamente próximos da realidade para vasta porção das seleções que se classificaram para a Copa. O Brasil obteve 100% de aproveitamento e ainda foi a seleção que menos gols sofreu, quando considerados o número de jogos disputados (vale observar que a Argentina foi o time menos vazado em termos absolutos, mas somente disputou a fase eliminatória da Copa).

A maior média de gols marcados foi atingida pela Espanha com a marca impressionante de 3.20 gols por jogo. Note também que a Espanha foi a seleção que mais sofreu gols, em termos absolutos, no torneio. Possivelmente, esse efeito foi provocado pelo modo ofensivo do time espanhol de jogar, favorecendo um jogo mais aberto e uma formação mais agressiva.

Tabela 1 – Participantes da Copa de 2002 e algumas estatísticas sobre o desempenho de cada seleção. **Índice FIFA:** Pontos no ranking FIFA em 15/05/2002¹ **GM:** Gols Marcados, **GS:** Gols Sofridos, **J:** Jogos disputados, **V:** Vitórias, **Média GM:** Média de Gols Marcados por jogo, **Média GS:** Média de Gols Sofridos por jogo, **% de Vitórias:** Percentual dos jogos vencidos.

Seleção	Índice					Média GM	Média GS	% de Vitórias
	FIFA	GM	GS	J	V			
Argentina	784	2	2	3	1	0,67	0,67	33
Belgium	653	6	7	4	1	1,50	1,75	25
Brazil	784	18	4	7	7	2,57	0,57	100
Cameroon	672	2	3	3	1	0,67	1,00	33
China PR	566	0	9	3	0	0,00	3,00	0
Costa Rica	643	5	6	3	1	1,67	2,00	33
Croatia	655	3	2	3	2	1,00	0,67	67
Denmark	657	5	5	4	2	1,25	1,25	50
Ecuador	624	2	4	3	1	0,67	1,33	33
England	694	6	3	5	2	1,20	0,60	40
France	802	0	3	3	0	0,00	1,00	0
Germany	695	14	3	7	5	2,00	0,43	71
Italy	717	5	5	4	1	1,25	1,25	25
Japan	634	5	3	4	2	1,25	0,75	50
Mexico	716	3	5	4	1	0,75	1,25	25
Nigeria	644	1	3	3	0	0,33	1,00	0
Paraguay	671	6	7	4	1	1,50	1,75	25
Poland	615	3	7	3	1	1,00	2,33	33
Portugal	726	6	4	3	1	2,00	1,33	33
Republic of Ireland	674	8	6	4	1	2,00	1,50	25
Russia	644	4	4	3	1	1,33	1,33	33
Saudi Arabia	627	0	12	3	0	0,00	4,00	0
Senegal	599	7	6	5	2	1,40	1,20	40
Slovenia	649	2	7	3	0	0,67	2,33	0
South Africa	623	5	5	3	1	1,67	1,67	33
South Korea	603	13	9	7	4	1,86	1,29	57
Spain	713	16	12	5	4	3,20	2,40	80
Sweden	665	5	5	4	1	1,25	1,25	25
Tunisia	635	1	5	3	0	0,33	1,67	0
Turkey	654	10	6	7	4	1,43	0,86	57
Uruguay	652	4	5	3	0	1,33	1,67	0
USA	690	7	7	5	2	1,40	1,40	40
Média	668,13	5,44	5,44	4,00	1,56	1,22	1,45	33
Desvio Padrão	53,07	4,44	2,46	1,32	1,60	0,71	0,74	24

¹ Última data que o ranking foi atualizado antes da Copa do Mundo de 2002

Tabela 2 – Correlações do número de gols marcados (y), índice FIFA do time em consideração (x_1) e o índice FIFA do oponente (x_2), para os jogos anteriores à Copa de 2002 (base de dados utilizados para realizar a regressão).

	Y	x_1	x_2
y	1,000	0,228	-0,257
x_1	0,228	1,000	-0,018
x_2	-0,257	-0,018	1,000

As correlações das variáveis número de gols marcados e de gols sofridos para os dados utilizados para realizar a regressão estão descritos na tabela anterior. Vale notar que a correlação entre o número de gols marcados (y) e o índice FIFA para o time considerado (x_1) é de 0.228. Esse resultado é positivo, indicando que quanto maior o índice FIFA para o time considerado, maior o número de gols marcados pelo time. Porém, sua intensidade é relativamente pequena, sugerindo que a construção do índice deixa a desejar como predictor para o número de gols marcados.

A correlação entre o número de gols marcados (y) e o índice FIFA para o adversário (x_2) é de -0.257. Esse resultado é negativo, indicando que quanto maior o índice FIFA para o oponente, menor o número de gols marcados pelo time considerado. Porém, sua intensidade é relativamente pequena, sugerindo que a construção do índice deixa a desejar como predictor para o número de gols marcados.

No entanto, ao utilizarmos uma regressão que incorpora tanto o índice FIFA do time em questão quanto do adversário, a explicação para o número de gols marcados pode aumentar significativamente.

Esses resultados confirmam a imperfeição do índice em mensurar a habilidade de um time. Conforme apontadas por Clarke e Dyte (2000), várias nações de federações de menor significância futebolística jogaram mais jogos dentro de suas áreas geográficas e poucos fora, viesando seu índice para cima.

Finalmente, a correlação entre o índice FIFA para o time em questão (x_1) e o índice FIFA para o adversário (x_2) é de -0.018. Esse resultado é negativo e próximo de zero, mais ainda, este valor está em linha com a expectativa de que o índice FIFA dos times não deveria ser correlacionado para a maioria das nações.

3.2. Análise Inferencial

Os parâmetros do modelo descrito em (2.2) foram estimados por máxima verossimilhança, utilizando o método numérico desenvolvido por Newton-Raphson e com função de ligação log para a média. (vide Apêndice B - Modelo 1). No entanto, no modelo estimado, observamos que o intercepto não foi estatisticamente significativo com 95% de confiança. Desta forma, estimamos um novo modelo de regressão sem intercepto (vide Apêndice B - Modelo 2).

Desta forma, o modelo final estimado é (entre parênteses estão os erros padrões dos estimadores):

$$\ln(\hat{y}_i) = 0,0022 x_{1i} - 0,0025 x_{2i} + 0,4680 x_{3i} + 0,2399 x_{4i} \quad (4)$$

$$(0,0002) \quad (0,0002) \quad (0,1003) \quad (0,0854)$$

Obviamente, nesse torneio, as únicas equipes a se beneficiarem da vantagem de jogar em casa foram o Japão e a Coréia do Sul. Note que a diferença entre as estimativas para um time, quando transitando entre um mando em casa para um neutro, é aproximadamente a mesma diferença para um time transitando de um mando neutro para um fora de casa. Isso indica que, pelo menos para a base de dados utilizada, a vantagem de jogar em casa dobra o a vantagem referente ao mando de jogo para o número esperado de gols marcados, quando comparado a seu adversário que joga fora de casa.

No modelo estimado, a única variável que afeta negativamente o número de gols marcados é o índice FIFA do time adversário. Todas as outras estimativas são positivas, indicando que o índice FIFA do time em questão, o mando de campo em casa ou em um campo neutro, em relação ao mando fora de casa, afetam positivamente o número de gols marcados. Ou seja, quanto maior o índice FIFA do time considerado e se ele joga em casa ou em campo neutro, em relação ao mando fora de casa, maior será o número de gols marcados pelo mesmo.

Por exemplo, se o índice FIFA do adversário aumentar em 100 pontos, espera-se aumento relativo no número de gols marcados pelo time de $\frac{e^{-0.002593 \cdot (x_2 + 100)}}{e^{-0.002593 \cdot (x_2)}} = e^{-0.2593} = 0.7716$. Ou seja, espera-se um decréscimo de 22.84% no número de gols marcados do time considerado.

Por outro lado, se o índice FIFA do time aumentar em 100 pontos, espera-se aumento relativo no número de gols marcados pelo time de $\frac{e^{0.002445 \cdot (x_1+100)}}{e^{0.002445 \cdot (x_1)}} = e^{0.2445} = 1.2770$. Ou seja, espera-se um aumento de 27.70% no número de gols marcados.

Notamos a proximidade do parâmetro para o número de gols marcados esperados para um dado time é afetado positivamente com um peso próximo ao peso negativo do parâmetro referente ao índice FIFA do seu adversário. Podemos testar as seguintes hipóteses com base num teste de Wald: $H_0 : \beta_1 = -\beta_2$ vs $H_A : \beta_1 \neq -\beta_2$. O teste apresenta nível descritivo igual a 0.2646 e comparando com um nível de significância de 5%, leva-nos a concluir acerca da não rejeição da hipótese nula. Portanto, isto indica que o número de gols marcados esperados para um dado time é impactado positivamente por seu índice FIFA com um peso semelhante ao impacto negativo do índice FIFA de seu adversário.

3.3. Análise de Resíduos

A função desvio, que é representa a diferença entre valores observados e esperados, de um modelo Poisson é dada por:

$$D(y; \hat{\mu}) = 2 \sum_{i=1}^n \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right],$$

em que y descreve o número de gols observados e $\hat{\mu}$ descreve o número de gols esperados. No entanto, foi utilizado o componente do desvio padronizado, sugerido por Paula (2004), para modelos com resposta de Poisson:

$$t_{D_i} = \frac{\pm \sqrt{2} \cdot \sqrt{y_i \log(y_i / \hat{\mu}_i) - (y_i - \hat{\mu}_i)}}{\sqrt{1 - \hat{h}_{ii}}}$$

em que

h_{ii} descreve o elemento da linha i e da coluna i a matriz hat, que é a matriz de projeção ortogonal de vetores \mathfrak{R}^n no subespaço gerado pelas colunas da matriz \tilde{X} , ou seja, $H = \tilde{X}(\tilde{X}'\tilde{X})^{-1}\tilde{X}'$, em que \tilde{X} é uma matriz de dimensão $n \times 5$, contendo as variáveis explicativas do modelo.

Estudos desenvolvidos por Williams (1984) mostram que em geral a distribuição de t_{D_i} é suficientement próxima da distribuição normal, podendo serem usadas nas análises

de resíduo as mesmas interpretações da regressão linear normal. Particularmente, a construção de envelopes, limites inferior e superior com base em percentis, via simulações é recomendada para t_{Di} .

A Figura 1, ilustrada a seguir, descreve a distribuição residual no formato de um histograma. Vale ressaltar que a assimetria residual é extremamente próxima de zero (0,468) e têm uma curtose aproximadamente igual a 3 (2,78). Por outro lado, o teste de Jarque-Bera (1981) rejeita a hipótese de normalidade dos erros para qualquer nível de significância usual.

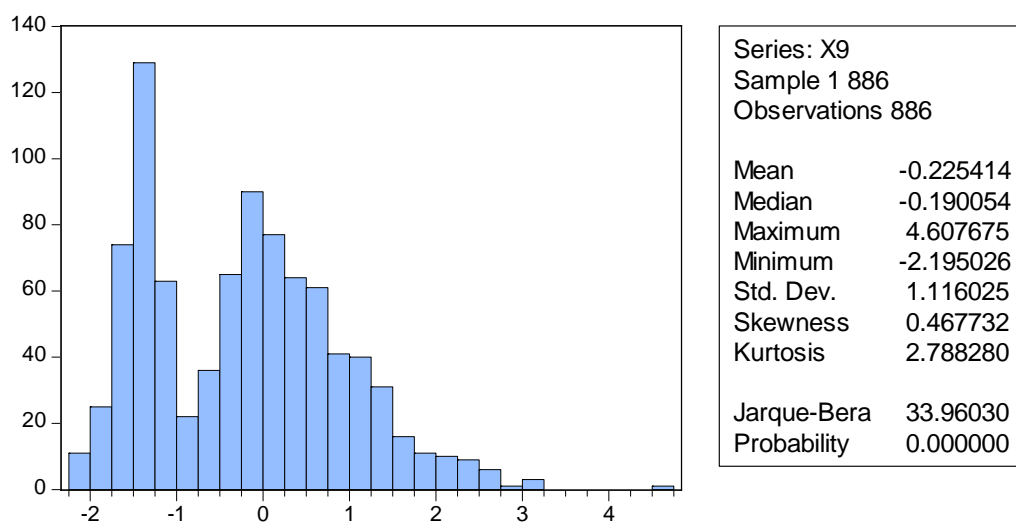


Figura 1 – Histograma e medidas resumo dos resíduos

Já a Figura 2 descreve o comportamento residual e seus limites inferior e superior construídos via simulação (envelopes). Na cauda superior, os maiores resíduos têm um leve desvio, ou seja, estão fora do limite superior do envelope, indicando uma leve falta de ajuste o que possivelmente explica a rejeição da normalidade pelo teste Jarque-Bera. Nos outros gráficos residuais (vide Apêndice C), não notamos grandes desvios em relação as suposições do modelo, pois as observações apresentam-se distribuídas aleatoriamente

Vale ressaltar o *outlier* evidenciado por ambas figuras. Esse resultado foi a partida entre a partida da Alemanha e a Inglaterra disputada em Setembro de 2001, em que a Inglaterra venceu por 5 a 1. O índice FIFA na época para a Inglaterra era de 675. Já a seleção alemã dentinha uma pontuação de 725. Esse hiato de índices é bastante significativo e somado ao fato do jogo ter ocorrido no *Olympiastadion*, na Alemanha, a previsão do modelo apontava para uma vitória Alemã. O resultado foi bastante

surpreendente, se considerarmos apenas os fatores de índice da FIFA e o mando de jogo. Outro fator interessante desta partida é que ela marcou o fim da supremacia Alemã de quase 25 anos sobre a seleção inglesa.

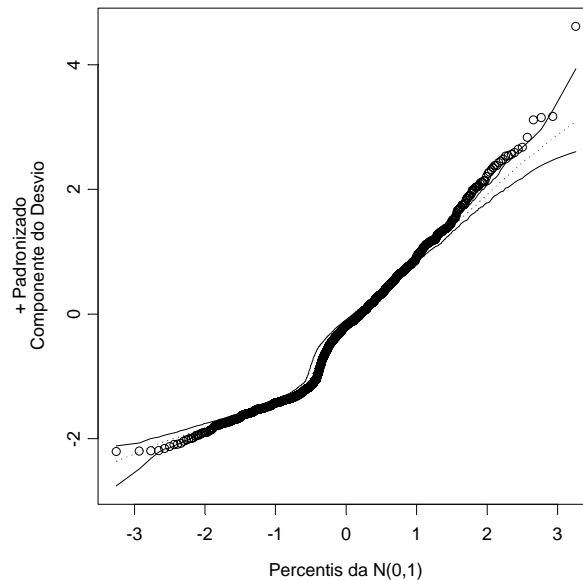


Figura 2 – Envelope dos resíduos

3.4. Previsão das Partidas

O jogo entre o Brasil e a Costa Rica providencia um exemplo para ilustrar como o método de previsão de partidas funciona.

$$FIFA_{Brasil} = 784$$

$$FIFA_{Costa Rica} = 643$$

$$\hat{y}_{Brasil} = e^{(784 \cdot 0.002445) + (-0.002593 \cdot 643) + 0.23992} = 1,6315$$

$$\hat{y}_{Costa Rica} = e^{(643 \cdot 0.002445) + (-0.002593 \cdot 784) + 0.23992} = 0,8018$$

Portanto, o valor esperado era de 1.6315 gols para o Brasil contra 0.8018 gols para a Costa Rica. Utilizando esses resultados como médias, foi possível gerar as probabilidades conjuntas e marginais para a distribuição de Poisson de gols marcados das seleções. Para calcular as probabilidades conjuntas usa-se a suposição de independência, então as probabilidades podem ser multiplicadas para gerar a probabilidade conjunta do resultado do jogo, conforme ilustrado na Tabela 3 (somente inclusos os eventos que tiveram

probabilidades de 0.001 ou maior). No caso do Brasil, a probabilidade de fazer 1 gol contra a Costa Rica é de 31.9% e de fazer 2 gols é de 26%. Para a Costa Rica, a probabilidade de fazer 1 gol é de 35.9% e de fazer 2 gols é de 14.4%.

Tabela 3 – Distribuição de probabilidades de resultados para o jogo Brasil versus Costa Rica.

		Brasil							Total
		0	1	2	3	4	5	6	
Costa Rica	0	0,088	0,143	0,117	0,064	0,026	0,008	0,002	0,448
	1	0,070	0,115	0,094	0,051	0,021	0,007	0,002	0,359
	2	0,028	0,046	0,038	0,020	0,008	0,003	0,001	0,144
	3	0,008	0,012	0,010	0,005	0,002	0,001	0,000	0,038
	4	0,002	0,002	0,002	0,001	0,000	0,000	0,000	0,008
Total		0,195	0,319	0,260	0,141	0,058	0,019	0,005	0,997

Ao somarmos os resultados da diagonal principal, os resultados acima da diagonal principal e os resultados abaixo da diagonal principal obterão a probabilidade de empate, de o Brasil vencer e da Costa Rica vencer, respectivamente. Ao computarmos esses resultados, a probabilidade de empate é de 0.246, a probabilidade de o Brasil vencer é de 0.570 e a probabilidade da Costa Rica vencer é de 0.182. O resultado exato mais provável é 1 a 0 para o Brasil, em que a probabilidade deste evento é de 14.3% (para calcular a probabilidade do jogo terminar 1 a 0 para o Brasil temos $P[B=1 \cap CR=0] = P[B=1] \times P[CR=0] = 0.319 \times 0.448 = 0.143$).

O resultado da partida foi 5 a 2 para o Brasil. Embora o placar absoluto previsto fora relativamente distante do placar observado, a razão dos placares é extremamente próxima. Se dividirmos os gols da Costa Rica pelos gols do Brasil, no placar observado essa razão é igual a 0.4, enquanto a razão do resultado previsto é aproximadamente 0.49.

A Tabela 3 serve também para ilustrar o problema básico de previsões para o futebol. A tabela inclui 99.7% dos resultados possíveis para essa partida, representados por 35 resultados possíveis. Embora fãz procurem um resultado com um alto grau de certeza, a variação natural dos placares impossibilita tal feito. Cada partida possui um elevado número de resultados possíveis, cujas probabilidades são relativamente baixas.

De forma análoga ao jogo entre Brasil e Costa Rica, calculamos a distribuição de probabilidades para a final entre Brasil e Alemanha, descrita pela Tabela 4 (somente inclusos os eventos que tiveram probabilidades de 0.001 ou maior).

$$FIFA_{Brasil} = 784$$

$$FIFA_{Alemanha} = 695$$

$$\hat{y}_{Brasil} = e^{(784-0.002445)+(-0.002593-695)+0.23992} = 1,4257$$

$$\hat{y}_{Alemanha} = e^{(695-0.002445)+(-0.002593-784)+0.23992} = 0,9105$$

Tabela 4 – Distribuição de probabilidades de resultados para o jogo Brasil versus Alemanha.

		Alemanha						Total
		0	1	2	3	4	5	
Brasil	0	0,0967	0,0880	0,0401	0,0122	0,0028	0,0005	0,2403
	1	0,1379	0,1255	0,0571	0,0173	0,0039	0,0007	0,3425
	2	0,0983	0,0895	0,0407	0,0124	0,0028	0,0005	0,2442
	3	0,0467	0,0425	0,0194	0,0059	0,0013	0,0002	0,1160
	4	0,0166	0,0152	0,0069	0,0021	0,0005	0,0001	0,0414
	5	0,0047	0,0043	0,0020	0,0006	0,0001	0,0000	0,0118
Total		0,4009	0,3650	0,1662	0,0504	0,0115	0,0021	

Portanto, o valor esperado era de 1,4257 gol para o Brasil contra 0,9105 gol para a Alemanha. No caso do Alemanha, a probabilidade de fazer 1 gol contra o Brasil é de 24,03% e de fazer 2 gols é de 34,25%. Já para a Brasil, a probabilidade de fazer 1 gol na Alemanha é de 40,09% e de fazer 2 gols é de 36,50%. A vantagem brasileira para esse jogo parece ser bastante absoluta devido principalmente a diferença considerável entre o índice das duas seleções.

Ao somarmos os resultados da diagonal principal, os resultados acima da diagonal principal e os resultados abaixo da diagonal principal obteremos a probabilidade de empate, do Brasil vencer e da Alemanha vencer, respectivamente. Ao computarmos esses resultados, a probabilidade de empate é de 0,2693, a probabilidade de o Brasil vencer é de 0,4867 e a probabilidade da Alemanha vencer é de 0,2401. Essas probabilidades reforçam a dificuldade de prever esta partida pois a margem de diferença entre a probabilidade esperada da Alemanha vencer e a probabilidade esperada do Brasil vencer é de apenas 1.1%. O resultado exato mais provável é a vitória brasileira por 1 a 0 sobre a Alemanha, em que a probabilidade deste evento é de 13,79% (para calcular a probabilidade do jogo terminar 1 a 0 temos $P[B=1 \cap A=0] = P[B=1] \times P[A=0] = 0,3426 \times 0,4002 = 0,1379$).

O placar observado na final da Copa foi 2 a 0 para o Brasil. Esse resultado está bastante próximo do previsto pelo modelo, ainda mais se considerarmos que a

probabilidade de vitória do Brasil era de quase metade, isso se aceitarmos que o empate era possível para a final. Rejeitando essa hipótese, a expectativa que a probabilidade de vitória brasileira seria ainda mais alta. Esse resultado aponta para que, no mínimo, utilizar o índice FIFA como previsor para partidas de Copa do Mundo é uma premissa razoável.

4. Conclusões

O índice Coca Cola-FIFA, aliado à informação do mando de jogo, formam um prognóstico útil devido a um bom poder preditivo quando assume-se que o número de gols marcados segue uma distribuição de Poisson. No modelo estimado, a vantagem trazida por jogar-se em casa é quase idêntica a desvantagem de jogar-se fora de casa, ou seja, a redução do número de gols marcados no time da casa é bastante próxima ao aumento no número de gols marcados pelo time mandante.

Quando aplicado à Copa do Mundo de 2002, o modelo demonstrou um bom desempenho e os resultados encontrados estavam alinhados à realidade observada, mesmo que nem sempre fosse possível atribuir um time favorito (vide o exemplo do jogo entre Brasil e Alemanha, em que o número esperado de gols marcados para esses times nesse jogo era muito próximo, dificultando a previsão da partida). No entanto, mesmo que em alguns casos o modelo não pudesse prever com um grau significativo de vantagem probabilística, ele demonstrou ser bastante coerente. A partida da final alocou uma vantagem à seleção brasileira em que o placar mais provável era da vitória do Brasil e a previsão de uma derrota era de apenas 24%. Esse resultado aponta para que, no mínimo, a utilização do índice FIFA como previsor para partidas de Copa do Mundo é uma premissa razoável.

Um dos elementos mais interessantes desse trabalho é a possibilidade da construção de modelos para previsão de partidas baseado no modelo estimado, seguindo em linha as simulações realizadas por Clarke e Dyte (2000). Ademais, visto que o modelo para a Copa de 2002 é razoável, seria útil para aqueles interessados em apostas futebolísticas estimar o modelo para tentar prever os jogos das próximas Copas. Caso seja utilizado uma versão mais atualizada do índice (já que ele é computado apenas até um mês antes da Copa e não reflete os jogos realizados durante a Copa), espera-se um modelo mais preciso e robusto.

No entanto, vale ressaltar que o pressuposto de que o número de gols marcados segue uma distribuição de Poisson tenha suas desvantagens, como a baixa probabilidade para

cada evento (o resultado observado da partida do Brasil e a Costa Rica na Copa de 2002 foi de 5 a 2, um evento que o modelo implicitamente alocou com uma probabilidade de apenas 0.3%). Porém, mesmo sendo difícil a previsão de um resultado com alta precisão, o modelo serviu para apontar um favorito para uma dada partida.

5. Referências bibliográficas

ANDERSON, D. R.; SWEENEY, D. J.; WILLIAMS, T. A. **Statistics for business and economics**. 7. ed. Cincinnati: Thomson - South-Western, 1998. 909 p.

BOLAND, P. J. **Statistical and probabilistic methods in actuarial science**. Boca Raton: Chapman & Hall/CRC, 2007. 351 p.

CLARKE, S. R.; DYTE D. S. Using Official Tennis Ratings to Estimate Tournament Chances. In: **Proceedings of the 28th Annual Meeting of the Western Decision Sciences Institute**, 1999, Mexico. Mexico: Massoud M(ed), p.777-779. 1999.

CLARKE, S. R.; DYTE, D. S. A Ratings Based Poisson Model for World Cup Soccer Simulation. **The Journal of Operational Research Society**, Vol.51, n.8, p.993-998, 2000.

DIXON, M. J.; ROBINSON, M. E. A Birth Process Model for Association Football Matches. **The Statistician**, Vol. 47, p. 523-538, 1998.

FIFA/COCA-COLA World Ranking. Disponível em: <<http://www.fifa.com/worldfootball/ranking/index.html>>. Acessado em 03 set. 2009

FIFA World Football: Fixtures and Results. Disponível em: <<http://www.fifa.com/worldfootball/results/index.html>>. Acessado em 03 set. 2009

JARQUE, C. M.; BERA, A. K. "Efficient tests for normality, homoscedasticity and serial independence of regression residuals: Monte Carlo evidence". **Economics Letters** 7 (4): p.313-318. 1981.

MAHER, M. J. Modeling Association Football Scores. **Statistica Neerlandica**, n.36, p.109-118. 1982.

MCCULLAGH, P.; NELDER, J. A. **Generalized Linear Models**, 2nd Edition. London: Chapman and Hall, 1989, 511 p.

PAULA, G. A. **Modelos de Regressão com Apoio Computacional**. IME-USP, 2004.

STEFANI, R. T. Survey of the major world spots rating systems. **Journal of Applied Statistics**, Vol. 24, p.635-646, 1997.

WILLIAMS, D. A. Residuals in generalized linear models. In: **Proceedings of the 12th International Biometrics Conference**, 1984, Tokyo, pp. 59-68.

Apêndice

Apêndice A – Programas e Macros

Macros (para VBA) desenvolvidas para auxílio na extração e compilação dos dados.

```

Sub extract()

Dim i As Integer
Dim n As Integer
Dim par As Double
Dim simp As Integer
Dim spar As Integer
Dim x(10000) As String
Dim z(10000) As String
Dim team(10000, 2) As String
Dim y(10000) As Date
Dim length As Integer

n = 1114
simp = 2
spar = 2

For i = 1 To n
par = (-1) ^ i
If par < 0 Then
Worksheets("Fixtures").Cells(simp, 3) = Trim(Worksheets("Raw Data").Cells(i, 1))
Worksheets("Fixtures").Cells(simp, 4) = (Worksheets("Raw Data").Cells(i, 4))
Worksheets("Fixtures").Cells(simp, 5) = Left(Worksheets("Raw Data").Cells(i, 3), 1)
Worksheets("Fixtures").Cells(simp, 6) = Mid(Worksheets("Raw Data").Cells(i, 3), 3, 1)
simp = simp + 1
Else
Worksheets("Fixtures").Cells(spar, 1) = Worksheets("Raw Data").Cells(i, 1)
Worksheets("Fixtures").Cells(spar, 9) = Worksheets("Raw Data").Cells(i, 2)
spar = spar + 1
End If
Next i

Worksheets("Fixtures").Select
Columns("A:A").Select
Application.CutCopyMode = False
Selection.NumberFormat = "dd/mm/yy;@"

For i = 2 To n
x(i) = Worksheets("Fixtures").Cells(i, 3)
y(i) = Worksheets("Fixtures").Cells(i, 1)

For j = 2 To 2100
If x(i) = Worksheets("Fifa Ratings").Cells(j, 1) And y(i) >= Worksheets("Fifa Ratings").Cells(j, 3)
And y(i) < Worksheets("Fifa Ratings").Cells(j + 155, 3) Then
Worksheets("Fixtures").Cells(i, 7) = Worksheets("Fifa Ratings").Cells(j, 2)
j = 2120
End If
Next j
Next i

```

```

For i = 2 To n
x(i) = Worksheets("Fixtures").Cells(i, 4)
y(i) = Worksheets("Fixtures").Cells(i, 1)
length = Len(x(i)) - 1
z(i) = Left(x(i), length)

For j = 2 To 2100

    If z(i) = (Worksheets("Fifa Ratings").Cells(j, 1)) And y(i) >= Worksheets("Fifa Ratings").Cells(j, 3)
And y(i) < Worksheets("Fifa Ratings").Cells(j + 155, 3) Then
        Worksheets("Fixtures").Cells(i, 8) = Worksheets("Fifa Ratings").Cells(j, 2)
        j = 2120
    End If

Next j
Next i

End Sub

Sub duplicates()

Dim i As Integer
Dim j As Integer
Dim x(10000) As Variant
Dim max As Variant

For i = 2 To 548
max = Cells(i, 12)
For j = 2 To 548
If x(j) = max Then
    Sheets("Fixtures").Select
    Rows(j & ":" & j).Select
    Selection.Delete Shift:=xlUp
j = 2
End If
Next j
Next i

End Sub

Sub finalfix()

Dim i As Integer
Dim n As Integer
Dim x(10000, 10) As Variant
Dim soma As Integer

n = 65

For i = 2 To n
'date
x(i, 1) = Worksheets("World Cup").Cells(i, 2)
'venue
x(i, 2) = Worksheets("World Cup").Cells(i, 3)
'First Team
x(i, 3) = Worksheets("World Cup").Cells(i, 4)

```

```

'Second Team
x(i, 4) = Worksheets("World Cup").Cells(i, 5)
'First Team gols
x(i, 5) = Worksheets("World Cup").Cells(i, 6)
'Second Team gols
x(i, 6) = Worksheets("World Cup").Cells(i, 7)
'First Team Rating
x(i, 7) = Worksheets("World Cup").Cells(i, 8)
'Second Team Rating
x(i, 8) = Worksheets("World Cup").Cells(i, 9)
Next i

soma = 2

For i = 2 To n
'Date
Worksheets("World Cup Final Data").Cells(soma, 1) = x(i, 1)
'Venue
Worksheets("World Cup Final Data").Cells(soma, 2) = x(i, 2)
'First Team
Worksheets("World Cup Final Data").Cells(soma, 3) = x(i, 3)
'Second Team
Worksheets("World Cup Final Data").Cells(soma, 4) = x(i, 4)
'Team Gols (First Team)
Worksheets("World Cup Final Data").Cells(soma, 5) = x(i, 5)
'First Team Rating
Worksheets("World Cup Final Data").Cells(soma, 6) = x(i, 7)
'Second Team Rating
Worksheets("World Cup Final Data").Cells(soma, 7) = x(i, 8)
soma = soma + 1
'Date
Worksheets("World Cup Final Data").Cells(soma, 1) = x(i, 1)
'Venue
Worksheets("World Cup Final Data").Cells(soma, 2) = (-1) * x(i, 2)
'First Team (Reversed Order)
Worksheets("World Cup Final Data").Cells(soma, 3) = x(i, 4)
'Second Team (Reversed Order)
Worksheets("World Cup Final Data").Cells(soma, 4) = x(i, 3)
'Team Gols (First Team)
Worksheets("World Cup Final Data").Cells(soma, 5) = x(i, 6)
'First Team Rating
Worksheets("World Cup Final Data").Cells(soma, 6) = x(i, 8)
'Second Team Rating
Worksheets("World Cup Final Data").Cells(soma, 7) = x(i, 7)
soma = soma + 1
Next i

End Sub

Sub analise_descritiva()

Dim i As Integer
Dim j As Integer
Dim na As Integer
Dim nb As Integer
Dim jogos As Integer
Dim v As Integer
Dim gm As Double
Dim gs As Double

```

```

Dim a As Worksheet
Dim b As Worksheet

Set a = Worksheets("Analise Descritiva")
Set b = Worksheets("World Cup")

na = 33
nb = 151

For i = 2 To na
v = 0
jogos = 0
gm = 0
gs = 0
For j = 2 To nb

If a.Cells(i, 1) = b.Cells(j, 3) Then
gm = gm + b.Cells(j, 5)
gs = gs + b.Cells(j, 6)
a.Cells(i, 2) = b.Cells(j, 7)
jogos = jogos + 1
End If

If b.Cells(j, 5) > b.Cells(j, 6) And a.Cells(i, 1) = b.Cells(j, 3) Then
v = v + 1
End If

If a.Cells(i, 1) = b.Cells(j, 4) Then
gm = gm + b.Cells(j, 6)
gs = gs + b.Cells(j, 5)
jogos = jogos + 1
End If

If b.Cells(j, 5) < b.Cells(j, 6) And a.Cells(i, 1) = b.Cells(j, 4) Then
v = v + 1
End If

Next j
a.Cells(i, 3) = gm
a.Cells(i, 4) = gs
a.Cells(i, 5) = jogos
a.Cells(i, 6) = v
a.Cells(i, 7) = gm / jogos
a.Cells(i, 8) = gs / jogos
a.Cells(i, 9) = v / jogos
Next i

End Sub

```

Programa para o S-PLUS utilizado para realizar a análise de resíduos e para a geração de envelopes e um histograma residual.

```

X _ model.matrix(fit.model)
n _ nrow(X)

```

```

p _ ncol(X)
w _ fit.model$weights
W _ diag(w)
H _ solve(t(X)%*%W%*%X)
H _ sqrt(W)%*%X%*%H%*%t(X)%*%sqrt(W)
h _ diag(H)
td _ resid(fit.model,type="deviance")/sqrt(1-h)

X1 _ model.matrix(fit.model1)

#
e _ matrix(0,n,100)
for(i in 1:100){
nresp _ rpois(n, fitted(fit.model))
fit _ glm(nresp ~ X1, family=poisson)
w _ fit$weights
W _ diag(w)
H _ solve(t(X1)%*%W%*%X1)
H _ sqrt(W)%*%X1%*%H%*%t(X1)%*%sqrt(W)
h _ diag(H)
e[,i] _ sort(resid(fit,type="deviance")/sqrt(1-h)) }
#
e1 _ numeric(n)
e2 _ numeric(n)
#
for(i in 1:n){
eo _ sort(e[,i])
e1[i] _ eo[5]
e2[i] _ eo[95] }
#
med _ apply(e,1,mean)
faixa _ range(td,e1,e2)
par(pty="s ")
qqnorm(td, xlab="Percentis da N(0,1)", ylab="Componente do Desvio
+ Padronizado", ylim=faixa)
par(new=T)
qqnorm(e1,axes=F,xlab="", ylab="", type="l ", ylim=faixa, lty=1)
par(new=T)
qqnorm(e2,axes=F,xlab="", ylab="", type="l ", ylim=faixa, lty=1)
par(new=T)
qqnorm(med,axes=F,xlab="", ylab="", type="l ", ylim=faixa, lty=2)

```

Apêndice B – Modelos ajustados

Regressões de Poisson utilizando como base os resultados das partidas 12 meses anteriores a Copa do Mundo de 2002. O primeiro modelo inclui a constante que é excluído da segunda regressão devido a sua insignificância (considerando $\alpha = 5\%$).

Modelo 1

Dependent Variable: GOLS_MARCADOS

Method: ML/QML - Poisson Count (Newton-Raphson)

Date: 09/11/09 Time: 13:12

Sample: 1 886

Included observations: 886

Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
C	-0.231229	0.279858	-0.826236	0.4087
TEAM_RATING	0.002653	0.000331	8.009639	0.0000
OPPONENT_RATING	-0.002461	0.000274	-8.996666	0.0000
VHOME	0.488929	0.104009	4.700827	0.0000
VNEUTRAL	0.263726	0.090781	2.905075	0.0037
R-squared	0.143366	Mean dependent var		1.247178
Adjusted R-squared	0.139477	S.D. dependent var		1.352490
S.E. of regression	1.254629	Akaike info criterion		2.900614
Sum squared resid	1386.777	Schwarz criterion		2.927627
Log likelihood	-1279.972	Hannan-Quinn criter.		2.910941
Restr. log likelihood	-1366.295	LR statistic		172.6467
Avg. log likelihood	-1.444664	Prob(LR statistic)		0.000000

Modelo 2

Dependent Variable: GOLS_MARCADOS

Method: ML/QML - Poisson Count (Newton-Raphson)

Date: 09/11/09 Time: 13:13

Sample: 1 886

Included observations: 886

Convergence achieved after 4 iterations

Covariance matrix computed using second derivatives

	Coefficient	Std. Error	z-Statistic	Prob.
TEAM_RATING	0.002445	0.000214	11.40900	0.0000
OPPONENT_RATING	-0.002593	0.000220	-11.76425	0.0000
VHOME	0.467997	0.100271	4.667321	0.0000
VNEUTRAL	0.239920	0.085415	2.808888	0.0050
Mean dependent var	1.247178	S.D. dependent var		1.352490
S.E. of regression	1.254924	Akaike info criterion		2.899131
Sum squared resid	1389.004	Schwarz criterion		2.920741
Log likelihood	-1280.315	Hannan-Quinn criter.		2.907393
Avg. log likelihood	-1.445051			

Apêndice C – Gráficos Residuais

