

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO**  
**DOUTORADO EM ENGENHARIA DE PRODUÇÃO**

**CELSO BILYNKIEVYCZ DOS SANTOS**

**PREVISÃO DO ÍNDICE DE DESENVOLVIMENTO HUMANO E DA  
EXPECTATIVA DE VIDA NA AMÉRICA LATINA POR MEIO DE  
TÉCNICAS DE MINERAÇÃO DE DADOS**

**TESE**

**PONTA GROSSA**

**DEZEMBRO - 2016**

**CELSO BILYNKIEVYCZ DOS SANTOS**

**PREVISÃO DO ÍNDICE DE DESENVOLVIMENTO HUMANO E DA  
EXPECTATIVA DE VIDA NA AMÉRICA LATINA POR MEIO DE  
TÉCNICAS DE MINERAÇÃO DE DADOS**

Tese apresentada como requisito parcial à obtenção do título de Doutor em Engenharia de Produção, do Programa de Pós-Graduação em Engenharia de Produção, da Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa.

Orientador: Prof. Luiz Alberto Pilatti, Dr.

Co-orientador: Prof. Bruno Pedroso, Dr.

**PONTA GROSSA**

**DEZEMBRO - 2016**

Ficha catalográfica elaborada pelo Departamento de Biblioteca  
da Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa  
n.14/17

S237 Santos, Celso Bilynkievycz dos

Previsão do índice de desenvolvimento humano e da expectativa de vida na América Latina por meio de técnicas de mineração de dados. / Celso Bilynkievycz dos Santos. -- 2016.

216 f. : il. ; 30 cm.

Orientador: Prof. Dr. Luiz Alberto Pilatti

Coorientador: Prof. Dr. Bruno Pedroso

Tese (Doutorado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção. Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2017.

1. Indicadores sociais. 2. Desenvolvimento social. 3. Qualidade de vida (América Latina). 4. Mineração de dados (Computação). I. Pilatti, Luiz Alberto. II. Pedroso, Bruno. III. Universidade Tecnológica Federal do Paraná. IV. Título.

CDD 670.42



**Universidade Tecnológica Federal do Paraná**  
**Campus Ponta Grossa**  
**Diretoria de Pesquisa e Pós-Graduação**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM**  
**ENGENHARIA DE PRODUÇÃO**



**FOLHA DE APROVAÇÃO**

Título de Tese Nº **6/2016**

**PREVISÃO DO ÍNDICE DE DESENVOLVIMENTO HUMANO E DA  
EXPECTATIVA DE VIDA NA AMÉRICA LATINA POR MEIO DE  
TÉCNICAS DE MINERAÇÃO DE DADOS**

**por**

**Celso Bilynkievycz dos Santos**

Esta Tese foi apresentada às **9 horas de 22 de dezembro de 2016** como requisito parcial para a obtenção do título de DOUTOR EM ENGENHARIA DE PRODUÇÃO, com área de concentração em Gestão Industrial, linha de pesquisa em Gestão de Recursos Humanos para o Ambiente Produtivo, Programa de Pós-Graduação em Engenharia de Produção. O candidato foi argüido pela Banca Examinadora composta pelos professores abaixo citados. Após deliberação, a Banca Examinadora considerou o trabalho aprovado.

**Profa. Dra. Alaine Margarete Guimarães**  
**(UEPG)**

**Profa. Dra. Deborah Ribeiro Carvalho**  
**(PUCPR)**

**Prof. Dr. Antonio Augusto de Paula Xavier**  
**(UTFPR-PG)**

**Prof. Dr. Gerson Ishikawa**  
**(UTFPR-PG)**

**Prof. Dr. Luiz Alberto Pilatti**  
**(UTFPR-PG)**  
**Orientador**

**Prof. Dr. Antonio Carlos de Francisco (UTFPR)**  
**Coordenador do PPGEP**

**A FOLHA DE APROVAÇÃO ASSINADA ENCONTRA-SE NO DEPARTAMENTO DE  
REGISTROS ACADÊMICOS DA UTFPR – CÂMPUS PONTA GROSSA.**



*Dedico esta Tese a minha  
esposa Morgana da Graças Procz dos Santos,  
filhos Bianca Bilynkievycz dos Santos,  
Camila Bilynkievycz dos Santos e  
Rômulo Bilynkievycz dos Santos.*

## **AGRADECIMENTOS**

Antes de tudo, agradeço a Deus por ter me acompanhado nesta jornada.

Agradeço ao meu orientador, Prof. Luiz Alberto Pilatti, e co-orientador, Bruno Pedroso, pela oportunidade, dedicação na condução da orientação deste trabalho e pela amizade.

Agradeço ainda ao meu orientador pela motivação e inspiração para o ingresso no mestrado desde os primeiros contatos na graduação. E ao meu co-orientador pela motivação em prosseguir até o doutorado.

Aos meus familiares, pelo incentivo e pela compreensão nos momentos de ausência.

Ao meu amigo e colega de trabalho, Cristiano Delabernarda, pelo apoio e por ter assumido muitas das minhas responsabilidades funcionais.

À Universidade Estadual de Ponta Grossa (UEPG), pela liberação concedida para cursar o doutorado.

Ao Conselho de Administração de 2012 da UEPG, principalmente ao professor Pró-Reitor de Pesquisa e Pós-Graduação (PROPEP) da UEPG, professor Benjamin de Melo Carvalho, pelos pareceres e debates favoráveis ao processo de liberação para cursar pós-graduação e pelo incentivo à qualificação do quadro de servidores da instituição.

Ao Magnífico Reitor da UEPG, Prof. Carlos Luciano Sant´Ana Vargas, pelo apoio.

Às Pró-Reitores de Recursos Humanos da UEPG, Ana Maria Salles Rosa Solak (2012) e Silviane Buss Tupich (2014), pelo apoio.

Aos servidores Luiz César dos Santos Lima e Tiago Correia Oliveira, da UTFPR, pelo apoio nos trâmites burocráticos.

Aos demais professores do PPGEP da UTFPR, Antônio Vanderley Herrero Sola, Antônio Augusto de Paula Xavier, Antônio Carlos de Francisco, Guataçara dos Santos Junior, João Luiz Kovaleski e Sérgio Mazurek Tebcherani; e à professora

Alaine Margarete Guimarães, do PPG em Computação Aplicada da UEPG, pela dedicação e pelo conhecimento transmitido durante o período de formação.

Aos professores da banca de qualificação e defesa, Luiz Alberto Pilatti, Alaine Margarete Guimarães, Antonio Augusto de Paula Xavier, Bruno Pedroso, Deborah Ribeiro Carvalho e Gerson Ishikawa, os quais ajudaram a nortear a tese.

Aos meus colegas e amigos de doutorado, que sofreram juntos as angústias do processo de pós-graduação, principalmente ao Fábio Sprada de Menezes, Eliana Aparecida Fagundes Queiroz Bortolozo e Dayana Carla de Macedo, pelas parcerias nas pesquisas e pelas palavras de incentivo, nos momentos de dificuldades.

À Universidade Tecnológica Federal do Paraná, por oportunizar esta titulação.

## RESUMO

SANTOS, Celso Bilynkievycz dos. **Previsão do Índice de Desenvolvimento Humano e da Expectativa de Vida na América Latina por Meio de Técnicas de Mineração de Dados**. 2016. 216f. Tese (Doutorado em Engenharia de Produção) - Universidade Tecnológica Federal do Paraná. Ponta Grossa, 2016.

A previsibilidade de indicadores de qualidade de vida pode contribuir na projeção de variáveis dependentes, auxiliar em tomadas de decisões para sustentar ou não políticas públicas e justificar o cenário vivido pelos países e o mundo. **Objetivo:** Prever o Índice de Desenvolvimento Humano (IDH) e a expectativa de vida (EV) nos países latino-americanos no período de 2015 a 2020, utilizando técnicas de Mineração de Dados. **Metodologia:** Foram percorridas as etapas do processo Descoberta de Conhecimento em Base Dados (DCBD). Durante a etapa de DCBD de Mineração de Dados, foi avaliado o desempenho de diferentes algoritmos com paradigma de aprendizado baseados em funções. A partir do algoritmo com melhor desempenho, foram desenvolvidos 748 modelos de previsão univariados e dois multivariados para previsão do IDH de 187 países do mundo e seus resultados, comparados com os últimos relatórios do *United Nations Development Programme* (UNDP), para definição do modelo mais eficiente. Os resultados desses testes de previsões ainda foram comparados com 44 modelos univariados *Autoregressive Integrated Moving Average* (ARIMA). A partir da definição do melhor algoritmo de Mineração de Dados e modelo, fez-se a previsão do IDH e da EV para os países da América Latina para o período de 2015 a 2020. **Resultados:** O algoritmo *SMOReg* e os modelos multivariados apresentaram melhor desempenho nos testes desenvolvidos durante o experimento. As médias de crescimento do IDH e EV previstas para os países latino-americanos tendem a aumentar no período analisado, respectivamente,  $4,99 \pm 3,90$  % e  $2,47 \pm 0,09$  anos. **Conclusão:** Experiências multivariadas possibilitam maior aprendizagem dos algoritmos, aumentando sua precisão. As técnicas de Mineração de Dados apresentaram melhor qualidade nas previsões em relação à técnica mais popular, ARIMA. As previsões sugerem média de crescimento do IDH e EV dos países latino-americanos maiores que a média mundial.

**Palavras chave:** Algoritmos. IDH. Qualidade de Vida. Países.

## ABSTRACT

SANTOS, Celso Bilynkievycz dos. **Prediction of the Human Development Index and Life Expectancy in Latin American using Data Mining Techniques**. 2016. 216f. Thesis (Doctorate in Production Engineering) - University of Technology of Paraná. Ponta Grossa, 2016.

The predictability of quality of life indicators can contribute to the projection of dependent variables, help decision-making processes to support public policies and justify the scenario experienced by the countries and the world. **Aim:** This study aimed to predict the Human Development Index (HDI) and life expectancy (LE) in Latin American countries in the period of 2015–2020 using data mining techniques. **Methodology:** The study followed the steps of Knowledge Discovery in Database (KDD) processes. During the data mining KDD step, the performance of different algorithms with function-based learning paradigms was analyzed. From the algorithm with the best performance, 748 prediction models of univariate and two multivariate were developed to predict the HDI of 187 countries and their results were compared with the last reports from the United Nations Development Program (UNDP) in order to define the most efficient model. The results of these prediction tests were compared with 44 univariate Autoregressive Integrated Moving Average (ARIMA) models. From the definition of the best algorithm of data mining and model, the prediction of HDI and LE for Latin American countries from 2015 and 2020 was done. **Results:** The SMOReg algorithm and the multivariate models presented the best performance in the tests during the experiment. The average growth in HDI and LE predicted for Latin American countries tend to increase in the period analyzed,  $4.99 \pm 3.90$  % and  $2.47 \pm 0.09$  years, respectively. **Conclusion:** Multivariate experiences allow better learning of algorithms, increasing their prediction. Mining data techniques present better quality in the predictions compared to Autoregressive Integrated Moving Average (ARIMA), which is the most popular technique. The predictions suggest an average growth in HDI and LE in Latin American countries compared to the world average.

**Keywords:** Algorithms. HDI. Quality of Life. Counties.

## LISTA DE FIGURAS

Figura 1 - Dimensões mencionadas nos relatórios de desenvolvimento humano da ONU.....	33
Figura 2 - Etapas do processo de construção do modelo de séries temporais usando o método ARIMA.....	45
Figura 3 - Etapas do Processo KDD.....	48
Figura 4 - Etapas de Desenvolvimento da Pesquisa.....	76
Figura 5 - Percentual de Ausência de Dados das Séries Temporais de IDH dos Países.....	77
Figura 6 - Séries Temporais das médias de IDH dos países filiados a ONU excluindo os países: Congo, República Democrática do Congo, Lesoto, Suazilândia e Zimbábue, apresentados separadamente.....	79
Figura 7 - Regressão Linear entre o Tempo (Ano) e o IDH dos Países Filiados a ONU e de países como baixa correlação temporal.....	81
Figura 8 - Correlação entre o Erro Absoluto das Previsões com duas defasagens de tempo e defasagens maiores (2 a 11).....	83
Figura 9 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2014.....	85
Figura 10 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2015.....	86
Figura 11 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2016.....	86
Figura 12 - Erro Absoluto dos Modelos (MMG, MME e MUs) para previsão do IDH 2013.....	93
Figura 13 - Média das DACs do último ponto de observação (2012) por modelos para previsão do IDH 2013.....	94
Figura 14 - Previsão do IDH 2013 x IDH 2013 Real a partir de Séries Temporais Multivariadas.....	95
Figura 15 - MAE acumulado por modelos, resultantes das previsões do IDH 2013-2014 dos países da América Latina.....	98
Figura 16 - Previsões e tendências do IDH de Cuba(a), Venezuela(b), Nicarágua(c), El Salvador(d).....	101
Figura 17 - Séries Temporais Históricas (1980 -2014) e Horizonte de Previsões (2015 a 2020) do IDH Mundial, da América Latina e de Países Latino-americanos com maior (Argentina) e menor (Haiti) índice.....	102
Figura 18 - Séries Temporais Históricas (1980 -2014) e Horizonte de Previsões (2015 a 2020) da EV Mundial, da América Latina e de Países Latino-americanos com maior (Chile) e menor (Haiti) índice.....	104
Figura 19 - DAC dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013..	141

Figura 20 - DAC dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014..	141
Figura 21 - MAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013 .	142
Figura 22 - MAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014 .	142
Figura 23 - MAPE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013.....	143
Figura 24 - MAPE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014.....	143
Figura 25 - MSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013 .	144
Figura 26 - MSE dos Modelos (MMGs, MMEs e MUs) para Previsão do IDH 2014.....	144
Figura 27 - RAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013..	145
Figura 28 - RAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014..	145
Figura 29 - RMSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013.....	146
Figura 30 - RMSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014.....	146
Figura 31 - RRSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013.....	147
Figura 32 - RRSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014.....	147
Figura 33 - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países Desenvolvidos com IDH muito alto .....	149
Figura 34 - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países em Desenvolvimento com IDH alto .....	150
Figura 35 - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países em Desenvolvimento com IDH médio .....	151
Figura 36 - Dados Reais - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países Subdesenvolvidos com IDH baixo ...	152
Figura 37 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 da Argentina (ARG) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	158
Figura 38 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Chile (CHL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	159
Figura 39 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Uruguai (URY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	160
Figura 40 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Panamá (PAN) e suas séries temporais:	

	original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	161
Figura 41	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Cuba (CUB) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	162
Figura 42	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Costa Rica (CRI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	163
Figura 43	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Venezuela (VEN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	164
Figura 44	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do México (MEX) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	165
Figura 45	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Brasil (BRA) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	166
Figura 46	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Peru (PER) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	167
Figura 47	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Equador (ECU) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	168
Figura 48	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Colômbia (COL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	169
Figura 49	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Belize (BLZ) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	170
Figura 50	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do República Dominicana (DOM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	171
Figura 51	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Paraguai (PRY) e suas séries temporais:	



	original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	172
Figura 52	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do El Salvador (SLV) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f).....	173
Figura 53	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Bolívia (BOL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	174
Figura 54	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Guiana (GUY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	175
Figura 55	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Nicarágua (NIC) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	176
Figura 56	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Guatemala (GTM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f).....	177
Figura 57	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Honduras (HND) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	178
Figura 58	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,1,0) para Previsão do IDH 2014 do Haiti (HTI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	179
Figura 59	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 da Argentina (ARG) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	181
Figura 60	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Chile (CHL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	182
Figura 61	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Uruguai (URY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	183
Figura 62	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Panamá (PAN) e suas séries temporais:	

	original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	184
Figura 63 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Cuba (CUB) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	185
Figura 64 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Costa Rica (CRI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	186
Figura 65 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Venezuela (VEN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	187
Figura 66 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do México (MEX) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	188
Figura 67 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Brasil (BRA) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	189
Figura 68 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Peru (PER) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	190
Figura 69 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Equador (ECU) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	191
Figura 70 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Colômbia (COL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	192
Figura 71 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Belize (BLZ) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	193
Figura 72 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do República Dominicana (DOM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	194
Figura 73 -	Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Paraguai (PRY) e suas séries temporais:	

	original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f) .....	195
Figura 74	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do El Salvador (SLV) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f).....	196
Figura 75	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Bolívia (BOL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	197
Figura 76	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Guiana (GUY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	198
Figura 77	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Nicarágua (NIC) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	199
Figura 78	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Guatemala (GTM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f).....	200
Figura 79	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Honduras (HND) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	201
Figura 80	- Correlogramas ACF(a) e PACF(b),do Modelo ARIMA(0,1,0) para Previsão do IDH 2014 do Haiti (HTI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f) .....	202
Figura 81	- IDH da Argentina: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	206
Figura 82	- IHD do Chile: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	206
Figura 83	- IDH do Uruguai: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	207
Figura 84	- IDH do Panamá: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	207
Figura 85	- IDH de Cuba: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	208

Figura 86 - IDH de Costa Rica: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	208
Figura 87 - IDH da Venezuela: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	209
Figura 88 - IDH do México: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	209
Figura 89 - IDH do Brasil: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	210
Figura 90 - IDH do Peru: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	210
Figura 91 - IDH do Equador: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	211
Figura 92 - IDH da Colômbia: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	211
Figura 93 - IDH de Belize: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	212
Figura 94 - IDH da República Dominicana: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	212
Figura 95 - IDH do Paraguai: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	213
Figura 96 - IDH de El Salvador: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	213
Figura 97 - IDH da Bolívia: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	214
Figura 98 - IDH da Guiana: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	214
Figura 99 - IDH da Nicarágua: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	215

Figura 100 - IDH de Guatemala: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	215
Figura 101 - IDH de Honduras: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	216
Figura 102 - IDH do Haiti: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020 .....	216

## LISTA DE TABELAS

Tabela 1 - População, Indicadores de Desenvolvimento Humano dos Países da América Latina e resumo estatístico dos indicadores mundiais e dos países latino-americanos .....	44
Tabela 2 - Estatística Descritiva das Séries Temporais do IDH Mundial no Período de 1980 a 2013 .....	77
Tabela 3 - Estatística de Resumo das Medidas de Qualidade das Previsões de IDH por Ano e Modelo .....	89
Tabela 4 - Correlação entre as Medidas de Qualidade das Previsões e as Medidas Estatísticas das Séries Temporais do Atributo Meta: Número de Elementos, Variância do IDH .....	90
Tabela 5 - MAE dos modelos de previsão desenvolvidos com algoritmo de Mineração de Dados com paradigma de aprendizado de funções .....	96
Tabela 6 - Medidas de qualidade dos modelos desenvolvidos para testar a previsão do IDH dos países da América Latina filiados ao UNDP .....	97
Tabela 7 - Últimos pontos de observações das séries temporais históricas do IDH dos países da América Latina, suas projeções para 2015 a 2020 e resumo estatístico do índice mundial e dos países latino-americanos .....	100
Tabela 8 - Últimos pontos de observações das séries temporais históricas da EV nos países da América Latina e projeções para 2015 a 2020 .....	103
Tabela 9 - Dados Reais (2012 e 2013) e Previsões dos Modelos Multivariados e Univariados para o IDH (2013 e 2014) dos países classificados por classe (Nível e IDH) e ranqueados em relação ao IDH 2013.....	153
Tabela 10 - Previsões do IHD 2013 E 2014 dos países da América Latina por modelos e suas tendências divulgadas pela UNDP.....	204

## LISTA DE QUADROS

Quadro 1 - Temas dos Relatórios do PNUD .....	32
Quadro 2 - Relação de Dependência entre as fórmulas utilizadas para o cálculo do IDH.....	38
Quadro 3 - Temas dos Relatórios do PNUD .....	39
Quadro 4 - Tarefas, Principais Técnicas e Algoritmos de Mineração de Dados mais Utilizados.....	52
Quadro 5 - Expressão de Busca de Palavras-chave em periódicos.....	64
Quadro 6 - Processo de Descoberta de Conhecimento em Bases de Dados .....	71
Quadro 7 - Séries Temporais Utilizadas.....	72
Quadro 8 - Configuração da Aplicação de Interface de Programação Forecast .....	84
Quadro 9 - Modelo Univariado de Previsão do IDH 2014 para o País Botswana....	137
Quadro 10 - Modelo Multivariado Específico de Previsão do IDH 2014 para o País New Zealand.....	138
Quadro 11 - Modelo Multivariado Global de Previsão do IDH 2014 para os Países filiados a ONU.....	139

## LISTA DE EQUAÇÕES

Equação 1 - Índice de Desenvolvimento Humano.....	38
Equação 2 - Expectativa de Vida ao Nascer .....	38
Equação 3 - Índice de Renda .....	38
Equação 4 - Índice de Educação.....	38
Equação 5 - Índice de Anos Médio de Estudo.....	38
Equação 6 - Índice de Anos Esperados de Escolaridade.....	38
Equação 7 - Índice de Industrialização.....	40
Equação 8 - Erro absoluto médio (MAE).....	73
Equação 9 - Erro quadrático médio (MSE).....	73
Equação 10 - Raiz quadrada do erro quadrático médio (RMSE) .....	73
Equação 11 - Erro percentual absoluto médio (MAPE).....	74
Equação 12 - Precisão direcional (DAC).....	74
Equação 13 - Erro absoluto relativo (RAE) .....	74
Equação 14 - Raiz quadrada do erro quadrático relativo (RRSE).....	74



## LISTA DE ABREVIATURAS E SIGLAS

API	- <i>Application Programming Interface</i>
ARFF	- <i>Attribute-Relation File Format</i>
DCBD	- <i>Descoberta de Conhecimento em Base de Dados</i>
EV	- <i>Expectativa de Vida</i>
HDI	- <i>Human Development Index</i>
IDH	- <i>Índice de Desenvolvimento Humano</i>
KDD	- <i>Knowledge Discovery in Databases</i>
LMS	- <i>Least Median Squared</i>
MD	- <i>Mineração de Dados</i>
ML	- <i>Machine Learning</i>
MLP	- <i>Multi-Layer Perceptron</i>
MM	- <i>Modelo Multivariado</i>
MME	- <i>Modelo Multivariado Específico</i>
MMG	- <i>Modelo Multivariado Global</i>
MSE	- <i>Mean squared error</i>
MTS	- <i>Multivariate Time Series</i>
MU	- <i>Modelo Univariado</i>
QV	- <i>Qualidade de Vida</i>
RBFNetwork	- <i>Radial Basis Function (RBF) Network</i>
RMSE	- <i>Root mean squared error</i>
RRSE	- <i>Root relative squared error</i>
SMO	- <i>Sequential Minimal Optimization</i>
SMOReg	- <i>Sequential Minimal Optimization for Regression</i>
SPSS	- <i>Statistical Package for the Social Sciences</i>
SQL	- <i>Structured Query Language</i>
SVM	- <i>Support Vector Machines</i>
SVR	- <i>Support Vector Regression</i>

## LISTA DE ACRÔNIMOS

ARIMA	- <i>Auto Regressive Integrated Moving Average</i>
CIA	- <i>Akaike Information Criterion</i>
DAC	- <i>Direction accuracy</i>
FAC	- <i>Função de Autocorrelação</i>
FACP	- <i>Função de Autocorrelação Parcial</i>
MAE	- <i>Mean absolute error</i>
NICs	- <i>Newly Industrialized Countries</i>
PNUD	- <i>Programa das Nações Unidas para o Desenvolvimento</i>
RAE	- <i>Relative absolute error</i>
UNDP	- <i>United Nations Development Programme</i>
UNIDO	- <i>United Nations Industrial Development Organization</i>
WEKA	- <i>Waikato Environment for Knowledge Analysis</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>25</b>
1.1 DEFINIÇÃO DO OBJETO .....	25
1.2 PROBLEMA DE PESQUISA .....	27
1.3 OBJETIVOS .....	28
1.3.1 Objetivo Geral .....	28
1.3.2 Objetivos Específicos .....	28
1.4 JUSTIFICATIVA .....	28
1.5 DISTRIBUIÇÃO DO TRABALHO .....	29
<b>2 REVISÃO DE BIBLIOGRAFIA</b> .....	<b>30</b>
2.1 REFERENCIAL TEÓRICO .....	30
2.1.1 Índice de Desenvolvimento Humano (IDH) .....	30
2.1.2 Expectativa de Vida ao Nascer .....	34
2.1.2.1 Índice de escolaridade (IE) .....	35
2.1.2.2 Rendimento nacional bruto (RNB) per capita .....	35
2.1.2.3 Cálculo do IDH .....	36
2.1.3 IDH dos Países da América Latina .....	41
2.1.4 Método de Previsão ARIMA .....	45
2.1.5 Processo de Descoberta de Conhecimento em Base de Dados .....	47
2.1.6 Mineração de Dados .....	50
2.1.6.1 Objetivos, tarefas e técnicas e algoritmos de mineração de dados .....	51
2.1.6.2 Algoritmos de Mineração de Dados para Previsão .....	53
2.1.6.3 Ambiente para análise de conhecimento Waikato (WEKA) .....	56
2.2 ESTADO DA ARTE .....	57
2.2.1 Índice de Desenvolvimento Humano e Expectativa de Vida .....	57
2.2.2 Mineração de Dados .....	60
2.2.3 Técnicas de Data Mining de Previsão .....	61
2.2.4 Originalidade do Tema de Pesquisa .....	63
<b>3 METODOLOGIA</b> .....	<b>65</b>
3.1 DESENHO DA PESQUISA .....	65
3.2 INSTRUMENTO, FONTE DE PESQUISA E INTERVENÇÃO EM AMBIENTES SOCIAIS E INSTITUCIONAIS .....	66
3.3 DADOS DA PESQUISA .....	66
3.4 FORMULAÇÃO DE HIPÓTESES .....	67

3.5 MATERIAIS E MÉTODOS.....	68
3.5.1 Medidas de Qualidade para Previsões.....	73
3.6 EXPERIMENTO: PREVISÃO DO IDH E DA EV DOS PAÍSES POR MEIO DE TÉCNICAS DE MINERAÇÃO DE DADOS.....	75
3.6.1 Pré-Processamento.....	75
3.6.2 Mineração de Dados .....	82
3.6.2.1 Seleção do algoritmo.....	84
3.6.3 Pós-Processamento .....	87
<b>4 RESULTADOS.....</b>	<b>96</b>
<b>5 DISCUSSÃO .....</b>	<b>105</b>
5.1 QUALIDADE DOS DADOS .....	106
5.2 MUDANÇA NA METODOLOGIA DE CÁLCULO DO IDH .....	107
5.3 ÍNDICE DE DESENVOLVIMENTO HUMANO.....	108
5.4 PREVISÃO DO IDH .....	108
5.5 PREVISÃO DA EXPECTATIVA DE VIDA .....	110
5.6 PROCESSO DE DCBD E MINERAÇÃO DE DADOS .....	112
5.7 CUSTOS OPERACIONAIS .....	114
5.8 MEDIDAS DE QUALIDADE DAS PREVISÕES .....	115
5.9 EFICIÊNCIA DOS MODELOS.....	116
<b>6 CONCLUSÃO .....</b>	<b>118</b>
<b>REFERÊNCIAS.....</b>	<b>120</b>
<b>APÊNDICE 1 – Exemplos de Modelos de Previsão .....</b>	<b>136</b>
<b>APÊNDICE 2 – Medidas de Qualidade das Previsões dos Modelos .....</b>	<b>140</b>
<b>APÊNDICE 3 – Resultados dos Testes Preliminares para Escolha do Modelo de Previsão desenvolvido pelo Algoritmo <i>SMOreg</i>.....</b>	<b>148</b>
<b>APÊNDICE 4 – Correlogramas e Séries Temporais Observadas para a Modelagem dos Modelos ARIMA para Previsão do IDH 2013 dos Países da América Latina.....</b>	<b>157</b>
<b>APÊNDICE 5 – Correlogramas e Séries Temporais Observadas para a Modelagem dos Modelos ARIMA para Previsão do IDH 2014 dos Países da América Latina.....</b>	<b>180</b>
<b>APÊNDICE 6 – Comparação das Previsões Realizadas com Metodos de Mineração de Dados e ARIMA.....</b>	<b>203</b>
<b>APÊNDICE 7 – IDH dos Países da América Latina: Testes de Previsões por Modelos (MMG, MME E MU) para 2013 e 2014, Tendência por Relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020.....</b>	<b>205</b>

# 1 INTRODUÇÃO

## 1.1 DEFINIÇÃO DO OBJETO

A Qualidade de Vida (QV) é um tema de estudo contemporâneo discutido em diferentes segmentos da ciência. Entre estes, a Engenharia de Produção. Pode ser avaliada por diferentes instrumentos ou indicadores sugeridos pela literatura para diferentes contextos.

Partindo da premissa de que a QV deve ser observada de forma global (FOX-RUSHBY e PARKER, 1995), busca-se neste trabalho adotar um indicador mundialmente utilizado.

Nesta perspectiva, procuraram-se indicadores que melhor avaliassem a QV de forma global, adotando-se, então, como objetos de estudo, o Índice de Desenvolvimento Humano (IDH) e a expectativa de vida (EV), considerados pela Organização Nações Unidas (ONU) (UNDP, 1990; ALKIRE, 2010; UNDP, 2014), respectivamente, como o seu indicador de desenvolvimento humano e a sua sub-componente do IDH mais associada à QV.

A ONU, por meio do *United Nations Development Programme* (UNDP), adota o IDH para avaliar os avanços no desenvolvimento econômico e na QV da população nos países membros (UNDP, 1990; ALKIRE, 2010; UNDP, 2014) a partir da média geométrica entre indicadores de educação, saúde e renda (KOVACEVIC, 2010).

Os resultados de estudos que utilizam o IDH geralmente têm como propósito orientar políticas públicas para a melhoria da QV da população de uma determinada região demográfica, como sugere o estudo de Ferreira (2012) e estudos na área de Engenharia de Produção (LIMA, 2006; MARIANO, 2012).

Por apresentar dependência da componente "saúde", medida por uma vida longa e saudável, o índice é também amplamente utilizado em pesquisas da área de saúde (GONZÁLEZ-ZAPATA *et al.*, 2011; KARIMINIA *et al.*, 2011; MARTINEZ *et al.*, 2011; CASTRO e RODRIGUES-JÚNIOR, 2012; SZUSTER *et al.*, 2012; TAVARES *et*

*al.*, 2014b; PERCIO, MEDINA e LUNA, 2015; SADOVSKY *et al.*, 2015), enquanto sua componente decomposta mais utilizada neste tipo de estudos é a EV (BRUNEKREEF *et al.*, 2012; MENDES *et al.*, 2012; CAMPOLINA *et al.*, 2013; CHIAVEGATTO FILHO e LAURENTI, 2013; KANSO *et al.*, 2013; MACKENBACH e LOOMAN, 2013; ANSTEY *et al.*, 2014; CERVANTES e BOTERO, 2014; MORTON, DE STAVOLA e LEON, 2014; CAMARGOS e GONZAGA, 2015; HENDI, 2015; LACEY *et al.*, 2015; LI *et al.*, 2015; STRINGHINI *et al.*, 2015), que também se encontra entre os indicadores mais utilizados para avaliar a QV e o desenvolvimento socioeconômico de um país.

A literatura apresenta um quantitativo muito grande de pesquisas recentes (GONZÁLEZ-ZAPATA *et al.*, 2011; KARIMINIA *et al.*, 2011; MARTINEZ *et al.*, 2011; BRUNEKREEF *et al.*, 2012; CASTRO e RODRIGUES-JÚNIOR, 2012; MENDES *et al.*, 2012; SZUSTER *et al.*, 2012; CAMPOLINA *et al.*, 2013; CHIAVEGATTO FILHO e LAURENTI, 2013; KANSO *et al.*, 2013; MACKENBACH e LOOMAN, 2013; ANSTEY *et al.*, 2014; CERVANTES e BOTERO, 2014; MORTON, DE STAVOLA e LEON, 2014; TAVARES *et al.*, 2014b; CAMARGOS e GONZAGA, 2015; HENDI, 2015; LACEY *et al.*, 2015; LI *et al.*, 2015; PERCIO, MEDINA e LUNA, 2015; SADOVSKY *et al.*, 2015; STRINGHINI *et al.*, 2015), que colocam a EV ou o IDH como variáveis balizadoras de estudos de QV e desenvolvimento humano.

A previsibilidade do IDH ou de suas componentes pode auxiliar em tomadas de decisões governamentais, apoiando ou não políticas públicas, caso os valores reais correspondam às expectativas das previsões. Suas projeções também podem ser utilizadas em estudos prospectivos de diferentes áreas, entre elas, a da Engenharia de Produção, para explicar o comportamento futuro de variáveis dependentes.

A Mineração de Dados (MD) também foi definida como objeto deste estudo por se tratar de uma área da Inteligência Artificial (IA) que permite descobrir conhecimento potencialmente implícito em bases de dados e que pode contribuir na previsão destes indicadores, auxiliando em tomadas de decisões dos gestores de QV e desenvolvimento humano.

Segundo Fayyad (1998), o ser humano não está preparado para interpretar

grandes volumes de dados ou espaços multidimensionais. Nesta conjuntura, a Tecnologia da Informação surge como a ferramenta necessária para exercer esta função, mais especificamente a IA, com suas técnicas de Mineração de Dados.

A função da Mineração de Dados é garimpar informações ocultas, geralmente em grandes bases de dados, ampliando ao máximo as comparações entre os dados e tornando explícito o conhecimento implícito.

As técnicas tradicionais de análise de dados exigem que, inicialmente, hipóteses sejam estabelecidas, para que, então, sejam construídas consultas, a fim de comprovar ou não estas suposições, enquanto as técnicas de Mineração de Dados têm maior independência do estabelecimento de hipóteses iniciais.

## 1.2 PROBLEMA DE PESQUISA

A literatura oferece uma variedade de técnicas de previsão, dentre as quais se destacam as previsões a partir de técnicas de Mineração de Dados aplicadas em diferentes áreas (XIE *et al.*; MANGALOVA e AGAFONOV, 2014; CORREA *et al.*, 2015; WINTERS-MINER *et al.*, 2015). No entanto, não foi possível identificar estudos de previsões do IDH ou a EV dos países da América Latina, principalmente por meio de técnicas de Mineração de Dados.

Diante desta abertura na literatura, se parte da pergunta norteadora da presente pesquisa:

- Como prever o Índice de Desenvolvimento Humano e a Expectativa de Vida por meio de técnicas de Mineração de Dados?

## 1.3 OBJETIVOS

### 1.3.1 Objetivo Geral

O presente estudo tem como objetivo prever o IDH e a EV nos países latino-americanos para o período de 2015 a 2020, a partir de seus dados históricos, utilizando técnicas de Mineração de Dados.

### 1.3.2 Objetivos Específicos

1. Avaliar o desempenho de algoritmos de Mineração de Dados com paradigma de aprendizagem de funções;
2. Avaliar o desempenho de modelos de previsão multivariados e univariados;
3. Comparar o desempenho dos modelos desenvolvidos com algoritmo de Mineração de Dados que apresentou melhor desempenho, com o modelo desenvolvido através do método popularmente conhecido: Modelo Auto-Regressivo Integrado de Média Móvel (ARIMA)<sup>1</sup>;
4. Definir uma unidade de medida de qualidade para avaliar previsões de IDH e EV;
5. Comparar o IDH e a EV da América Latina e o mundo.

## 1.4 JUSTIFICATIVA

As previsões seguem tendências matemáticas baseadas em padrões e comportamentos comuns ou específicos de cada país, de acordo com o modelo adotado. Então, a ocorrência de mudanças econômicas ou políticas que fujam aos padrões e à tendência implícita nos modelos pode interferir nas previsões. No entanto, seus resultados podem auxiliar outros tipos de leituras, principalmente quando o real não for o esperado. As previsões com tendências de crescimento do

---

<sup>1</sup> ARIMA - do Inglês - *Autoregressive Integrated Moving Average*



IDH ou da EV, por exemplo, podem ser entendidas como parâmetros mínimos a serem superados pelos países. Possíveis contradições de valores entre o predito e o real podem justificar o êxito ou o insucesso de políticas públicas ou econômicas, como, também, podem ser consequência da interferência de outros fatores, como catástrofes naturais, mudanças climáticas, conflitos armados ou guerras. Esses resultados podem contribuir em reflexões para o planejamento e a gestão da promoção de QV e desenvolvimento humano.

O presente estudo contribui na projeção destes indicadores utilizados em pesquisas de QV e, também, contribui na avaliação de algoritmos, modelos empregados, a partir das comparações entre as previsões e as tendências divulgadas pelo UNDP para o IDH, em períodos anteriores às projeções e entre suas medidas de qualidade de previsões.

Os procedimentos metodológicos aplicados neste estudo podem ser utilizados para solução de outros problemas de previsão da engenharia de produção.

## 1.5 DISTRIBUIÇÃO DO TRABALHO

Este trabalho está dividido em seis capítulos, incluindo a Introdução, apresentada na seção anterior. O Capítulo 2 apresenta a revisão de bibliografia, que está subdividida em dois segmentos. No primeiro apresenta o referencial teórico e no segundo o estado da arte. Os fundamentos metodológicos são apresentados no Capítulo 3, bem como os materiais e métodos e a descrição do Processo de Descoberta de Conhecimento em Base de Dados (DCBD), utilizado para o desenvolvimento do experimento. No Capítulo 4, são apresentados os resultados do DCBD. A discussão se faz no Capítulo 5. E, finalmente, no Capítulo 6 concluir-se-á o trabalho.

## 2 REVISÃO DE BIBLIOGRAFIA

Neste capítulo, será apresentado o referencial teórico básico, necessário para a compreensão do tema de pesquisa proposto e o estado da arte.

### 2.1 REFERENCIAL TEÓRICO

#### 2.1.1 Índice de Desenvolvimento Humano (IDH)

A necessidade da criação de um indicador de desenvolvimento humano surgiu a partir da crítica de que o desenvolvimento de um país não é apenas a mera expansão do seu crescimento econômico. As abordagens anteriores de desenvolvimento de um país se concentravam quase que exclusivamente na eficiência e no crescimento (ALKIRE, 2010).

Um novo conceito de desenvolvimento humano propôs uma mudança no sentido de sua perspectiva, que passou a basear-se no ser humano, na justiça de distribuição dos benefícios econômicos, como o bem-estar, a liberdade e a QV (UNDP, 1990; NUSSBAUM e SEN, 1993; ALKIRE, 2010).

Desde os anos 1970, essa percepção entre os profissionais de desenvolvimento humano cresceu, observando-se que os esforços investidos na industrialização e no crescimento econômico não conduziram a uma redução considerável da pobreza e das desigualdades nos países em desenvolvimento, haja vista que somente os desenvolvimentos econômicos não conseguiram atender às necessidades básicas das camadas mais pobres da população, como água, eletricidade, saúde e educação.

Observou-se também que, em algumas áreas, os indicadores sociais pioraram, enquanto o PIB global tem apresentado taxas de crescimento significativas.

Em 1990, um grupo de economistas formado por Mahbub ul Haq, Amartya Sen, Paul Streeten e Keith Griffin criou o IDH (UNDP, 1990). Este indicador combina o PIB per capto com níveis de educação e EV de uma população para delinear um conjunto mais amplo e mais abrangente de desenvolvimento, com foco nas escolhas sociais e QV.

O desenvolvimento humano é definido como um processo de ampliação das escolhas das pessoas, de reforçar as capacidades humanas (a gama de coisas que as pessoas podem fazer) e de liberdade, permitindo-lhes viver uma vida longa e saudável, ter acesso ao conhecimento e um padrão de vida decente, além de participar a vida de sua comunidade e as decisões que afetam suas vidas (UNDP, 1990).

O IDH é um índice composto que mede o desempenho em três dimensões básicas do desenvolvimento humano: uma vida longa e saudável, conhecimento e um padrão decente de vida. O primeiro se mede pela EV, o segundo pela média de anos de escolaridade e o último pelo rendimento per capto.

Segundo a UNDP (2014), o IDH é utilizado no planejamento e nas tomadas de decisões para orientação e defesa de políticas de desenvolvimento do índice.

A ONU produziu, ao longo de 25 anos, mais de 700 relatórios. Nos relatórios dos últimos 20 anos, a avaliação do IDH tem proporcionado alternativa para agentes de desenvolvimento promoverem o crescimento econômico de sua região. Seus relatórios explicam como e por que as decisões políticas devem orientar os processos do crescimento econômico, industrialização e prestação de serviços com liberdade humana (ALKIRE, 2010), e apresentam diferentes temáticas, para contribuir no desenvolvimento dos países (Quadro 1).

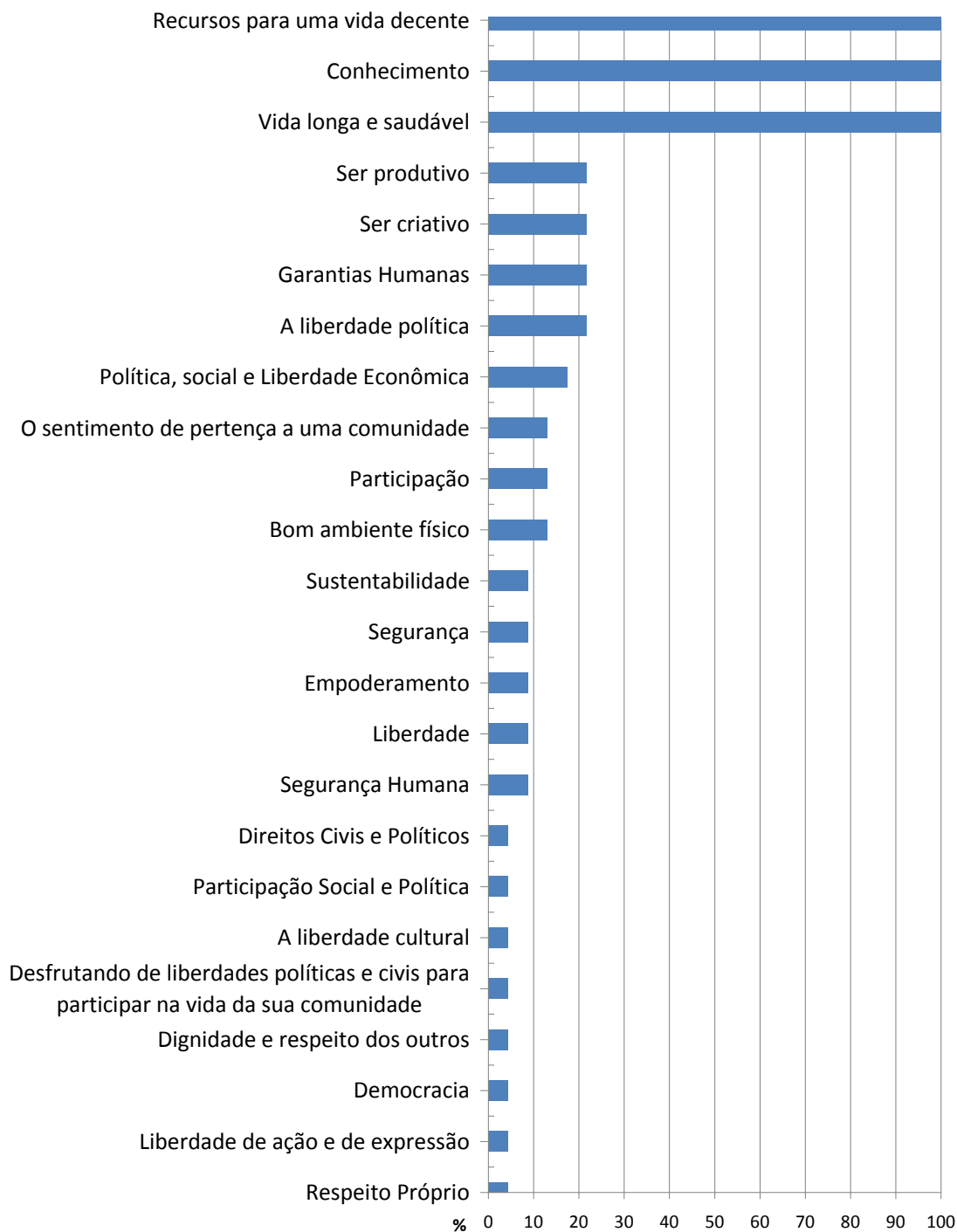
**Quadro 1 - Temas dos Relatórios do PNUD**

<b>Relatório (Ano)</b>	<b>Tema</b>	<b>Fonte</b>
2015	O Trabalho como Motor do Desenvolvimento Humano	(UNDP, 2015)
2014	Sustentar o Progresso Humano: Reduzir as Vulnerabilidades e Reforçar a Resiliência	(UNDP, 2014)
2013	A Ascensão do Sul: o Progresso Humano num Mundo Diversificado	(UNDP, 2013)
2011/2012	Futuro Sustentável que Queremos	(UNDP, 2011-2014)
2011	Sustentabilidade e Equidade: Um Futuro Melhor para Todos	(UNDP, 2011)
2010	A Verdadeira Riqueza das Nações: Vias para o Desenvolvimento Humano	(UNDP, 2010)
2009	Ultrapassar Barreiras: Mobilidade e Desenvolvimento Humanos	(UNDP, 2009)
2007/2008	Combater as Alterações Climáticas: Solidariedade Humana num Mundo Dividido	(UNDP, 2007/2008)
2006	A Água para lá da Escassez: Poder, Pobreza e a Crise Mundial da Água	(UNDP, 2006)
2005	Cooperação Internacional numa Encruzilhada: Ajuda, Comércio e Segurança num Mundo Desigual	(UNDP, 2005)
2004	Liberdade Cultural num Mundo Diversificado	(UNDP, 2004)
2003	Objetivos de Desenvolvimento do Milênio: Um Pacto Entre Nações para Eliminar a Pobreza Humana	(UNDP, 2003)
2002	Aprofundar a Democracia num Mundo Fragmentado	(UNDP, 2002)
2001	Fazer as Novas Tecnologias Trabalharem para o Desenvolvimento Humano	(UNDP, 2001)
2000	Direitos Humanos e Desenvolvimento Humano	(UNDP, 2000)
1999	Globalização com Uma Face Humana	(UNDP, 1999)
1998	Padrões de Consumo para o Desenvolvimento Humano	(UNDP, 1998)
1997	Desenvolvimento Humano para Erradicar a Pobreza	(UNDP, 1997)
1996	Crescimento Econômico e Desenvolvimento Humano	(UNDP, 1996)
1995	Gênero e Desenvolvimento Humano	(UNDP, 1995)
1994	Novas Dimensões da Segurança Humana	(UNDP, 1994)
1993	Participação das Pessoas	(UNDP, 1993)
1992	Dimensões Globais do Desenvolvimento Humano	(UNDP, 1992)
1991	Financiamento do Desenvolvimento Humano	(UNDP, 1991)
1990	Conceito e Medição do Desenvolvimento Humano	(UNDP, 1990)

O último relatório a UNDP (2015) debate o tema "O Trabalho como Motor do Desenvolvimento Humano", destacando que o trabalho é fundamental para a geração de riqueza das economias e das pessoas, mas deve ser conceituado principalmente em termos de desenvolvimento humano e não apenas econômico.

Alkire (2010) apresenta as dimensões de IDH mencionadas nos relatórios da ONU, desde a sua primeira publicação, em 1990, até 2010 (Figura 1).

**Figura 1 - Dimensões mencionadas nos relatórios de desenvolvimento humano da ONU**



Fonte: O autor, a partir dos dados de Alkire (2010).

As três dimensões fundamentais do IDH são discutidas em todos os relatórios da UDNP (Figura 1). Outras variáveis intervenientes são apresentadas em diferentes contextos históricos, culturais, nacionais e internacionais, merecedores de destaques em seus relatórios.

O desenvolvimento humano também abrange outras liberdades de valor, associadas ao bem-estar humano em países em desenvolvimento e industrializados. A ênfase e as particularidades variam, mas geralmente incluem a subsistência segura e significativa; relações de cuidado e dignidade, proteção contra crime e violência; atividades artísticas, culturais e espirituais, participação em atividades políticas e comunitárias; respeito por si próprio e bem-estar emocional (ALKIRE, 2010).

O penúltimo relatório da UNDP (2014) apresenta alterações nos limites máximos e mínimos utilizados na normalização das componentes e mudança na agregação de indicadores de educação.

### 2.1.2 Expectativa de Vida ao Nascer

A expectativa de vida ao nascer representa o número de anos que um recém-nascido poderá viver, se os padrões das taxas de mortalidade específicas por idade vigente, no momento do seu nascimento, permanecerem os mesmos durante toda sua vida.

Os valores máximos e mínimos utilizados como limites para o cálculo da EV adotado pela ONU, em seu programa de desenvolvimento, tiveram seus valores, em anos, reduzidos (UNDP, 2010).

O valor mínimo foi fixado em 20 anos, com base em Riley (2005) e Maddison (2010), que em seus estudos apresentam evidências históricas de que nenhum país do século XX teve EV menor do que 20 anos. E o valor máximo foi fixado em 85 anos (UNDP, 2014).

A EV é utilizada para obtenção do Índice de Saúde (IS) a partir do cálculo de normalização do valor, utilizando como limites os valores máximo e mínimo apresentados para variável.

#### 2.1.2.1 Índice de escolaridade (IE)

O índice de escolaridade é a média aritmética entre as subcomponentes índice de anos médio de estudo (IAME) e índice de anos esperados de escolaridade (IAEE).

O primeiro é o número de anos de educação, normalizado, recebido por uma pessoa adulta com idade acima de 25 anos. Para a normalização, fixou-se o valor máximo de 15 anos, correspondente ao máximo projetado para o ano de 2025, e o valor mínimo zero, pois uma sociedade pode subsistir sem educação formal.

O segundo é o índice de anos esperados de escolaridade, normalizado, que uma criança, ao entrar na idade escolar, pode esperar receber se os padrões prevalentes de taxas de matrículas específicas por idade persistirem por toda sua vida. O valor máximo foi fixado em 18 anos de escolaridade esperada, equivalente à obtenção do grau de mestrado na maioria dos países. Em alguns casos, o ano de escolaridade esperado é estimado (UNDP, 2014).

#### 2.1.2.2 Rendimento nacional bruto (RNB) per capita

O rendimento nacional bruto é o índice de rendimento nacional bruto per capita anual normalizado e com paridade do poder aquisitivo internacional.

Para paridade do poder aquisitivo, o Banco Mundial – por meio do seu Programa de Comparação Internacional (IPC) – recolhe periodicamente milhares de preços de bens e serviços em muitos países para gerar taxas de conversão oficiais de paridade de poder de compras (PPC) entre os países. A partir desta base de

dados, é definida uma moeda internacional padrão, o dólar PPC (PPC \$), que melhor reflete o poder aquisitivo em diferentes países.

Para a normalização do índice, adotou-se, em 2014, o valor máximo de 75,000 dólares per capita, pois segundo Kahneman e Deaton (2010), acima deste valor não ocorrem ganhos no desenvolvimento humano e bem-estar. E como valor mínimo definiu-se o valor de 100 dólares de rendimento per capita.

### 2.1.2.3 Cálculo do IDH

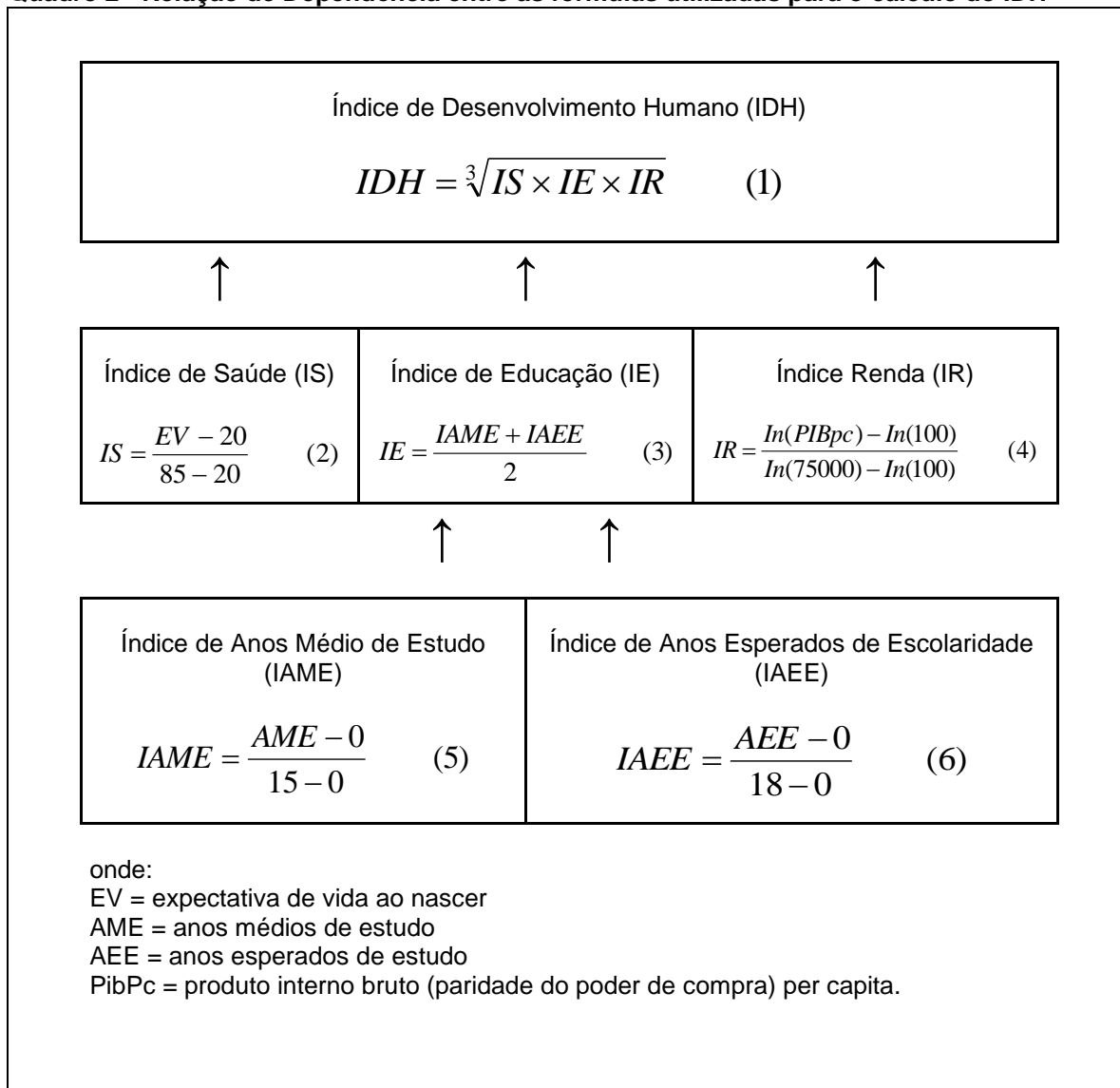
Em 2010, a metodologia de cálculo do índice sofreu alterações. Substituiu-se a média ponderada de suas componentes pela média geométrica. No entanto, a nova fórmula foi criticada por Ravallion (2012b), principalmente porque não permite indicar onde se deve investir e que certos países serão prejudicados com a nova fórmula. Em seu estudo, também demonstra que países com baixo IDH e baixa EV não poderão visualizar mudanças significativas no desenvolvimento do índice com aumento da EV, que deveria ser entendida como uma faceta que merece maior ponderação.

Apesar das críticas sobre o novo tipo de cálculo do IDH, observa-se que a média geométrica privilegia países com menor desigualdade entre seus componentes, pois a média geométrica tende a ser cada vez menor que a média aritmética, à medida que se aumenta a variância entre as componentes.

A metodologia de cálculo padroniza pesos iguais em todas as dimensões, considerando justificativas estatísticas de estudos como o de Noorbakhsh (1998) e Decanq e Lugo (2009), pois consideram que todas as dimensões são igualmente importantes. A média geométrica reduz o nível de substituíbilidade entre as dimensões, em virtude de que o baixo desempenho em uma determinada dimensão não pode ser mais compensado pelo maior desempenho em outra dimensão (UNDP, 2014).



O Quadro 2 apresenta as fórmulas para o cálculo do IDH, suas relações de dependências e os novos valores limites divulgados no último relatório do UNDP (2014).

**Quadro 2 - Relação de Dependência entre as fórmulas utilizadas para o cálculo do IDH**

**Fonte: Adaptado de UNDP (2014).**

Os valores máximos e mínimos utilizados como limites para o cálculo da EV, adotado pela Organização das Nações Unidas (ONU) em seu programa de desenvolvimento, tiveram os anos atualizados em relação aos modelos matemáticos anteriores.

Em seu estudo, Kovacevic (2010) apresenta diferentes limites utilizados ao longo do tempo na normalização das componentes do IDH.

O Programa das Nações Unidas (UNDP) consulta diferentes agências internacionais para obtenção de dados referentes à expectativa de vida ao nascer (UNDESA, 2013); média de anos de escolaridade (BARRO e LEE, 2013; UNESCO, 2013b; a); anos de escolaridade esperados (UNESCO, 2013b), e PIB per capita

(UNDESA, 2013; FMI, 2014; UNSD, 2014; WORLDBANK, 2014). Essas agências recebem dados de agências nacionais dos países filiados.

Alguns valores de IDH ou de seus componentes são estimados ou baseados em estudos de regressão do crescimento dos países ou são referentes aos anos anteriores mais recentes (UNDP, 2013). O índice apresenta uma robustez, com possibilidade de aplicação de metodologias alternativas (AGUNA e KOVACEVIC, 2010). Em termos estatísticos, a diferença do IDH não é significativa a partir da quarta casa decimal (AGUNA e KOVACEVIC, 2010; HØYLAND, MOENE e WILLUMSEN, 2012).

O IDH permite a classificação dos países por grau de desenvolvimento humano e classe de IDH a partir dos intervalos do índice apresentados no Quadro 3.

**Quadro 3 - Temas dos Relatórios do PNUD**

Nível de Desenvolvimento do País	IDH	
	Classe	Índice
Desenvolvido	IDH muito elevado	> 0,799
Em desenvolvimento	IDH elevado	> 0,699 e < 0,800
	IDH médio	> 0,550 e < 0,700
Subdesenvolvido	IDH baixo	< 0,550

**Fonte: O autor (2016).**

O índice é uma medida universal e resumida que, apesar de ampliar a perspectiva sobre o desenvolvimento humano, não contempla e nem esgota todas as facetas de desenvolvimento (UNDP, 2014).

Embora o IDH seja uma variável dependente, das três dimensões, o índice ainda apresenta uma relação implícita como a industrialização e desindustrialização. Estas podem ser medidas através do índice de industrialização, que é a relação entre duas dimensões: emprego e população, cujo cálculo se dá a partir da seguinte fórmula (UNIDO, 2013):

$$II = \frac{\frac{E_u}{P_u}}{\frac{E_t}{P_t}} \quad (7)$$

Onde:

II é o índice de industrialização.

E é o número de empregos.

P é a população.

U se refere à unidade.

T se refere ao total de unidades.

Pesquisas como de Murphy, Shleifer e Vishny (1988) já apontavam que praticamente todos os países desenvolvidos passaram por um processo de industrialização.

Storbacka (2011) apresenta algumas soluções para os processos de industrialização e comercialização em diferentes cenários que envolvem identificação de capacidade de negócio e práticas de gestão.

Os países de maior poder econômico impõem pressões de concorrência que podem abafar a diversificação econômica e a industrialização nas economias menores, influenciando no seu desenvolvimento humano (UNIDO, 2013).

O processo de industrialização depende de competências industriais que podem ser desenvolvidas neste período. Vários países, como Turquia, República da Coreia e Tailândia iniciaram períodos de substituição de importações, desenvolvendo competências industriais que resultaram no fornecimento aos mercados internacionais (UNIDO, 2013).

Nos últimos 200 anos, praticamente todos os países que tiveram crescimento rápido de produção, renda e QV se deram pela sua industrialização (MURPHY, SHLEIFER e VISHNY, 1988).

Segundo Hartmann e Pyka (2013), a política de desenvolvimento humano precisa estar sintonizada com uma política industrial que promova a diversificação econômica qualitativa. Esta diversificação pode ser alcançada através da variedade de investimento em setores relacionados e não relacionados, como, por exemplo, a

urbanização, a qual leva a um aumento na demanda por produtos manufaturados, tornando a industrialização autossustentável, o que pode proporcionar aumento da renda e do bem-estar (MURPHY, SHLEIFER e VISHNY, 1988).

Políticas de desenvolvimento a partir de uma melhor compreensão do desenvolvimento co-evolutivo da variedade, liberdade de escolha e bem-estar podem contribuir no desenvolvimento econômico e humano (HARTMANN e PYKA, 2013).

Murphy, Shleifer e Vishny (1988) destacam, em seu estudo, a importância do papel do governo no processo de desenvolvimento da industrialização, como uma coordenação de investimento em diversos setores, com subsídios temporários e no investimento em infraestrutura compartilhada, como estradas, ferrovia e portos.

Alguns países conhecidos pela expressão *Newly Industrialized Countries* (NICs), classificados como em desenvolvimento, estão passando por esse processo de industrialização. Entre eles, pode-se citar: México, Brasil, Coréia do Sul, Taiwan, Cingapura, Hong Kong, África do Sul. Nestes países, as empresas devem dominar a ciência básica e aplicar conhecimentos e tecnologias específicas (ROBERTSON e PATEL, 2007).

O Brasil é considerado um dos NICs, pois atingiu elevada competência em setores de média e alta tecnologia (FREITAS, MARQUES e E SILVA, 2013).

Segundo dados da agência especializada da Organização das Nações Unidas para o Desenvolvimento Industrial (UNIDO), o índice de industrialização do Brasil foi de 0,42 no ano de 2010, apresentando-se acima da média mundial, de  $0,34 \pm 0,18$ , para o mesmo período. Assim, apresentando uma leve correlação com o IDH ( $r=0,6$ ;  $p<0,01$ ). Este índice vem se mantendo constante desde 2005.

### 2.1.3 IDH dos Países da América Latina

A maioria dos países latino-americanos passa por um processo de desenvolvimento humano muito semelhante, possivelmente em função do contexto

histórico de suas emancipações políticas e características sociais e culturais. Este processo de desenvolvimento pode ser avaliado a partir do Índice de Desenvolvimento Humano, adotado pelo United Nations Development Programme (UNDP) para medir os avanços na qualidade de vida de um país a partir da média geométrica entre indicadores de educação, saúde e renda, que classifica maior parte dos países da América Latina como países em desenvolvimento com IDH alto.

A América Latina é composta de 22 países, onde habitam aproximadamente 8,54% da população mundial, estimada para 2014 em 7,16 bilhões de habitantes, segundo dados da UNDP (2015).

A Tabela 1 apresenta a população e os principais indicadores de desenvolvimento humano dos países da América Latina, comparados com a média mundial, extraídos do UNDP (2015). A partir destes indicadores, observa-se que os países latino-americanos apresentam a média populacional de  $27,82 \pm 47,15$  milhões de habitantes, enquanto que a média dos países filiados a ONU é de  $38,51 \pm 42,38$  milhões de habitantes. O Brasil (202 milhões de habitantes) e Belize (300 mil habitantes) representam os países com a maior e menor população da América Latina, respectivamente. A Argentina apresenta a maior renda nacional bruta (RNB) per capita da América Latina, equivalente a 22.050 dólares, enquanto o Haiti (\$1.669) representa o país mais pobre.

Os países da América Latina apresentam maiores médias em relação à média mundial do IDH, de suas componentes e da maioria de suas subcomponentes de desenvolvimento humano, exceto na renda nacional bruta (RNB) per capita e nos anos médios de estudo (AME). A média mundial da renda nacional bruta (RNB) per capita é de  $\$16.882,4 \pm 18.676,29$ , enquanto que a média dos países latino-americanos é de  $\$11.197,41 \pm 5.849,26$ . O tempo médio de estudos mundial é de  $8,09 \pm 3,10$  anos, contra  $8,06 \pm 1,66$  anos na América Latina. Já o índice de anos médio de estudo (IAME) dos países latino-americanos se iguala com a média mundial (0,54), no entanto, apresentando menor variabilidade ( $\pm 0,11$ ) que a variância mundial ( $\pm 0,21$ ) neste índice, como também nos demais indicadores de desenvolvimento humano.

Dadas essas características, principalmente em relação às baixas variabilidades destes indicadores e os tempos de emancipações políticas muito próximos, os países da América Latina foram escolhidos como objeto de estudo para previsão do IDH e de suas expectativas de vida.

**Tabela 1 - População, Indicadores de Desenvolvimento Humano dos Países da América Latina e resumo estatístico dos indicadores mundiais e dos países latino-americanos**

Rank do IDH	País (código internacional)	População (milhões de habitantes)	IDH	Componentes do IDH			Sub-componentes do IDH						
				IS	IE	IR	EV	IAEE	IAME	RNB	AEE	AME	
40	Argentina (ARG)	41,8	0,836	0,866154	0,82389	0,815081	76,3	0,99444	0,65333	22050	17,9	9,8	
42	Chile (CHL)	17,8	0,832	0,949231	0,74889	0,809783	81,7	0,84444	0,65333	21290	15,2	9,8	
52	Uruguai (URY)	3,4	0,793	0,88	0,71389	0,794826	77,2	0,86111	0,56667	19283	15,5	8,5	
60	Panamá (PAN)	3,9	0,78	0,886154	0,67944	0,786029	77,6	0,73889	0,62	18192	13,3	9,3	
67	Cuba (CUB)	11,3	0,769	0,913846	0,76667	0,648119	79,4	0,76667	0,76667	7301	13,8	11,5	
69	Costa Rica (CRI)	4,9	0,766	0,913846	0,66611	0,739993	79,4	0,77222	0,56	13413	13,9	8,4	
71	Venezuela (VEN)	30,9	0,762	0,833846	0,69111	0,768128	74,2	0,78889	0,59333	16159	14,2	8,9	
74	México (MEX)	123,8	0,756	0,873846	0,64722	0,767162	76,8	0,72778	0,56667	16056	13,1	8,5	
75	Brasil (BRA)	202	0,755	0,838462	0,67889	0,758637	74,5	0,84444	0,51333	15175	15,2	7,7	
84	Peru (PER)	30,8	0,734	0,84	0,66389	0,71024	74,6	0,72778	0,6	11015	13,1	9	
88	Equador (ECU)	16	0,732	0,86	0,64778	0,70451	75,9	0,78889	0,50667	10605	14,2	7,6	
97	Colômbia (COL)	48,9	0,72	0,830769	0,61833	0,723681	74	0,75	0,48667	12040	13,5	7,3	
101	Belize (BLZ)	0,3	0,715	0,769231	0,72778	0,65446	70	0,75556	0,7	7614	13,6	10,5	
101	República Dominicana (DOM)	10,5	0,715	0,823077	0,61722	0,721698	73,5	0,72778	0,50667	11883	13,1	7,6	
158	Paraguai (PRY)	6,9	0,679	0,813846	0,58722	0,655034	72,9	0,66111	0,51333	7643	11,9	7,7	
116	El Salvador (SLV)	6,4	0,666	0,815385	0,55833	0,649109	73	0,68333	0,43333	7349	12,3	6,5	
119	Bolívia (BOL)	10,8	0,662	0,743077	0,64	0,612308	68,3	0,73333	0,54667	5760	13,2	8,2	
124	Guiana (GUY)	0,8	0,636	0,713846	0,56944	0,631076	66,4	0,57222	0,56667	6522	10,3	8,5	
125	Nicarágua (NIC)	6,2	0,631	0,844615	0,51944	0,573568	74,9	0,63889	0,4	4457	11,5	6	
128	Guatemala (GTM)	15,9	0,627	0,796923	0,48389	0,64022	71,8	0,59444	0,37333	6929	10,7	5,6	
131	Honduras (HND)	8,3	0,606	0,816923	0,49167	0,554867	73,1	0,61667	0,36667	3938	11,1	5,5	
163	Haiti (HTI)	10,5	0,483	0,658462	0,405	0,425193	62,8	0,48333	0,32667	1669	8,7	4,9	
Estatística de Resumo	América Latina	μ	27,82	0,71	0,83	0,63	0,69	74,01	0,73	0,54	11197,41	13,15	8,06
		±	47,15	0,08	0,07	0,1	0,1	4,36	0,11	0,11	5849,26	1,99	1,66
	Mundial	μ	38,51	0,69	0,79	0,63	0,68	71,03	0,72	0,54	16882,4	12,87	8,09
		±	142,38	0,15	0,13	0,17	0,18	8,37	0,16	0,21	18676,29	2,88	3,1

AEE - Anos esperado de estudo  
AME - Anos médios de estudo  
EV - Expectativa de vida ao nascer (em anos)  
IAEE - Índice de Anos Esperados de Escolaridade  
IAME - Índice de Anos Médio de Estudo  
IE - Índice de Educação  
IR - Índice Renda  
IS - Índice de Saúde  
Renda Nacional Bruta (RNB) per capita- (em \$)

Fonte: O autor (2016).

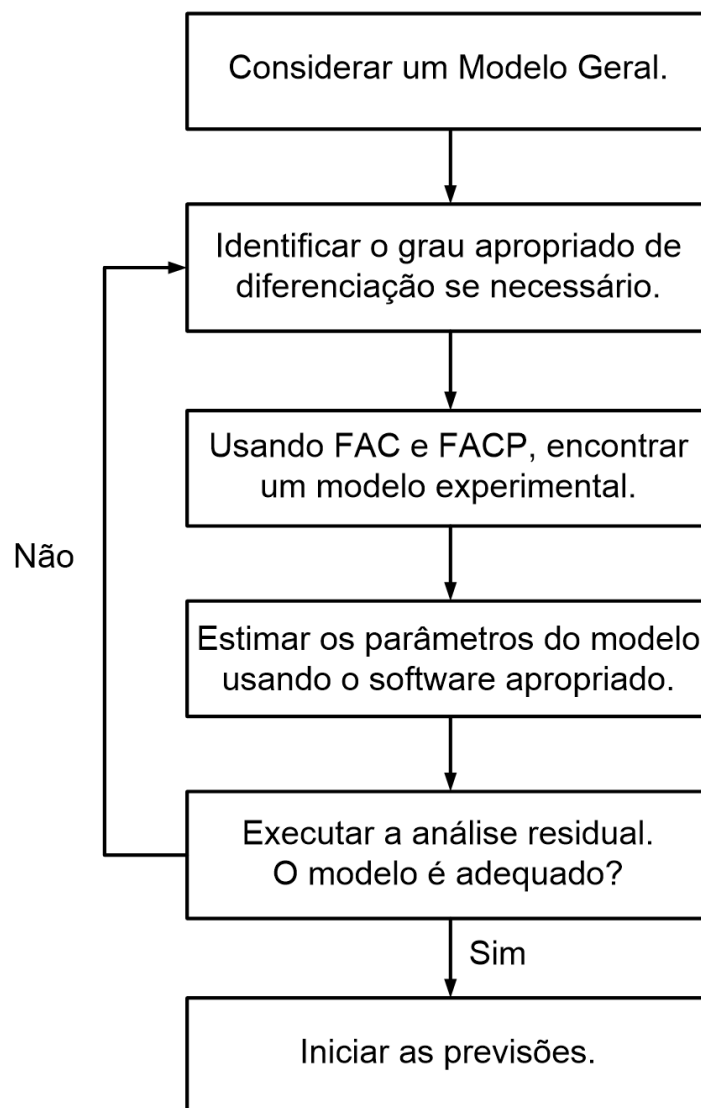


#### 2.1.4 Método de Previsão ARIMA

O método ARIMA, popularmente conhecido em estatística e econometria, foi criado por George Box e Gwilym Jenkins em 1976 para previsões de séries temporais não estacionárias.

A modelagem é interativa e dependente da interpretação do analista durante a definição dos parâmetros e seu refinamento, exigindo experiência e bom senso (BISGAARD e KULAHCI, 2011). A Figura 2 apresenta as etapas do processo de construção do modelo de séries temporais usando o método ARIMA.

**Figura 2 - Etapas do processo de construção do modelo de séries temporais usando o método ARIMA**



Fonte: Bisgaard e Kulahci (2011)

Durante o desenvolvimento do modelo, devem-se definir seus parâmetros, também conhecidos como fatores ou ordens de auto-regressão ( $p$ ), diferenciação/integração ( $d$ ) e média móvel ( $q$ ), descritos como:

AR ( $p$ ) - corresponde à ordem da componente auto-regressiva;

I ( $d$ ) ( $i$ ) - corresponde ao número de diferenças necessárias entre os valores da série para torná-la estacionária;

MA ( $q$ ) - corresponde à ordem da componente média móvel.

Estes parâmetros são definidos a partir da observação de características das séries temporais, através de testes estatísticos das funções de autocorrelações (FAC) e autocorrelações parciais (FACP). Estes testes verificam se existem correlações entre os dados no momento  $t$ ,  $t-1$ ,  $t-2$  ...  $t-k$ .

O modelo inicial pode ser modificado e o processo repetido até que as análises dos erros (resíduos) mostre que a alteração trouxe melhorias ao modelo. Maiores detalhes sobre os modelos ARIMA podem ser obtidos em Montgomery, Jennings e Kulahci (2015) e Box et al. (2015).

### 2.1.5 Processo de Descoberta de Conhecimento em Base de Dados

A Mineração de Dados<sup>2</sup> compreende apenas uma das etapas do processo de Descoberta de Conhecimento em Base de Dados (DCBD), referente à aplicação de algoritmos para extração de padrões em um processo de prospecção de conhecimento em bases de dados, envolvendo a automação da identificação e do reconhecimento de padrões (CARVALHO, 1999). Esse processo é um conjunto de procedimentos contínuos que compartilham o conhecimento descoberto a partir de bases de dados (CARVALHO, 1999).

Historicamente, noções de descoberta de conhecimento receberam várias nomenclaturas, incluindo mineração de dados, extração de conhecimento, descoberta de informação, arqueologia de dados, colheita de informação (AMARAL, 2001; VASCONCELOS, 2002a).

Esse processo surgiu em 1989, com o propósito de encontrar o conhecimento existente em uma base de dados e enfatizar o alto nível das aplicações dos métodos de prospecção de dados (AMARAL, 2001).

Em 1995, depois da realização de uma série de *workshops* sobre o assunto (FAYYAD e UTHURUSAMY, 1994), foi realizada, na cidade de Montreal, no Canadá, a primeira Conferência Internacional de Prospecção de Dados (FAYYAD e UTHURUSAMY, 1995), quando se padronizou o termo “Descoberta de Conhecimento em Base de Dados” como KDD - *Knowledge Discovery in Databases*.

Esse processo tem por objetivo a extração do conhecimento implícito previamente desconhecido e a busca de informações potencialmente úteis sobre os dados (FAYYAD, 1998).

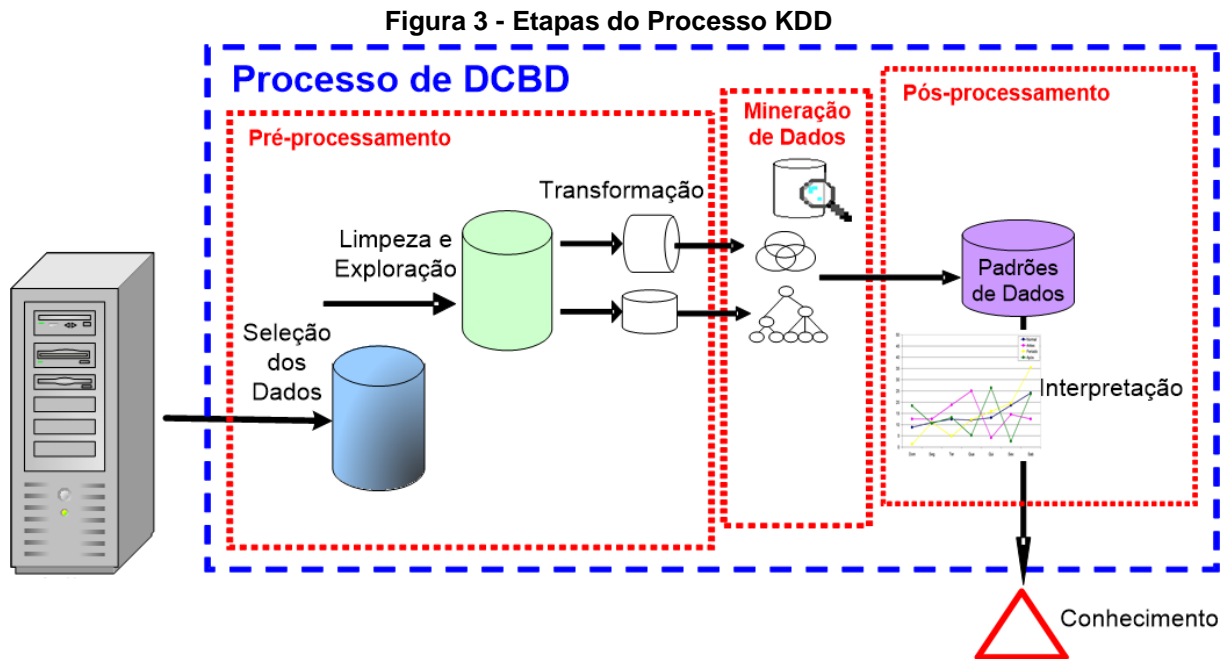
Segundo Fayyad (1998), são quatro os procedimentos básicos para a implantação do deste processo de DCBD: seleção dos dados, pré-processamento (limpeza, exploração e transformação), Mineração de Dados, interpretação dos resultados.

---

<sup>2</sup> Mineração de Dados - do inglês - *Data Mining*.

Em alto nível de abstração, pode-se dizer que as etapas de DCBD fazem parte de três grandes grupos: pré-processamento, aplicação de um algoritmo de Mineração de Dados e pós-processamento (MICHALSKI; KAUFMAN, 1998).

A Figura 3 incorpora as duas definições das etapas Fayyad (1998) e (MICHALSKI; KAUFMAN, 1998).



### a) Seleção dos dados

Etapa de compreensão do domínio e dos objetivos da tarefa de Mineração de Dados, em que é realizada a seleção do conjunto de origem, de um subconjunto de variáveis ou, ainda, de uma amostra (CARVALHO, 1999). Os dados normalmente são extraídos de um banco de dados ou de um *data warehouse*<sup>3</sup>, criado para servir às diversas necessidades de análise.

### b) Limpeza e Exploração

Nesta etapa, são decididas as estratégias e são realizadas a identificação e a limpeza dos dados, a fim de remover ruídos<sup>4</sup>, dados incompletos, repetições de registros, problemas de tipagem e tratar a ausência de dados de alguns atributos, visando adequá-los aos algoritmos.

Essas etapas não necessitam ser executadas nessa ordem, algumas dessas operações podem ser parcialmente executadas durante a construção de um Banco de Dados ou *Data Warehouse* (LOPES, 1999).

### **c) Transformação**

Ainda no pré-processamento, os dados pré-processados devem ainda passar por uma transformação que os armazena adequadamente, visando facilitar o uso das técnicas de Mineração de Dados, pois existem diversos tipos de algoritmos e cada um necessita de uma entrada específica, além das conversões de dados, criação de novas variáveis e categorização de variáveis contínuas (FIGUEIRA, 1998).

Normalmente, os algoritmos de mineração não podem acessar os dados em seu formato nativo, seja em razão da forma como são armazenados, seja pela normalização adotada na modelagem do banco; então, é necessária a conversão desses para um formato específico. Pode-se, ainda, sumarizar os dados, a fim de reduzir o número de variáveis sob consideração (CARVALHO, 1999).

### **d) Mineração de Dados**

Esta etapa caracteriza-se pela aplicação do algoritmo que, diante da tarefa adequada, será capaz de extrair conhecimento implícito e útil de um banco de dados (LOPES, 1999).

Consiste na efetiva aplicação do algoritmo escolhido sobre os dados a serem analisados, de acordo com a tarefa a ser cumprida, com o objetivo de localizar os padrões desejados (CARVALHO, 1999).

Esta etapa do processo de DCBD será detalhada na seção 2.1.5.

### **e) Interpretação dos resultados**

---

<sup>3</sup> *data warehouse* - do inglês - armazém de dados.

<sup>4</sup> ruídos – dados estranhos e/ou inconsistentes

As informações resultantes das etapas anteriores são interpretadas e avaliadas de forma que se selecione o conhecimento resultante de todo o processo (CARVALHO, 1999), com possível retorno aos passos anteriores (FAYYAD *et al.*, 1996).

#### 2.1.6 Mineração de Dados

Mineração de Dados é um método de extrair informação válida, previamente desconhecida e de máxima abrangência, geralmente de grandes bases de dados, usando-as para tomada de decisões, que permitem aos usuários explorar e inferir informação útil a partir dos dados, descobrindo relacionamentos escondidos no banco de dados (ADRIAANS e ZABTINGE, 1996).

A Mineração de Dados é uma área de estudo da Inteligência Artificial (IA), que descende de três conceitos: da estatística, da IA e da Aprendizagem de Máquina (ML<sup>5</sup>). Do primeiro, herdou seus conceitos clássicos de: distribuição normal, variância, desvio simples, análise de regressão, análise de discriminante, análise de conjuntos e intervalos de confiança. Do segundo, recebeu conceitos heurísticos, de como o homem pensa para a solução de seus problemas; e do último, ganhou seus algoritmos, desenvolvidos com a união dos conceitos de IA e ML (FAYYAD *et al.*, 1996).

Segundo Lemos (2003), o uso da Mineração de Dados para construção de modelos traz as seguintes vantagens:

1. Fácil compreensão;
2. Possibilita a análise de grandes bases de dados;
3. Descobre informações surpreendentes;
4. Os modelos são relativamente mais precisos.

---

<sup>5</sup>ML - do inglês - *Machine Learning*.

### 2.1.6.1 Objetivos, tarefas e técnicas e algoritmos de mineração de dados

As diversas tarefas desenvolvidas em Mineração de Dados têm como objetivo elementar a predição ou descrição (FAYYAD *et al.*, 1996).

A predição utiliza atributos para prognosticar o desconhecido ou os valores futuros de outras variáveis de interesse, enquanto a descrição contempla o que foi descoberto nos dados, sob o ponto de vista da interpretação humana (CARVALHO, 1999).

Em outras palavras, os objetivos da Mineração de Dados são de natureza (YIN, 2001):

1. Explanatória: explicar algum evento ou medida observada. Questionamentos, como, por exemplo: Por que a venda de determinado produto caiu na região sul?
2. Confirmatória: confirmar uma hipótese. Uma companhia seguro, por exemplo, pode verificar os registros de seus clientes para determinar se famílias com mais de um provedor de renda têm mais probabilidade de contratar um seguro residencial que famílias com apenas um provedor de renda.

As tarefas de Mineração de Dados são dependentes do domínio da aplicação e do interesse do usuário. Cada tarefa de Mineração de Dados extrai um tipo diferente de conhecimento do banco de dados, podendo requerer um algoritmo diferente para a extração de conhecimento (LOPES, 1999).

Segundo Viana (2004), as principais classes de tarefas de Mineração de Dados são: associação, classificação, agrupamento.

Técnicas de Mineração de Dados são procedimentos diferentes de exploração dos dados, utilizados para retirar diferentes formas de conhecimento.

Rodrigues (2000) apresenta um grande número de técnicas de Mineração de Dados. No entanto, cinco delas englobam as demais (CARVALHO, 1999). São elas: técnicas de redes neurais, regras de indução, árvores de decisão, análise de séries temporais e visualização (CARVALHO, 1999; CARVALHO, 2001).

O Quadro 4 apresenta as técnicas mais utilizadas de acordo com as principais tarefas de Mineração de Dados.

**Quadro 4 - Tarefas, Principais Técnicas e Algoritmos de Mineração de Dados mais Utilizados**

Problema	Tarefa	Técnica	Exemplos de Algoritmos mais utilizados
Predição	Classificação	Árvore de Decisão	<i>J48</i>
	Regressão	Redes Neurais Artificiais	<i>Multi-Layer Perceptron</i> <i>RBF Network</i>
Previsão de Series Temporais		<i>Least Median Squared</i> <i>Linear Regression</i> <i>SMOReg</i>	
Previsão			
Descrição	Associação	Regras de Associação	<i>Apriori</i>
	Agrupamento	Particionamento	<i>K-Means</i>
		Hierarquia	Hac

Fonte: Elaborada a partir das definições de Fayyad et al. (1996), Carvalho (1999), Lopes (1999), Viana (2004), Hall et al. (2009) e Pentaho (2014).

As técnicas de Mineração de Dados são de caráter universal e podem ser praticadas por meio de diferentes algoritmos. A escolha da técnica mais apropriada depende dos objetivos da pesquisa, do tipo de extração desejada e das características dos dados.

No presente estudo, destacamos a técnica de Previsões de Séries Temporais, que tem como objetivo inferir valores que a variável da série deverá assumir no futuro considerando como base valores passados dessa série. A previsão se diferencia das técnicas de classificação e regressão por trabalhar com conjuntos de observações de variáveis com dependência serial, ordenadas em função do tempo.



### 2.1.6.2 Algoritmos de Mineração de Dados para Previsão

A literatura apresenta uma quantidade significativa de algoritmos. Neste referencial teórico, são destacados os algoritmos de previsão de séries temporais, com paradigma de aprendizagem de funções, que serão utilizados no experimento apresentado na seção 3.6. Estes algoritmos, apresentados a seguir, podem ser utilizados em tarefas de classificação e regressão para solução de problemas de predição e previsão.

#### **a) *Gaussian Processes***

Este algoritmo aplica o processo Gaussiano para regressão sem ajuste de hiperparâmetro, substituído os valores em falta pelos valores de média ou moda (WITTEN *et al.*, 2016).

Maiores informações sobre o método podem ser obtidas em Mackay (1998).

#### **b) *Multi-Layer Perceptron***

O algoritmo de percepção multicamada (MLP<sup>6</sup>) foi introduzido por Williams e Hinton (1986) e utiliza conceito de Redes Neurais Artificiais (RNAs) com retropropagação dos erros para ajustar pesos das camadas, denominada *Backpropagation*.

Sua rede neural pode ser construída manualmente, por um algoritmo ou por ambos. A rede também pode ser monitorada e modificada durante o tempo de treinamento (NAPAGODA, 2013).

#### **c) *RBFNetwork***

O algoritmo de Redes de Funções de Base Radial (RBFNetwork<sup>7</sup>) foi introduzido por Broomhead e Lowe em 1988 (BROOMHEAD e LOWE, 1988).

---

<sup>6</sup> MLP - do inglês - *Multi-Layer Perceptron*.

<sup>7</sup> RBFNetwork - do inglês - *Radial Basis Function (RBF) Network*.

O *RBFNetwork* implementa uma base de função radial baseada em base gaussiana normalizada e aplica o algoritmo de agrupamento k-means para fornecer as funções de base e aprende uma regressão logística, em caso de classes discretas ou regressão linear, em caso de classe numericas (WITTEN *et al.*, 2016).

#### **d) *Least Median Squared***

Introduzido em 1984 por Rousseeuw (SIMPSON, 1997), o algoritmo dos Mínimos Quadrados Médios (LMS<sup>8</sup>) gera funções de regressão de menor quadrado a partir de subamostras aleatórias dos dados. A regressão ao quadrado mínimo com o erro quadrático mínimo mais baixo é escolhida como o modelo final (WITTEN *et al.*, 2016).

Maiores informações sobre o algoritmo podem ser obtidas em Leroy e Rousseeu (1987).

#### **e) *Linear Regression***

O *Linear Regression* aplica conceitos de regressão linear, com o critério de informação de Akaike (CIA<sup>9</sup>) (AKAIKE, 1974), uma medida da qualidade relativa para seleção do modelo de regressão. É um algoritmo adequado para trabalhar em caso de variáveis ponderadas.

#### **f) *SMOReg***

O algoritmo de Otimização Mínima Sequencial para a Regressão (*SMOReg*<sup>10</sup>) foi introduzido por Smola e Schölkopf (1998) e corresponde à extensão do algoritmo de Otimização Mínima Seqüencial (SMO<sup>11</sup>) de vetores de suporte proposto por Platt (1998).

---

<sup>8</sup> LMS - do inglês *Least Median Squared*.

<sup>9</sup> CIA - do Inglês - *Akaike Information Criterion*.

<sup>10</sup> SMOReg - do inglês - *Sequential Minimal Optimization for Regression*.

<sup>11</sup> SMO - do inglês - *Sequential Minimal Optimization*.

O *SMOReg* aplica um motor de Vetores de Suporte para Regressão (SVR<sup>12</sup>). Esse método, segundo Muller et al. (1997), apresenta significativo desempenho em estudos de previsão comparativos e é discutido no estudo de Smola e Schölkopf (2004).

O SVR é uma extensão de Máquinas de Vetores Suporte (SVM<sup>13</sup>), criada originalmente para predição de classes (MÜLLER *et al.*, 1997), que acrescenta a função de regressão numérica.

Durante a aplicação do *SMOReg*, os parâmetros de entrada podem ser aprendidos utilizando vários algoritmos. Na interface de programação de aplicativos (API<sup>14</sup>) *Forecast* do WEKA, para séries temporais, o algoritmo é selecionado configurando o *RegOptimizer*. O algoritmo mais popular utilizado como padrão é o *RegSMOImproved* (SHEVADE *et al.*, 2000), utilizado neste experimento.

Segundo Zhu et al. (2014), em uma abordagem de linguagem de alto nível, o *SMOReg* resolve o problema de programação quadrática (QP) durante o treinamento de SVM e implementa um motor de MVR. Seleciona um multiplicador de Lagrange ( $\alpha_1$ ), que permite encontrar extremos máximos e mínimos, que não atendem as condições de Karush-Kuhn-Tucker (KKT). Escolhe outro multiplicador ( $\alpha_2$ ) utilizado para otimizar o par de multiplicadores ( $\alpha_1$ ;  $\alpha_2$ ) e repete até todos satisfazerem as condições KKT, empregando uma estrutura de repetição de dois laços: externo e interno. O laço externo escolhe  $\alpha_2$  e para cada  $\alpha_2$  escolhido o laço interno escolhe  $\alpha_1$ . O laço externo repete sobre todos os padrões que violam as condições de otimalidade, primeiro somente sobre aqueles com multiplicadores de Lagrange, e uma vez que todos eles são satisfeitos.

Os índices correspondentes aos multiplicadores de Lagrange não-limite são mantidos atualizados em cache, garantindo uma implementação eficiente. Os demais vetores são computados quando necessário até o processo de regressão ser concluído.

---

<sup>12</sup> SVR - do inglês - *Support Vector Regression*. O SVM adota uma abordagem de aprendizado não paramétrico para tarefas de classificação e regressão de alta performance, que utilizam funções de *kernel*, as quais amenizam o efeito da maldição da dimensionalidade (DE MORAES LIMA, 2004).

<sup>13</sup> SVM - do inglês - *Support Vector Machines*.

<sup>14</sup> API - do inglês - *Application Programming Interface*.

### 2.1.6.3 Ambiente para análise de conhecimento Waikato (WEKA15)

O software WEKA possui um ambiente gráfico, multiplataforma, desenvolvido em linguagem Java pelo Departamento de Ciência da Computação da Universidade de Waikato, na Nova Zelândia (WITTEN e FRANK, 2005).

O WEKA compreende um conjunto de implementações de algoritmos de pré-processamento de dados e de diversas técnicas de Mineração de Dados, derivados de diferentes paradigmas ou abordagens.

Neste ambiente, após a etapa de transformação dos dados para formato *arff*<sup>16</sup>, um único arquivo de entrada pode ser utilizado por diferentes algoritmos nas demais etapas do processo de DCBD.

Durante a etapa de conhecimento do domínio, é possível gerar informações estatísticas básicas (média, máximo, mínimo e desvio padrão) de cada atributo e, também, gerar gráficos da frequência dos dados, correlacionando-os com um atributo de referência.

O software permite percorrer as etapas do processo de DCBD de forma versátil. Seu ambiente oferece ferramentas de pré-processamento e técnicas de Mineração de Dados de Classificação, Agrupamento, Associação e Visualização. A partir da versão 3.7, é possível incluir a técnica de Previsão de Séries Temporais, utilizada neste estudo, que foi disponibilizada para a comunidade científica no segundo semestre de 2013.

Segundo Napagoda (2013), esta ferramenta remove a ordenação temporal de exemplos de entrada individuais e, através de campos de entrada adicionais, codifica a variável em relação à sua dependência do tempo. Tendências e sazonalidade também são computadas automaticamente através de outras entradas adicionais. Após o pré-processamento dos dados, diferentes algoritmos de regressão do WEKA podem ser aplicados para aprendizagem de um modelo de previsão, entre eles o de regressões lineares múltiplas. A ferramenta ainda permite aplicação de métodos capazes de prever um alvo contínuo, como por exemplo, a máquina de vetores de suporte para árvores de regressão e árvores de modelo, que apresentam regressão linear em suas folhas.

---

<sup>15</sup> WEKA - do inglês - *Waikato Environment for Knowledge Analysis*.

<sup>16</sup> *arff* - do inglês *attribute relation file format*.

O sistema WEKA é mantido atualizado com diversos algoritmos para Mineração de Dados já divulgados na literatura. Pode ser baixado gratuitamente pelo sítio eletrônico<sup>17</sup>, onde podem ser encontradas várias informações sobre seu projeto, seu código aberto e de projetos desenvolvidos utilizando-se o sistema.

## 2.2 ESTADO DA ARTE

Nesta seção, será apresentado o estado da arte, demonstrando-se os principais trabalhos, seus pesquisadores e aspectos publicados sobre os assuntos do tema, evidenciando-se a existência de lacunas e a singularidade do trabalho proposto.

Com base em Luna (1999), esta revisão de literatura classifica-se como empírica, pois procura explicar como o problema vem sendo investigado do ponto de vista metodológico.

Os assuntos do tema desta pesquisa – IDH, EV e Mineração de Dados – são objetos de estudo da Engenharia de Produção. Porém, os dois primeiros estão mais presentes na área de Saúde, enquanto que o terceiro na área de Ciências da Computação/Tecnologias da Informação. A seguir, são apresentados trabalhos envolvendo os temas.

### 2.2.1 Índice de Desenvolvimento Humano e Expectativa de Vida

Diferentes estudos correlacionam o IDH com o tema de QV, saúde e desenvolvimento humano. Mariano (2012) mede a eficiência social dos países a partir da sua capacidade de converter sua riqueza produzida em QV, e identifica fatores que possam explicá-la. Verifica que as ex-repúblicas soviéticas e os países de passado socialista foram os que mais se destacaram na eficiência social, enquanto que os países desenvolvidos, mesmo apresentando indicadores sociais elevados, não se sobressaíram na eficiência social. Já os países do sul do continente africano, além de possuírem as piores condições sociais, foram

---

<sup>17</sup> sítio eletrônico WEKA - <http://www.cs.waikato.ac.nz/ml/weka/>

considerados os mais ineficientes.

Lima (2006) utiliza o IDH para avaliar e mapear a QV das populações em municípios brasileiros, aplicando cartografia temática estatística para possibilitar aos gestores territoriais e ambientais adoção de medidas que levem à mitigação ou extinção da desigualdade e exclusão social.

Sadovsky et al. (2015) analisaram a relação da proporção de exames de mamografia e papanicolaou com o IDH em localidades brasileiras, encontrando alta correlação positiva do índice com este tipos de exame. Concluem que a promoção de políticas públicas que incentivam o desenvolvimento socioeconômico favorece o aumento da cobertura desses exames.

Tavares et al. (2014a) identificaram os padrões alimentares de adolescentes brasileiros e sua associação com o Índice de Desenvolvimento Humano Municipal (IDH-M), encontrando correlações positivas – como o padrão saudável, e negativas – com padrões mistos.

Szuster et al. (2012) analisaram a sobrevida de pacientes em diálise no Sistema Único de Saúde (SUS), entre 2002 e 2004, concluindo que os pacientes que residem em cidades com melhor IDH correm menor risco de vida.

Castro e Rodrigues Júnior (2012) estimaram a influência de óbitos por causas externas na faixa de fronteira brasileira entre 2000 e 2005, associada ao IDH, concluindo que existe um gradual aumento destes tipos de óbitos no sentido norte-sul correlacionado com o índice.

González-Zapata et al. (2011) avaliaram a relação entre a prevalência de peso e o IDH, concluindo que o índice está diretamente associado com a prevalência do maior de excesso de peso da população, com diferentes comportamentos de gênero.

Martinez et al. (2011) buscaram padrões espaciais da associação entre os percentuais de gravidez na adolescência e características socioeconômicas dos municípios do Estado de São Paulo, Brasil, verificando que a ocorrência de gravidez precoce tem uma estreita associação com indicadores econômicos e sociais, entre eles o IDH.

Neumayer (2001) utilizou o IDH de maneira preditiva. Em seu estudo, relacionou o IDH de um país com a sua sustentabilidade, qualificando o índice através deste fator. Analisando 155 países, alertou que 42 destes serão potencialmente insustentáveis em um período de 15 anos, e que a maioria possui baixo IDH.

Ravallion (2011) discute a importância do conhecimento dos índices para adoção de políticas para o seu desenvolvimento. Em outro estudo, Ravallion (2012b) analisa as vantagens e desvantagens da nova versão IDH, apresentada em PNUD (2010) e discutida em Klugman, Rodríguez e Choi (2011).

Ravallion (2012b) resume a forma como o índice foi alterado em relação às avaliações do IDH de longevidade e escolaridade, questionando os valores dos limites adotados para a confecção no novo IDH, que acaba privilegiando alguns países, além de não explicar com maior precisão onde se deve investir.

Estudos envolvendo o IDH vêm auxiliando o Programa das Nações Unidas para o Desenvolvimento (PNUD) na promoção de ajuda externa e auxiliando os países membros no planejamento de políticas públicas para melhoria do seu IDH.

A maioria dos estudos sobre IDH baseia-se nos relatórios da ONU. Entre eles, pode-se citar o estudo de Alkire (2010), que sintetiza discussões sobre o conceito de desenvolvimento humano, informando a definição do Relatório de 2010 e propondo relações entre o conceito de desenvolvimento humano e assuntos de importância fundamental, como: i) Objetivos de Desenvolvimento do Milênio; ii) Direitos Humanos (segurança humana e felicidade); iii) Desigualdade (a duração dos resultados ao longo do tempo); iv) Sustentabilidade ambiental.

Outra variável utilizada em pesquisas da área de QV e desenvolvimento humano são a EV. Em relação ao tema, Chiavegatto Filho e Laurenti (2013) fazem a sua comparação em três estados do sudeste brasileiro: São Paulo, Rio de Janeiro e Minas Gerais, nos anos de 1996, 2000 e 2010, encontrando diferenças que podem ser explicadas pelo grupo etário dos idosos e as doenças crônicas. Concluem, ainda, que as diferenças no estilo de vida e prática de atividades físicas, melhores em Minas Gerais, podem ajudar a explicar a inconsistência entre os níveis socioeconômicos e a situação de saúde.

Existe um quantitativo significativo de pesquisas que colocam o IDH e a EV como variáveis balizadoras de estudos de QV e desenvolvimento humano (GONZÁLEZ-ZAPATA *et al.*, 2011; KARIMINIA *et al.*, 2011; MARTINEZ *et al.*, 2011; BRUNEKREEF *et al.*, 2012; CASTRO e RODRIGUES-JÚNIOR, 2012; MENDES *et al.*, 2012; SZUSTER *et al.*, 2012; CAMPOLINA *et al.*, 2013; CHIAVEGATTO FILHO e LAURENTI, 2013; KANSO *et al.*, 2013; MACKENBACH e LOOMAN, 2013; ANSTEY *et al.*, 2014; CERVANTES e BOTERO, 2014; MORTON, DE STAVOLA e LEON, 2014; TAVARES *et al.*, 2014b; CAMARGOS e GONZAGA, 2015; HENDI, 2015; LACEY *et al.*, 2015; LI *et al.*, 2015; PERCIO, MEDINA e LUNA, 2015; SADOVSKY *et al.*, 2015; STRINGHINI *et al.*, 2015).

### 2.2.2 Mineração de Dados

O tema Mineração de Dados começou a ser mais difundido a partir de 1995, quando a comunidade científica reuniu-se na Cidade Montreal – Canadá, para discutir a Temática Prospecção de Dados (FAYYAD e UTHURUSAMY, 1995), e após vários Workshops (FAYYAD e UTHURUSAMY, 1994), definiu, entre outros termos, o “*Knowledge Discovery in Databases - KDD*”, para processo de Aquisição de Conhecimento em Base de Dados, no qual a Mineração de Dados está presente como uma das etapas do processo. Já o tema IDH recebeu destaque a partir de 1990, quando foi apresentado pela primeira vez pelos cientistas Mahbub ul Haq, Amartya Sen, Paul Streeten e Keith Griffin.

Foram encontrados na literatura diversos trabalhos com a temática Mineração de Dados, em sua maioria apresentando comparação do desempenho de algoritmos de Mineração de Dados, havendo também os que consistem em aplicações das técnicas em problemas diversificados.

Observou-se grande contribuição de trabalhos envolvendo a arquitetura de algoritmos de Mineração de Dados (CHEN e BAI, 2010; MARX, 2010; PIQUER *et al.*, 2010; AHN *et al.*, 2011; CHEN e HUANG, 2011; HENDERSON e GREY, 2011; RAVISANKAR *et al.*, 2011; NIKOLOVA e JECHEVA, 2012), sistemas baseados em conhecimento: representação do conhecimento (HOONAKKER *et al.*, 2011); gestão



do conhecimento (FESHARAKI, SHIRAZI e BAKHSHI, 2011) e aquisição de conhecimento (GAJZLER, 2010); descoberta de padrões (TREMBLAY, DUTTA e VANDERMEER, 2010); gestão de relacionamento com o cliente (WANG *et al.*, 2010).

Trabalhos de revisão, como os de Liao, Chu e Hsiao (2012), apontam lacunas na literatura e possíveis tendências para as técnicas de Mineração de Dados. Outros trabalhos de revisão (STROHMEIER e PIAZZA, 2013) apresentam tendências para domínios mais específicos, como a gestão de recursos humanos.

Em seus estudos, Liao, Chu e Hsiao (2012) sugerem que o desenvolvimento das técnicas de Mineração de Dados tende a torná-las mais utilizáveis e orientadas para aplicações em problemas mais pontuais. Diferentes metodologias das ciências sociais, como a psicologia, a ciência cognitiva e do comportamento humano poderão fazer uso das técnicas de Mineração de Dados como uma metodologia alternativa e promissora. A integração de métodos qualitativos, quantitativos e científicos e a integração de estudos de metodologias de técnicas irão aumentar a compreensão do assunto. A capacidade de evolução contínua das técnicas de Mineração de Dados e de fornecer uma nova compreensão para os dados são as principais vantagens das metodologias de Mineração de Dados para o futuro.

### 2.2.3 Técnicas de Data Mining de Previsão

A literatura oferece uma variedade de técnicas de previsão, dentre as quais se destacam as previsões a partir de técnicas de Mineração de Dados. Estudos contemporâneos de previsão utilizando estas técnicas foram desenvolvidos em diferentes áreas, entre elas: energia eólica (MANGALOVA e AGAFONOV, 2014; SILVA, 2014); mercado financeiro (RODRIGUES e STEVENSON, 2013); engenharia (PALIT e POPOVIC, 2006a); agronegócios (CORREA *et al.*, 2015); meio ambiente (SOUSA *et al.*, 2013); saúde (WINTERS-MINER *et al.*, 2015; XIE *et al.*, 2016).

Estudos anteriores (ARINZE, 1994) já recomendavam algoritmos de Mineração de Dados para aquisição de conhecimento, reduzindo a necessidade de especialistas. No entanto, estudos subsequentes fazem julgamentos desfavoráveis à

utilização de técnicas de Mineração de Dados (CHATFIELD, 1995; KEOGH e KASSETTY, 2003; ARMSTRONG, 2006). Keogh e Kasetty (2003) criticam a falta de pesquisadores da área de Mineração de Dados para testar métodos alternativos em problemas reais.

Diferentes estudos contemporâneos de previsão utilizando técnicas de Mineração de Dados foram desenvolvidos. Entre eles, estudos para previsão de energia eólica (MANGALOVA e AGAFONOV, 2014; SILVA, 2014), previsões do mercado financeiro (RODRIGUES e STEVENSON, 2013) e na engenharia (PALIT e POPOVIC, 2006b).

Os trabalhos de Silva (2014), Mangalova e Agafonov (2014) foram premiados, respectivamente, com o primeiro e segundo lugares no *Global Energy Forecasting Competition* (GEFCOM, 2012). A equipe de Silva (2014) desenvolveu um modelo de previsão de potência em usinas de energia eólica, algoritmos de Mineração de Dados de regressão linear e árvores de decisão. Para o mesmo propósito, Mangalova e Agafonov (2014) combinam métodos heurísticos e formais, e, na seleção de fatores, utiliza o algoritmo de Mineração de Dados de árvore de classificação e regressão.

No estudo de Rodrigues e Stevenson (2013), foram utilizados modelos de Redes Neurais Artificiais (RNA), Regressão Linear (RL) e combinações entre ambos para identificar na bolsa de valores, de forma precoce, empresas-alvo de aquisições ou fusão. Seus resultados apontaram melhor desempenho individual dos modelos RNA em relação aos modelos RL. Mas as combinações de modelos superaram todos os resultados.

Palit e Popovic (2006b) realizaram análise das séries temporais, apresentando modelo e aplicações de previsão em engenharia a partir de métodos matemáticos não convencionais da inteligência artificial. Também, apresentam taxonomia para séries temporais, de acordo com suas características.

Napagoda (2013) aplica técnicas de Mineração de Dados em séries temporais em dados financeiros para o cálculo das taxas de câmbio do dólar para Rúpias. Compara quatro modelos desenvolvidos em diferentes algoritmos e ambientes de mineração de Dados: Regressão Linear Múltipla no MS Excel,

Regressão Linear Múltipla de Séries Temporais através no Weka, modelo de Vetor Auto regressivo no R e modelo de Redes Neurais implementado em Neural Works. Compara seus erros de previsão, adotando as mesmas medidas de qualidade das previsões: o erro médio (ME), o erro médio absoluto (MAE), o erro quadrático médio (MSE), o erro quadrático médio de raiz (RMSE), o erro de porcentagem média (MPE) e o erro médio de porcentagem absoluta (MAPE). Seus resultados apontam precisão em todos os modelos, mas o modelo desenvolvido no ambiente WEKA apresentou os melhores resultados.

Silva (2016) utiliza técnicas de mineração de dados para caracterização geoespacial e previsão da incidência de pneumonia em Portugal, testando diferentes algoritmos para o desenvolvimento dos modelos de previsão para o número de casos de pneumonia, destacando o *SMOReg* como o algoritmo que apresentou melhor desempenho entre os modelos de previsão em séries temporais.

#### 2.2.4 Originalidade do Tema de Pesquisa

Apesar da existência de vários trabalhos envolvendo os temas, observa-se a falta de trabalhos de previsão de indicadores de IDH e EV através de técnicas de Mineração de Dados. Todos os trabalhos mencionados assemelham-se aos temas propostos, embora de formas independentes, pois os assuntos envolvidos estão inseridos em diferentes áreas, incluindo as de Engenharia de Produção.

Menezes e Silva (2001) mencionam que a revisão de literatura é fundamental para fornecer informações e impedir a duplicação de pesquisas sobre a mesma perspectiva do tema, além de favorecer definições adjacentes mais pontuais do problema a ser estudado.

Para o desenvolvimento desta revisão de literatura, buscou-se, em diferentes periódicos, estudos semelhantes com as seguintes palavras-chave: mineração de dados, previsão, índice de desenvolvimento humano, expectativa de vida e seus sinônimos, e abreviaturas em língua portuguesa e inglesa. Durante as buscas em títulos, palavras-chave e resumos, aplicou-se a expressão apresentada

no Quadro 5, não identificando nenhum estudo. Portanto, esta revisão de literatura é finalizada, confirmando a originalidade deste estudo.

**Quadro 5 - Expressão de Busca de Palavras-chave em periódicos**

```
{  
    ("Data Mining") + (Forecast)  
  
    AND  
  
    ( "Human Development Index" OU "Life Expectancy ")  
}
```

### 3 METODOLOGIA

Neste capítulo, estão descritos: o desenho da pesquisa; o instrumento e fonte de pesquisa e intervenção em ambientes sociais e institucionais; dados da pesquisa; a formulação de hipóteses; os materiais e métodos utilizados na pesquisa, bem como a descrição de todo experimento.

#### 3.1 DESENHO DA PESQUISA

Segundo Vasconcelos (2002b), a pesquisa pode ser classificada sob cinco aspectos principais, quanto: ao enquadramento do objeto; a natureza dos dados e da análise; os objetivos, aplicações; e tipo de conhecimento implicado.

Quanto ao objeto, esta pesquisa é classificada como Temática ou Focal simples, pois eleger como objeto uma temática específica, com foco preciso para produção de um novo conhecimento, com contribuição para os debates teóricos, científicos e técnicos, e com subsídio para tomada de decisões de caráter gerencial.

Quanto à natureza dos dados e da análise, é classificada como quantitativa. Quanto aos objetivos, aplicações e tipo de conhecimento implicado, é considerada Aplicada, pois se articula dentro de diferentes ciências (Engenharia de Produção, Economia e Ciências da Computação).

Considerando a definição da lógica que determina as possibilidades de generalização dos resultados encontrados e a combinação de métodos quantitativos e qualitativos, a pesquisa também pode ser classificada como empírica, com lógica de replicação, pois a solidificação desse conhecimento se faz por meio das avaliações posteriores da comunidade científica e de novos testes através de replicações integrais ou parciais dos mesmos procedimentos de investigação (VASCONCELOS, 2002b).

A relação entre o tempo em que se realiza a pesquisa e o tempo do fenômeno observado a classifica como ex-post facto, pois o processo principal de

investigação é construído após a ocorrência do fenômeno (VASCONCELOS, 2002b).

Considerando os tipos de corte no tempo do fenômeno e o significado teórico e intrínseco das marcas temporais feitas no tempo do fenômeno, será utilizado na pesquisa um corte transversal para focar o objeto de investigação, pois se busca a análise do fenômeno em um momento específico (VASCONCELOS, 2002b).

### 3.2 INSTRUMENTO, FONTE DE PESQUISA E INTERVENÇÃO EM AMBIENTES SOCIAIS E INSTITUCIONAIS

Segundo Vasconcelos (2002b), os recursos e as fontes desta pesquisa se classificam como primários e as fontes bibliográficas como secundárias. A primeira classificação se dá pela falta de tratamento analítico em relação às informações e dados. A segunda classificação é atribuída ao fato de a fonte de pesquisa contemplar, em tese, todos os tipos de investigação.

### 3.3 DADOS DA PESQUISA

Uma variável consiste em qualquer atributo e um fenômeno classificado em duas ou mais categorias, pois os acontecimentos sociais e humanos sempre apresentam predicados que descrevem suas características (VASCONCELOS, 2002b).

As variáveis podem ser classificadas em simples, quantitativas, descontínuas ou discretas, contínuas e multidimensionais complexas (VASCONCELOS, 2002b).

Em uma concepção de causalidade mecanicista, as variáveis ainda podem ser classificadas em: variáveis independentes, dependentes e outras

(VASCONCELOS, 2002b). A segunda sofre interferência da primeira e as duas juntas sofrem influências da terceira.

### 3.4 FORMULAÇÃO DE HIPÓTESES

Geralmente, as informações demográficas ou informações adicionais sobre a amostra não participam automaticamente da análise de dados, a sua utilização depende da visão crítica e analítica do pesquisador. A sua percepção conduz o experimento, a partir do levantamento de hipóteses.

Neste experimento, definiram-se as seguintes hipóteses deduzidas da revisão de bibliográfica e reformuladas a partir dos resultados da pesquisa, conforme sugerem Koche (1997):

a) Hipótese Básica:

Utilizando-se de Técnicas de Mineração de Dados, é possível prever o IDH e a EV dos países da America Latina, porque essa tecnologia permite trabalhar com grandes volumes de dados de maneira mais automatizada e com eficiência.

b) Hipóteses Secundárias:

- i) Algoritmos de Mineração de Dados com paradigma de aprendizagem de funções são mais eficientes que o método ARIMA para prever o IDH, porque são mais independentes da inferência humana durante o processo de modelagem.
- ii) Modelos multivariados apresentam melhor qualidade nas previsões em relação aos modelos univariados, porque oferecem maior experiência de observações aos algoritmos durante a fase de treinamento;
- iii) Modelos multivariados globais apresentam melhor qualidade nas previsões em relação aos modelos multivariados específicos, porque

durante a fase de treinamento oferecem maior experiência aos algoritmos;

- iv) O algoritmo *SMOReg* apresenta melhor qualidade nas previsões entre os algoritmos com paradigma de aprendizagem baseada em funções (*Least Median Squared, Linear Regression, Multi-Layer Perceptron, RBF Network e Gaussian Processes*), porque utiliza um motor de Vetores de Suporte para Regressão, que, segundo Muller et al. (1997), é um método que apresenta significativo desempenho em estudos de previsão comparativa.

Apesar das técnicas Mineração de Dados não depender da formulação de hipóteses em seus diferentes métodos, incluindo-se os de previsões, as escolhas entre os diferentes algoritmos ou modelos podem ser formalizadas através de hipóteses para análise estatística. Então, adotaram-se as seguintes proposições para as análises de variância das medidas de qualidades, com um nível de significância de  $\alpha = 0,05$ , tanto para escolha do algoritmo como para escolha dos modelos de previsão, como, também, para comparar as previsões com valores já divulgados:

- i) Hipótese nula ( $H_0$ ): as médias ou medianas são iguais;
- ii) Hipótese alternativa ( $H_1$ ): as médias ou medianas são diferentes.

### 3.5 MATERIAIS E MÉTODOS

A partir dos dados históricos do IDH, iniciais de 187 países filiados ao UNDP, referentes ao período de 1990 a 2013, atualizados em 24 de julho de 2014, foram desenvolvidas todas as etapas do processo de Descoberta de Conhecimento em Base Dados (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996), apresentadas detalhadamente nas próximas subseções. O processo foi percorrido em dois momentos. Na primeira passagem pelo processo, foi definido o algoritmo e modelo com melhor desempenho a partir da previsão do IDH 2013 e 2014. No segundo momento, foi efetivada a previsão do IDH e da EV nos 22 países latino-americanos



filiados ao UNDP para o período de 2015 a 2020, a partir do melhor algoritmo e modelo, aplicados em uma nova base de dados, atualizada em 14 de dezembro de 2015, composta de 188 países.

A maioria das etapas do processo de DCBD foi realizada no ambiente de Mineração de Dados WEKA (HALL *et al.*, 2009), utilizando a técnica de *Forecast*, através da API do WEKA. Este software foi escolhido por oferecer código aberto disponível gratuitamente na internet e, também, por ser mantido atualizado com diversos algoritmos para Mineração de Dados já publicados na literatura.

Durante o processo, foi avaliado o desempenho de diferentes algoritmos com paradigma de aprendizado baseados em funções, indicados por Pentaho (2014) para previsões de séries temporais. A partir do algoritmo com melhor desempenho, foram desenvolvidos modelos de previsão, e seus resultados comparados com os últimos relatórios do UNDP, para definição dos modelos mais eficientes.

Para avaliar os resultados, utilizaram-se medidas de qualidade das previsões das séries temporais, apresentadas no item 3.5.1. Também, foram realizados em diferentes momentos do processo de DCBD testes estatísticos de análise de variância e correlação, partindo-se do teste de normalidade para orientar na escolha de testes paramétricos ou não paramétricos, adotando-se com nível de significância,  $\alpha = 0,05$ .

Paralelo ao DCBD, foi desenvolvido, através do pacote estatístico para as ciências sociais (SPSS<sup>18</sup>) (PESTANA e GAGEIRO, 2003), modelos de previsão utilizando-se a técnica de previsão mais popular, ARIMA, para, ao final dos testes, compará-la com modelos desenvolvidos a partir de técnicas de Mineração de Dados. O método não permitiu o desenvolvimento de modelos multivariados. Então, foram desenvolvidos 88 modelos univariados para previsão do IDH dos países da América Latina. Seus parâmetros utilizados na modelagem são apresentados nos Apêndices 4 e 5.

---

<sup>18</sup> SPSS - do inglês - *Statistical Package for the Social Sciences*

O hardware utilizado foi composto de um processador de 2.40 GHz e 10 GByte de memória RAM, dedicada ao processamento do Software WEKA.

Os parâmetros experimentais utilizados no processo de DCBD estão apresentados no Quadro 6, assim como a síntese das séries temporais utilizadas encontra-se no Quadro 7, para possíveis replicações ou comparações entre pesquisas.

Buscou-se incluir dados de industrialização em modelos multivariados, consultados na base de dados da UNIDO. No entanto, verificou-se que suas séries temporais mais atualizadas (1960-2011) se encontram defasadas em relação às do IDH (1980-2014) e com elevada ausência de dados.

**Quadro 6 - Processo de Descoberta de Conhecimento em Bases de Dados**

Procedimentos: Etapas do Processo de DCBD

Fonte de Dados: Relatório da UNDP (2014, 2015) e site da UNDATA (2014) com séries temporais

Softwares: WEKA, MS ACCESS, MS Excel, GraphPad InStat. e SPSS

API: Forecast para WEKA

Problema de DCBD: Previsão

Tipo de Dados: Séries Temporais Univariadas e Multivariada de IDH e EV

Técnicas de Mineração de Dados: Previsão (predição numérica) e Seleção de Atributos

Algoritmos Testados: *Least Median Squared, Linear Regression, Multi-Layer Perceptron, RBF Network, SMOReg, Gaussian Processes*

Paradigma de Aprendizagem dos Algoritmos: Aprendizagem baseada em Função

Algoritmo Selecionado: *SMOReg*

Avaliação dos Resultados: Medidas de Qualidade para Previsão

Técnicas Estatísticas: Kolmogorov-Smirnov, Análise de Variância, Regressão Linear, Correlação

Testes Estatísticos Paramétricos: Correlação de Pearson, Teste T pareado, ANOVA

Testes Estatísticos Não-Paramétricos: Spearman Rank Correlation, Wilcoxon matched-pairs, Nonparametric Repeated Measures ANOVA

Quadro 7 - Séries Temporais Utilizadas

Classe	Quant.	Nome	Unidades	Período	Previsão	Registros		Características predominantes	Descrição	Fonte	
						N	%				
Multivariada Global	1	EV nos países	ano	Anual (1980-2014)	EV 2015-2020	2357	89,23	Não estacionárias; Não sazonais; Não Linear	IDH dos 188 Países filiados a ONU	UNDATA (2015)	14/12/2015
	1	IDH dos países	IDH 2015-2020		UNDP (2015)						
	1	IDH dos países	Anual (1980-2012) intermitente	IDH 2013	1982	88,32	IDH dos 187 Países filiados a ONU				
	1		Anual (1980-2013) intermitente	IDH 2014	2169	89,22					
Multivariada Específica	187	IDH dos países Previsores	Índice	Anual (1980-2012) intermitente	IDH 2013	$\mu=55,45$ $\pm 5,76$ (8-286)	$\mu=2,80$ $\pm 0,29$ (0,40-14,43)	Não estacionárias; Não sazonais; Não Linear	IDH de grupos de país filiado a ONU com capacidade explicativa do IDH de determinado países	UNDP (2014)	24/07/2014
	187			Anual (1980-2013) intermitente	IDH 2014	$\mu=54,56$ $\pm 2,80$ (10-112)	$\mu=2,52$ $\pm 0,13$ (0,46-5,16)				
Univariada	187	IDH por país		Anual (1980-2012) intermitente	IDH 2013	$\mu=11,60$ $\pm 2,41$ (4-13)	$\mu=89,23$ $\pm 18,54$ (30,77-100)	Não estacionárias; Não sazonais; Não Linear	IDH de cada país filiado a ONU		
	187			Anual (1980-2013) intermitente	IDH 2014	$\mu=12,60$ $\pm 2,41$ (5-14)	$\mu=90,00$ $\pm 18,54$ (35,71-100)				

### 3.5.1 Medidas de Qualidade para Previsões

A literatura oferece diversas medidas estatísticas para avaliar previsões, conhecidas como medidas de qualidade de previsões de séries temporais.

Neste estudo, foram utilizadas somente as medidas disponíveis na API do *Forecast do WEKA*, que englobam as mais utilizadas na literatura. Entre os estudos que utilizam as mesmas medidas de qualidade de previsão, destaca-se o de Napagoda (2013), que compara modelos de previsão utilizando diferentes técnicas e ambientes de mineração de dados. São elas:

Erro absoluto médio (MAE<sup>19</sup>):

$$MAE = \frac{1}{N} \sum |(previsto - real)| \quad (13)$$

Erro quadrático médio (MSE<sup>20</sup>):

$$MSE = \frac{1}{N} \sum (previsto - real)^2 \quad (14)$$

Raiz quadrada do erro quadrático médio (RMSE<sup>21</sup>)

$$RMSE = \sqrt{\frac{1}{N} \sum (previsto - real)^2} \quad (15)$$

Erro percentual absoluto médio (MAPE<sup>22</sup>):

---

<sup>19</sup> MAE – do inglês - *Mean absolute error*.

<sup>20</sup> MSE - do inglês - *Mean squared error*.

<sup>21</sup> RMSE - do inglês - *Root mean squared error*.

<sup>22</sup> MAPE - do inglês - *Mean absolute percentage error*.

$$MAPE = \frac{1}{N} \sum \left| \frac{(\text{previsto} - \text{real})}{\text{real}} \right| \quad (16)$$

Precisão direcional (DAC<sup>23</sup>)

$$DAC = \frac{100}{N} \sum a_t \% \quad a_t = \begin{cases} 1, & (\text{real}_{(t)} - \text{real}_{(t-1)})(\text{previsto}_{(t)} - \text{previsto}_{(t-1)}) \geq 0 \\ 0, & \text{caso contrário} \end{cases} \quad (17)$$

Erro absoluto relativo (RAE<sup>24</sup>):

$$RAE = \frac{\frac{1}{N} \sum |(\text{previsto} - \text{real})|}{\frac{1}{N} \sum |(\text{alvo previsto} - \text{real})|} \quad (18)$$

Raiz quadrada do erro quadrático relativo (RRSE<sup>25</sup>):

$$RRSE = \sqrt{\frac{\frac{1}{N} \sum (\text{previsto} - \text{real})^2}{\frac{1}{N} \sum (\text{alvo previsto} - \text{real})^2}} \quad (19)$$

---

<sup>23</sup> DAC - do inglês - *Directional accuracy*.

<sup>24</sup> RAE - do inglês - *Relative absolute error*.

<sup>25</sup> RRSE - do inglês - *Root relative squared error*.

### 3.6 EXPERIMENTO: PREVISÃO DO IDH E DA EV DOS PAÍSES POR MEIO DE TÉCNICAS DE MINERAÇÃO DE DADOS

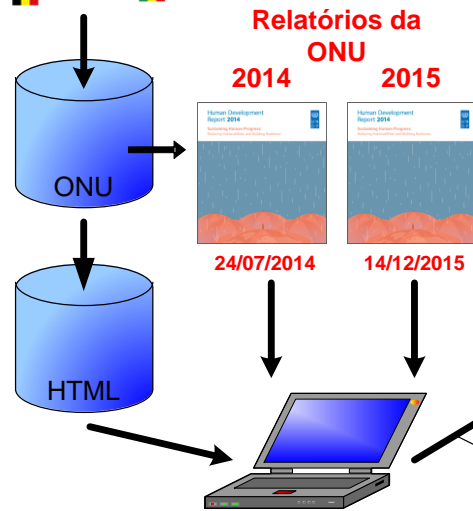
A previsão foi desenvolvida seguindo as etapas do processo de DCBD (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996), que são apresentadas na Figura 4 e estão divididas, segundo definições de Michalski e Kaufman (1998), em três macros etapas: (i) Pré-processamento, (ii) Mineração de Dados: etapa onde se aplicam os algoritmos mineradores; (iii) Pós-Processamento: todas as subetapas utilizadas para se consolidar o conhecimento.

#### 3.6.1 Pré-Processamento

O pré-processamento iniciou-se com a obtenção dos dados de IDH e EV da base do UNDP (UNDATA, 2014) de todos 187 países filiados, atualizada em 24 de julho de 2014 e do seu relatório referente ao IDH 2013 (UNDP, 2014). Esta fonte de dados pode sofrer atualizações pontuais a qualquer momento, e a cada divulgação de um novo relatório de IDH anual, as séries temporais podem sofrer atualizações mais significativas.

A partir destas fontes, foi desenvolvida uma base de dados específica, através do Microsoft Access, com as séries temporais referentes ao período intermitente de tempo de 1980 a 2013. Após a implementação desta base de dados, foi realizada a etapa de DCBD de "exploração da base de dados", por meio da Linguagem de Consultas Estruturadas (SQL). Os dados foram organizados em planilhas de cálculo, onde foram realizados procedimentos estatísticos, resultando na estatística descritiva das séries temporais, apresentada na Tabela 2 e incorporadas suas estatísticas de resumo às Tabelas 7 e 8, apresentada na Seção 5 - Resultados.

Informações dos Países



Processo de KDD

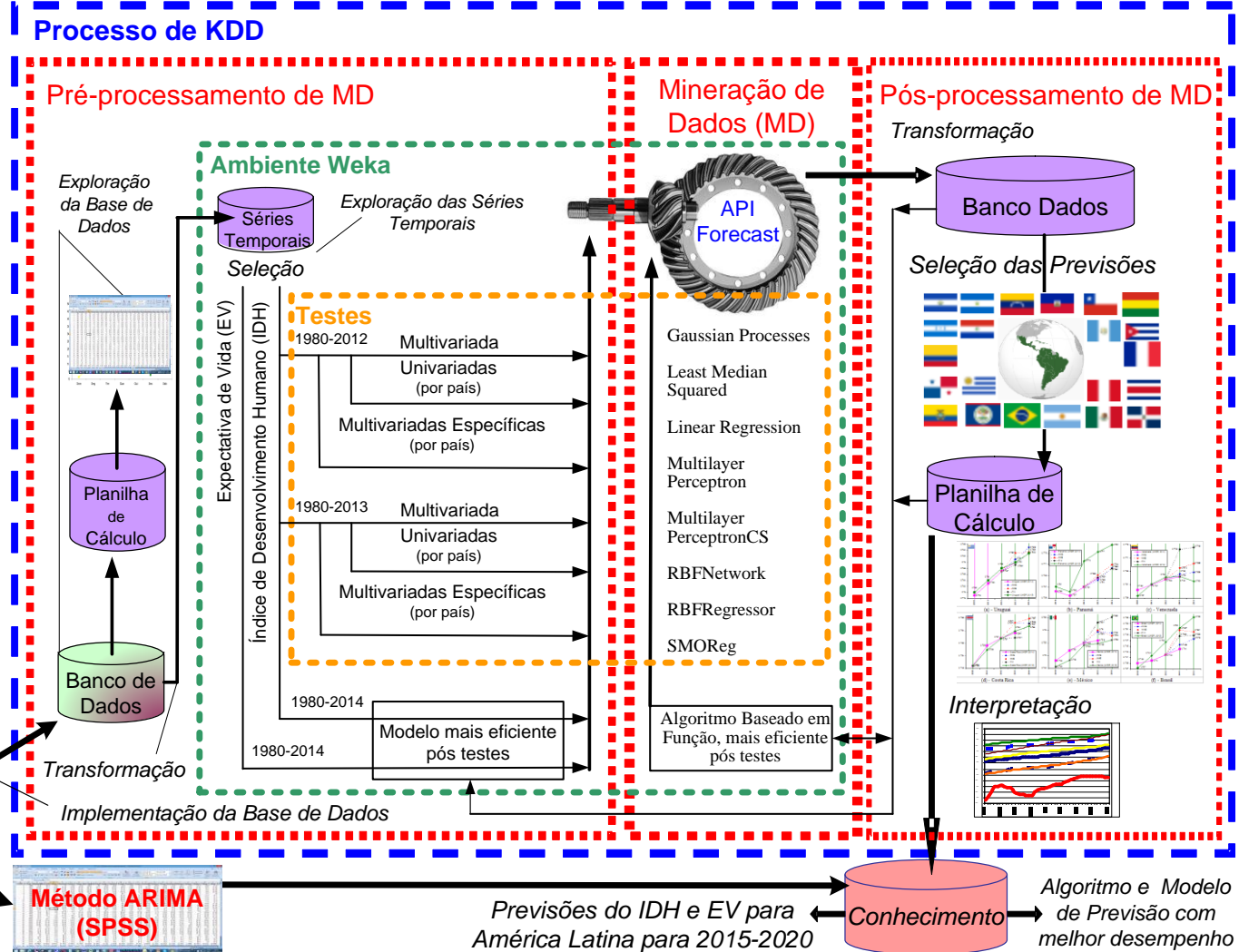


Figura 4 - Etapas de Desenvolvimento da Pesquisa

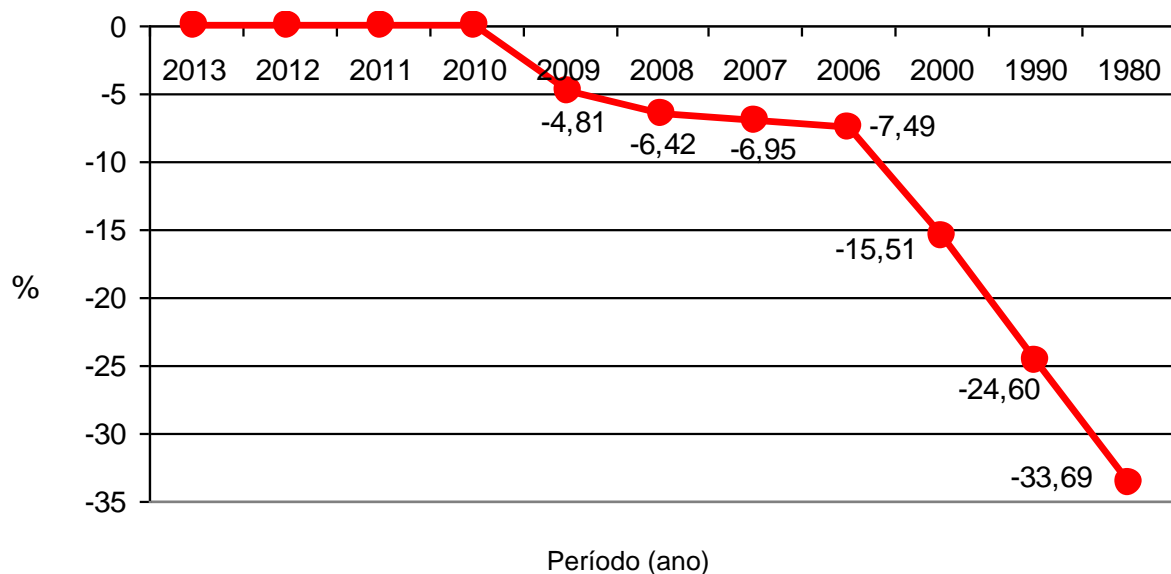


**Tabela 2 - Estatística Descritiva das Séries Temporais do IDH Mundial no Período de 1980 a 2013**

Estatística Descritiva	p/Período												p/País
	1980	1990	2000	2005	2006	2007	2008	2009	2010	2011	2012	2013	1980-2013
Média	0,538	0,587	0,620	0,642	0,648	0,656	0,662	0,665	0,670	0,672	0,675	0,686	0,649
Erro padrão	0,018	0,016	0,015	0,014	0,014	0,014	0,014	0,013	0,013	0,013	0,012	0,011	0,012
Mediana	0,556	0,615	0,652	0,683	0,689	0,695	0,700	0,705	0,709	0,710	0,712	0,717	0,665
Desvio padrão	0,185	0,181	0,188	0,185	0,184	0,182	0,180	0,175	0,172	0,172	0,171	0,156	0,042
Var. amostra	0,034	0,033	0,035	0,034	0,034	0,033	0,032	0,031	0,030	0,029	0,029	0,024	0,002
Curtose	-1,128	-0,867	-0,943	-1,000	-0,985	-0,958	-0,926	-0,898	-0,853	-0,846	-0,848	-0,792	0,512
Assimetria	-0,175	-0,409	-0,338	-0,342	-0,354	-0,378	-0,406	-0,407	-0,421	-0,422	-0,427	-0,433	-0,978
Intervalo	0,681	0,682	0,688	0,690	0,689	0,674	0,668	0,661	0,657	0,656	0,651	0,607	0,122
Mínimo	0,176	0,198	0,234	0,258	0,262	0,278	0,282	0,289	0,295	0,297	0,304	0,337	0,564
Máximo	0,857	0,880	0,922	0,948	0,951	0,952	0,950	0,950	0,952	0,953	0,955	0,944	0,686
Contagem	111	132	152	169	164	166	167	175	187	187	187	187	11,606
N.C (95,0%)	0,035	0,031	0,030	0,028	0,028	0,028	0,027	0,026	0,025	0,025	0,025	0,023	0,023

Observa-se, a partir da Tabela 2, que o IDH dos países filiados a ONU, ao longo do período, apresenta média crescente, e que o desvio padrão é decrescente. Os dados apresentam alta homogeneidade (baixa dispersão ou variabilidade). Constatou-se, também, que 65,78% dos países apresentavam dados completos nas suas séries temporais. A ausência de dados é retroativa a 2010 (Figura 5). A ausência de dados não foi tratada, pois ocorria no início das séries temporais de alguns países e afastadas do último ponto de observação.

**Figura 5 - Percentual de Ausência de Dados das Séries Temporais de IDH dos Países**

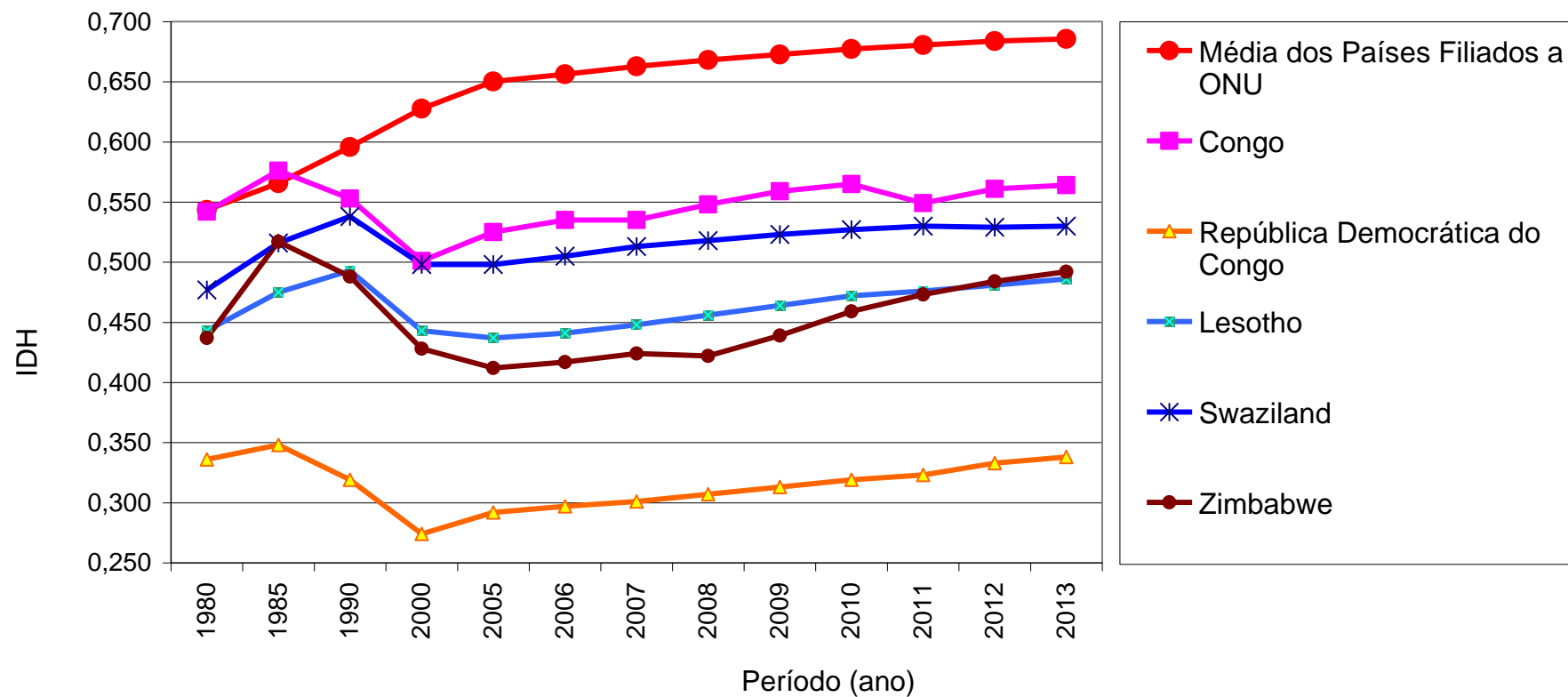


Em seguida, os dados foram transformados no formato arff (*Attribute-Relation File Format*) e prosseguiu-se com a exploração das séries temporais no ambiente WEKA.

A partir da mineração visual, da estatística descritiva e da análise de correlação, as séries temporais foram caracterizadas, segundo as definições de Palit e Popovic (2006b), em: (i) Não estacionárias: apresentam comportamento de crescimento; (ii) Não sazonais: não apresentam padrões de comportamento em períodos regulares de tempo; e (iii) Não lineares: não apresentam média e variância constante.

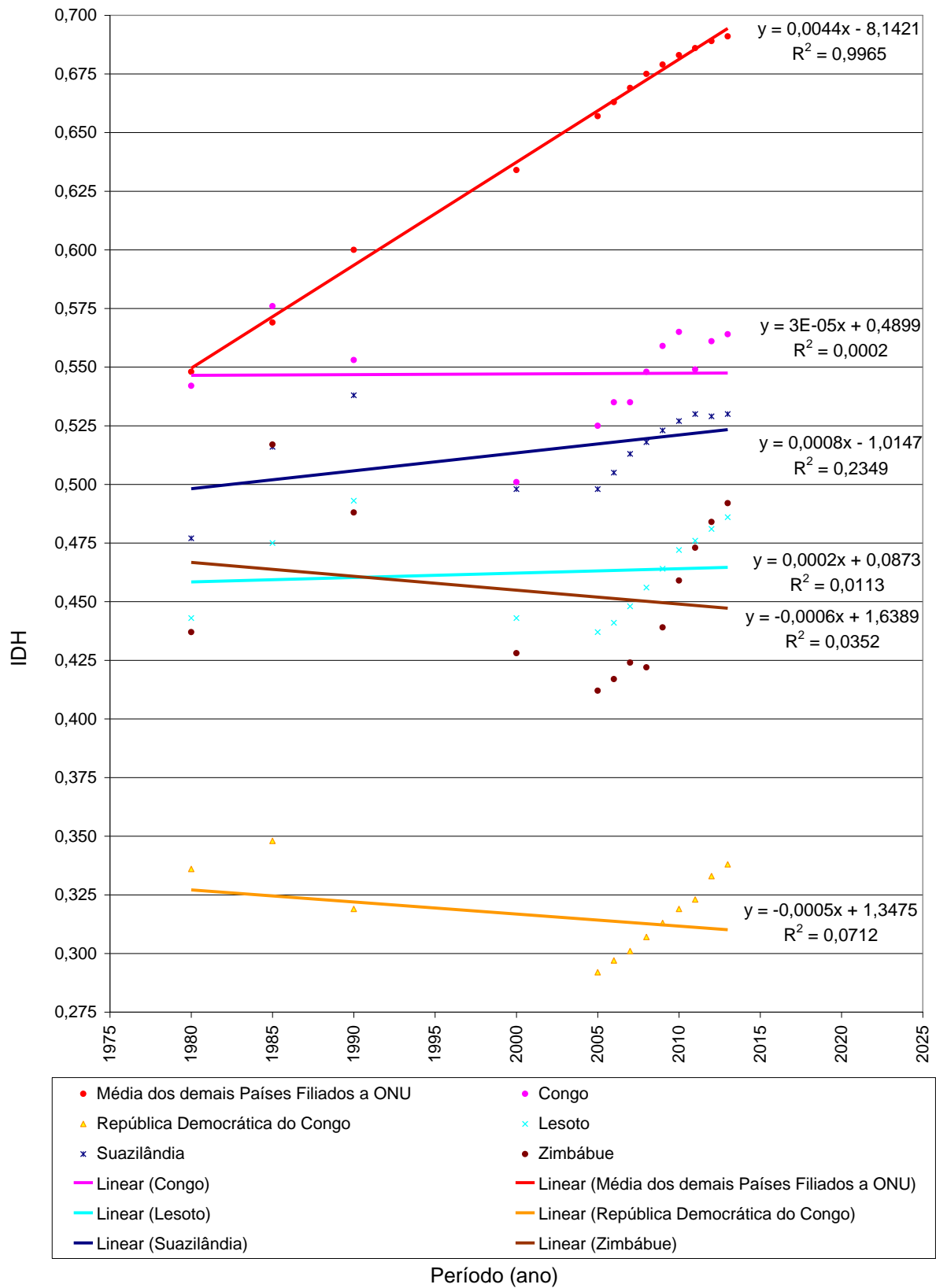
Observou-se, também, que as séries multivariadas apresentavam alta correlação entre si, exceto com as séries temporais de cinco países que se apresentaram como *outliers* (Congo, República Democrática do Congo, Lesoto, Suazilândia e Zimbábue) e que também apresentaram correlação com o tempo (Figura 6):

Figura 6 - Séries Temporais das médias de IDH dos países filiados a ONU excluindo os países: Congo, República Democrática do Congo, Lesoto, Suazilândia e Zimbábue, apresentados separadamente



A Figura 7 apresenta Regressões Lineares Temporais do IDH, em que se observa que o tempo tem capacidade explicativa de 99,65% da média do IDH dos países filiados a ONU, excluindo-se os países: Congo, República Democrática do Congo, Lesoto, Suazilândia e Zimbábue. Para estes países, o tempo apresenta baixa capacidade explicativa, variando de 0,002 a 23,49%.

**Figura 7 - Regressão Linear entre o Tempo (Ano) e o IDH dos Países Filiados a ONU e de países como baixa correlação temporal**



Apesar da detecção de possíveis anomalias de dados, foram mantidos todos os registros dos IDH de todos os países, mesmos os aparentes *outliers*, pois estes não foram considerados erros, mas valores surpreendentes e corretos, díspares do padrão das demais séries temporais.

A API *Forecast* não requer uma pré-análise detalhada das séries temporais, como ocorre durante o desenvolvimento de um modelo utilizando o método ARIMA. No entanto, foram realizados, através do software SPSS, outros testes para melhor caracterizar os dados para possíveis comparações com estudos futuros. Entre eles, destacam-se os testes de autocorrelação e correlação cruzada nas séries temporais de um a cinco períodos de defasagem. Nos resultados dos testes, observou-se moderada autocorrelação ( $0,33 < r < 0,66$ ) em um período de defasagem nas séries temporais univariadas e, a partir de dois períodos de defasagem, a maioria das séries temporais apresentava autocorrelação baixa ( $r < 0,33$ ). Enquanto que, nos testes de correlação cruzada, observa-se baixa correlação entre os IDH dos países para a maioria das defasagens.

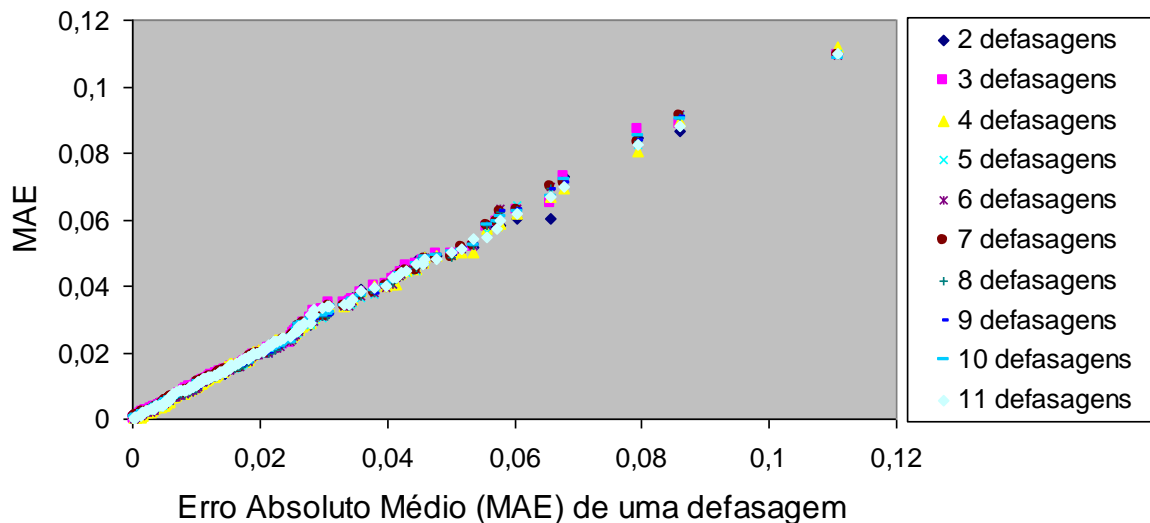
Ao final do Pré-Processamento de Mineração de Dados, foram selecionadas 750 séries temporais do IDH para testes, separadas em dois lotes de dados, sendo o primeiro para testar a previsão do IDH 2013 e o segundo para previsão do IDH 2014, com dados anteriores ao período de previsão. Cada lote de dados foi utilizado para o desenvolvimento de: i) um modelo multivariado global (MMG), treinado com séries multivariadas correspondentes aos 187 países filiados à UNDP; ii) 187 modelos multivariados específicos (MME), treinados com grupos que variavam de dois a 45 países com capacidade explicativa do índice de cada país filiado a UNDP; iii) 187 modelos univariados (MU), treinados com séries correspondentes a cada país, resultando no total de 375 modelos por lote de dados.

### 3.6.2 Mineração de Dados

A etapa de Mineração de Dados consiste na aplicação de um ou vários algoritmos de aprendizado de máquina. Nesta etapa, foi configurada a API e também foram selecionados, testados e configurados os algoritmos disponíveis para mineração.

A API foi configurada considerando-se testes em relação ao número máximo de defasagens (lag do inglês) e o ajuste ou não da variância. Primeiramente, foram realizados testes de previsão nos Modelos Multivariados Globais (MMG) com diferentes configurações de máxima defasagem (1-11). Por meio do teste de análise de variância de Friedman (*Nonparametric Repeated Measures ANOVA*), observou-se a inexistência de diferenças significativas ( $p=0,61$ ) em relação ao MAE para diferentes configurações. E com o teste de correlação de *Sperman*, verificou-se alta colinearidade (Figura 8) entre o MAE das diferentes defasagens ( $0,96 < r > 1$ ;  $p < 0,001$ ). A partir da interpretação destes resultados, decidiu-se dispensar a definição de valores máximos de defasagem e adotou-se a configuração padrão estabelecida pelo painel de configuração básica da API para uma periodicidade anual de cinco anos.

**Figura 8 - Correlação entre o Erro Absoluto das Previsões com duas defasagens de tempo e defasagens maiores (2 a 11)**



Em seguida, foram testados os modelos com e sem o ajuste de variância. Segundo Pentaho (2014), o ajuste de variância pode, ou não, melhorar o desempenho dos algoritmos. No caso específico deste trabalho, verificou-se que o ajuste aumentou o MAE, então se decidiu não o adotar. Ao final destes testes preliminares, foi definida a configuração para a API *Forescast* do WEKA, conforme apresentado no Quadro 8:

**Quadro 8 - Configuração da Aplicação de Interface de Programação Forecast**

Configurações do API Forecast				
Básica	Avançada			
Target Selection	Base Learner	Lag creation	Evaluation	Output
set - Contry(s) Parametros Number of time unists to forecast =1 Time stamp <Ano> Periodicity <Yearly> <On> Intervalo de confiança <On> Perform evaluation	<i>Gaussian Processes;</i> <i>Least Median Squared ;</i> <i>Linear Regression;</i> <i>Multi-Layer Perceptron;</i> <i>RBFNetwork;</i> <i>SMOReg;</i>	<Off> Ajust for variance <Off> Use custom lag lengths <Off> Fine tune lag selection	<On> evaluate on training	<On> Output future predictions beyout end of series

### 3.6.2.1 Seleção do algoritmo

Para a seleção do algoritmo mais adequado ao estudo, foram testados os algoritmos pertencentes ao grupo com paradigma de aprendizagem baseada em funções: *Least Median Squared*, *Linear Regression*, *Multi-Layer Perceptron*, *RBF Network*, *SMOReg*, *Gaussian Processes*. Mantiveram-se suas configurações padrão definidas no ambiente WEKA.

Para poupar custos operacionais, os testes preliminares foram realizados apenas para previsão do IDH 2013 e 2014.

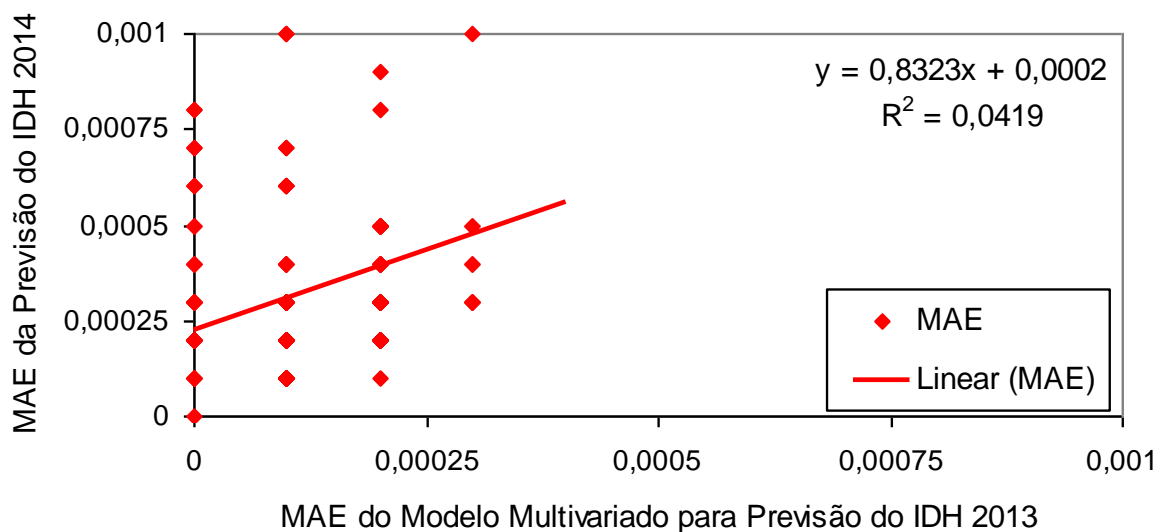
Apenas os algoritmos *Gaussian Processes* e *SMOReg* apresentaram custo operacional satisfatório, com tempo de resposta menor que 10 minutos nos MMGs. Já nos MMEs e MUs, todos os algoritmos testados apresentaram tempo de resposta satisfatório, menor que 5 segundos. Os algoritmos *Multi-Layer Perceptron*, *Linear Regression* e *Least Median Squared Linear Regression* não permitiram a utilização de todos os 187 países simultaneamente nos MMs (MMG e MME). Ainda os algoritmos *Linear Regression* e *Least Median Squared Linear Regression*, também, não permitiram a utilização de conjuntos de países nos MMEs. (Tabela 6 - Seção 5 - Resultados).

Ao final dos testes preliminares, foi selecionado o algoritmo *SMOReg*, por apresentar os melhores resultados (Tabela 6 - Seção 5 - Resultados), tanto nos modelos multivariados como nos univariados, além de custo operacional significativamente reduzido em relação aos demais.

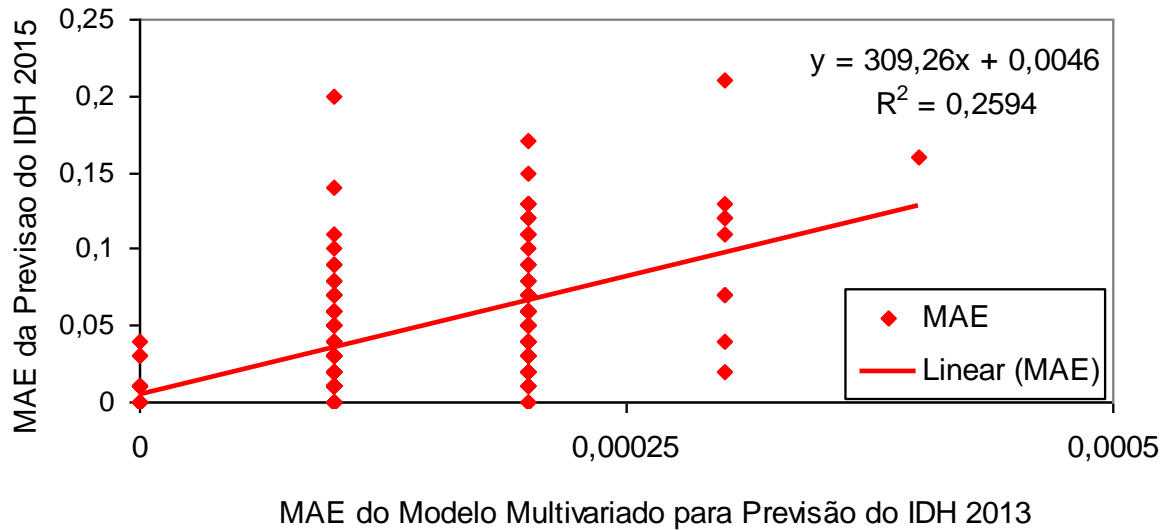


O algoritmo *SMOReg* ainda foi testado considerando parâmetros de avaliação da eficiência de modelos sugeridos por Putsis (1998), Lawrence, O'Connor e Edmundson (2000), respectivamente em relação correlação e distribuição dos erros. Primeiramente, o algoritmo foi testado na evolução de um modelo multivariado para verificar a correlação temporal dos erros, que segundo Putsis (1998) pode caracterizar ineficiência de um modelo de previsão. Então, com retroalimentação do valor de IDH de 2013 nas séries temporais inicialmente utilizadas para a previsão do IDH 2013, foram realizadas previsões de três *steps* subsequentes (IDH 2014 a 2016). E através do teste de *Spearman Rank Correlation*, verificou-se baixa correlação do MAE no horizonte de previsão de três pontos de observações subsequentes (Figura 9, Figura 10 e Figura 11).

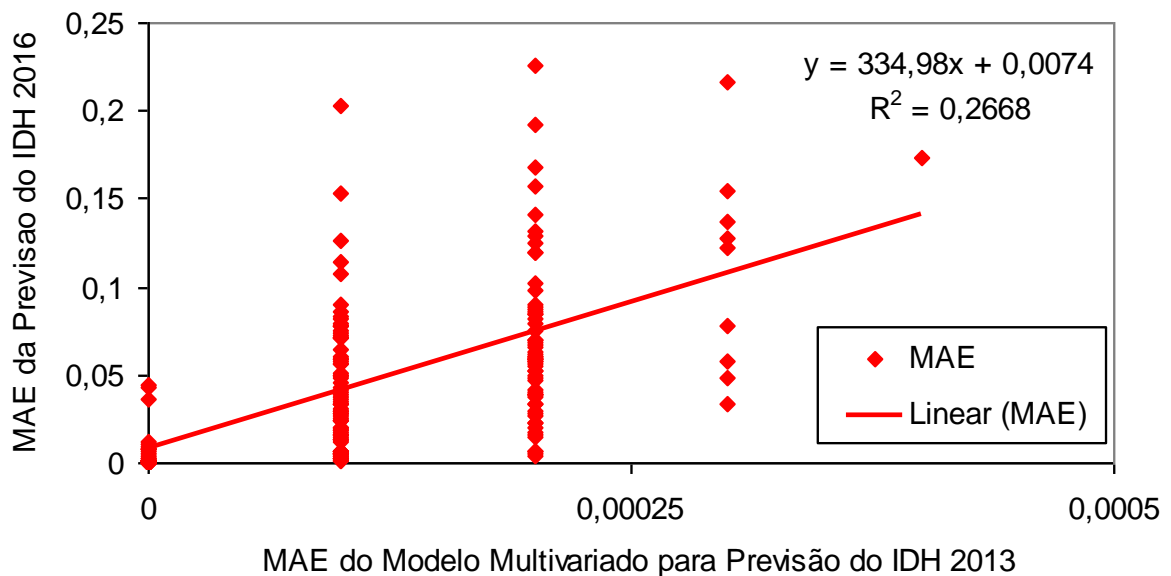
**Figura 9 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2014**



**Figura 10 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2015**



**Figura 11 - Correlação entre MAE do Modelo de Previsão do IDH 2013 com o MAE da Previsão do IDH 2016**



A baixa correlação entre os erros apresentados nas Figuras 9, 10 e 11 indica que o algoritmo tem capacidade de aprender com os erros do passado.

O algoritmo ainda foi testado nas séries temporais univariadas para verificar se a distribuição dos erros dos treinos tendia a uma distribuição de Gaus, através do teste de normalidade de *Kolmogorov-Smirnov* (KS). Isto se confirmou na maioria dos modelos, principalmente daqueles com as séries temporais completas. Segundo

Lawrence, O'Connor e Edmundson (2000), em modelos eficientes, a distribuição dos erros devem ter uma forma próxima da normal.

Para o desenvolvimento dos MMEs, foram selecionados conjuntos de dados de IDH de países candidatos a previsores de cada país da filiado a UNDP (atributo meta). Os conjuntos de dados foram escolhidos através do algoritmo de seleção baseada em correlação: *Correlation-based Feature Selection (CFS)* (HALL, 1999), por meio do método de validação cruzada. Este algoritmo prioriza conjuntos de atributos (variáveis independentes) mais relacionados com o atributo meta (variável dependente) e pouco relacionados entre si. O algoritmo selecionou, em média,  $4,88 \pm 3,68$  países previsores por modelo. Os MMEs desenvolvidos para os países Andorra e Turkmenistan apresentaram o maior número de países previsores selecionados ( $n=22$ ), enquanto que os países Antigua and Barbuda, Bahrain, Gabon e Ghana foram desenvolvidos com o menor número de países previsores ( $n=2$ ).

Finalizando a etapa de Mineração de Dados, 750 modelos foram desenvolvidos com o uso do algoritmo *SMOReg*: dois MMGs, 374 MMEs e 374 MUs, para a previsão do IDH de 187 países nos períodos de 2013 e 2014.

Os MMGs foram treinados com os dados de 187 países para aumentar a experiência do algoritmo.

Exemplos destes modelos podem ser observados nos quadros apresentados no Apêndice 1. Neste Apêndice, o Quadro 9 apresenta um MU para previsão do IDH do país Botswana (BWA). O Quadro 10 apresenta um MME de países previsores do IDH do país New Zealand (NZL). O Quadro 11 apresenta parcialmente um MMG, que na sua forma original é composto de 356.186 linhas.

### 3.6.3 Pós-Processamento

Os resultados dos modelos alimentaram uma base de dados e, através de SQLs específicas, esses resultados foram organizados por modelos e pareados por países. Isto permitiu comparações entre os valores reais e as previsões, assim como entre as medidas de qualidade dos modelos. Os valores reais referentes ao IDH 2013 foram consultados no UNDP (2014), em 24 de julho de 2014, enquanto os

valores referentes ao IDH 2014 foram consultados após sua atualização e divulgação realizada pelo UNDP (2015), em 14 de dezembro de 2015. Também, alimentaram a base de dados medidas de qualidades calculadas por meio de planilhas eletrônicas, a partir dos últimos pontos de observação para comparação com o valor real.

As Figuras 19 a 32 (Apêndice 2) apresentam as medidas de qualidade das previsões dos MMGs, MMEs e MUs para o IDH de 2013 e 2014. Os MMGs apresentam as melhores medidas de qualidade das previsões.

Os testes de análise de variância Friedman Test (*Nonparametric Repeated Measures ANOVA*) apontaram diferenças significativas entre as medidas de qualidade dos modelos ( $p < 0,001$ ). Os MMGs apresentaram as maiores médias da DAC e as menores médias de erros que os MMEs e MUs (Tabela 3):

Tabela 3 - Estatística de Resumo das Medidas de Qualidade das Previsões de IDH por Ano e Modelo

Previsão		2013						2014					
Modelo		MMG		MME		MU		MMG		MME		MU	
Estatística		$\mu$	$\pm$	$\mu$	$\pm$	$\mu$	$\pm$	$\mu$	$\pm$	$\mu$	$\pm$	$\mu$	$\pm$
Medida de Qualidade	DAC	96,52	7,45	93,73	8,55	91,57	10,10	96,29	6,63	93,52	7,85	91,15	9,46
	MAE	0,00013	0,00008	0,00066	0,00048	0,00147	0,00099	0,00013	0,00008	0,00065	0,00047	0,00160	0,00122
	MAPE	0,02344	0,01707	0,11435	0,09967	0,25847	0,21078	0,02419	0,01862	0,11334	0,09977	0,28800	0,31703
	MSE	0,00000	0,00000	0,00000	0,00000	0,00000	0,00002	0,00000	0,00000	0,00000	0,00000	0,00000	0,00002
	RAE	2,55	1,18	12,97	9,45	27,17	15,85	2,67	1,23	13,06	9,16	29,32	18,30
	RMSE	0,00014	0,00009	0,00121	0,00099	0,00244	0,00175	0,00014	0,00009	0,00119	0,00087	0,00265	0,00197
	RRSE	2,50	1,19	19,98	14,06	37,38	20,12	2,62	1,25	19,94	13,23	40,09	22,09

Verificou-se, através do teste *Spearman Rank Correlation*, correlações entre as medidas de qualidade das previsões e características estatísticas do IDH, como o número de observações e a variância do índice (Tabela 4):

**Tabela 4 - Correlação entre as Medidas de Qualidade das Previsões e as Medidas Estatísticas das Séries Temporais do Atributo Meta: Número de Elementos, Variância do IDH**

Medida de Qualidade das Previsões	Número de Observações			Variância do IDH		
	MMG	MME	MU	MMG	MME	MU
DAC	0,17	-	-	0,23	0,26	0,30
MAE	0,67	0,52	0,34	0,90	0,48	0,48
MAPE	0,66	0,51	0,35	0,86	0,50	0,40
MSE	-	-	-	-	-	-
RAE	0,84	0,53	0,37	0,72	0,28	0,12
RMSE	0,67	0,43	0,29	0,92	0,29	0,37
RRSE	0,81	0,48	0,33	0,76	0,10	0,10

A quantidade observações (número de exemplos) das séries temporais dos atributos metas apresenta baixa correlação como a DAC, com significância ( $p < 0,05$ ) apenas nos MMG ( $r = 0,17$ ). E, também, apresenta correlação com os tipos de erro, exceto com o MSE. Estas correlações são classificadas com alta ( $r > 0,66$ ) nos MMGs; moderada ( $0,33 < r < 0,67$ ) nos MMEs; e, moderada e baixa ( $r < 0,33$ ) nos MUs.

A variância do IDH das séries temporais apresenta baixa correlação como a DAC, alta correlação como os tipos de erros nos MMGs e varia entre moderada e baixa correlação nos MMEs e MUs. Isto sugere que, quanto maior o número de elementos e/ou a variância das séries temporais do atributo meta, piores as medidas de qualidade relativas aos erros, principalmente nos MMGs.

Especificamente em relação aos MMEs, buscaram-se possíveis correlações através do teste *Spearman Rank Correlation*, que envolvessem o número de países previsores, com características das séries temporais (variância e ausência de dados), medidas de qualidades das previsões e erro absoluto da previsão do IDH 2013. No entanto, não se observaram correlações significativas entre estas variáveis.

As previsões do IDH dos modelos utilizando o algoritmo *SMOReg*, realizadas durante os testes para escolha do melhor modelo, são apresentadas nas Figuras 33 a 36, e Tabela 9, no Apêndice 3. Nestas figuras, os resultados dos modelos e previsões estão organizados por ordem decrescente do IDH 2013 e foram elaborados antes da divulgação do IDH 2014. Cada figura agrupa uma categoria de país (desenvolvido, em desenvolvimento e subdesenvolvido). Suas linhas correspondem aos valores de IDH de 2013 divulgados em 24/07/2015 e os valores previstos pelos modelos, enquanto que os pontos correspondem aos valores previstos para IDH 2014, até então não divulgados.

A Tabela 9 apresenta os dados reais referentes ao IDH 2012 e 2013, pareados com as previsões dos modelos multivariados (globais e específicos) e univariados para os IDH 2013 e 2014, onde estão também classificados por classes (nível e tipos de IDH) e ranqueados por país, em relação ao IDH 2013.

Observa-se, através da Figura 33, que no grupo de países desenvolvidos, Reino Unido (GBR), Andorra (AND) e Cuba (CUB) apresentam valores de IDH 2013 acima das expectativas dos modelos. Já no grupo de países em desenvolvimento com IDH elevado (Figura 34), 16,98% do grupo apresentaram valores acima das expectativas das previsões, são eles: Bahamas (BHS), Montenegro (MNE), Romênia (ROU), Omã (PSE), Antígua e Barbuda (ATG), Seychelles (SYC), Sérvia (SRB), São Vicente e Granadinas (VCT), Suriname (SUR). No mesmo grupo, mas com IDH médio (Figura 35), os países Turquemenistão (TKM), República da Moldávia Republic of Moldova (MDA) e Quirguistão (KGZ), o que equivale a 7,14% do grupo, apresentaram valores maiores que as previsões.

No grupo de países subdesenvolvidos (Figura 36), apenas o país Lesoto (LSO) apresentou valor superior às previsões dos modelos.

A partir da Tabela 9, observa-se que a Líbia (LBY) e República Árabe Síria (SYR) apresentaram a maior queda no IDH (-0,005) e de cinco posições no ranking. As expectativas dos modelos que consideram apenas os dados históricos do IDH eram de que a Líbia aumentaria o índice, ficando entre 0,7974 (MMG) a 0,8123 (MU), enquanto que o índice da República Árabe Síria (SYR) também aumentaria, ficando entre 0,6634 (MMG) a 0,6674 (MME), o que não ocorreu nestes países. Estas contradições possivelmente ocorreram em consequência da guerra civil que os países vêm sofrendo.

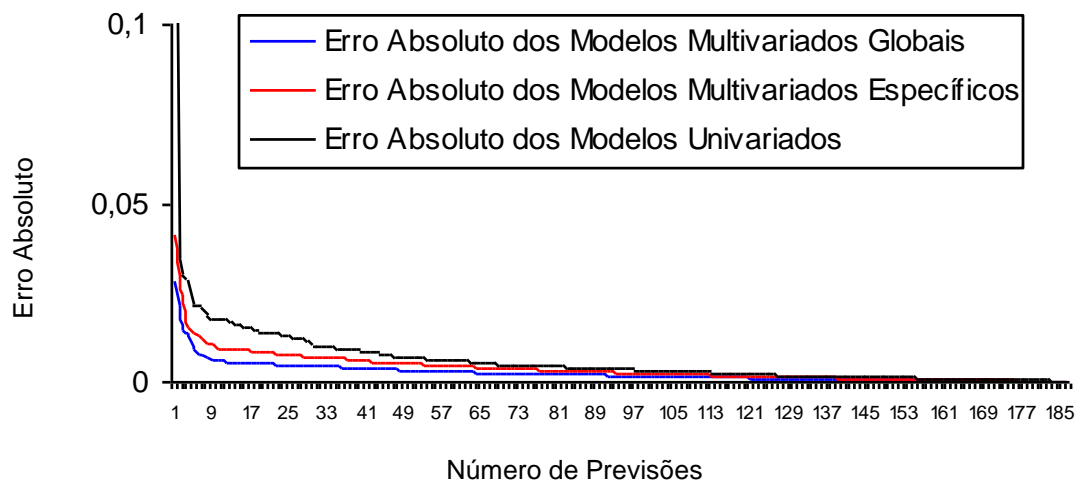
Zimbábue (ZWE) foi o país que mais subiu posições (quatro postos). Do total dos 187 países avaliados, 26 desceram e 31 subiram no ranking, enquanto que os demais, 130 países, se mantiveram no mesmo posto.

Em relação ao valor absoluto do IDH 2013, observou-se que 33,16% dos valores de IDH ficaram entre os intervalos das previsões dos modelos multivariados e univariados. Ainda, que 58,29% foram menores simultaneamente que as previsões dos três tipos de modelos e que 8,55% foram simultaneamente maiores que os previstos pelos modelos. A maioria dos valores absolutos não apresentou diferenças significativas com as previsões, principalmente em relação ao MMG.



A partir do teste de análise de variância ANOVA pareado por país, utilizando como controle o IDH 2013, observou-se que apenas o MMG não apresentou diferenças significativas ( $p > 0,05$ ). E a partir do Friedman Test (*Nonparametric Repeated Measures ANOVA*), observou-se que os modelos apresentaram diferenças significativas ( $p < 0,05$ ) entre seus erros absolutos. O MMG ( $\mu = 0,0026 \pm 0,0031$ ) apresentou os menores erros absolutos que os MMEs ( $\mu = 0,0039 \pm 0,0047$ ) e MUs ( $\mu = 0,0066 \pm 0,0016$ ) apresentados na Figura 12.

**Figura 12 - Erro Absoluto dos Modelos (MMG, MME e MUs) para previsão do IDH 2013**



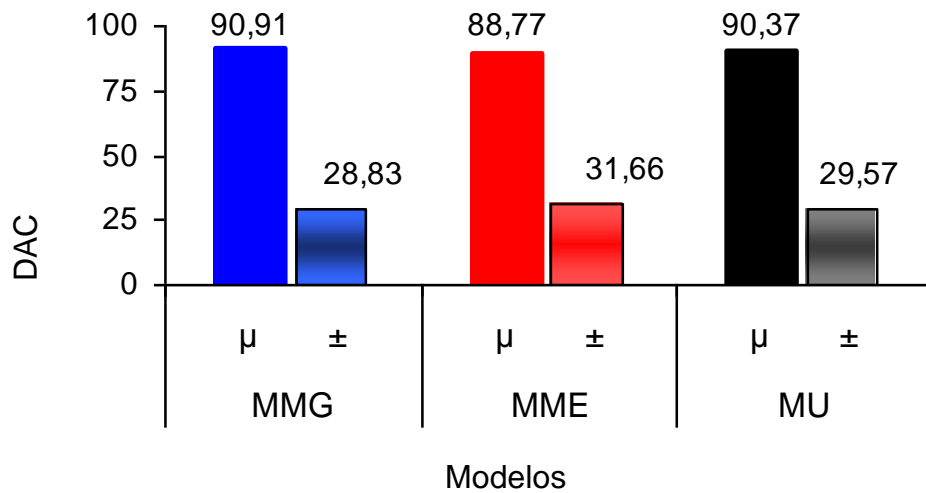
Verificou-se que 49,73% das previsões com menor erro absoluto envolviam os MMGs, enquanto que 33,16% envolviam o MME e 20,32% os MUs. O empate dos erros absolutos ocorridos em mais de um modelo representou 3,21% das previsões.

O menor erro absoluto dos MMG confirma suas melhores medidas de qualidades das previsões entre os modelos.

Verificou-se, através do teste de *Spearman Rank Correlation*, baixa correlação entre o erro absoluto e a ausência de dados, e moderada correlação entre a variância das séries temporais nos diferentes modelos, levando ao entendimento de que a ausência de dados, principalmente longe dos últimos pontos de observações, não prejudicou os modelos, e que a variabilidade das séries temporais tende a diminuir, mesmo que moderadamente, a precisão das previsões.

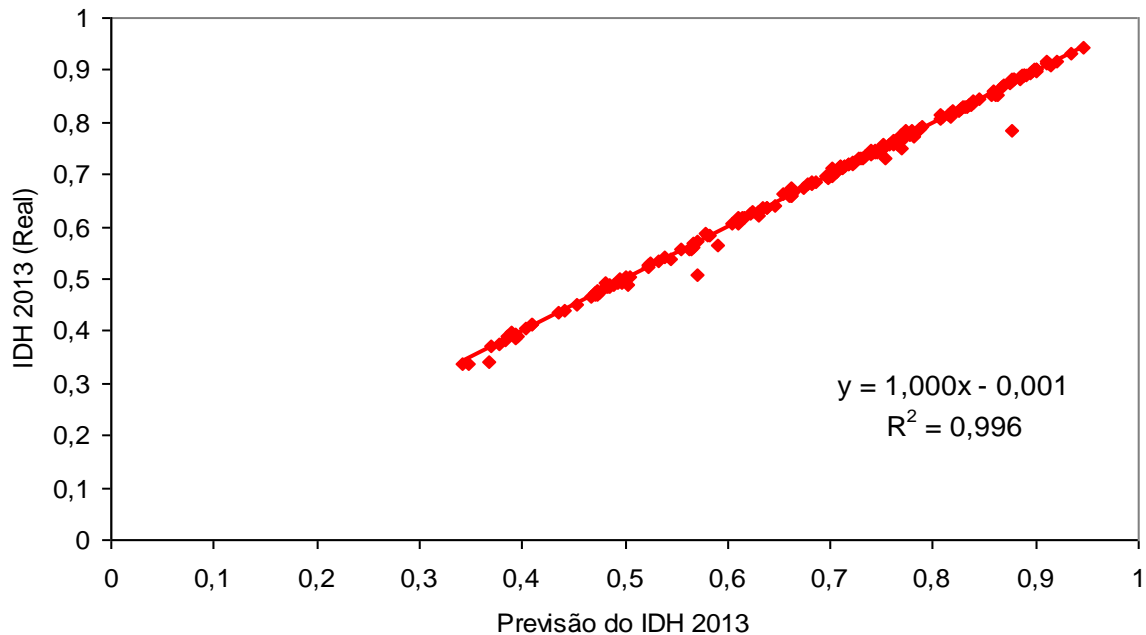
Após os cálculos das DACs dos IDHs dos países, realizou-se o teste de Análise de Variância de Friedman (*Nonparametric Repeated Measures ANOVA*) e não se verificaram diferenças significativas entre os modelos (Figura 13).

**Figura 13 - Média das DACs do último ponto de observação (2012) por modelos para previsão do IDH 2013**



Verificou-se, também, a capacidade explicativa do modelo com o melhor desempenho através do teste de correlação de Pearson. Verificou-se alta correlação ( $r=0.99$ ;  $p<0,0001$ ) das previsões do MMG para o IDH 2013 com as tendências reais divulgadas pela UNDP (2014). As previsões deste modelo apresentam capacidade explicativa de 99,95% da variabilidade da tendência do IDH divulgado pela ONU. A partir da Regressão Linear (Figura 14), observa-se acréscimo de 0,29% e diferença de 0,0039 pontos entre o valor previsto pelo modelo e a tendência do IDH divulgada pela UNDP (2014).

Figura 14 - Previsão do IDH 2013 x IDH 2013 Real a partir de Séries Temporais Multivariadas



Depois de percorridas todas as etapas de DCBD para testar algoritmos e modelos, retornaram-se as etapas anteriores, repetindo-se o processo para previsão do IDH e EV no período de 2015 a 2020, aplicando-se apenas o algoritmo e o modelo com melhor desempenho, *SMOReg* e *MMG*, respectivamente. Antes da efetivação da previsão, as séries temporais de 1980 a 2014 foram atualizadas em 14 de dezembro de 2015, agora com dados de 188 países, pois, a cada divulgação de um novo relatório, a base de dados da UNDP (UNDATA, 2014) pode sofrer atualizações mais significativas (UNDP, 2014).

As Figuras 81 a 102, no Apêndice 7, apresentam direções e previsões dos modelos (linhas tracejadas), tendências (linhas contínuas) divulgadas pelo UNDP (2014); 2015) e previsões do *MMG* para IDH 2015 a 2020 dos países da América Latina filiados ao UNDP.

A última etapa do Processo de DCBD, segundo Fayyad et al. (1996) chamada de "Conhecimento", que exige a interpretação dos padrões descobertos, será apresentada como resultados, na seção 5.

## 4 RESULTADOS

Nesta seção, são apresentados os principais resultados do processo de DCBD.

A Tabela 5 apresenta estatísticas de resumo do MAE dos testes realizados para escolha do algoritmo de melhor desempenho, aferindo ao *SMOReg* esta classificação.

**Tabela 5 - MAE dos modelos de previsão desenvolvidos com algoritmo de Mineração de Dados com paradigma de aprendizado de funções**

Previsão		IDH 2013 - 2014					
Algoritmo	Modelo	MMG		MME		MU	
	Estatística	$\mu$	$\pm$	$\mu$	$\pm$	$\mu$	$\pm$
	<i>SMOReg</i>	0,0002 <sup>a</sup>	0,00005	0,0008 <sup>a</sup>	0,0005	0,0014 <sup>a</sup>	0,0007
	<i>Gaussian Processes</i>	0,0011 <sup>b</sup>	0,0008	0,0117 <sup>c</sup>	0,0057	0,0174 <sup>f</sup>	0,0088
	<i>RBF Network</i>	0,0165 <sup>d</sup>	0,0079	0,0161 <sup>d</sup>	0,0062	0,0160 <sup>e</sup>	0,0070
	<i>Multi-Layer Perceptron</i>	**	**	0,0021 <sup>b</sup>	0,005	0,0020 <sup>b</sup>	0,0007
	<i>Linear Regression</i>	**	**	*	*	0,0025 <sup>c</sup>	0,0028
	<i>Least Median Squared</i>	**	**	*	*	0,0044 <sup>d</sup>	0,0046

$p < 0,05$  em  $a < b < c < d < e < f$  (comparados por coluna)

\* não permitiu testes com conjuntos de países

\*\* não permitiu testes com todos os países

A Tabela 6 apresenta estatísticas de resumo das medidas de qualidade dos testes de previsão do IDH 2013-2014 dos países da América Latina, para seleção do melhor modelo de previsão:

**Tabela 6 - Medidas de qualidade dos modelos desenvolvidos para testar a previsão do IDH dos países da América Latina filiados ao UNDP**

Horizonte de Previsão		IDH 2013 - 2014					
Modelo		MMG <sup>a</sup>		MME <sup>b</sup>		MU <sup>c</sup>	
Estatística		μ	±	μ	±	μ	±
Medida de Qualidade	DAC <sup>**</sup>	98,61	4,23	96,46	5,54	96,11	5,72
	MAE <sup>*</sup>	0,0002	0,00005	0,0008	0,0005	0,0014	0,0007
	MAPE <sup>*</sup>	0,026	0,006	0,12	0,07	0,21	0,08
	MSE	0,0	0,0	0,0	0,0	0,0	0,0
	RAE <sup>*</sup>	3,53	0,57	17,49	11,28	29,58	15,11
	RMSE <sup>*</sup>	0,0002	0,00005	0,001	0,001	0,002	0,001
	RRSE <sup>*</sup>	3,51	0,65	25,15	13,80	40,20	17,38

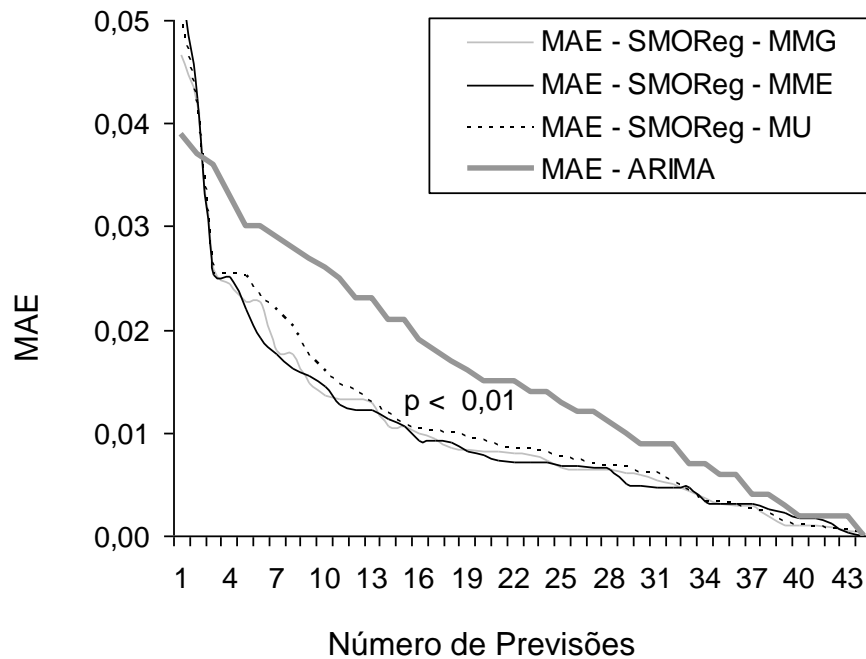
\*  $p < 0,01$  em  $a < b < c$

\*\*  $p < 0,05$  em  $a > b > c$

Verificou-se, por meio da Tabela 6, que os MMGs apresentaram as melhores medidas de qualidade das previsões, correspondendo aos maiores valores de DAC e menores tipos de erros (MAE, RMSE, MAPE, RAE e RRSE), em relação aos MMEs e MUs.

A Figura 15 apresenta o MAE de testes realizados com modelos de previsões adotados neste estudo e comparados com o método ARIMA, para previsão do IDH 2013 e 2014 dos países da América Latina.

**Figura 15 - MAE acumulado por modelos, resultantes das previsões do IDH 2013-2014 dos países da América Latina**



Observou-se que os modelos desenvolvidos através de técnicas de Mineração de Dados apresentaram os menores erros absolutos em relação ao modelo ARIMA. Os correlogramas e séries temporais utilizados para definição inicial dos modelos ARIMA são apresentados nos Apêndices 4 e 5, enquanto que os resultados detalhados dos modelos testados e ajustados são apresentados no Apêndice 6.

A Tabela 7 apresenta os cinco últimos pontos de observações das séries temporais históricas do IDH nos países latino-americanos (UNDATA, 2014), as previsões do índice (2015 a 2020) estimadas pelo algoritmo *SMOReg* em MMGs, o resumo estatístico do índice mundial e na América Latina e seu o percentual de crescimento para o período de projeção.

A Figura 16 apresenta direções e previsões dos modelos (linhas tracejadas), tendências (linhas contínuas) divulgadas pelo UNDP (2014, 2015) e previsões do MMGs do IDH 2015-2020 de alguns países da América Latina, que apresentaram as melhores e piores medidas de qualidade das previsões, apesar de não apresentarem diferenças estatísticas significativas entre os valores nominais das previsões e os valores das tendências já divulgadas (2014 e 2015).

A Figura 17 apresenta a curva de crescimento do IDH mundial, com os valores médio, máximo, mínimo e de variâncias, registrados ao longo do período, além da média da América Latina e seus países com o maior e menor IDH.

A Tabela 8 apresenta os cinco últimos pontos de observações das séries temporais históricas da EV nos países latino-americanos (UNDATA, 2014), as previsões para variável (2015 a 2020) estimadas pelo algoritmo *SMOReg* em MMGs, resumo estatístico da variável no mundo e na América Latina e o seu percentual de crescimento para o período de projeção.

A Figura 18 apresenta a curva de crescimento da EV mundial, com os valores médio, máximo, mínimo e de variâncias, registrados ao longo do período, além da média dos países desenvolvidos, da América Latina e seus países como a maior e menor EV.

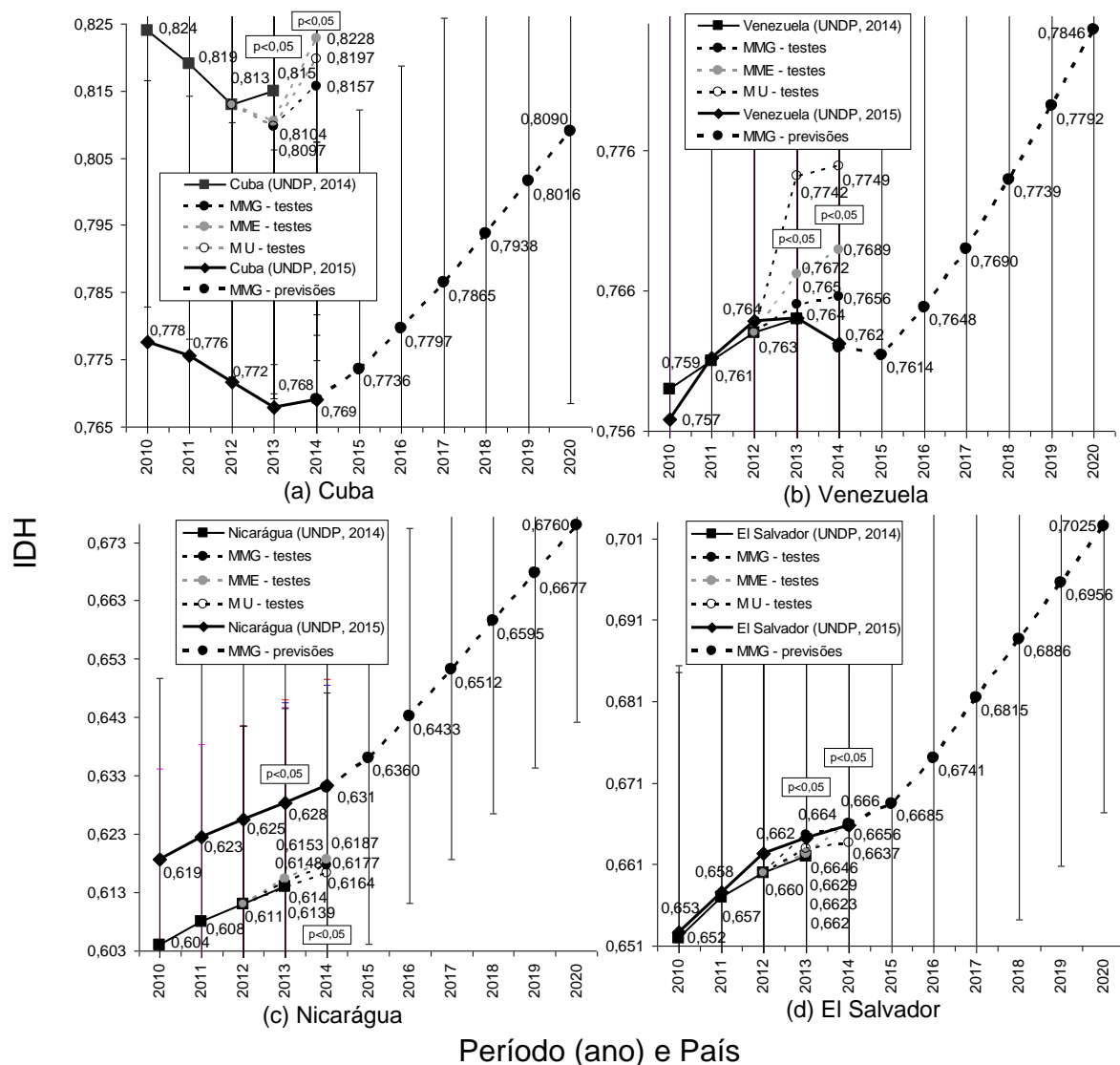
**Tabela 7 - Últimos pontos de observações das séries temporais históricas do IDH dos países da América Latina, suas projeções para 2015 a 2020 e resumo estatístico do índice mundial e dos países latino-americanos**

País*	Últimos Pontos de Observação					Horizonte de Previsão						Variação %	
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2015-2020	
ARG	0,811	0,818	0,831	0,833	0,836	0,843	0,85	0,858	0,865	0,874	0,882	5,50	
CHL	0,814	0,821	0,827	0,83	0,832	0,836	0,844	0,852	0,86	0,869	0,877	5,43	
URY	0,78	0,784	0,788	0,79	0,793	0,799	0,805	0,812	0,819	0,826	0,833	4,98	
PAN	0,761	0,759	0,772	0,777	0,78	0,786	0,793	0,8	0,807	0,815	0,823	5,49	
CUB	0,778	0,776	0,772	0,768	0,769	0,774	0,78	0,787	0,794	0,802	0,809	5,21	
CRI	0,75	0,756	0,761	0,764	0,766	0,77	0,776	0,783	0,789	0,796	0,803	4,78	
VEN	0,757	0,761	0,764	0,764	0,762	0,761	0,765	0,769	0,774	0,779	0,785	2,97	
MEX	0,746	0,748	0,754	0,755	0,756	0,76	0,766	0,772	0,778	0,784	0,791	4,63	
BRA	0,737	0,742	0,746	0,752	0,755	0,761	0,769	0,778	0,786	0,795	0,804	6,45	
PER	0,718	0,722	0,728	0,732	0,734	0,739	0,745	0,752	0,759	0,766	0,773	5,26	
ECU	0,717	0,723	0,727	0,73	0,732	0,734	0,739	0,743	0,748	0,754	0,759	3,68	
COL	0,706	0,713	0,715	0,718	0,72	0,724	0,73	0,736	0,743	0,75	0,757	5,19	
BLZ	0,709	0,711	0,716	0,715	0,715	0,715	0,717	0,72	0,723	0,726	0,728	1,86	
DOM	0,701	0,704	0,708	0,711	0,715	0,72	0,726	0,733	0,74	0,747	0,754	5,41	
PRY	0,668	0,671	0,669	0,677	0,679	0,681	0,686	0,691	0,696	0,701	0,707	4,05	
SLV	0,653	0,658	0,662	0,664	0,666	0,669	0,674	0,682	0,689	0,696	0,703	5,48	
BOL	0,641	0,647	0,654	0,658	0,662	0,667	0,671	0,677	0,683	0,689	0,695	4,96	
GUY	0,624	0,63	0,629	0,634	0,636	0,637	0,64	0,644	0,648	0,651	0,655	3,00	
NIC	0,619	0,623	0,625	0,628	0,631	0,636	0,643	0,651	0,66	0,668	0,676	7,13	
GTM	0,611	0,617	0,624	0,626	0,627	0,632	0,639	0,647	0,655	0,663	0,671	7,05	
HND	0,61	0,612	0,607	0,604	0,606	0,61	0,616	0,622	0,629	0,635	0,641	5,84	
HTI	0,471	0,475	0,479	0,481	0,483	0,487	0,492	0,497	0,502	0,507	0,513	6,13	
América Latina	μ	0,699	0,703	0,707	0,71	0,712	0,716	0,721	0,727	0,734	0,74	0,747	4,99
	±	0,081	0,081	0,083	0,083	0,083	0,083	0,083	0,084	0,084	0,085	0,086	3,90
Mundial	μ	0,679	0,683	0,687	0,689	0,691	0,695	0,7	0,705	0,711	0,717	0,723	4,63
	±	0,157	0,156	0,156	0,155	0,155	0,154	0,155	0,155	0,155	0,155	0,155	0,2

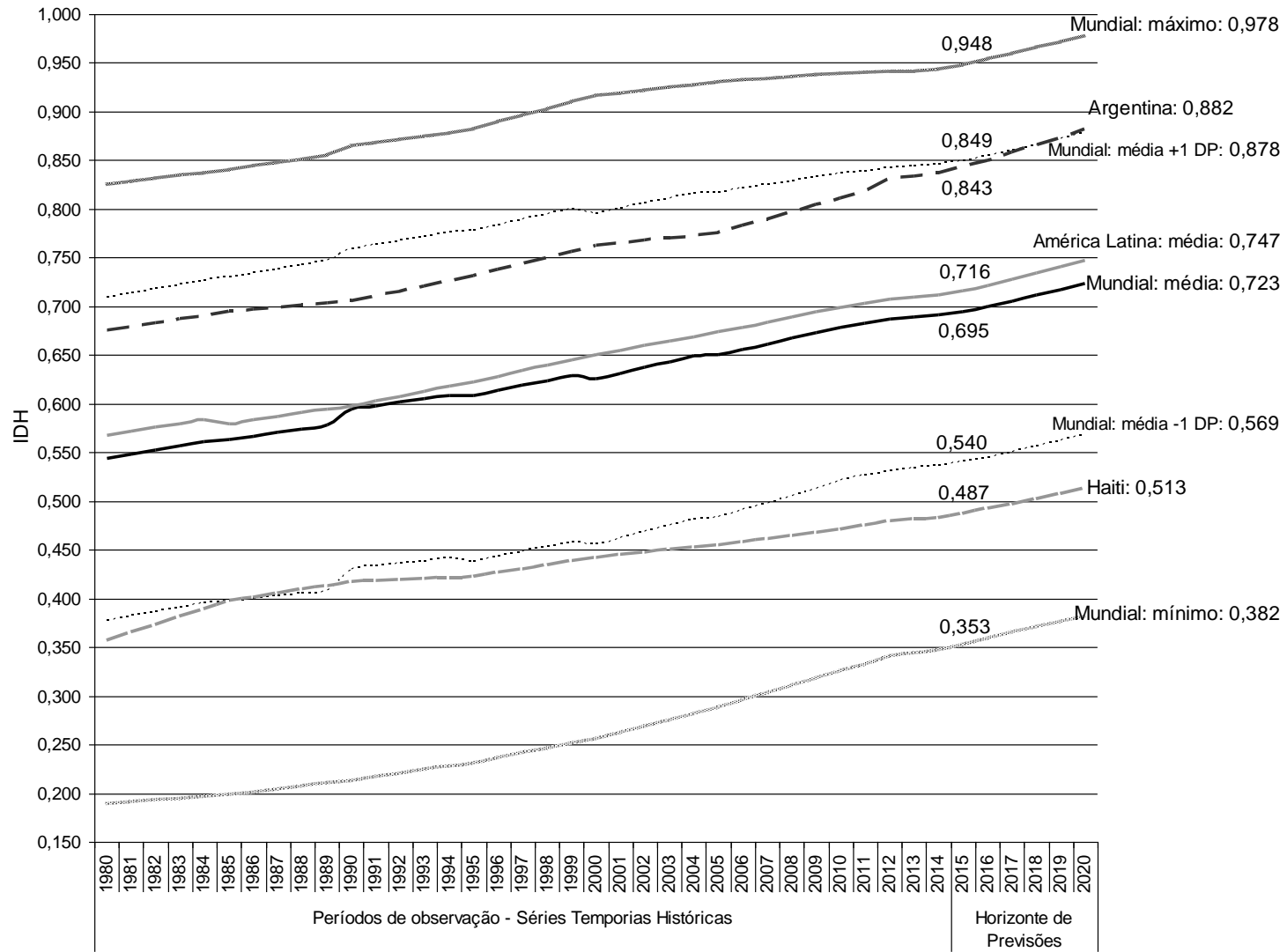
\* código internacional de países com três letras



Figura 16 - Previsões e tendências do IDH de Cuba(a), Venezuela(b), Nicarágua(c), El Salvador(d)



**Figura 17 - Séries Temporais Históricas (1980 -2014) e Horizonte de Previsões (2015 a 2020) do IDH Mundial, da América Latina e de Países Latino-americanos com maior (Argentina) e menor (Haiti) índice**

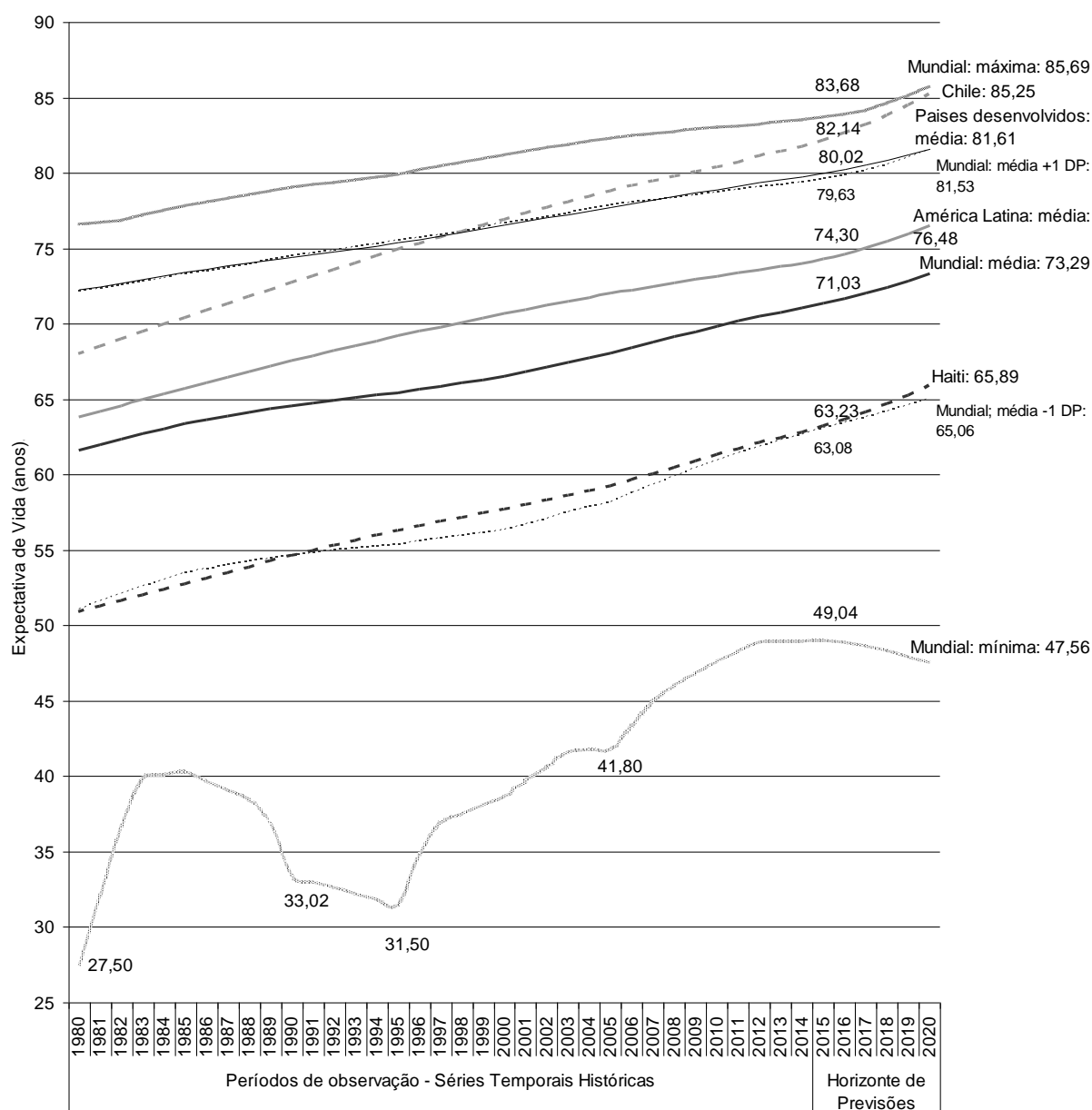


**Tabela 8 - Últimos pontos de observações das séries temporais históricas da EV nos países da América Latina e projeções para 2015 a 2020**

País*	Últimos Pontos de Observação					Horizonte de Previsão						Variação %	
	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2015-2020	
CHL	80,4	80,7	81,1	81,4	81,7	82,14	82,63	83,19	83,81	84,50	85,25	4,35	
CRI	78,8	78,9	79,1	79,2	79,4	79,58	79,82	80,09	80,40	80,75	81,13	2,18	
CUB	79	79,1	79,2	79,3	79,4	79,53	79,64	79,78	79,95	80,17	80,41	1,27	
PAN	76,8	77	77,2	77,4	77,6	77,86	78,12	78,43	78,76	79,14	79,56	2,52	
URY	76,6	76,7	76,9	77	77,2	77,37	77,59	77,84	78,13	78,47	78,83	2,11	
MEX	76,1	76,2	76,4	76,6	76,8	77,04	77,33	77,70	78,13	78,63	79,18	3,10	
ARG	75,6	75,8	75,9	76,1	76,3	76,54	76,79	77,09	77,43	77,80	78,20	2,49	
ECU	75	75,2	75,4	75,7	75,9	76,25	76,65	77,15	77,72	78,35	79,06	4,16	
NIC	73,7	74	74,3	74,6	74,9	75,36	75,90	76,56	77,30	78,13	79,05	5,55	
PER	73,7	73,9	74,1	74,3	74,6	74,93	75,36	75,89	76,51	77,22	77,99	4,55	
BRA	73,3	73,6	73,9	74,2	74,5	74,91	75,35	75,86	76,43	77,08	77,79	4,41	
VEN	73,6	73,7	73,9	74	74,2	74,38	74,61	74,85	75,14	75,46	75,81	2,17	
COL	73,3	73,5	73,7	73,9	74	74,22	74,44	74,71	75,00	75,34	75,72	2,33	
DOM	72,7	72,9	73,1	73,3	73,5	73,79	74,11	74,51	74,96	75,47	76,05	3,47	
HND	72,4	72,6	72,8	72,9	73,1	73,37	73,69	74,09	74,58	75,16	75,81	3,71	
SLV	71,9	72,2	72,5	72,8	73	73,41	73,94	74,59	75,30	76,11	77,03	5,52	
PRY	72,3	72,5	72,6	72,8	72,9	73,09	73,25	73,47	73,69	73,97	74,27	1,88	
GTM	70,9	71,1	71,4	71,6	71,8	72,14	72,56	73,07	73,66	74,34	75,11	4,61	
BLZ	69,7	69,7	69,8	69,9	70	70,07	70,14	70,22	70,28	70,32	70,34	0,48	
BOL	66,4	66,9	67,5	67,9	68,3	68,87	69,44	70,09	70,80	71,63	72,55	6,23	
GUY	66	66,1	66,2	66,3	66,4	66,54	66,72	66,91	67,12	67,34	67,57	1,77	
HTI	61,3	61,7	62,1	62,4	62,8	63,23	63,66	64,13	64,65	65,24	65,89	4,91	
América Latina	μ	73,16	73,36	73,60	73,80	74,01	74,30	74,63	75,01	75,44	75,94	76,48	3,33
	±	4,47	4,43	4,39	4,38	4,36	4,34	4,33	4,33	4,35	4,37	4,42	2,05
Mundial	μ	69,85	70,17	70,47	70,76	71,03	71,36	71,67	72,02	72,40	72,82	73,29	3,18
	±	8,89	8,73	8,59	8,47	8,37	8,27	8,22	8,19	8,19	8,21	8,24	2,20

\* código internacional de países.

**Figura 18 - Séries Temporais Históricas (1980 -2014) e Horizonte de Previsões (2015 a 2020) da EV Mundial, da América Latina e de Países Latino-americanos com maior (Chile) e menor (Haiti) índice**



## 5 DISCUSSÃO

No presente estudo, foi proposta uma metodologia para prever o IDH e a EV utilizando-se de técnicas de Mineração de Dados em modelos univariados e multivariados, totalizando 750 modelos, com evidências de bom desempenho, operacionalidade e flexibilidade, que superam o método tradicional, ARIMA.

Utilizando-se de técnicas Mineração de Dados, o analista precisa fornecer apenas as séries temporais e especificar quais os períodos o algoritmo deve prever. Enquanto que para o desenvolvimento de um modelo ARIMA, o analista precisa definir parâmetros baseados em suas análises a partir das observações dos resultados dos testes de AFC e FACP das séries temporais, tornado o processo mais interativo e aumentando os custos operacionais.

O problema da ausência de dados nas séries temporais é tratado automaticamente com o uso das técnicas de Mineração de Dados, permitindo trabalhar nos modelos multivariados com diferentes intervalos de dados ausentes em suas variáveis, contanto que estes ocorram distantes dos últimos pontos de observação. O mesmo não ocorre nos modelos ARIMA, que não permite trabalhar com modelos multivariados nestas condições.

A metodologia apresentada neste estudo pode ser utilizada para solução de outros problemas de previsão, principalmente na área de Engenharia de Produção.

O experimento fez a previsão do IDH 2015 a 2020 para os países da América Latina e, também, fez comparações das previsões do IDH 2013 e 2014 dos países filiados a ONU, com as tendências divulgadas pelo UNDP em 24/07/2014 e 14/12/2015.

O estudo empírico, utilizando dados reais, contribui com a difusão de novos métodos de previsão e complementa o rol de experimento que atende a carência apontada anteriormente há mais de uma década, por Keogh e Kasetty (2003), e que ainda persiste, da falta de pesquisas de Mineração de Dados para testar novos métodos, à medida que surgem novos algoritmos. Contribui principalmente na projeção dos indicadores IDH e EV, utilizados em pesquisas de QV, e também na

avaliação de algoritmos, modelos e medidas de qualidade, a partir das comparações entre as previsões e as tendências divulgadas pelo UNDP para o IDH em períodos anteriores às projeções.

Atualizações significativas dos índices de alguns países podem limitar o estudo. Segundo a UNDP (2013), as estimativas internacionais e nacionais de dados podem apresentar inconsistência, uma vez que as agências de dados internacionais consultam os dados nacionais e, eventualmente, estimam dados inexistentes para comparação entre países.

## 5.1 QUALIDADE DOS DADOS

A má qualidade dos dados é um problema que prejudica a Mineração de Dados. Em grandes bases de dados, a ocorrência de erros e dados incompletos é comum (WITTEN e FRANK, 2005). Intervalos de previsões são muitas vezes sensíveis a *outliers*, principalmente se ocorrer perto da origem da previsão (CHEN e LIU, 1993b). Técnicas de Mineração de Dados às vezes podem solucionar alguns desses problemas (WITTEN e FRANK, 2005). A redução de dimensionalidade é uma delas, que pode ser utilizada para eliminação de ruídos ou dados irrelevantes (TAN, STEINBACH e KUMAR, 2005). O estudo fez a previsão com 34,22% dos países apresentando dados incompletos retroativos a 2010. Esta ausência representava 10,78 % das séries temporais (1980-2013) nos MMGs, e até 69,23% nos MMEs e MUs. A ausência de dados foi tratada de forma não supervisionada. Os algoritmos utilizados superam essas faltas, através de interpolação de dados, sem comprometer as previsões, porque as ausências não ocorreram próximas às origens das previsões.

A literatura apresenta diferentes técnicas de detecção de anomalias em séries temporais. Fox (1972) introduz técnicas estatisticamente rigorosas para tratar questões de elementos estranhos em séries temporais. Outros estudos apresentam diferentes tipos de elementos estranhos em séries temporais (MUIRHEAD, 1986; CHEN e LIU, 1993a; CHEN e LIU, 1993b), nos quais são propostas técnicas para detectá-los e para obtenção de boas estimativas. Ainda, Abraham e Chuang (1989)

propõem a utilização de redes bayesianas para o tratamento de elementos estranhos.

A presença de *outliers* afeta a regressão, porque o quadrado da distância mínima se acentua com a influência dos pontos mais distantes a partir da linha de regressão (WITTEN e FRANK, 2005). A prática sugere eliminação desses *outliers* de forma supervisionada ou não supervisionada durante a etapa de limpeza dos dados.

Neste estudo, foram mantidos todos os registros dos IDH de todos os países, mesmos os aparentes *outliers*, pois estes não foram considerados erros, mas valores surpreendentes e corretos, díspares do padrão das demais séries temporais.

## 5.2 MUDANÇA NA METODOLOGIA DE CÁLCULO DO IDH

No último relatório da ONU (UNDP, 2013), foram apresentadas atualizações das séries temporais dos IDH dos países. Nestas atualizações, foi aplicada a mesma metodologia de cálculo do IDH de 2013, retroativo a cinco anos. Assim, a mudança no cálculo do IDH, ocorrida em 2011, foi antecipada em dois anos. Isto afastou a mudança de metodologia do ponto de previsão. Antes desta alteração, verificou-se que as previsões do IDH de 2013 apresentavam diferenças significativas em relação ao valor real, possivelmente pela proximidade entre os pontos de transição de metodologia de cálculo e de previsão.

O distanciamento entre esses pontos contribuiu para a melhoria da qualidade das previsões, pois mudanças inesperadas de valores, principalmente em alguns países em que o índice se apresenta como *outlier*, podem, segundo Chen e Liu (1993b), prejudicar as previsões.

### 5.3 ÍNDICE DE DESENVOLVIMENTO HUMANO

As altas correlações entre as séries temporais do IDH dos países observadas neste estudo sustentam, de forma implícita, a constatação do UNDP (2013) de que a interdependência entre os países não só é maior como também mais profunda, pois decisões e acontecimentos em uma determinada região podem afetar diretamente outras partes do mundo, entre elas a América Latina.

A baixa variância pode impossibilitar a verificação de diferenças significativas entre os modelos de previsão e os valores reais. A partir da quarta casa decimal, a diferença do IDH não é mais significativa, em concordância com Aguna e Kovacevic (2010), Høyland, Moene e Willumsen (2012), tornando-se necessária a avaliação a partir de medidas de qualidade de previsões específicas.

A baixa quantidade de exemplos (número de observações) de dados temporais não diminuiu a qualidade da previsão. Ela apresenta baixa correlação com a DAC e variando relativamente ao modelo de moderada (MME e MU) a alta (MMG) correlação com alguns tipos de erros, pois quanto maior o número de elementos nas séries temporais de treinamento, maior o erro acumulado.

A baixa variância do IDH entre os períodos não garante melhoria na taxa de acerto das tendências direcionais, pois se observou baixa correlação entre DAC e o índice.

Estas características observadas nas séries temporais reforçam a afirmativa de que o IDH é um índice robusto, com possibilidade de aplicação de metodologias alternativas.

### 5.4 PREVISÃO DO IDH

Em relação às previsões do IDH, ressalta-se que atualizações significativas dos índices de alguns países podem limitar o estudo. Segundo o UNDP (1990), as estimativas internacionais e nacionais de dados podem apresentar inconsistência, uma vez que as agências de dados internacionais consultam os dados nacionais e,



eventualmente, estimam dados inexistentes para comparação entre países. Em relação a estas atualizações, observaram-se diferenças significativas entre os valores de IDH, divulgados em 24 de julho de 2014 (UNDP, 2014) e 14 de dezembro de 2015 (UNDP, 2015).

Algumas previsões deste estudo, decorrentes dos testes para seleção do algoritmo e modelo, de previsão, apontaram contradições com as direções das tendências do IDH de outros estudos (UNDP, 2014; 2015) em relação a todos os resultados dos modelos para Cuba, em 2013 (Figura 16a), e Venezuela, em 2014 (Figura 16b). Cuba também apresentou o maior MAE em relação à previsão do IDH 2014, além de apresentar as maiores diferenças entre relatórios da UNDP (2014, 2015). Já as previsões para a Nicarágua, em 2013 (Figura 16c), e El Salvador, em 2014 (Figura 16d), apresentaram os menores MAE. Estas comparações, quando favoráveis às previsões, no caso da Nicarágua e El Salvador, podem sustentar políticas públicas ou econômicas adotadas por estes países para o desenvolvimento do índice e, quando desfavoráveis às estimativas, no caso da Venezuela, podem levantar questionamentos de suas políticas ou de seus dados, que podem encontrar-se desatualizados ou inconsistentes, como foi caso de Cuba.

Bolívia (+0,61%) e Cuba (+0,13%) apresentaram, respectivamente, o maior e menor ganho percentual do IDH no último período (2013-2014) em relação aos países da América Latina, enquanto a Venezuela (-0,61%) apresentou perda no índice (UNDP, 2014; 2015).

A partir das previsões apresentadas neste estudo (Tabela 7), observa-se que o Uruguai poderá atingir, até 2016, o mesmo nível de desenvolvimento da Argentina e Chile, atualmente classificados (UNDP, 2015) como países desenvolvidos com desenvolvimento humano muito elevado ( $IDH > 0,799$ ). O mesmo poderá ocorrer com o Panamá em 2017, Cuba em 2019, e Costa Rica em 2020. Os países em desenvolvimento, Paraguai e El Salvador, poderão passar da classe de desenvolvimento humano médio ( $0,599 < IDH < 0,7$ ) para desenvolvimento humano elevado ( $0,699 < IDH < 0,8$ ), respectivamente em 2019 e 2020.

O IDH médio dos países da América Latina ( $4,99 \pm 3,90\%$ ) nas projeções deste estudo, referentes ao período de 2015 a 2020, apresentam uma expectativa

de crescimento acima da média mundial ( $4,63 \pm 0,20\%$ ), mantendo a mesma tendência (UNDP, 2015) que destaca a América Latina e o Caribe com os maiores IDH, classificados como alto e com índices superiores às regiões da Europa, Ásia, Pacífico, Estados Árabes e África subsaariana.

Nicarágua (7,13%) e Guatemala (7,02%) tendem a apresentar maior crescimento do índice para o mesmo período, enquanto Belize (1,86%) tende a apresentar o menor crescimento. O Haiti (6,13%), apesar de tender a um crescimento acima da média mundial, permanecerá como único país da América Latina classificado (UNDP, 2015) como subdesenvolvido ( $IDH < 0,55$ ). Os demais países tendem a manter-se na mesma classe de desenvolvimento humano, mesmo apresentando crescimento do índice.

O Brasil, que atualmente passa por um momento de crise econômica (WATTS, 2016), não perceberá mudanças tão expressivas no IDH, mesmo no caso de avanços significativos na educação ou na EV, como se prevê, pois com a nova metodologia de cálculo, estes avanços tendem a ser atenuados pela renda baixa, em função do déficit do PIB previsto. A média geométrica empregada no cálculo do índice reduz o nível de substitutibilidade entre as dimensões, em virtude de que o baixo desempenho em uma determinada dimensão não pode ser mais compensado pelo maior desempenho em outra dimensão (UNDP, 2015). Apesar das críticas (RAVALLION, 2012a) ao novo tipo de cálculo do IDH, observa-se que este privilegia países com menor desigualdade entre suas componentes (KOVACEVIC, 2010), pois a média geométrica tende a ser cada vez menor que a média aritmética à medida que se aumenta a variância entre as componentes.

A maioria dos países latino-americanos passa por um processo de desenvolvimento humano muito semelhante, possivelmente em função do contexto histórico de suas emancipações políticas e características sociais e culturais.

## 5.5 PREVISÃO DA EXPECTATIVA DE VIDA

Foi possível confrontar as previsões com estudos recentes de outras agências internacionais (AGENCY, 2016; WORLD BANK, 2016), que já

disponibilizam a EV de 2015 de seus países filiados. No entanto, observam-se diferenças em suas séries temporais em relação à fonte de dados (UNDATA, 2014; UNDP, 2014; 2015) utilizada no treinamento dos modelos desenvolvidos neste estudo, que o limita até que se divulguem novos valores de EV ou outros estudos para possibilitar comparações.

Este indicador também pode apresentar inconsistência, pois muitos óbitos não são registrados corretamente (CHIAVEGATTO FILHO e LAURENTI, 2013).

Analisando as séries históricas mundiais de EV, antes das previsões para a América Latina, através da base de dados do UNDP (UNDATA, 2014), observa-se que, nos últimos 34 anos, a média mundial foi de  $67,84 \pm 2,89$  anos. A menor EV registrada no período foi a da população do Camboja, em 1980, com média de 27,5 anos, enquanto a da população mundial nesse período era de  $61,62 \pm 10,5$  anos. Em 1995, Ruanda apresentou a menor EV (31,50 anos), bem abaixo dos padrões mundiais da época, de  $65,44 \pm 10,18$  anos.

No último relatório (UNDP, 2015), Hong Kong - China registrou a maior EV (84 anos em 2014) e a Suazilândia apresentou a menor EV (49 anos), enquanto os padrões mundiais apresentaram a EV média de  $71,03 \pm 8,37$  anos. O Haiti é o país da América Latina que sempre apresentou a menor EV, a qual era, em 2014, de 62,8 anos, com perspectivas, segundo as projeções deste estudo, de chegar à EV de 65,89 anos até 2020, acima de -1DP (65,06 anos) da previsão da média mundial. Outros países, como Belize (70 anos), Bolívia (68,3 anos) e Guiana (66,4 anos), também apresentam EV abaixo da média mundial, enquanto a maioria, correspondente a 81,82% dos países da América Latina, apresentam EV acima da média mundial. A EV média da América Latina historicamente (UNDATA, 2014; AGENCY, 2016; WORLD BANK, 2016) sempre se apresentou acima da média mundial.

As projeções deste estudo estimam que, nos próximos seis anos, a média da EV da América Latina cresça de  $74,01 \pm 4,36$  para  $76,48 \pm 4,42$  anos, enquanto a mundial é estimada em  $73,29 \pm 8,24$  anos, e que em 2015, a EV da América Latina já seria de  $74,30 \pm 4,34$  anos, o que se confirmou em outro estudo (AGENCY, 2016).

Atualmente, segundo dados de UNDATA (2014), o Chile (81,7 anos) apresenta a maior EV da América Latina, com média superior aos demais países desenvolvidos ( $79,9 \pm 2,81$ ), e tende a chegar até 2020 à EV de 85,25 anos, também superior à prevista neste estudo para os países desenvolvidos ( $81,61 \pm 3,12$  anos).

Apesar das EV se apresentarem crescentes, Kanso et al. (2013) apontam que a EV aos 60 anos aumentaria em 20% se esses óbitos causados por morte evitáveis não ocorressem e que a sobremortalidade masculina foi observada em quase todas as causas de morte evitáveis analisadas, o que pode estar relacionado à maior exposição a fatores de risco e à menor utilização dos serviços de saúde pelos homens. Por outro lado, estudos (CAMARGOS e GONZAGA, 2015) apontam uma desvantagem notável do sexo feminino em relação à EV saudável.

Os dados de EV, principalmente analisados de maneiras decompostas, com ou sem saúde, por sexo e regiões, podem ser utilizados em políticas públicas, como referências para cálculo de planos de saúde e de contribuições previdenciárias (INOUE, RODRIGUES e AFONSO), decursos para pagamentos de pensões (LU, HE e PIGGOTT, 2014; BRASIL, 2015), como justificativas para reformas previdenciárias (ROCHA, 2016), no planejamento do futuro da assistência à saúde (MENDES *et al.*, 2012) e QV dos idosos (MINAYO, HARTZ e BUSS, 2000; VECCHIA *et al.*, 2005; ROSA, 2016), bem como para prever o aumento de doenças relacionadas ao envelhecimento (SALGADO FILHO e BRITO, 2006).

## 5.6 PROCESSO DE DCBD E MINERAÇÃO DE DADOS

Em relação ao método utilizado, pesquisas empíricas de Keogh e Kasetty (2003) encontraram pouco ganho com a Mineração de Dados na época. Para Armstrong (2006), os métodos promissores precisam ser replicados para se identificar em que condições eles podem falhar. Em seu estudo, também menciona que as técnicas de Mineração de Dados oferecem pouca promessa e que, talvez, a grande falha desses métodos esteja na falta de conhecimento do domínio.

Já no presente estudo, foram percorridas todas as etapas do processo de DCBD, definidas por Fayyad, Piatetsky-Shapiro e Smyth (1996), e isto corrobora

com seu estudo e de Michalski e Kaufman (1998), que mencionam que a eficácia do método depende do rigor deste processo e que todas as suas etapas são importantes para o produto final, e, ainda, que a Mineração de Dados é apenas uma das suas etapas.

Apesar das técnicas de Mineração de Dados não exigirem o conhecimento prévio do domínio, como menciona Armstrong (2006), observou-se, nesta pesquisa, que se pode conhecer muito sobre o domínio na etapa de DCBD de Pré-processamento que antecede a de Mineração de Dados, principalmente na sub-etapa de "Exploração da Base de Dados", como sugere Fayyad, Piatetsky-Shapiro e Smyth (1996). E, também, reafirma-se o estudo dos autores de que essas etapas são de fundamental importância para a condução do processo e para o estabelecimento da técnica mais adequada para o tipo de problema e natureza dos dados (FAYYAD, PIATETSKY-SHAPIRO e SMYTH, 1996).

Apesar dos algoritmos de aprendizado de máquina serem recomendados para aquisição de conhecimento, por reduzirem a necessidade de especialistas (ARINZE, 1994), a literatura recomenda interação entre especialistas em Mineração de Dados e do domínio investigado (GARGANO e RAGGAD, 1999; HONG e HAN, 2002; KOPANAS, AVOURIS e DASKALAKI, 2002; NEMATI *et al.*, 2002; HOFMANN e TIERNEY, 2003; DUBEY, PANDEY e GAUTAM, 2014; KADHIM, ALAM e KAUR, 2014). Esta interação foi possível neste estudo, tendo contribuído para uma melhor compreensão dos dados, bem como dos resultados obtidos.

Apesar de este trabalho testar exclusivamente algoritmos de aprendizagem baseada em funções e comparar o desempenho do mais promissor em modelos multivariados e univariados, considerando-se as características específicas dos dados, observa-se grande avanço na qualidade das previsões obtidas, contrariando estudos anteriores de Keoch e Kasetty (2003) e Armstrong (2006), que não verificaram vantagens no uso das técnicas de Mineração de Dados. O mesmo se observa em outros estudos recentes, como de LLoyd (2014), que, apesar de não deixar explícito o uso do processo de DCBD, destaca algumas etapas do processo e o uso de técnicas de Mineração de Dados utilizadas para a solução do problema de previsão. Ainda, Hong, Pinson e Fan (2014b) apresentam diversos aspectos da Competição Global de Previsão de Energia de 2012 (GEFCom2012), incluindo os

métodos utilizados pelos participantes, confirmando que algumas técnicas de Mineração de Dados têm vantagens sobre outras populares, como a ARIMA.

O algoritmo de Mineração de Dados, *SMOReg* apresentou as melhores medidas de qualidade durante os testes de previsão em relação a outros do mesmo paradigma de aprendizagem, confirmando estudos anteriores (HONG, PINSON e FAN, 2014a; SILVA, 2016), e também reafirmado vantagens no uso das técnicas de Mineração de Dados em relação a outras técnicas mais populares de previsão, como ARIMA, já apontadas em outros estudos (HONG, PINSON e FAN, 2014a).

## 5.7 CUSTOS OPERACIONAIS

As maiores dificuldades enfrentadas neste método de previsão se referem ao custo operacional. As etapas de pré e pós-processamento de Mineração de Dados consumiram aproximadamente 80% do custo operacional, como sugerem estudos anteriores (MANNILA, 1996). Falta de acesso a dados totalmente consistentes foi outro problema, frequente em grandes bases de dados como sugeriram WITTEN e FRANK (2005), pois atualizações nos dados de observação já publicadas limitaram o estudo, diminuindo a previsibilidade.

A Maldição da dimensionalidade (*Curse of dimensionality*), termo introduzido por Bellman (1961), pode afetar os custos operacionais e o desempenho de algoritmos de Mineração de Dados. Então, a redução de dimensionalidade pode ser empregada para diminuir custos operacionais e eliminar ruídos (TAN, STEINBACH e KUMAR, 2005), aumentando a taxa de acerto das previsões. No entanto, a alta dimensionalidade, neste estudo em específico, não aumentou o custo operacional, pois se testou a possibilidade de redução dimensionalidade, mas verificou-se que não houve melhorias significativas do tempo de resposta e nas medidas de qualidade das previsões. Já a definição do algoritmo foi determinante no custo operacional, pois alguns algoritmos aplicados em MMGs ultrapassavam muito o tempo médio de processamento esperado nos MMEs e MUs. Os MMEs desenvolvidos com técnicas de redução de dimensionalidade apresentaram resultados intermediários, entre os MMGs e MUs.

## 5.8 MEDIDAS DE QUALIDADE DAS PREVISÕES

Entre as medidas de qualidade das previsões, geralmente o MSE é mais utilizado por resultar em valores na mesma escala dos dados. O RMSE e MSE são muito populares, principalmente porque são muito empregados em modelagens estatísticas (HYNDMAN e KOEHLER, 2006), mas são mais sensíveis a *outliers* que outras medidas como o MAE (PASSARI, 2003; HYNDMAN e KOEHLER, 2006).

Armstrong (2001) apresenta uma lista de 32 princípios para avaliar sistematicamente o método de previsão, não recomendando medidas sensíveis a *outliers*. Abraham e Chuang (1989) e Passari (2003) sugerem a MAPE como a melhor medida neste caso, por ser uma medida absoluta em porcentagem do valor previsto, além de possibilitar uma visão da amplitude do erro. Já no caso de modelos que respeitem limites de erro máximo, o MAE é o mais indicado. Tanto no MAE como no MSE, durante o somatório, um erro positivo não é anulado por um erro negativo ou vice-versa (PASSARI, 2003). Peña e Sánchez (2007) utilizam o MSE para apresentar vantagens nos modelos multivariados em relação aos modelos univariados, criando uma equação de previsibilidade da série temporal na adoção de preditores multivariados ao invés de univariados.

Estudos como o de Greer (2003) para previsões direcionais de taxas de juros de longo prazo e de Tang e He (2014) utilizam a DAC, a qual fornece a correção da direção prevista e também pode ser utilizada para avaliar a precisão da previsão. Quanto maior o seu valor, melhores serão as previsões (WANG *et al.*, 2012). MAPE permite comparar modelos com dados diferentes (PASSARI, 2003). A competição de Energia Global de Previsão de 2012 (GEFCOM, 2012) utiliza o RMSE para avaliar os melhores modelos apresentados na competição. O RMSE permite retornar a medida original dos dados a partir da raiz do MSE (PASSARI, 2003).

Neste estudo, foram utilizadas as medidas de avaliação das previsões disponíveis na API *Forecast* do WEKA, observando-se suas propriedades discutidas na literatura. No entanto, dado às características do experimento como utilização de variáveis com uma única unidade de medida, observa-se que apenas as medidas de

qualidade DAC e MAE são suficientes para avaliação das previsões do IDH ou EV, dispensando-se a necessidade da análise de variância, pois apesar de não existirem diferenças estatísticas significativas entre previsões dos modelos e os valores reais, o MAE permitiu identificar os melhores modelos, confirmando estudos (ARMSTRONG, 2001) que discutem o uso das medidas de qualidade específicas para previsão. A DAC permite avaliar a direção dos indicadores (crescente ou decrescente), enquanto o MAE permite avaliar a diferença absoluta entre a previsão e o valor real.

Observou-se, ainda, através de testes de correlação, que, quanto maior o número de exemplos ou variância das séries temporais, maiores os erros acumulados com as experiências das previsões, principalmente nos MMGs.

## 5.9 EFICIÊNCIA DOS MODELOS

Segundo Rodrigues e Stevenson (2013), expressiva parte da literatura sugere que previsões combinadas podem melhorar as previsões individuais. Isto foi visível nos modelos multivariados que apresentam melhores resultados que os modelos univariados. Nos modelos multivariados, o algoritmo apreende com o comportamento de histórico das séries temporais de todos (MMGs) ou de grupos (MMEs) de países, enquanto que nos MUs a aprendizagem se restringe às séries temporais do país-alvo.

Em relação à eficiência de um modelo, Putsis (1998), Lawrence, O'Connor e Edmundson (2000) apontam que certas características do erro devem ser observadas. Segundo Putsis (1998), em um modelo eficiente não deve existir correlação entre os erros de um período para outro, o que indica que o modelo aprende com os erros do passado. Esta premissa foi observada nas três categorias de modelos apresentados neste estudo. Apesar da alta correlação entre as séries temporais, observou-se que não existem correlações entre os erros das previsões nos pontos de observações subsequentes.

Para Lawrence, O'Connor e Edmundson (2000), a distribuição dos erros deve ter uma forma próxima da normal. Isto também foi verificado neste estudo, pois



os valores dos erros no treinamento dos modelos foram submetidos e aprovados no Teste KS. No entanto, esta condição apontada por Lawrence, O'Connor e Edmundson (2000), também dependerá do número de exemplos da série temporal.

Os modelos multivariados apresentaram melhor desempenho que os modelos univariados. No entanto, esta vantagem relativa do preditor multivariado pode ser muito diferente em cada país.

Peña e Sánchez (2007) também destacam vantagens dos modelos multivariados, principalmente se existirem fortes relações entre as séries temporais, o que também ocorreu no presente estudo, evidenciado a partir do número médio de países previsores selecionados nos MMEs.

Os testes de análise de variância sugerem evidências da inexistência de diferenças significativas entre as previsões dos modelos e as tendências divulgadas pelo UNDP (2015) para o IDH 2013 e 2014. No entanto, os MMGs apresentaram as melhores medidas de qualidade acumuladas durante todo o período de treinamento e previsão, com a maior DAC e os menores tipos de erro em relação aos demais modelos.

Os modelos não apresentaram diferenças significativas em relação à DAC do último ponto de observação, apesar de os resultados do período completo apontarem melhor rendimento do MMG. Resta aguardar a divulgação do IDH 2014 para confirmar as mesmas tendências.

Os MMG apresentam os melhores resultados, tanto nas medidas de qualidade das previsões como nas comparações como os valores divulgados do IDH em UNDP (2014, 2015). A eficiência dos MMGs pode ser explicada implicitamente pelas interdependências e vulnerabilidades dos países apontadas pelo UNDP (2013).

## 6 CONCLUSÃO

As previsões de eventos com probabilidades de ocorrerem com base em históricos de séries temporais multivariadas ou univariadas são cada vez mais comuns em diferentes áreas.

Utilizando-se de Técnicas de Mineração de Dados, é possível prever o IDH e a EV de países, como por exemplo da América Latina, porque essa tecnologia permite trabalhar com grandes volumes de dados de maneira mais automatizada.

Os procedimentos metodológicos apresentados neste estudo podem ser utilizados para a solução de outros problemas de previsão da engenharia de produção.

Modelos desenvolvidos a partir de séries temporais multivariadas, apesar de mais complexos, apresentaram melhor precisão que os modelos desenvolvidos a partir de séries univariadas, principalmente se existir alta correlação entre as séries temporais.

As séries temporais multivariadas possibilitam maior aprendizagem dos algoritmos com o aumento de diferentes experiências históricas univariadas.

Modelos multivariados globais apresentam melhor qualidade nas previsões em relação aos modelos multivariados específicos, porque oferecem maior experiência de observações aos algoritmos durante a fase de treinamento.

A execução da Mineração de Dados respeitando todas as etapas do processo de DCBD resultou em previsões de séries temporais com precisão satisfatória.

As técnicas de Mineração de Dados apresentaram evidências de melhor qualidade nas previsões em relação à técnica mais popular, ARIMA, possivelmente por serem mais independentes da inferência humana durante o processo de modelagem.

O algoritmo de Otimização Mínima Sequencial para a Regressão, *SMOReg*,

apresentou evidências de melhor desempenho entre o conjunto de algoritmos com paradigma de aprendizagem de funções testados durante o experimento.

Entre as diferentes medidas de qualidade de previsão apresentadas, este estudo destaca a DAC e MAE como medidas mais adequadas para avaliação das previsões do IDH e da EV. A DAC permite avaliar o acerto da tendência de direção dos indicadores, enquanto que a MAE permite avaliar a diferença absoluta da previsão em relação ao valor real.

O IDH é um índice robusto, com grande previsibilidade e vulnerabilidade, amplamente utilizado em pesquisas de QV e desenvolvimento humano, principalmente como parâmetro demográfico ou comparativo.

As previsões apontam que as médias de crescimento do IDH e da EV nos países da América Latina tendem a permanecer maiores que a média mundial nos próximos seis anos (2015 a 2020).

As contradições entre a previsão e os valores reais do IDH ou de suas componentes, se comparados, futuramente poderão desencadear reflexões e auxiliar em tomadas de decisões para sustentação ou não de políticas públicas voltadas ao planejamento e gestão da promoção de QV e desenvolvimento humano e, também, justificar o cenário vivido pelos países e mundo.

## REFERÊNCIAS

ABRAHAM, B.; CHUANG, A. Outlier detection and time series modeling. **Technometrics**, v. 31, n. 2, p. 241-248, 1989.

ADRIAANS, P.; ZABTINGE, D. Data Mining. Harlow: Addison-Wesley, 1996. 176p.

AGENCY, C. I. Central Intelligence Agency. **The World Factbook**. CIA. <https://www.cia.gov/library/publications/resources/the-world-factbook/rankorder/2102rank.html>. Cons. 19-set-2016.

AGUNA, C.; KOVACEVIC, M. Uncertainty and sensitivity analysis of the human development index. **Human Development Research Paper**, v. 11, 2010.

AHN, H. et al. Facilitating cross-selling in a mobile telecom market to develop customer classification model based on hybrid data mining techniques. **Expert Systems with Applications**, v. 38, n. 5, p. 5005-5012, 2011.

AKAIKE, H. A new look at the statistical model identification. **IEEE transactions on automatic control**, v. 19, n. 6, p. 716-723, 1974.

ALKIRE, S. Human development: Definitions, critiques, and related concepts. **UNDP-HDRO Occasional Papers**, n. 2010/1, 2010.

AMARAL, F. C. N. D. Data Mining: técnicas e aplicações para o marketing direto. São Paulo: Berkeley, 2001.

ANSTEY, K. J. et al. The influence of smoking, sedentary lifestyle and obesity on cognitive impairment-free life expectancy. **International Journal of Epidemiology**, v. 43, n. 6, p. 1874-1883, dez 2014.

ARINZE, B. Selecting appropriate forecasting models using rule induction. **Omega**, v. 22, n. 6, p. 647-658, 1994.

ARMSTRONG, J. S. Evaluating forecasting methods. In: ARMSTRONG, J. S. (Ed.). **Principles of forecasting**: Springer, 2001. p.443-472.

\_\_\_\_\_. Findings from evidence-based forecasting: Methods for reducing forecast error. **International Journal of Forecasting**, v. 22, n. 3, p. 583-598, 2006.

BARRO, R. J.; LEE, J. W. A new data set of educational attainment in the world, 1950–2010. **Journal of development economics**, v. 104, p. 184-198, 2013.

BELLMAN, R. E. **Adaptive control processes: a guided tour.** Princeton university press Princeton, 1961.

BISGAARD, S.; KULAHCI, M. Time series analysis and forecasting by example. John Wiley & Sons, 2011.

BOX, G. E. et al. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

BRASIL. **Lei nº 13.183.** Altera as Leis nº 8.212, de 24 de julho de 1991, e nº 8.213, de 24 de julho de 1991; e dá outras providências. Brasília: Diário Oficial da União 2015.

BROOMHEAD, D. S.; LOWE, D. **Radial basis functions, multi-variable functional interpolation and adaptive networks.** DTIC Document. 1988

BRUNEKREEF, B. et al. Early life exposure to farm animals and symptoms of asthma, rhinoconjunctivitis and eczema: an ISAAC Phase Three Study. **International Journal of Epidemiology**, v. 41, n. 3, p. 753-761, jun 2012.

CAMARGOS, M. C. S.; GONZAGA, M. R. Viver mais e melhor? Estimativas de expectativa de vida saudável para a população brasileira. **Cadernos de Saúde Pública**, v. 31, p. 1460-1472, 2015.

CAMPOLINA, A. G. et al. A transição de saúde e as mudanças na expectativa de vida saudável da população idosa: possíveis impactos da prevenção de doenças crônicas. **Cadernos de Saúde Pública**, v. 29, p. 1217-1229, 2013.

CARVALHO, D. R. **Data Mining através de indução de Regras e Algoritmos Genéticos**, 1999. 142f. Dissertação (Mestrado em Informática Aplicada) - Pontifícia Universidade Católica do Paraná, Curitiba, 1999.

CARVALHO, L. A. V. **Datamining: A Mineração de Dados no Marketing, Medicina, Economia, Engenharia e Administração.** São Paulo: Érica, 2001. 237p.

CASTRO, J. M. D.; RODRIGUES-JÚNIOR, A. L. A influência da mortalidade por causas externas no desenvolvimento humano na Faixa de Fronteira brasileira. **Cadernos de Saúde Pública**, v. 28, p. 195-200, 2012.

CERVANTES, C. A. D.; BOTERO, M. A. Average years of life lost due to breast and cervical cancer and the association with the marginalization index in Mexico in 2000 and 2010. **Cadernos de Saúde Pública**, v. 30, p. 1093-1102, 2014.

CHATFIELD, C. Model Uncertainty, Data Mining and Statistical Inference. **Journal of the Royal Statistical Society. Series A (Statistics in Society)**, v. 158, n. 3, p. 419-466, 1995.

CHEN, C.; LIU, L.-M. Joint estimation of model parameters and outlier effects in time series. **Journal of the American Statistical Association**, v. 88, n. 421, p. 284-297, 1993a.

CHEN, C.; LIU, L. M. Forecasting time series with outliers. **Journal of Forecasting**, v. 12, n. 1, p. 13-35, 1993b.

CHEN, S.-M.; BAI, S.-M. Using data mining techniques to automatically construct concept maps for adaptive learning systems. **Expert Systems with Applications**, v. 37, n. 6, p. 4496-4503, 2010.

CHEN, S. C.; HUANG, M. Y. Constructing credit auditing and control & management model with data mining technique. **Expert Systems with Applications**, v. 38, n. 5, p. 5359-5365, 2011.

CHIAVEGATTO FILHO, A. D. P.; LAURENTI, R. Decomposição da diferença da expectativa de vida de Minas Gerais em relação ao Rio de Janeiro e São Paulo, Brasil. **Cadernos de Saúde Pública**, v. 29, p. 1131-1140, 2013.

CORREA, F. E. et al. Data mining frequent temporal events in agrieconomic time series. **Latin America Transactions, IEEE (Revista IEEE America Latina)**, v. 13, n. 7, p. 2329-2334, 2015.

DECANQ, K.; LUGO, M. A. **Setting Weights in Multidimensional Indices of Well-being and Deprivation**. OPHI Working Paper. 2009

DUBEY, S.; PANDEY, R. K.; GAUTAM, S. S. Development of multimedia fuzzy based diagnostic expert system for integrated disease management in chickpea. **International Journal of Science and Modern Engineering**, v. 2, n. 2, p. 16-20, 2014.

FAYYAD, U. M. Diving into databases: SQL is helpless in the face of massive, accumulating data stores. **Database Programming and Design**, v. 11, p. 24-31, mar. 1998.

FAYYAD, U. M.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to knowledge discovery in databases. **AI magazine**, v. 17, n. 3, p. 37, 1996.

FAYYAD, U. M. et al. **Advances in Knowledge Discovery and Data Mining**: Menlo Park, CA: AAAI Press / The MIT Press, 1996.

FAYYAD, U. M.; UTHURUSAMY, R. In: WORKSHOP TECHNICAL REPORT KNOWLEDGE DISCOVERY IN DATABASES, 1994, Seattle, WA. **Proceedings ...** Seattle, WA: AAAI Press, 1994. 473p.

\_\_\_\_\_. In: KDD-95 INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, 1, 1995, Montréal Québec, Canada. **Proceedings ...** Montréal Québec, Canada: AAAI, 1995. 348p.

FERREIRA, A. **Desenvolvimento Regional: Limites e Possibilidades Institucionais—Um Estudo de Caso da Região do Vale do Paraíba—RJ**, 2012. 239f. Tese (Doutorado) - Programa de Pós-Graduação em Políticas Públicas, Estratégias e Desenvolvimento, Universidade Federal do Rio de Janeiro, Rio de Janeiro,

FESHARAKI, M.; SHIRAZI, H.; BAKHSHI, A. knowledge Acquisition from Database of Information Management and Documentation Softwares by DataMining Techniques. **Journal of Information processing and Management**, v. 26, n. 2, p. 260-283, 2011.

FIGUEIRA, R. **Mineração de Dados e Bancos de Dados Orientados a Objetos**, 1998. 96f. Dissertação (Mestrado em Ciência da Computação) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 1998.

FMI. International Monetary Fund. **World Economic Outlook database**, 2014. Disponível em: <  
<http://www.imf.org/external/pubs/ft/weo/2014/01/weodata/index.aspx> >. Acesso em: 02 nov 2014.

FOX-RUSHBY, J.; PARKER, M. Culture and the measurement of health-related quality of life. **European review of applied psychology**, v. 45, n. 4, p. 257-264, 1995.

FOX, A. J. Outliers in time series. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 350-363, 1972.

FREITAS, I. M. B.; MARQUES, R. A.; E SILVA, E. M. D. P. University–industry collaboration and innovation in emergent and mature industries in new industrialized countries. **Research Policy**, v. 42, n. 2, p. 443-453, 2013.

GAJZLER, M. Text and data mining techniques in aspect of knowledge acquisition for decision support system in construction industry. **Technological and Economic Development of Economy**, n. 2, p. 219-232, 2010.

GARGANO, M. L.; RAGGAD, B. G. Data mining-a powerful information creating tool. **OCLC Systems & Services**, v. 15, n. 2, p. 81-90, 1999.

GEFCOM. Wind Forecasting. 2012. Disponível em: < <http://www.kaggle.com/c/GEF2012-wind-forecasting/details/evaluation> >. Acesso em: 02/11/2014.

GONZÁLEZ-ZAPATA, L. I. et al. Exceso de peso, aspectos económicos, políticos y sociales en el mundo: un análisis ecológico. **Cadernos de Saúde Pública**, v. 27, p. 1746-1756, 2011.

GREER, M. Directional accuracy tests of long-term interest rate forecasts. **International Journal of Forecasting**, v. 19, n. 2, p. 291-298, 2003.

HALL, M. et al. The WEKA data mining software: an update. **ACM SIGKDD explorations newsletter**, v. 11, n. 1, p. 10-18, 2009.

HALL, M. A. **Correlation-based feature selection for machine learning**, 1999. Thesis - The University of Waikato 1999.

HARTMANN, D.; PYKA, A. **Innovation, economic diversification and human development**. FZID Discussion Papers. 2013

HENDERSON, M.; GREY, V. Establishing and evaluating pediatric thyroid reference intervals on the Roche Modular Analytics E 170 using computational statistics and data-mining techniques. **Clinical biochemistry**, v. 44, n. 10, p. 767-770, 2011.

HENDI, A. S. Trends in U.S. life expectancy gradients: the role of changing educational composition. **International Journal of Epidemiology**, v. 44, n. 3, p. 946-955, jun 2015.

HOFMANN, M.; TIERNEY, B. The involvement of human resources in large scale data mining projects. In: The involvement of human resources in large scale data mining projects 1ST INTERNATIONAL SYMPOSIUM ON INFORMATION AND COMMUNICATION TECHNOLOGIES, 2003 **Proceedings ...** Trinity College Dublin. 103-109p.

HONG, T.; HAN, I. Knowledge-based data mining of news information on the Internet using cognitive maps and neural networks. **Expert Systems with Applications**, v. 23, n. 1, p. 1-8, 2002.

HONG, T.; PINSON, P.; FAN, S. Global energy forecasting competition 2012. **Int J Forecasting**, v. 30, n. 2, p. 357-363, 2014a.

\_\_\_\_\_. Global energy forecasting competition 2012. **International Journal of Forecasting**, v. 30, n. 2, p. 357-363, 2014b.



HOONAKKER, F. et al. A Representation to Apply Usual Data Mining Techniques to Chemical Reactions - Illustration On The Rate Constant of Sn 2 Reactions In Water. **International Journal on Artificial Intelligence Tools**, v. 20, n. 02, p. 253-270, 2011.

HØYLAND, B.; MOENE, K.; WILLUMSEN, F. The tyranny of international index rankings. **Journal of Development Economics**, v. 97, n. 1, p. 1-14, 2012.

HYNDMAN, R. J.; KOEHLER, A. B. Another look at measures of forecast accuracy. **International Journal of Forecasting**, v. 22, n. 4, p. 679-688, 2006.

INOUE, J. T.; RODRIGUES, C. G.; AFONSO, L. E. Tábua de Mortalidade e Expectativa de Vida Saudável: Uma Aplicação à População Beneficiária de Planos de Saúde Privados no Brasil em 2008.

KADHIM, M. A.; ALAM, M. A.; KAUR, H. A Multi-Intelligent Agent for Knowledge Discovery in Database (MIAKDD): Cooperative Approach with Domain Expert for Rules Extraction. In: HUANG, D. S. (Ed.). **Intelligent Computing Methodologies**. Switzerland, Suíça: Springer, v.8589, 2014. p.602-614.

KAHNEMAN, D.; DEATON, A. High income improves evaluation of life but not emotional well-being. **Proceedings of the National Academy of Sciences**, v. 107, n. 38, p. 16489-16493, 2010.

KANSO, S. et al. A evitabilidade de óbitos entre idosos em São Paulo, Brasil: análise das principais causas de morte. **Cadernos de Saúde Pública**, v. 29, p. 735-748, 2013.

KARIMINIA, A. et al. Cohort Profile: The TREAT Asia Pediatric HIV Observational Database. **International Journal of Epidemiology**, v. 40, n. 1, p. 15-24, fev 2011.

KEOGH, E.; KASETTY, S. On the need for time series data mining benchmarks: a survey and empirical demonstration. **Data Mining and knowledge discovery**, v. 7, n. 4, p. 349-371, 2003.

KLUGMAN, J.; RODRÍGUEZ, F.; CHOI, H.-J. The HDI 2010: new controversies, old critiques. **The Journal of Economic Inequality**, v. 9, n. 2, p. 249-288, 2011.

KÖCHE, J. C. Fundamentos de metodologia científica. **Petrópolis: Vozes**, 1997.

KOPANAS, I.; AVOURIS, N. M.; DASKALAKI, S. The role of domain knowledge in a large scale data mining project. In: VLAHAVAS, I. P. e SPYROPOULOS, C. D., The role of domain knowledge in a large scale data mining project SECOND HELLENIC CONFERENCE ON AI: METHODS AND APPLICATIONS OF ARTIFICIAL INTELLIGENCE, 2002, Berlin Heidelberg. **Proceedings ...** Springer-Verlag. 288-299p.

KOVACEVIC, M. Review of HDI critiques and potential improvements. **Human development research paper**, v. 33, 2010.

LACEY, R. E. et al. Work-family life courses and markers of stress and inflammation in mid-life: evidence from the National Child Development Study. **International Journal of Epidemiology**, October 14, 2015 2015.

LAWRENCE, M.; O'CONNOR, M.; EDMUNDSON, B. A field study of sales forecasting accuracy and processes. **European Journal of Operational Research**, v. 122, n. 1, p. 151-160, 2000.

LE MOS, E. P. **Análise de crédito bancário com o uso de data mining: redes neurais e árvores de decisão**, 2003. 147f. Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2003.

LEROY, A. M.; ROUSSEEUW, P. J. Robust regression and outlier detection. **Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1987**, 1987.

LI, L. et al. Life-course body mass index trajectories and blood pressure in mid life in two British birth cohorts: stronger associations in the later-born generation. **International Journal of Epidemiology**, v. 44, n. 3, p. 1018-1026, jun 2015.

LIAO, S.-H.; CHU, P.-H.; HSIAO, P.-Y. Data mining techniques and applications—A decade review from 2000 to 2011. **Expert Systems with Applications**, v. 39, n. 12, p. 11303-11311, 2012.

LIMA, R. F. P. D. **Espacialização dos Índices de Desenvolvimento Humano-IDH por cartogramas corocromáticos estatísticos**, 2006. 186f. Tese (Doutorado em Engenharia de Produção) - Programa de Pós-Graduação em Engenharia de Produção e Sistemas, Universidade Federal de Santa Catarina, Florianópolis, 2006.

LLOYD, J. R. GEFCom2012 hierarchical load forecasting: Gradient boosting machines and Gaussian processes. **International Journal of Forecasting**, v. 30, n. 2, p. 369-374, 2014.

LOPES, C. H. P. **Classificação de registros em banco de dados por evolução de regras de associação: utilizando algoritmos genéticos**, 1999. 136f. Dissertação (Mestrado em Engenharia Elétrica e Sistemas de Computação) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1999.

LU, B.; HE, W.; PIGGOTT, J. Should China introduce a social pension? **The Journal of the Economics of Ageing**, v. 4, p. 76-87, 12// 2014.

LUNA, S. V. Planejamento de pesquisa: uma introdução. São Paulo: Educ, 1999.

MACKAY, D. J. Introduction to Gaussian processes. **NATO ASI Series F Computer and Systems Sciences**, v. 168, p. 133-166, 1998.

MACKENBACH, J. P.; LOOMAN, C. W. Life expectancy and national income in Europe, 1900-2008: an update of Preston's analysis. **International Journal of Epidemiology**, v. 42, n. 4, p. 1100-1110, ago 2013.

MADDISON, A. **Historical Statistics of the World Economy: 1–2030 AD**. Paris: Organisation for Economic Co-operation and Development., 2010.

MANGALOVA, E.; AGAFONOV, E. Wind power forecasting using the k-nearest neighbors algorithm. **International Journal of Forecasting**, v. 30, n. 2, p. 402-406, 2014.

MANNILA, H. Data mining: machine learning, statistics, and databases. In: Data mining: machine learning, statistics, and databases INTERNATIONAL CONFERENCE ON SCIENTIFIC AND STATISTICAL DATABASE MANAGEMENT, 1996 **Proceedings ...** IEEE Computer Society. 2-2p.

MARIANO, E. B. **Crescimento econômico e desenvolvimento humano: uma análise mundial da eficiência social de Estados-nação**, 2012. 277f. Tese (Doutorado em Engenharia de Produção) - Programa de Doutorado em Engenharia de Produção, Universidade de São Paulo, São Carlos, 2012.

MARTINEZ, E. Z. et al. Gravidez na adolescência e características socioeconômicas dos municípios do Estado de São Paulo, Brasil: análise espacial. **Cadernos de Saúde Pública**, v. 27, p. 855-867, 2011.

MARX, A. Detection and classification of bark beetle infestation in pure norway spruce stands with multi-temporal RapidEye imagery and data mining techniques. **Photogrammetrie-Fernerkundung-Geoinformation**, v. 2010, n. 4, p. 243-252, 2010.

MENDES, A. D. C. G. et al. Assistência pública de saúde no contexto da transição demográfica brasileira: exigências atuais e futuras. **Cadernos de Saúde Pública**, v. 28, p. 955-964, 2012.

MENEZES, E. M.; SILVA, E. L. **Metodologia da pesquisa e elaboração de dissertação**: 3.ed. Florianópolis: Laboratório de Ensino à Distância da UFSC, 2001.

MICHALSKI, R. S.; KAUFMAN, K. A. Data mining and knowledge discovery: A review of issues and a multistrategy approach. **Machine Learning and Data Mining: Methods and Applications**, p. 71-112, 1998.

MINAYO, M. C. S.; HARTZ, Z. M. A.; BUSS, P. M. Qualidade de vida e saúde: um debate necessário. **Ciênc Saúde Coletiva**, v. 5, n. 1, p. 7-18, 2000.

MONTGOMERY, D. C.; JENNINGS, C. L.; KULAHCI, M. Introduction to time series analysis and forecasting. 2 Ed..ed. John Wiley & Sons, 2015.

MORTON, S. M.; DE STAVOLA, B. L.; LEON, D. A. Intergenerational determinants of offspring size at birth: a life course and graphical analysis using the Aberdeen Children of the 1950s Study (ACONF). **International Journal of Epidemiology**, v. 43, n. 3, p. 749-759, jun 2014.

MUIRHEAD, C. R. Distinguishing outlier types in time series. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 39-47, 1986.

MÜLLER, K.-R. et al. Time series with support vector machines. In: Time series with support vector machines INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, 1997**Proceedings ...** Springer. 999-1004p.

MURPHY, K. M.; SHLEIFER, A.; VISHNY, R. W. **Industrialization and the big push**. National Bureau of Economic Research. 1988

NAPAGODA, C. Web site visit forecasting using data mining techniques. **International Journal of Scientific & Technology Research**, v. 2, n. 12, 2013.

NEMATI, H. R. et al. Knowledge warehouse: an architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. **Decision Support Systems**, v. 33, n. 2, p. 143-161, 2002.

NEUMAYER, E. The human development index and sustainability—a constructive proposal. **Ecological Economics**, v. 39, n. 1, p. 101-114, 2001.

NIKOLOVA, E.; JECHEVA, V. Some similarity coefficients and application of data mining techniques to the anomaly-based IDS. **Telecommunication Systems**, v. 50, n. 2, p. 127-135, 2012.

NOORBAKHS, F. The human development index: some technical issues and alternative indices. **Journal of International Development**, v. 10, n. 5, p. 589-605, 1998.

NUSSBAUM, M.; SEN, A. **The quality of life**: New York: Oxford University Press, 1993. 453p.

PALIT, A. K.; POPOVIC, D. Computational intelligence in time series forecasting: theory and engineering applications. Springer, 2006a.

\_\_\_\_\_. **Computational intelligence in time series forecasting: theory and engineering applications.** Londres: Springer Science & Business Media, 2006b. 372p.

PASSARI, A. F. L. **Exploração de dados atomizados para previsão de vendas no varejo utilizando redes neurais,** 2003. 143f. (Mestrado em Administração) - Departamento de Administração, Universidade de São Paulo, São Paulo, 2003.

PEÑA, D.; SÁNCHEZ, I. Measuring the advantages of multivariate vs. univariate forecasts. **Journal of Time Series Analysis,** v. 28, n. 6, p. 886-909, 2007.

PENTAHO. **Time Series Analysis and Forecasting with Weka,** 2014. Disponível em: <http://wiki.pentaho.com/display/DATAMINING/Time+Series+Analysis+and+Forecasting+with+Weka> >. Acesso em: 02-nov-2014.

PERCIO, J.; MEDINA, N. H.; LUNA, E. A. Visual Impairment and Human Development in Brazil. **International Journal of Epidemiology,** v. 44, n. suppl 1, p. i157, out 2015.

PESTANA, M. H.; GAGEIRO, J. N. Análise de dados para ciências sociais: a complementaridade do SPSS. 2003.

PIQUER, Á. G. et al. Assessment of competences in university degrees using data mining techniques, Validación de Competencias en Titulaciones Universitarias Usando Minería de Datos **Revista Iberoameri,** v. 5, n. 1, p. 23-29, fev 2010.

PLATT, J. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.

PUTSIS, W. P. Parameter variation and new product diffusion. **Journal of Forecasting,** v. 17, n. 3- 4, p. 231-257, 1998.

RAVALLION, M. The human development index: a response to Klugman, Rodriguez and Choi. **The Journal of Economic Inequality,** v. 9, n. 3, p. 475-478, 2011.

\_\_\_\_\_. Troubling tradeoffs in the human development index. **J Dev Econ,** v. 99, n. 2, p. 201-209, 2012a.

\_\_\_\_\_. Troubling tradeoffs in the human development index. **Journal of Development Economics,** v. 99, n. 2, p. 201-209, 2012b.

RAVISANKAR, P. et al. Detection of financial statement fraud and feature selection using data mining techniques. **Decision Support Systems,** v. 50, n. 2, p. 491-500, 2011.

RILEY, J. C. **Poverty and life expectancy: The Jamaica paradox.** Cambridge: Cambridge University Press, 2005. 250p.

ROBERTSON, P. L.; PATEL, P. R. New wine in old bottles: Technological diffusion in developed economies. **Research Policy**, v. 36, n. 5, p. 708-721, 2007.

ROCHA, F. R. F. A previdência social no Brasil: uma política em reestruturação. **Temporalis**, v. 2, n. 30, p. 453-473, 2016.

RODRIGUES, A. M. **Técnicas de Data Mining classificadas do ponto de vista do usuário**, 2000. 116f. Dissertação (Mestrado em Engenharia de Produção) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2000.

RODRIGUES, B. D.; STEVENSON, M. J. Takeover prediction using forecast combinations. **International Journal of Forecasting**, v. 29, n. 4, p. 628-641, 2013.

ROSA, V. D. Atividade física e a qualidade de vida de mulheres idosas. **FACES**, 2016.

SADOVSKY, A. D. I. D. et al. Índice de Desenvolvimento Humano e prevenção secundária de câncer de mama e colo do útero: um estudo ecológico. **Cadernos de Saúde Pública**, v. 31, p. 1539-1550, 2015.

SALGADO FILHO, N.; BRITO, D. J. A. Doença renal crônica: a grande epidemia deste milênio. **J Bras Nefrol**, v. 28, n. 2, p. 1-5, 2006.

SHEVADE, S. K. et al. Improvements to the SMO algorithm for SVM regression. **Neural Networks, IEEE Transactions on**, v. 11, n. 5, p. 1188-1193, 2000.

SILVA, L. A feature engineering approach to wind power forecasting: GEFCom 2012. **International Journal of Forecasting**, v. 30, n. 2, p. 395-401, 2014.

SILVA, R. F. G. D. **Data mining na caracterização geo-espacial e previsão da incidência de pneumonia em Portugal**, 2016. -

SIMPSON, D. Introduction to Rousseeuw (1984) Least Median of Squares Regression. In: (Ed.). **Breakthroughs in Statistics**: Springer, 1997. p.433-461.

SMOLA, A. J.; SCHÖLKOPF, B. A tutorial on support vector regression. **Statistics and computing**, v. 14, n. 3, p. 199-222, 2004.

SOUSA, W. R. N. et al. Evaluation of desertification processes in ouricuri-pe through trend estimates of times series. **Latin America Transactions, IEEE (Revista IEEE America Latina)**, v. 11, n. 1, p. 602-606, 2013.

STORBACKA, K. A solution business model: Capabilities and management practices for integrated solutions. **Industrial Marketing Management**, v. 40, n. 5, p. 699-711, 2011.

STRINGHINI, S. et al. Life-course socioeconomic status and DNA methylation of genes regulating inflammation. **International Journal of Epidemiology**, v. 44, n. 4, p. 1320-1330, ago 2015.

STROHMEIER, S.; PIAZZA, F. Domain driven data mining in human resource management: A review of current research. **Expert Systems with Applications**, v. 40, n. 7, p. 2410-2420, jun 2013.

SZUSTER, D. A. C. et al. Sobrevida de pacientes em diálise no SUS no Brasil. **Cadernos de Saúde Pública**, v. 28, p. 415-424, 2012.

TAN, P.-N.; STEINBACH, M.; KUMAR, V. **Introduction to Data Mining**: 1.ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 2005.

TANG, L.; YU, L.; HE, K. A novel data-characteristic-driven modeling methodology for nuclear energy consumption forecasting. **Applied Energy**, v. 128, p. 1-14, 2014.

TAVARES, L. F. et al. Dietary patterns of Brazilian adolescents: results of the Brazilian National School-Based Health Survey (PeNSE). **Cadernos de Saúde Pública**, v. 30, p. 2679-2690, 2014a.

\_\_\_\_\_. Dietary patterns of Brazilian adolescents: results of the Brazilian National School-Based Health Survey (PeNSE). **Cadernos de Saúde Pública**, v. 30, n. 12, p. 2679-2690, 2014b.

TREMBLAY, M. C.; DUTTA, K.; VANDERMEER, D. Using data mining techniques to discover bias patterns in missing data. **Journal of Data and Information Quality (JDIQ)**, v. 2, n. 1, p. 2, 2010.

UNDATA. United Nations Development Programme. **Human Development Index trends, 1980–2013**. Undata. <http://data.un.org/DocumentData.aspx?id=364>. Cons. 02-nov-2014.

UNDESA. United Nations Department of Economic and Social Affairs. **World Population Prospects: The 2012 Revision**, New York, 2013. Disponível em: < <http://esa.un.org/unpd/wpp> >. Acesso em: 02 november.

UNDP. **Human Development Report (HDR) 1990: Concept and Measurement of human development**. United Nations Development Programme. New York:PRESS, O. U., p.141. 1990

\_\_\_\_\_. **Human Development Report (HDR) 1991: Financing human development.** United Nations Development Programme. New York:PRESS, O. U. 1991

\_\_\_\_\_. **Human Development Report (HDR)1992: Global dimensions of human development.** United Nations Development Programme. New York:PRESS, O. U. 1992

\_\_\_\_\_. **Human Development Report (HDR) 1993: People's participation.** United Nations Development Programm. New York:PRESS, O. U. 1993

\_\_\_\_\_. **Human Development Report (HDR) 1994: New dimensions of human security.** United Nations Development Programme. New York:PRESS, O. U. 1994

\_\_\_\_\_. **Human Development Report (HDR) 1995: Gender and human development** United Nations Development Programme. New York:PRESS, O. U. 1995

\_\_\_\_\_. **Human Development Report (HDR) 1996: Economic growth and human development.** United Nations Development Programme. New York:PRESS, O. U. 1996

\_\_\_\_\_. **Human Development Report (HDR) 1997: Human development to eradicate poverty.** United Nations Development Programme. New York:PRESS, O. U. 1997

\_\_\_\_\_. **Human Development Report (HDR) 1998: Consumption for human development.** United Nations Development Programme. New York:PRESS, O. U. 1998

\_\_\_\_\_. **Human Development Report (HDR) 1999: Globalization with a human face.** United Nations Development Programme. New York:PRESS, O. U. 1999

\_\_\_\_\_. **Human Development Report (HDR) 2000: Human rights and human development.** United Nations Development Programme. New York:PRESS, O. U. 2000

\_\_\_\_\_. **Human Development Report (HDR) 2001: Making new technologies work for human development.** United Nations Development Programme. New York:PRESS, O. U. 2001

\_\_\_\_\_. **Human Development Report (HDR) 2002: Deepening democracy in a fragmented world.** United Nations Development Programme. New York:PRESS, O. U. 2002



\_\_\_\_\_. **Human Development Report (HDR) 2003: Millennium Development Goals: A compact among nations to end human poverty.** United Nations Development Programme. New York:PRESS, O. U. 2003

\_\_\_\_\_. **Human Development Report (HDR) 2004: Cultural Liberty in today's diverse world.** United Nations Development Programme. New York:CO, H. P. 2004

\_\_\_\_\_. **Human Development Report (HDR) 2005: International cooperation at a crossroads: aid, trade and security in an unequal world.** United Nations Development Programme. New York:CO, H. P. 2005

\_\_\_\_\_. **Human Development Report (HDR) 2006: Beyond scarcity: power, poverty and the global water crisis.** United Nations Development Programme. New York:PALGRAVE. 2006

\_\_\_\_\_. **Human Development Report (HDR) 2007/2008: Fighting climate change: human solidarity in a divided world.** United Nations Development Programme. New York:PALGRAVE. 2007/2008

\_\_\_\_\_. **Human Development Report (HDR) 2009: Overcoming barriers: human mobility and development.** United Nations Development Programme. New York:PALGRAVE. 2009

\_\_\_\_\_. **Human Development Report (HDR) 2010: The Real Wealth of Nations. Palgrave Macmillan for the UNDP.** United Nations Development Programme. New York:PRESS, O. U. 2010

\_\_\_\_\_. **Human Development Report (HDR) 2011: Sustainability and Equity: A Better Future for All.** United Nations Development Programme. New York:PRESS, O. U., p.185. 2011

\_\_\_\_\_. **Human Development Report (HDR) 2011/2012: The Sustainable Future We Want.** United Nations Development Programme. New York:PRESS, O. U. 2011-2014

\_\_\_\_\_. **Human Development Report (HDR) 2013: The Rise of the South: Human Progress in a Diverse World.** United Nations Development Programme. New York:PRESS, O. U., p.239. 2013

\_\_\_\_\_. **Human Development Report (HDR) 2014. Sustaining Human Progress: Reducing Vulnerabilities and Building Resilience.** United Nations Development Programme. New York:PRESS, O. U. 2014

\_\_\_\_\_. **Human Development Report (HDR) 2015. Rethinking Work for Human Development.** United Nations Development Programme. New York:PRESS, O. U. 2015

UNESCO. Institute for Statistics. **Education: Duration by level of education.** Data Centre. <http://data.uis.unesco.org/index.aspx?queryid=219>.Cons. 02-nov-2014.

\_\_\_\_\_. Institute for Statistics. **Education: Mean years of schooling.** Data Centre. <http://data.uis.unesco.org/index.aspx?queryid=218>.Cons. 02-nov-2014.

UNIDO. **United Nations Industrial Development Organization.** <http://www.unido.org/resources/statistics/statistical-databases.htm>.Cons. 28-jun-2013.

UNSD. United Nations Statistics Division. **GDP and its breakdown at current prices in US Dollars.** National Accounts Main Aggregates Database. <http://unstats.un.org/unsd/snaama/dnlList.asp>.Cons. 02-nov-2014.

VASCONCELOS, B. S. **Mineração de regras de classificação com sistemas de banco de dados objeto - relacional. Estudo de caso: regras de classificação de litofácies de poços de petróleo,** 2002a. 128f. Dissertação (Mestrado em Informática) - Universidade Federal de Campina Grande, Campina Grande, 2002.

VASCONCELOS, E. M. **Complexidade e pesquisa interdisciplinar:** Epistemologia e Metodologia Operativa. Petrópolis: Vozes, 2002b. 1000p.

VECCHIA, R. D. et al. Qualidade de vida na terceira idade: um conceito subjetivo. **Rev Bras Epidemiol**, v. 8, n. 3, p. 246-252, 2005.

VIANA, R. Mineração de dados: teoria e prática. **Revista SQL Magazine**, n. 10, 2004.

WANG, J.-J. et al. Stock index forecasting based on a hybrid model. **Omega**, v. 40, n. 6, p. 758-766, 2012.

WANG, Y. et al. Applying data mining techniques to WIFLY in customer relationship management. **Information Technology Journal**, v. 9, n. 3, p. 488-493, 2010.

WATTS, J. Brazil's health system woes worsen in economic crisis. **The Lancet**, v. 387, n. 10028, p. 1603-1604, 2016.

WILLIAMS, D.; HINTON, G. Learning representations by back-propagating errors. **Nature**, v. 323, n. 6088, p. 533-538, 1986.

WINTERS-MINER, L. A. et al. Chapter 3 - Biomedical Informatics. In: (Ed.). **Practical Predictive Analytics and Decisioning Systems for Medicine**: Academic Press, 2015. p.42-59.

WITTEN, I. H.; FRANK, E. **Data Mining**: Practical machine learning tools and techniques. 2.ed. San Francisco, USA: Morgan Kaufmann, 2005. 558p.

WITTEN, I. H. et al. **Data Mining**: Practical machine learning tools and techniques. Morgan Kaufmann, 2016.

WORLDBANK. The World Bank. **World Development Indicators**. World DataBank. <http://databank.worldbank.org/data/views/reports/tableview.aspx>. Cons. 02-nov-2014.

\_\_\_\_\_. The World Bank. **World Development Indicators**. World DataBank. [http://databank.worldbank.org/data/reports.aspx?Code=SP.DYN.LE00.IN&id=af3ce82b&report\\_name=Popular\\_indicators&populartype=series&ispopular=y](http://databank.worldbank.org/data/reports.aspx?Code=SP.DYN.LE00.IN&id=af3ce82b&report_name=Popular_indicators&populartype=series&ispopular=y). Cons. 19-set-2016.

XIE, Y. et al. Analyzing health insurance claims on different timescales to predict days in hospital. **Journal of Biomedical Informatics**,

\_\_\_\_\_. Analyzing health insurance claims on different timescales to predict days in hospital. **J Biomed Inform**, v. 60, p. 187-196, 2016.

YIN, R. K. **Estudo de caso**: Planejamento e Métodos. 2.ed. Porto Alegre: Bookman, 2001.

ZHU, N. et al. Regression-Based Microblogging Influence Detection Framework for Stock Market. **JNW**, v. 9, n. 8, p. 2129-2136, 2014.

## **APÊNDICE 1 – Exemplos de Modelos de Previsão**

### Quadro 9 - Modelo Univariado de Previsão do IDH 2014 para o País Botswana

<pre> === Run information === Scheme:   SMOReg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P   1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"  Lagged and derived variable options:   -F [Botswana] -L 1 -M 5 -G Ano  Relation:IDH_Temporal_ultimo- weka.filters.unsupervised.attribute.Remove-R1 Instances: 13 Attributes: 188   [list of attributes omitted]  Transformed training data:   Ano   Botswana   Ano-remapped   Lag_Botswana-1   Lag_Botswana-2   Lag_Botswana-3   Lag_Botswana-4   Lag_Botswana-5   Ano-remapped^2   Ano-remapped^3   Ano-remapped*Lag_Botswana-1   Ano-remapped*Lag_Botswana-2   Ano-remapped*Lag_Botswana-3   Ano-remapped*Lag_Botswana-4   Ano-remapped*Lag_Botswana-5 ----- Instances were inserted in the taining data for the following time-stamps (target values set by interpolation):  1981-01-01T00:00:00, 1982-01-01T00:00:00, 1983-01-01T00:00:00, 1984-01-01T00:00:00, 1986-01-01T00:00:00, 1987-01-01T00:00:00, 1988-01-01T00:00:00, 1989-01-01T00:00:00, 1991-01-01T00:00:00, 1992-01-01T00:00:00, 1993-01-01T00:00:00, 1994-01-01T00:00:00, 1995-01-01T00:00:00, 1996-01-01T00:00:00, 1997-01-01T00:00:00, 1998-01-01T00:00:00, 1999-01-01T00:00:00, 2001-01-01T00:00:00, 2002-01-01T00:00:00, 2003-01-01T00:00:00, 2004-01-01T00:00:00 ----- The following training instances had missing values imputed via interpolation. Check source data as this may affect forecasting performance:                  2,3,4,5,7,8,9,10,12,13,14,15,16,17,18,19,20,22,23,24,25  Botswana: SMOReg  weights (not support vectors): + 0.2135 * (normalized) Ano-remapped + 0.6045 * (normalized) Lag_Botswana-1 + 0.1694 * (normalized) Lag_Botswana-2 + 0.0831 * (normalized) Lag_Botswana-3 + 0.0047 * (normalized) Lag_Botswana-4 - 0.1526 * (normalized) Lag_Botswana-5 + 0.2042 * (normalized) Ano-remapped^2 + 0.2234 * (normalized) Ano-remapped^3 - 0.1336 * (normalized) Ano-remapped*Lag_Botswana-1 - 0.2317 * (normalized) Ano-remapped*Lag_Botswana-2 - 0.121 * (normalized) Ano-remapped*Lag_Botswana-3 - 0.0293 * (normalized) Ano-remapped*Lag_Botswana-4 + 0.0654 * (normalized) Ano-remapped*Lag_Botswana-5 + 0.1362 </pre>	<pre> Number of kernel evaluations: 595 (97.803% cached) === Future predictions from end of training data === Time Botswana 1980 0.47 1981 0.4816 1982 0.4932 1983 0.5048 1984 0.5164 1985 0.528 1986 0.539 1987 0.55 1988 0.561 1989 0.572 1990 0.583 1991 0.5807 1992 0.5784 1993 0.5761 1994 0.5738 1995 0.5715 1996 0.5692 1997 0.5669 1998 0.5646 1999 0.5623 2000 0.56 2001 0.57 2002 0.58 2003 0.59 2004 0.6 2005 0.61 2006 0.625 2007 0.643 2008 0.656 2009 0.662 2010 0.672 2011 0.678 2012 0.681 2013 0.683 2014* 0.697  === Evaluation on training data === Target                1-step-ahead ===== Botswana N                        29 Mean absolute error   0.0037 Root relative squared error   58.4752 Direction accuracy   85.7143 Relative absolute error   50.0986 Mean absolute percentage error   0.6291 Root mean squared error                0.0051 Mean squared error   0 Total number of instances:                34 </pre>
--	---

### Quadro 10 - Modelo Multivariado Específico de Previsão do IDH 2014 para o País New Zealand

```

=== Run information ===

Scheme:
  SMOReg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001
-V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"

Lagged and derived variable options:
  -F "[Germany, Ireland, Israel, Mauritania,
Switzerland, Tonga, New Zealand]" -L 1 -M 5 -G Ano

Relation: IDH_Temporal_ultimo-
weka.filters.unsupervised.attribute.Remove-R1-          weka.
filters.supervised.attribute.AttributeSelection-Eweka.attribute
Selection .CfsSubsetEval -P 1 -E 1-weka.
attributeSelection.BestFirst -P 1 -D 1 -N 5

Instances: 12
Attributes: 8
  Ano
  Germany
  Ireland
  Israel
  Mauritania
  Switzerland
  Tonga
  New Zealand

Transformed training data:
  Ano
  Germany
  Ireland
  Israel
  Mauritania
  Switzerland
  Tonga
  New Zealand
  Ano-remapped
  Lag_Germany-1
  Lag_Germany-2
  Lag_Germany-3
  Lag_Germany-4
  Lag_Germany-5
  Lag_Ireland-1
  ...
  Ano-remapped^2
  Ano-remapped^3
  Ano-remapped^Lag_Germany-1
  ...
  Ano-remapped^Lag_New Zealand-4
  Ano-remapped^Lag_New Zealand-5

SMOReg
weights (not support vectors):
+ 0.5171 * (normalized) Ano-remapped
+ 0.0899 * (normalized) Lag_Germany-1
+ 0.0208 * (normalized) Lag_Germany-2
- 0.0105 * (normalized) Lag_Germany-3
- 0.0202 * (normalized) Lag_Germany-4
+ 0.029 * (normalized) Lag_Germany-5
+ 0.0916 * (normalized) Lag_Ireland-1
+ 0.0063 * (normalized) Lag_Ireland-2
+ 0.0145 * (normalized) Lag_Ireland-3
+ 0.0149 * (normalized) Lag_Ireland-4
+ 0.0124 * (normalized) Lag_Ireland-5
+ 0.0419 * (normalized) Lag_Israel-1
- 0.0035 * (normalized) Lag_Israel-2
+ 0.0227 * (normalized) Lag_Israel-3
+ 0.0386 * (normalized) Lag_Israel-4
+ 0.0528 * (normalized) Lag_Israel-5
- 0.0065 * (normalized) Lag_Mauritania-1
- 0.0851 * (normalized) Lag_Mauritania-2
- 0.1313 * (normalized) Lag_Mauritania-3
- 0.0847 * (normalized) Lag_Mauritania-4
- 0.0509 * (normalized) Lag_Mauritania-5
+ 0.0946 * (normalized) Lag_Switzerland-1
+ 0.0519 * (normalized) Lag_Switzerland-2
+ 0.022 * (normalized) Lag_Switzerland-3
+ 0.0297 * (normalized) Lag_Switzerland-4
+ 0.0264 * (normalized) Lag_Switzerland-5
+ 0.1542 * (normalized) Lag_Tonga-1
+ 0.1185 * (normalized) Lag_Tonga-2
+ 0.08 * (normalized) Lag_Tonga-3
+ 0.0149 * (normalized) Lag_Tonga-4
- 0.0213 * (normalized) Lag_Tonga-5
+ 0.0714 * (normalized) Lag_New Zealand-1
+ 0.0405 * (normalized) Lag_New Zealand-2
+ 0.0316 * (normalized) Lag_New Zealand-3
+ 0.0141 * (normalized) Lag_New Zealand-4
- 0.0036 * (normalized) Lag_New Zealand-5
+ 0.169 * (normalized) Ano-remapped^2
- 0.0536 * (normalized) Ano-remapped^3
- 0.0229 * (normalized) Ano-remapped^Lag_Germany-1
- 0.0264 * (normalized) Ano-remapped^Lag_Germany-2
- 0.0123 * (normalized) Ano-remapped^Lag_Germany-3
- 0.0082 * (normalized) Ano-remapped^Lag_Germany-4
- 0.0051 * (normalized) Ano-remapped^Lag_Germany-5
- 0.0203 * (normalized) Ano-remapped^Lag_Ireland-1
- 0.0264 * (normalized) Ano-remapped^Lag_Ireland-2
- 0.0063 * (normalized) Ano-remapped^Lag_Ireland-3
- 0.0013 * (normalized) Ano-remapped^Lag_Ireland-4
+ 0.0009 * (normalized) Ano-remapped^Lag_Ireland-5
- 0.0198 * (normalized) Ano-remapped^Lag_Israel-1
- 0.0183 * (normalized) Ano-remapped^Lag_Israel-2
+ 0.0047 * (normalized) Ano-remapped^Lag_Israel-3
+ 0.0119 * (normalized) Ano-remapped^Lag_Israel-4
+ 0.015 * (normalized) Ano-remapped^Lag_Israel-5
- 0.0486 * (normalized) Ano-remapped^Lag_Mauritania-1
- 0.0574 * (normalized) Ano-remapped^Lag_Mauritania-2
- 0.0537 * (normalized) Ano-remapped^Lag_Mauritania-3
- 0.0374 * (normalized) Ano-remapped^Lag_Mauritania-4
- 0.0265 * (normalized) Ano-remapped^Lag_Mauritania-5

Instances were inserted in the training data for the following time-stamps (target values set by
interpolation):
1981-01-01T00:00:00, 1982-01-01T00:00:00, 1983-01-01T00:00:00, 1984-01-01T00:00:00, 1986-01-
01T00:00:00, 1987-01-01T00:00:00, 1988-01-01T00:00:00, 1989-01-01T00:00:00, 1991-01-
01T00:00:00, 1992-01-01T00:00:00, 1993-01-01T00:00:00, 1994-01-01T00:00:00, 1995-01-
01T00:00:00, 1996-01-01T00:00:00, 1997-01-01T00:00:00, 1998-01-01T00:00:00, 1999-01-
01T00:00:00, 2001-01-01T00:00:00, 2002-01-01T00:00:00, 2003-01-01T00:00:00, 2004-01-
01T00:00:00

The following training instances had missing values imputed via interpolation. Check source data as
this may affect forecasting performance: 2,3,4,5,7,8,9,10,12,13,14,15,16,17,18,19,20,22,23,24,25

Germany:
SMOReg
...
Ireland:
...
Israel:
...
Mauritania:
...
Switzerland:
...
Tonga:
...
New Zealand:
...
Number of kernel evaluations: 561 (99.824% cached)

=== Future predictions from end of training data ===
Time      Germany      Ireland      Israel      Mauritania      Switzerland      Tonga      New Zealand
1980      0.739          0.734       0.749       0.347           0.806           0.602     0.793
1981      0.7416         0.7374      0.7534      0.3486          0.8074          0.605     0.7956
1982      0.7442         0.7408      0.7578      0.3502          0.8088          0.608     0.7982
1983      0.7468         0.7442      0.7622      0.3518          0.8102          0.611     0.8008
1984      0.7494         0.7476      0.7666      0.3534          0.8116          0.614     0.8034
1985      0.752          0.751       0.771       0.355           0.813           0.617     0.806
1986      0.758          0.7558      0.7738      0.3574          0.8162          0.6198    0.809
1987      0.764          0.7606     0.7766     0.3598          0.8194          0.6226    0.812
1988      0.777          0.7654     0.7794     0.3622          0.8226          0.6254    0.815
1989      0.776         0.7702     0.7822     0.3646          0.8258          0.6282    0.818
1990      0.782          0.775       0.785       0.367           0.829           0.631     0.821
1991      0.7892         0.7837     0.7914     0.3736          0.8347          0.6351    0.8262
1992      0.7964         0.7924     0.7978     0.3802          0.8404          0.6392    0.8314
1993      0.8036         0.8011     0.8042     0.3868          0.8461          0.6433    0.8366
1994      0.8108         0.8098     0.8106     0.3934          0.8518          0.6474    0.8418
1995      0.818          0.8185     0.817       0.4             0.8575          0.6515    0.847
1996      0.8252         0.8272     0.8234     0.4066          0.8632          0.6556    0.8522
1997      0.8324         0.8359     0.8298     0.4132          0.8689          0.6597    0.8574
1998      0.8396         0.8446     0.8362     0.4198          0.8746          0.6638    0.8626
1999      0.8468         0.8533     0.8426     0.4264          0.8803          0.6679    0.8678
2000      0.854          0.862       0.849       0.433           0.886           0.672     0.873
2001      0.8606         0.8676     0.853       0.4374          0.889           0.6766    0.8772
2002      0.8672         0.8732     0.857       0.4418          0.892           0.6812    0.8814
2003      0.8738         0.8788     0.861       0.4462          0.895           0.6858    0.8856
2004      0.8804         0.8844     0.865       0.4506          0.898           0.6904    0.8898
2005      0.887          0.89       0.869       0.455           0.901           0.695     0.894
2006      0.896          0.895     0.872       0.465           0.905           0.696     0.896
2007      0.899          0.901     0.877       0.467           0.905           0.697     0.899
2008      0.902          0.902     0.877       0.466           0.903           0.696     0.899
2009      0.901          0.898     0.878       0.474           0.909           0.698     0.903
2010      0.904          0.899     0.881       0.475           0.915           0.701     0.903
2011      0.908          0.9        0.885       0.475           0.914           0.702     0.904
2012      0.911          0.901     0.886       0.485           0.916           0.704     0.908
2013*     0.9148         0.9023     0.8895     0.4889          0.919           0.7064    0.9101

=== Evaluation on training data ===
Target      1-step-ahead
=====
New Zealand
  N          28
  Mean absolute error          0.0004
  Root relative squared error  16.3461
  Direction accuracy           92.5926
  Relative absolute error      10.7522
  Mean absolute percentage error 0.0456
  Root mean squared error      0.0007
  Mean squared error           0
Total number of instances: 33

```



## **APÊNDICE 2 – Medidas de Qualidade das Previsões dos Modelos**



Figura 19 - DAC dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

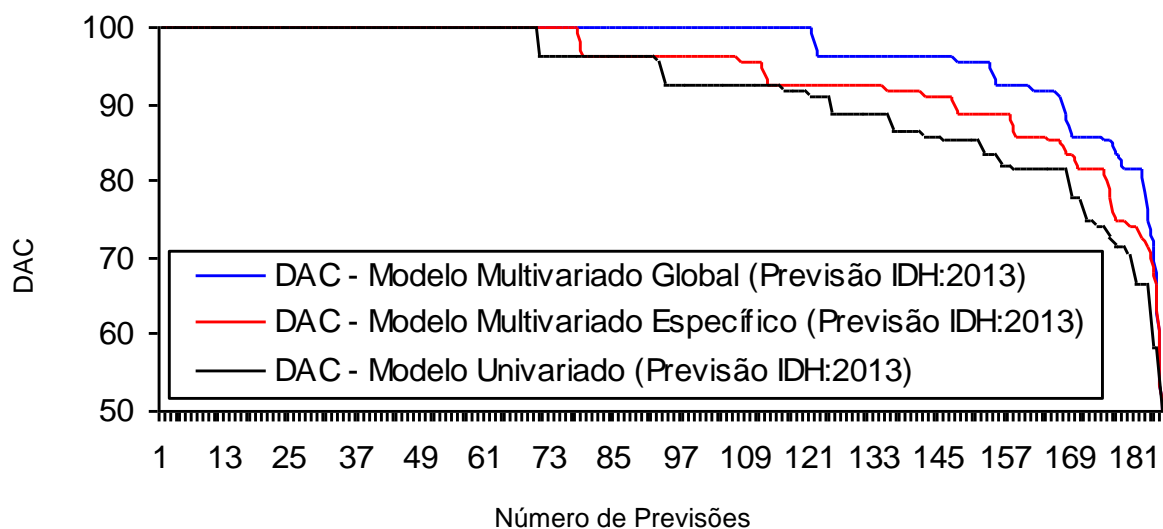


Figura 20 - DAC dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014

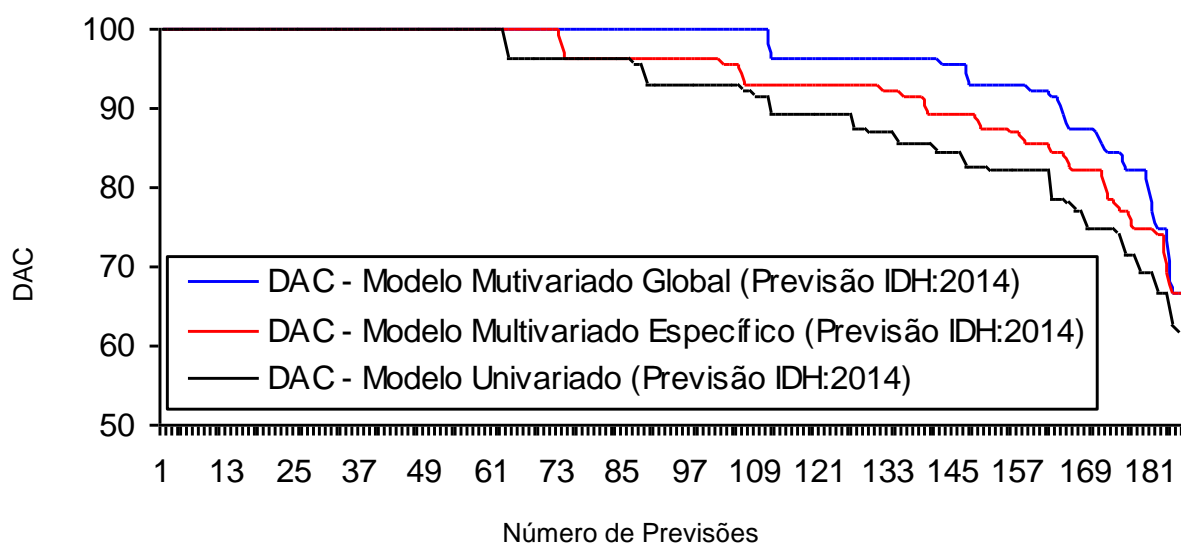


Figura 21 - MAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

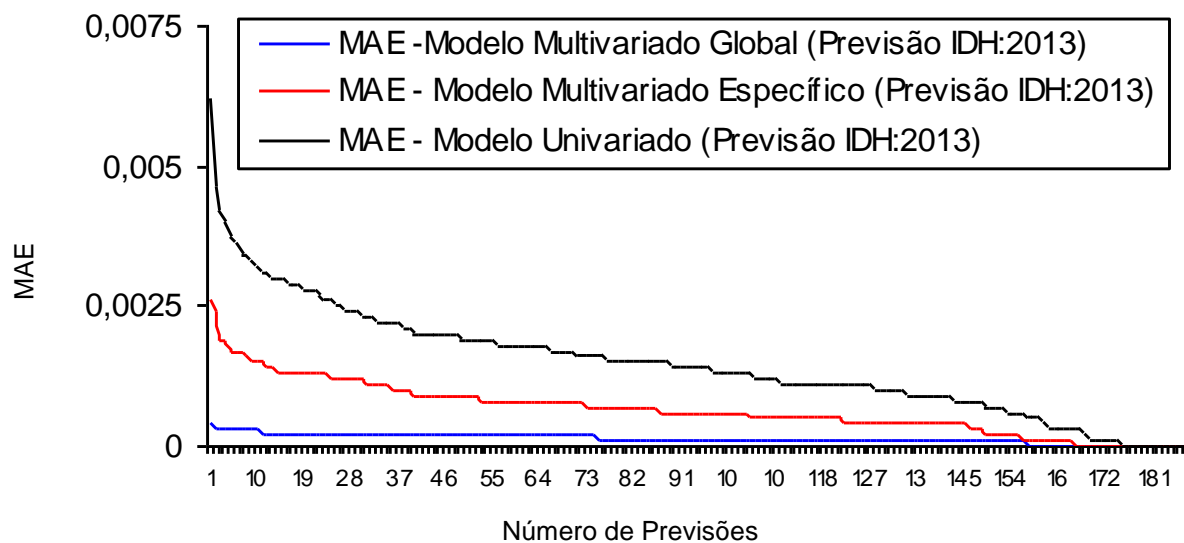


Figura 22 - MAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014

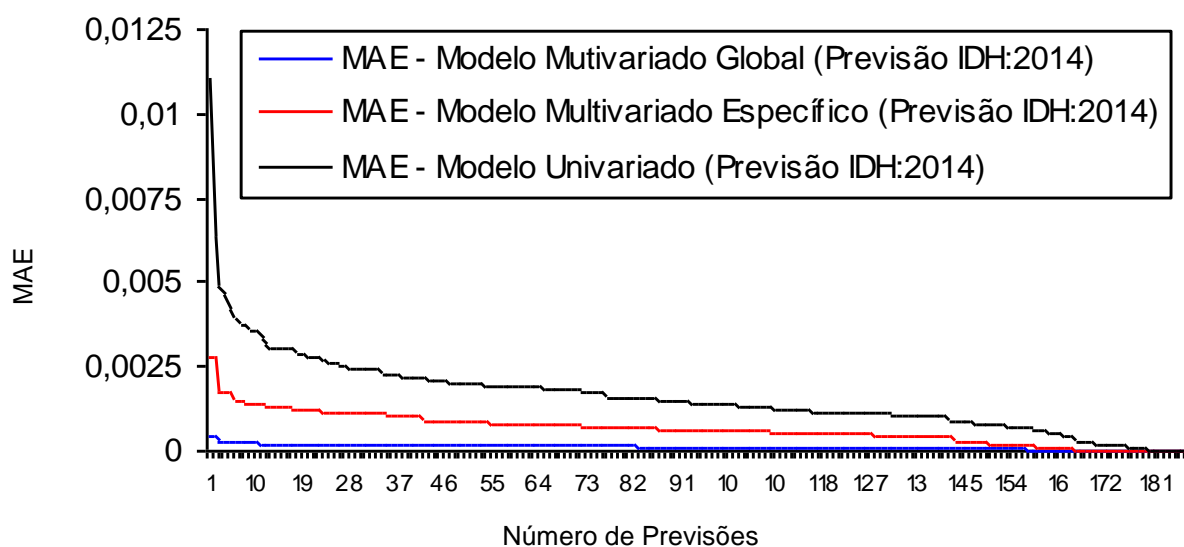


Figura 23 - MAPE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

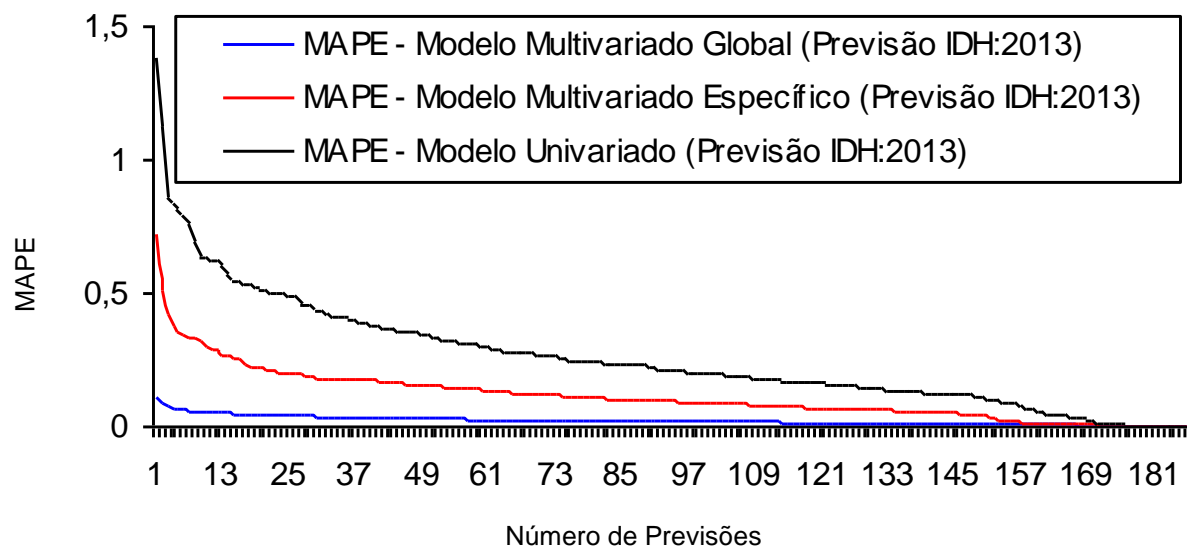


Figura 24 - MAPE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014

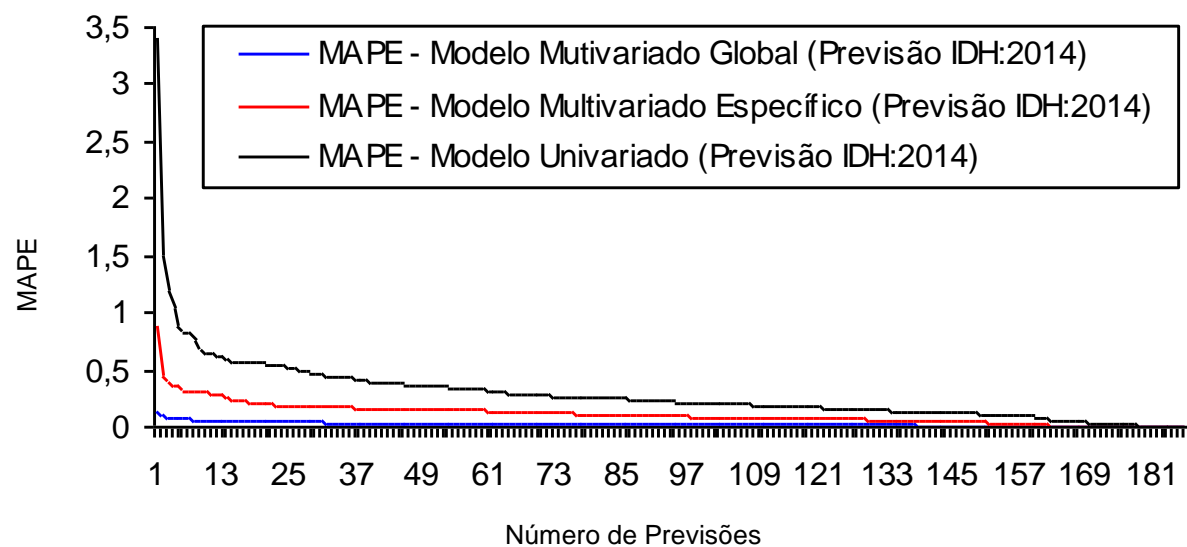


Figura 25 - MSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

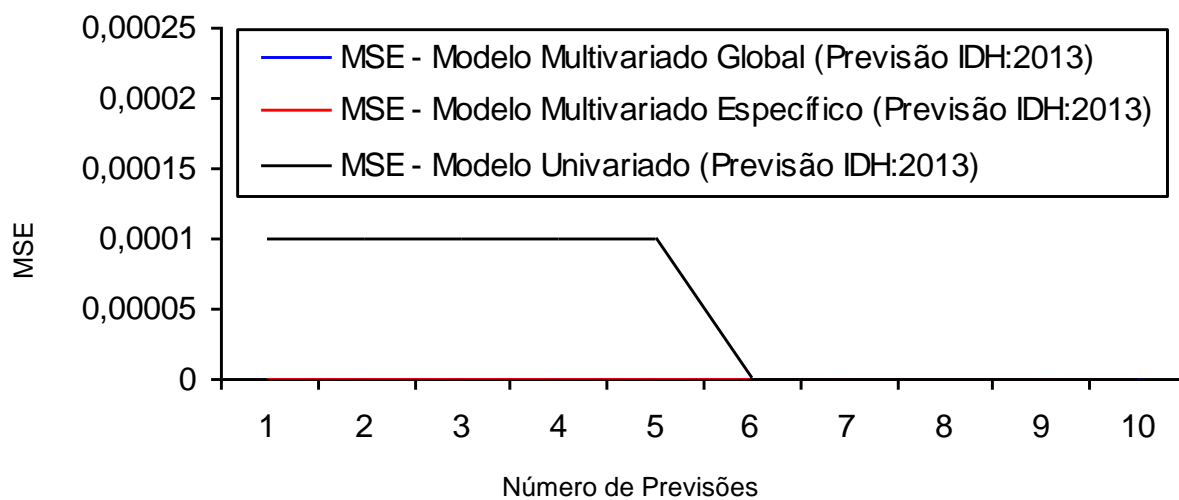


Figura 26 - MSE dos Modelos (MMGs, MMEs e MUs) para Previsão do IDH 2014

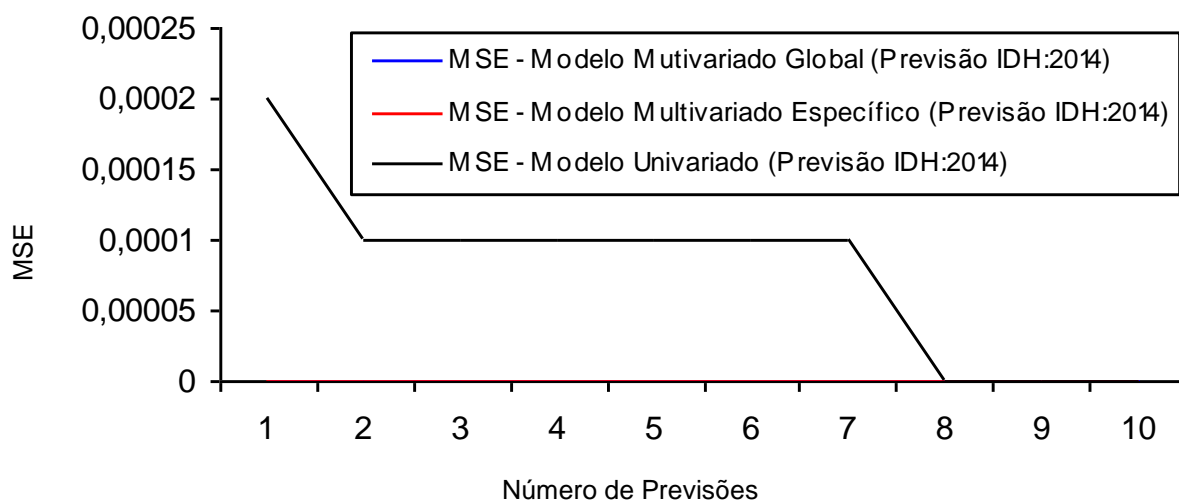


Figura 27 - RAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

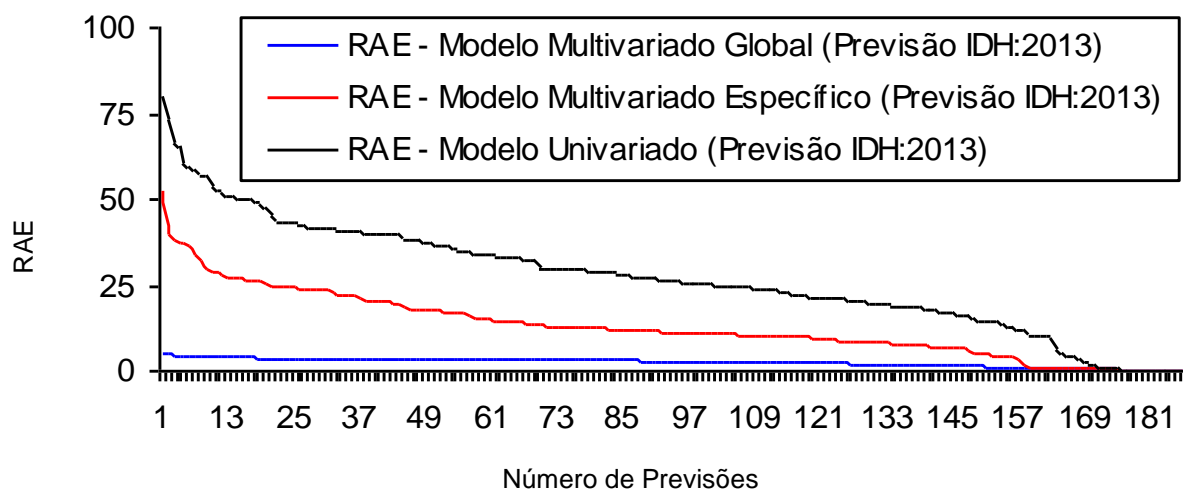


Figura 28 - RAE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014

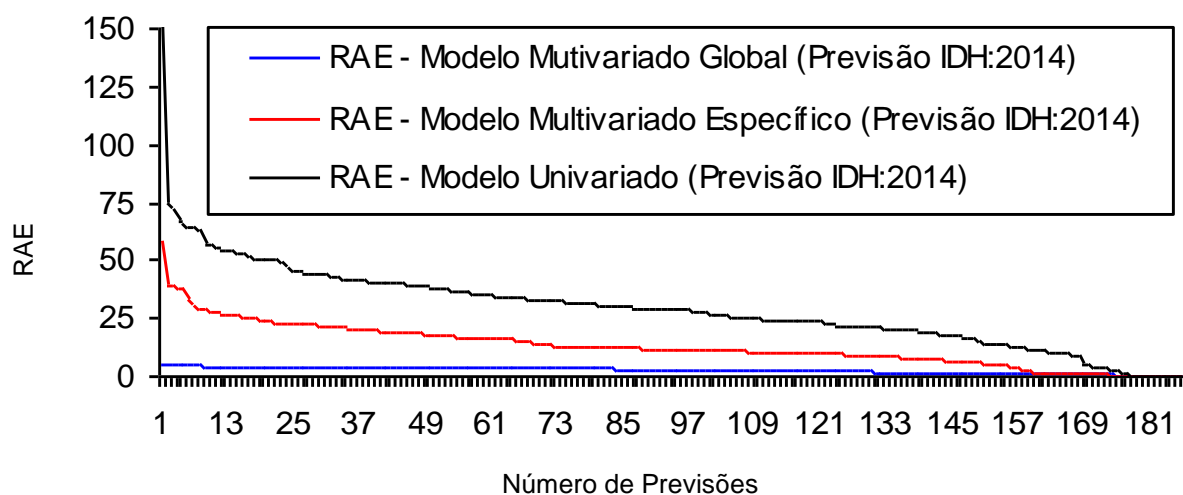


Figura 29 - RMSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

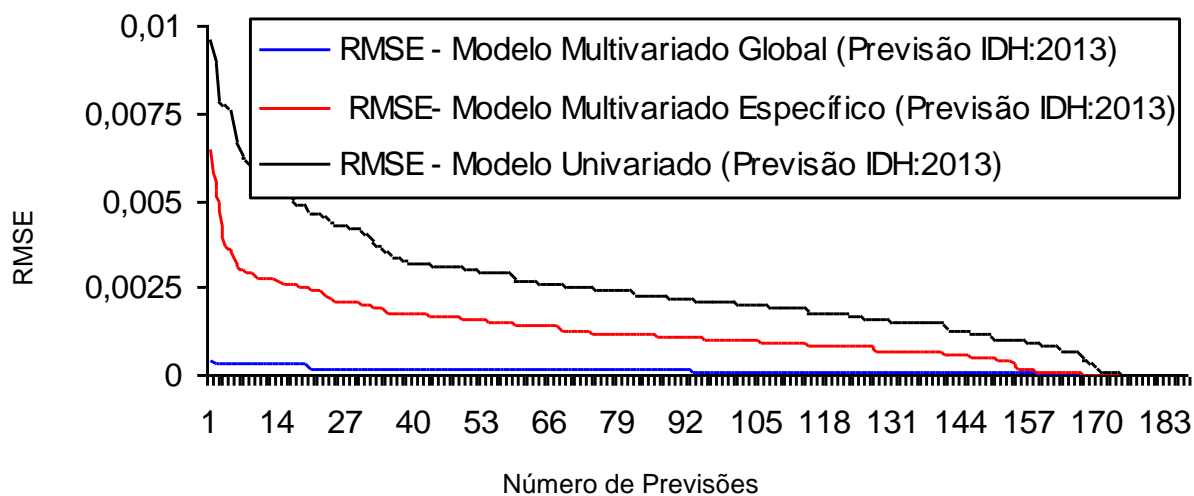


Figura 30 - RMSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014

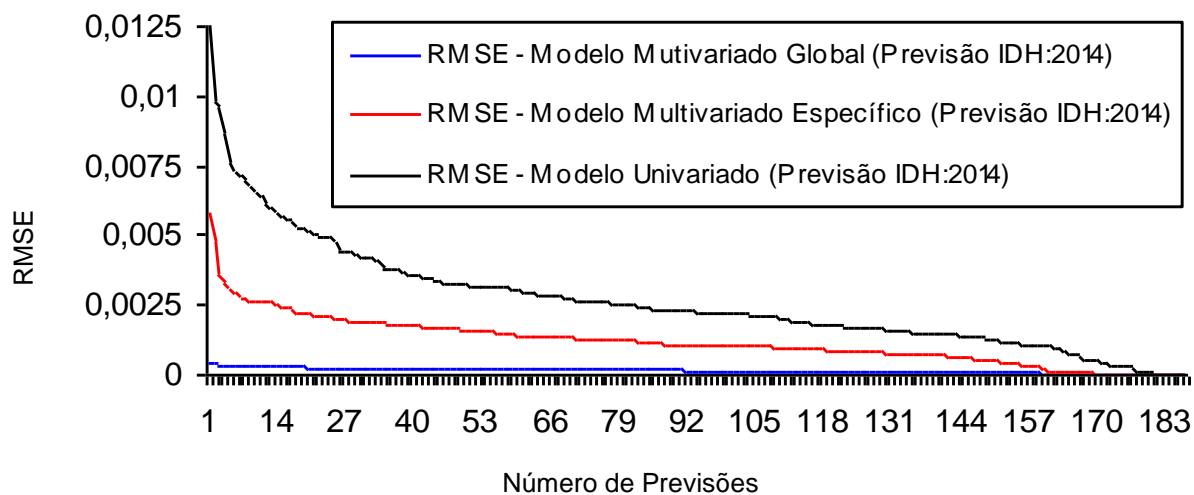


Figura 31 - RRSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2013

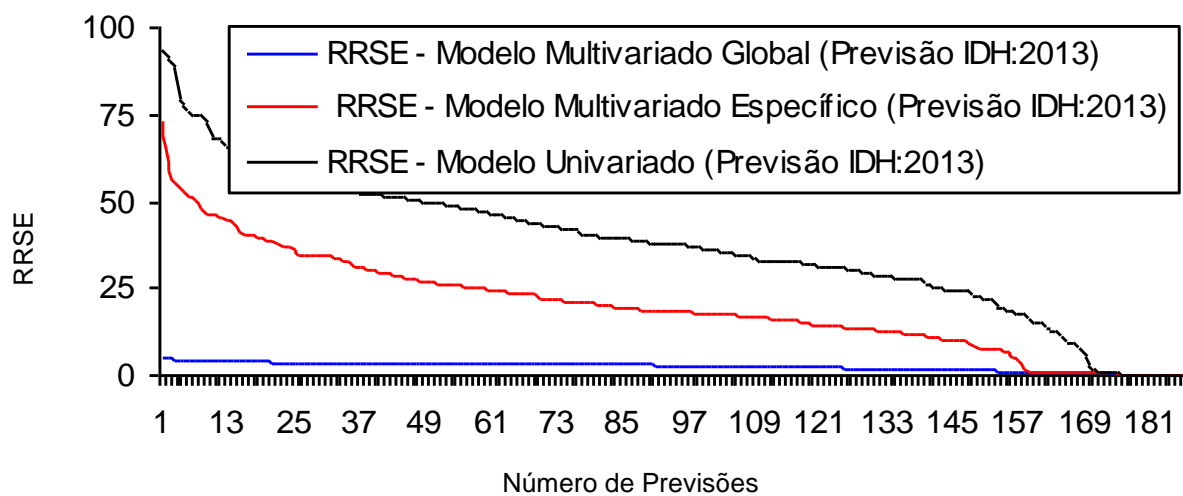
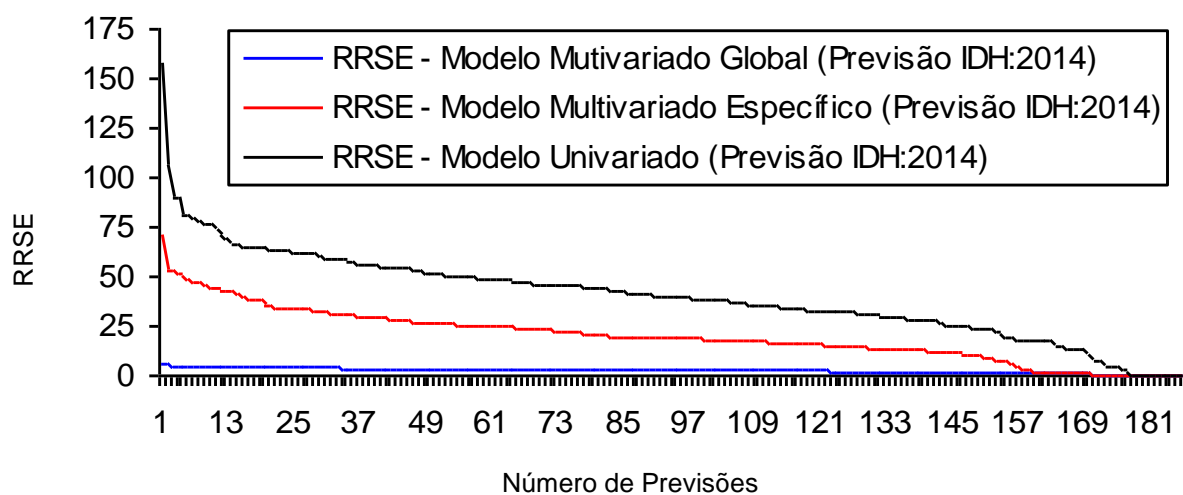


Figura 32 - RRSE dos Modelos (MMG, MMEs e MUs) para Previsão do IDH 2014



**APÊNDICE 3 – Resultados dos Testes Preliminares para Escolha do  
Modelo de Previsão desenvolvido pelo Algoritmo *SMOreg***



Figura 33 - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países Desenvolvidos com IDH muito alto

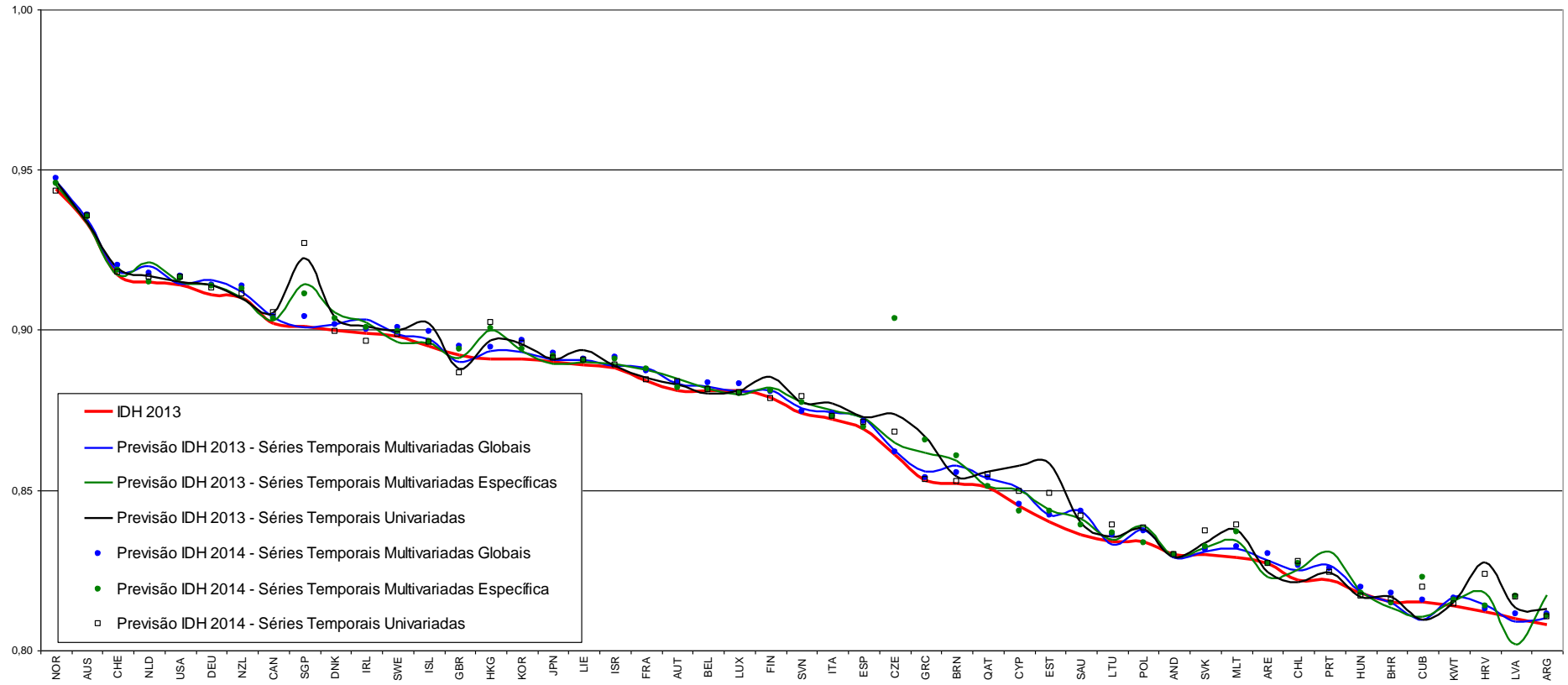
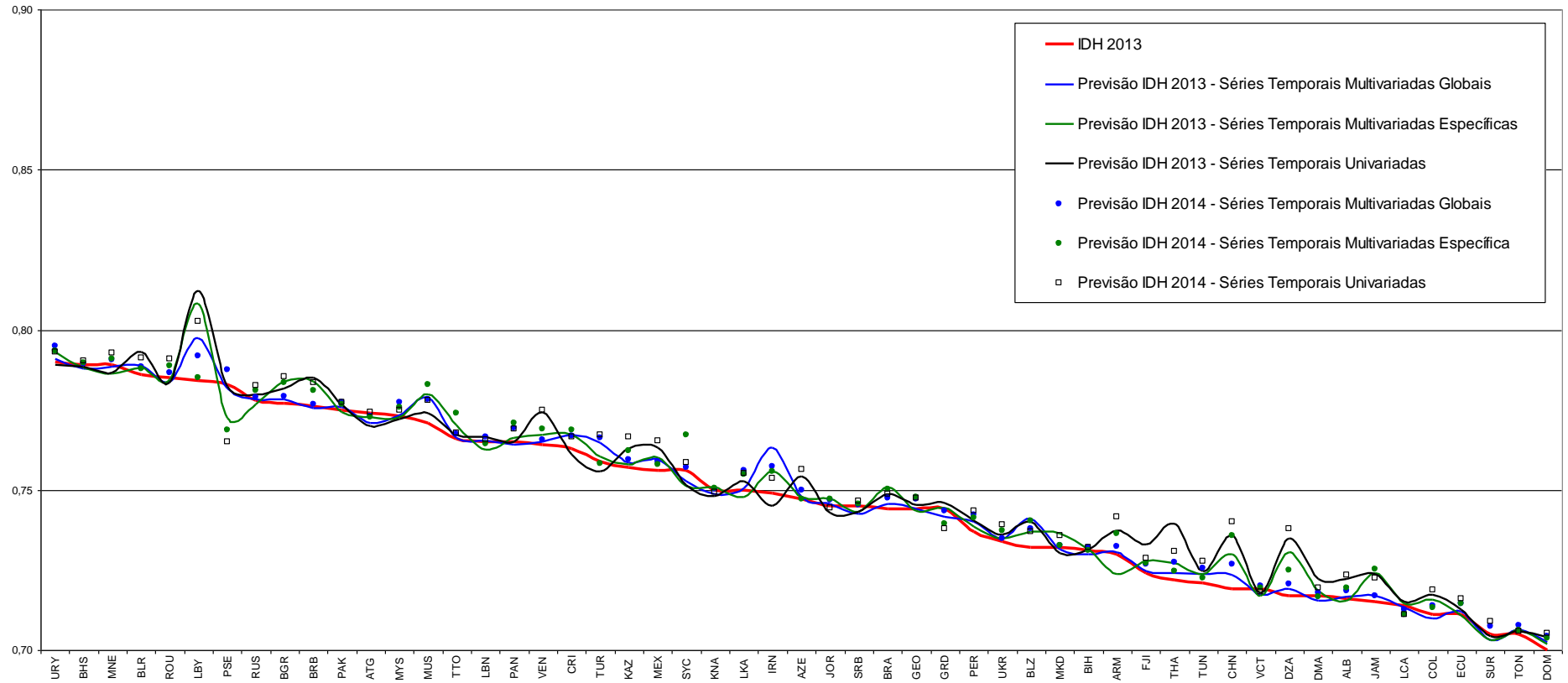


Figura 34 - IDH 2013 e Previsões dos MMGs, MMEs e MUs para o IDH (2013 e 2014) dos Países em Desenvolvimento com IDH alto







**Tabela 9 - Dados Reais (2012 e 2013) e Previsões dos Modelos Multivariados e Univariados para o IDH (2013 e 2014) dos países classificados por classe (Nível e IDH) e ranqueados em relação ao IDH 2013**

Classe	País (código internacional)	2012		2013		2013						2014		
		IDH	Posição	IDH	Posição	Modelos								
						Multivariado				Univariado				
						Global		Específico		Global		Específico		Univariado
Nível IDH	IDH	Posição	IDH	Posição	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	IDH(P)	IDH(P)	
Desenvolvido	NOR	0,943	1	0,944	1	0,9465	-0,0025	0,9453	-0,0013	0,9464	-0,0024	0,9474	0,9457	0,9432
	AUS	0,931	2	0,933	2	0,9344	-0,0014	0,9332	-0,0002	0,9335	-0,0005	0,9359	0,9352	0,9356
	CHE	0,916	3	0,917	3	0,9182	-0,0012	0,9172	-0,0002	0,9193	-0,0023	0,9202	0,9185	0,9179
	NLD	0,915	4	0,915	4	0,9198	-0,0048	0,9211	-0,0061	0,9169	-0,0019	0,9178	0,9148	0,9166
	USA	0,912	5	0,914	5	0,9144	-0,0004	0,9148	-0,0008	0,9148	-0,0008	0,9167	0,9161	0,9164
	DEU	0,911	6	0,911	6	0,9156	-0,0046	0,9141	-0,0031	0,914	-0,003	0,9141	0,9138	0,9132
	NZL	0,908	7	0,91	7	0,9118	-0,0018	0,9101	-0,0001	0,9096	0,0004	0,9138	0,9129	0,9113
	CAN	0,901	8	0,902	8	0,9039	-0,0019	0,903	-0,001	0,9055	-0,0035	0,9041	0,9036	0,9054
	SGP	0,899	11	0,901	9	0,9009	1E-04	0,9143	-0,0133	0,9224	-0,0214	0,9043	0,9113	0,9271
	DNK	0,9	10	0,9	10	0,9018	-0,0018	0,9055	-0,0055	0,904	-0,004	0,9016	0,9036	0,8996
	IRL	0,901	9	0,899	11	0,9032	-0,0042	0,9023	-0,0033	0,9011	-0,0021	0,9003	0,9009	0,8964
	SWE	0,897	12	0,898	12	0,8987	-0,0007	0,896	0,002	0,8997	-0,0017	0,9007	0,8996	0,8986
	ISL	0,893	13	0,895	13	0,897	-0,002	0,8959	-0,0009	0,9021	-0,0071	0,8996	0,8965	0,8961
	GBR	0,89	14	0,892	14	0,8899	0,0021	0,8913	0,0007	0,8877	0,0043	0,8949	0,8941	0,8866
	HKG	0,889	15	0,891	15	0,8935	-0,0025	0,8999	-0,0089	0,8967	-0,0057	0,8947	0,9005	0,9024
	KOR	0,888	16	0,891	16	0,893	-0,002	0,8933	-0,0023	0,8954	-0,0044	0,8968	0,8941	0,8958
	JPN	0,888	17	0,89	17	0,8907	-0,0007	0,8893	0,0007	0,8908	-0,0008	0,8929	0,8919	0,8911
	LIE	0,888	18	0,889	18	0,8906	-0,0016	0,8899	-0,0009	0,8937	-0,0047	0,8909	0,8904	0,8905
	ISR	0,886	19	0,888	19	0,8887	-0,0007	0,8893	-0,0013	0,8889	-0,0009	0,8915	0,8908	0,889
	FRA	0,884	20	0,884	20	0,8881	-0,0041	0,8875	-0,0035	0,8852	-0,0012	0,8871	0,8879	0,8843
	AUT	0,88	21	0,881	21	0,8831	-0,0021	0,8847	-0,0037	0,8829	-0,0019	0,8835	0,8819	0,8838
	BEL	0,88	22	0,881	22	0,8824	-0,0014	0,8817	-0,0007	0,8802	0,0008	0,8834	0,8817	0,8813
	LUX	0,88	23	0,881	23	0,8807	0,0003	0,8798	0,0012	0,881	0	0,8831	0,8802	0,8805
	FIN	0,879	24	0,879	24	0,8809	-0,0019	0,882	-0,003	0,8854	-0,0064	0,8807	0,8812	0,8786
	SVN	0,874	25	0,874	25	0,8754	-0,0014	0,8773	-0,0033	0,8772	-0,0032	0,8745	0,8775	0,8791
	ITA	0,872	26	0,872	26	0,8744	-0,0024	0,875	-0,003	0,8769	-0,0049	0,8741	0,8729	0,8731
	ESP	0,869	27	0,869	27	0,8725	-0,0035	0,8725	-0,0035	0,8728	-0,0038	0,8716	0,8696	0,8713
	CZE	0,861	28	0,861	28	0,8623	-0,0013	0,8646	-0,0036	0,8735	-0,0125	0,8618	0,9036	0,868
	GRC	0,854	29	0,853	29	0,8557	-0,0027	0,8616	-0,0086	0,8665	-0,0135	0,8538	0,8656	0,8533
	BRN	0,852	30	0,852	30	0,8576	-0,0056	0,8591	-0,0071	0,8542	-0,0022	0,8554	0,8606	0,8528
	QAT	0,85	31	0,851	31	0,8537	-0,0027	0,8511	-0,0001	0,8559	-0,0049	0,8539	0,8512	0,855
	CYP	0,848	32	0,845	32	0,8504	-0,0054	0,8499	-0,0049	0,8576	-0,0126	0,8457	0,8433	0,8497
	EST	0,839	33	0,84	33	0,8423	-0,0023	0,8437	-0,0037	0,8583	-0,0183	0,8422	0,8433	0,8489
	SAU	0,833	34	0,836	34	0,8435	-0,0075	0,8409	-0,0049	0,8397	-0,0037	0,8436	0,8391	0,842
	LTU	0,831	36	0,834	35	0,8329	0,0011	0,8346	-0,0006	0,8353	-0,0013	0,8362	0,8366	0,8392
	POL	0,833	35	0,834	36	0,8375	-0,0035	0,8388	-0,0048	0,8378	-0,0038	0,8373	0,8336	0,8382
	AND	0,83	37	0,83	37	0,8291	0,0009	0,8292	0,0008	0,829	0,001	0,8295	0,8298	0,8299
	SVK	0,829	38	0,83	38	0,8309	-0,0009	0,8322	-0,0022	0,8335	-0,0035	0,8313	0,8325	0,8374
	MLT	0,827	39	0,829	39	0,8316	-0,0026	0,8341	-0,0051	0,8376	-0,0086	0,8325	0,8371	0,8392
	ARE	0,825	40	0,827	40	0,8279	-0,0009	0,8227	0,0043	0,8242	0,0028	0,8303	0,827	0,827
	CHL	0,819	42	0,822	41	0,8249	-0,0029	0,8253	-0,0033	0,8213	0,0007	0,8265	0,8272	0,8276
	PRT	0,822	41	0,822	42	0,8264	-0,0044	0,8308	-0,0088	0,8244	-0,0024	0,8257	0,6631	0,8243
	HUN	0,817	43	0,818	43	0,8183	-0,0003	0,8184	-0,0004	0,8165	0,0015	0,8198	0,8178	0,8168
	BHR	0,813	44	0,815	44	0,8148	0,0002	0,8134	0,0016	0,8167	-0,0017	0,8179	0,8149	0,816
	CUB	0,813	45	0,815	45	0,8097	0,0053	0,8104	0,0046	0,8097	0,0053	0,8157	0,8228	0,8197

Classe	País (código internacional)	2012		2013		2013						2014		
		IDH	Posição	IDH	Posição	Modelos								
Nível	IDH	IDH	Posição	IDH	Posição	Multivariado				Univariado		Multivariado		Univariado
						Global		Específico		IDH(P)	Erro *	IDH(P)	Erro *	
						IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)		
Em desenvolvimento	KWT	0,813	46	0,814	46	0,8165	-0,0025	0,8155	-0,0015	0,815	-0,001	0,8162	0,8157	0,8145
	HRV	0,812	47	0,812	47	0,8142	-0,0022	0,8179	-0,0059	0,8274	-0,0154	0,8129	0,8139	0,8236
	LVA	0,808	48	0,81	48	0,809	0,001	0,8018	0,0082	0,8134	-0,0034	0,8113	0,8168	0,8166
	ARG	0,806	49	0,808	49	0,8102	-0,0022	0,8174	-0,0094	0,8128	-0,0048	0,8115	0,8109	0,8106
	URY	0,787	52	0,79	50	0,791	-0,001	0,7931	-0,0031	0,7891	0,0009	0,7949	0,7934	0,7932
	BHS	0,788	51	0,789	51	0,7879	0,0011	0,7882	0,0008	0,7883	0,0007	0,7895	0,7897	0,7904
	MNE	0,787	53	0,789	52	0,7885	0,0005	0,7862	0,0028	0,7865	0,0025	0,7907	0,791	0,7929
	BLR	0,785	54	0,786	53	0,7887	-0,0027	0,7882	-0,0022	0,7932	-0,0072	0,7883	0,7877	0,7913
	ROU	0,782	55	0,785	54	0,7837	0,0013	0,7844	0,0006	0,7836	0,0014	0,7865	0,7888	0,7909
	LBY	0,789	50	0,784	55	0,7974	-0,0134	0,8081	-0,0241	0,8123	-0,0283	0,7917	0,7849	0,8027
	PSE	0,781	56	0,783	56	0,7818	0,0012	0,7724	0,0106	0,7822	0,0008	0,7874	0,7687	0,765
	RUS	0,777	57	0,778	57	0,7782	-0,0002	0,7768	0,0012	0,7797	-0,0017	0,779	0,7812	0,7827
	BGR	0,776	58	0,777	58	0,7784	-0,0014	0,7839	-0,0069	0,7816	-0,0046	0,7793	0,7834	0,7855
	BRB	0,776	59	0,776	59	0,7755	0,0005	0,7838	-0,0078	0,785	-0,009	0,7766	0,7811	0,7835
	PAK	0,773	60	0,775	60	0,7757	-0,0007	0,7743	0,0007	0,7768	-0,0018	0,7772	0,7766	0,7772
	ATG	0,773	61	0,774	61	0,7709	0,0031	0,7726	0,0014	0,7698	0,0042	0,7731	0,7726	0,7742
	MYS	0,77	62	0,773	62	0,7733	-0,0003	0,7726	0,0004	0,7722	0,0008	0,7773	0,7758	0,7748
	MUS	0,769	63	0,771	63	0,779	-0,008	0,7798	-0,0088	0,774	-0,003	0,7782	0,7829	0,7781
	TTO	0,765	64	0,766	64	0,7662	-0,0002	0,7704	-0,0044	0,7671	-0,0011	0,7676	0,774	0,7677
	LBN	0,764	65	0,765	65	0,7654	-0,0004	0,7626	0,0024	0,7666	-0,0016	0,7665	0,7644	0,7657
	PAN	0,761	67	0,765	66	0,764	0,001	0,7663	-0,0013	0,7651	-0,0001	0,7693	0,7708	0,7691
	VEN	0,763	66	0,764	67	0,765	-0,001	0,7672	-0,0032	0,7742	-0,0102	0,7656	0,7689	0,7749
	CRI	0,761	68	0,763	68	0,7671	-0,0041	0,7671	-0,0041	0,7611	0,0019	0,7668	0,7687	0,7667
	TUR	0,756	69	0,759	69	0,7647	-0,0057	0,7604	-0,0014	0,7557	0,0033	0,7662	0,7581	0,7672
	KAZ	0,755	70	0,757	70	0,7587	-0,0017	0,7579	-0,0009	0,7628	-0,0058	0,7594	0,7623	0,7667
	MEX	0,755	71	0,756	71	0,7594	-0,0034	0,76	-0,004	0,7631	-0,0071	0,759	0,7578	0,7652
	SYC	0,755	72	0,756	72	0,7528	0,0032	0,7512	0,0048	0,7515	0,0045	0,757	0,7672	0,7587
	KNA	0,749	73	0,75	73	0,7487	0,0013	0,7508	-0,0008	0,7482	0,0018	0,7503	0,7504	0,7494
	LKA	0,745	75	0,75	74	0,7505	-0,0005	0,7479	0,0021	0,7528	-0,0028	0,7561	0,7548	0,7551
	IRN	0,749	74	0,749	75	0,7632	-0,0142	0,7559	-0,0069	0,7449	0,0041	0,7574	0,7557	0,7537
	AZE	0,745	76	0,747	76	0,7477	-0,0007	0,7478	-0,0008	0,7542	-0,0072	0,7499	0,7471	0,7563
	JOR	0,744	77	0,745	77	0,7456	-0,0006	0,7473	-0,0023	0,7427	0,0023	0,7467	0,7471	0,7444
	SRB	0,743	78	0,745	78	0,7425	0,0025	0,7431	0,0019	0,7432	0,0018	0,7454	0,7456	0,7466
	BRA	0,742	80	0,744	79	0,7455	-0,0015	0,7507	-0,0067	0,7486	-0,0046	0,7476	0,7501	0,7488
	GEO	0,741	81	0,744	80	0,7441	-0,0001	0,7433	0,0007	0,7453	-0,0013	0,7471	0,7477	0,7476
	GRD	0,743	79	0,744	81	0,7416	0,0024	0,7444	-0,0004	0,7458	-0,0018	0,7435	0,7395	0,7378
	PER	0,734	82	0,737	82	0,7402	-0,0032	0,7385	-0,0015	0,7404	-0,0034	0,7421	0,7413	0,7436
	UKR	0,733	83	0,734	83	0,7349	-0,0009	0,7348	-0,0008	0,7362	-0,0022	0,7348	0,7373	0,7391
	BLZ	0,731	84	0,732	84	0,7409	-0,0089	0,7371	-0,0051	0,7402	-0,0082	0,7378	0,7404	0,7369
	MKD	0,73	85	0,732	85	0,7315	0,0005	0,7365	-0,0045	0,7303	0,0017	0,7327	0,7328	0,7358
BIH	0,729	86	0,731	86	0,7299	0,0011	0,7315	-0,0005	0,7314	-0,0004	0,732	0,731	0,7322	
ARM	0,728	87	0,73	87	0,7305	-0,0005	0,7237	0,0063	0,7374	-0,0074	0,7325	0,7364	0,7415	
FJI	0,722	88	0,724	88	0,7245	-0,0005	0,7278	-0,0038	0,7329	-0,0089	0,7267	0,7272	0,7286	
THA	0,72	89	0,722	89	0,724	-0,002	0,727	-0,005	0,7393	-0,0173	0,7273	0,7246	0,7308	
TUN	0,719	90	0,721	90	0,7236	-0,0026	0,7236	-0,0026	0,7245	-0,0035	0,7257	0,7224	0,7276	
CHN	0,715	93	0,719	91	0,7235	-0,0045	0,7299	-0,0109	0,7359	-0,0169	0,7268	0,7358	0,7402	
VCT	0,717	91	0,719	92	0,7173	0,0017	0,7168	0,0022	0,7179	0,0011	0,7199	0,7194	0,7181	
DZA	0,715	94	0,717	93	0,719	-0,002	0,7304	-0,0134	0,7347	-0,0177	0,7207	0,7251	0,7379	

Classe	País (código internacional)	2012		2013		2013						2014		
		IDH	Posição	IDH	Posição	Modelos								
						Multivariado				Univariado				
						Global		Específico		Global		Específico		Univariado
Nível	IDH	Posição	IDH	Posição	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Global	Específico	Univariado
	DMA	0,716	92	0,717	94	0,7154	0,0016	0,7186	-0,0016	0,7222	-0,0052	0,7178	0,7165	0,7194
	ALB	0,714	97	0,716	95	0,7166	-0,0006	0,7153	0,0007	0,7223	-0,0063	0,7185	0,7194	0,7233
	JAM	0,715	95	0,715	96	0,717	-0,002	0,7239	-0,0089	0,7236	-0,0086	0,7168	0,7253	0,7226
	LCA	0,715	96	0,714	97	0,7133	0,0007	0,7146	-0,0006	0,7151	-0,0011	0,7127	0,7112	0,7112
	COL	0,708	98	0,711	98	0,7098	0,0012	0,7157	-0,0047	0,7174	-0,0064	0,714	0,7132	0,7188
	ECU	0,708	99	0,711	99	0,712	-0,001	0,7107	0,0003	0,7125	-0,0015	0,7144	0,7144	0,7161
	SUR	0,702	101	0,705	100	0,7032	0,0018	0,703	0,002	0,7044	0,0006	0,7075	0,8712	0,7088
	TON	0,704	100	0,705	101	0,7064	-0,0014	0,7061	-0,0011	0,7058	-0,0008	0,7077	0,7062	0,706
	DOM	0,698	102	0,7	102	0,7026	-0,0026	0,7017	-0,0017	0,7041	-0,0041	0,7043	0,7036	0,7051
	MDV	0,695	103	0,698	103	0,6982	-0,0002	0,6983	-0,0003	0,6956	0,0024	0,7026	0,7016	0,7067
	MNG	0,692	106	0,698	104	0,7025	-0,0045	0,7021	-0,0041	0,7037	-0,0057	0,7058	0,7088	0,7134
	TKM	0,693	104	0,698	105	0,6956	0,0024	0,6955	0,0025	0,6961	0,0019	0,7012	0,7024	0,7032
	WSM	0,693	105	0,694	106	0,6955	-0,0015	0,6948	-0,0008	0,6948	-0,0008	0,6963	0,6957	0,6973
	PLW	0,683	107	0,686	107	0,6874	-0,0014	0,6863	-0,0003	0,6855	0,0005	0,6893	0,691	0,6934
	IDN	0,681	108	0,684	108	0,687	-0,003	0,6893	-0,0053	0,6905	-0,0065	0,6897	0,69	0,6898
	BWA	0,681	109	0,683	109	0,6862	-0,0032	0,682	0,001	0,6865	-0,0035	0,6883	0,6861	0,697
	EGY	0,681	110	0,682	110	0,6848	-0,0028	0,6881	-0,0061	0,6844	-0,0024	0,6863	0,6839	0,6871
	PRY	0,67	111	0,676	111	0,6706	0,0054	0,6703	0,0057	0,6797	-0,0037	0,68	0,6824	0,6824
	GAB	0,67	112	0,674	112	0,6746	-0,0006	0,6735	0,0005	0,6763	-0,0023	0,6791	0,6778	0,6794
	BOL	0,663	113	0,667	113	0,6675	-0,0005	0,6658	0,0012	0,6653	0,0017	0,672	0,6691	0,6697
	MDA	0,657	116	0,663	114	0,6582	0,0048	0,6594	0,0036	0,6595	0,0035	0,6653	0,6638	0,668
	SLV	0,66	115	0,662	115	0,6646	-0,0026	0,6623	-0,0003	0,6629	-0,0009	0,6656	0,666	0,6637
	UZB	0,657	117	0,661	116	0,6608	0,0002	0,6632	-0,0022	0,6617	-0,0007	0,6643	0,6648	0,6656
	PHL	0,656	118	0,66	117	0,659	0,001	0,6607	-0,0007	0,6612	-0,0012	0,6642	0,6631	0,6635
	ZAF	0,654	119	0,658	118	0,6619	-0,0039	0,6608	-0,0028	0,6625	-0,0045	0,6642	0,6602	0,6633
	SYR	0,662	114	0,658	119	0,6634	-0,0054	0,6674	-0,0094	0,6739	-0,0159	0,6578	0,6699	0,6692
	IRQ	0,641	120	0,642	120	0,6458	-0,0038	0,6447	-0,0027	0,6424	-0,0004	0,645	0,6436	0,6427
	GUY	0,635	121	0,638	121	0,637	0,001	0,6388	-0,0008	0,6392	-0,0012	0,6422	0,6431	0,6444
	VNM	0,635	122	0,638	122	0,6389	-0,0009	0,6408	-0,0028	0,6415	-0,0035	0,6422	0,6442	0,6416
	CPV	0,635	123	0,636	123	0,6408	-0,0048	0,6384	-0,0024	0,6418	-0,0058	0,6383	0,639	0,6469
	FSM	0,629	124	0,63	124	0,6299	1E-04	0,6295	0,0005	0,6307	-0,0007	0,6309	0,6301	0,6307
	GTM	0,626	125	0,628	125	0,6327	-0,0047	0,6332	-0,0052	0,6322	-0,0042	0,6335	0,6392	0,6374
	KGZ	0,621	126	0,628	126	0,6236	0,0044	0,6237	0,0043	0,6237	0,0043	0,6319	0,6363	0,6305
	NAM	0,62	127	0,624	127	0,6244	-0,0004	0,6236	0,0004	0,6259	-0,0019	0,6278	0,6242	0,6287
	TLS	0,616	129	0,62	128	0,6223	-0,0023	0,6236	-0,0036	0,6404	-0,0204	0,6255	0,632	0,6335
	HND	0,616	130	0,617	129	0,6189	-0,0019	0,6185	-0,0015	0,6273	-0,0103	0,6197	0,6223	0,6313
	MAR	0,614	131	0,617	130	0,6197	-0,0027	0,632	-0,015	0,6311	-0,0141	0,6219	0,6289	0,6356
	VUT	0,617	128	0,616	131	0,618	-0,002	0,6183	-0,0023	0,6139	0,0021	0,6165	0,6164	0,6162
	NIC	0,611	132	0,614	132	0,6148	-0,0008	0,6153	-0,0013	0,6139	1E-04	0,6177	0,6187	0,6164
	KIR	0,606	133	0,607	133	0,6091	-0,0021	0,6077	-0,0007	0,6097	-0,0027	0,6092	0,606	0,606
	TJK	0,603	134	0,607	134	0,6051	0,0019	0,6145	-0,0075	0,6054	0,0016	0,6088	0,6185	0,6104
	IND	0,583	135	0,586	135	0,5893	-0,0033	0,5858	0,0002	0,5998	-0,0138	0,5911	0,591	0,6076
	BTN	0,58	136	0,584	136	0,5847	-0,0007	0,5843	-0,0003	0,3785	0,2055	0,5882	0,5888	0,5873
	KHM	0,579	137	0,584	137	0,5852	-0,0012	0,5916	-0,0076	0,5921	-0,0081	0,5915	0,6022	0,6124
	GHA	0,571	138	0,573	138	0,5785	-0,0055	0,5759	-0,0029	0,5796	-0,0066	0,5783	0,5748	0,5822
	LAO	0,565	139	0,569	139	0,5731	-0,0041	0,5738	-0,0048	0,5784	-0,0094	0,5753	0,5758	0,5797
	COG	0,561	140	0,564	140	0,5634	0,0006	0,5552	0,0088	0,5669	-0,0029	0,5669	0,5713	0,5705
	ZMB	0,554	143	0,561	141	0,5657	-0,0047	0,5679	-0,0069	0,5702	-0,0092	0,5672	0,5697	0,5786

Classe	País (código internacional)	2012		2013		2013						2014			
		IDH	Posição	IDH	Posição	Modelos									
Nível	IDH	IDH	Posição	IDH	Posição	Multivariado				Univariado		Multivariado		Univariado	
						Global		Específico		IDH(P)	Erro *	Global	Específico		
		IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *	IDH(P)	Erro *		
	BGD	0,554	144	0,558	142	0,562	-0,004	0,5635	-0,0055	0,5683	-0,0103	0,5649	0,5715	0,5723	
	STP	0,556	141	0,558	143	0,5616	-0,0036	0,5647	-0,0067	0,5639	-0,0059	0,562	0,5613	0,5649	
	GNQ	0,556	142	0,556	144	0,5566	-0,0006	0,5597	-0,0037	0,564	-0,008	0,5571	0,5603	0,5654	
Subdesenvolvido	Baixo	NPL	0,537	145	0,54	145	0,5443	-0,0043	0,5483	-0,0083	0,5492	-0,0092	0,5469	0,5405	0,5506
		OMN	0,535	146	0,537	146	0,5422	-0,0052	0,5365	0,0005	0,547	-0,01	0,5412	0,5507	0,5521
		KEN	0,531	147	0,535	147	0,5347	0,0003	0,5375	-0,0025	0,5351	-0,0001	0,5393	0,5383	0,5392
		SWZ	0,529	148	0,53	148	0,5297	0,0003	0,5306	-0,0006	0,5331	-0,0031	0,5315	0,5331	0,5344
		AGO	0,524	149	0,526	149	0,5319	-0,0059	0,5273	-0,0013	0,5557	-0,0297	0,5299	0,5292	0,5492
		MMR	0,52	150	0,524	150	0,5241	-0,0001	0,528	-0,004	0,5342	-0,0102	0,5287	0,5289	0,5377
		RWA	0,502	151	0,506	151	0,527	-0,021	0,5476	-0,0416	0,5116	-0,0056	0,5171	0,5293	0,5225
		CMR	0,501	152	0,504	152	0,5049	-0,0009	0,5069	-0,0029	0,5177	-0,0137	0,5082	0,5083	0,5156
		NGA	0,5	153	0,504	153	0,5035	0,0005	0,5037	0,0003	0,5046	-0,0006	0,5078	0,5074	0,5084
		YEM	0,499	154	0,5	154	0,5046	-0,0046	0,502	-0,002	0,5114	-0,0114	0,5025	0,5005	0,5115
		MDG	0,496	155	0,498	155	0,4978	0,0002	0,4994	-0,0014	0,4999	-0,0019	0,4995	0,5002	0,5013
		ZWE	0,484	160	0,492	156	0,493	-0,001	0,4937	-0,0017	0,5	-0,008	0,5006	0,495	0,5054
		PNG	0,49	156	0,491	157	0,4977	-0,0067	0,4965	-0,0055	0,4944	-0,0034	0,4947	0,4994	0,5005
		SLB	0,489	157	0,491	158	0,4883	0,0027	0,4915	-0,0005	0,4947	-0,0037	0,4897	0,4947	0,4978
		COM	0,486	158	0,488	159	0,489	-0,001	0,4894	-0,0014	0,4896	-0,0016	0,4904	0,491	0,4912
		TZA	0,484	161	0,488	160	0,4915	-0,0035	0,4938	-0,0058	0,5009	-0,0129	0,4916	0,506	0,503
		MRT	0,485	159	0,487	161	0,4926	-0,0056	0,4886	-0,0016	0,4887	-0,0017	0,4929	0,4936	0,4903
		LSO	0,481	163	0,486	162	0,485	0,001	0,4857	0,0003	0,4855	0,0005	0,4907	0,4923	0,4909
		SEN	0,484	162	0,485	163	0,4857	-0,0007	0,4876	-0,0026	0,497	-0,012	0,4878	0,4893	0,5029
		UGA	0,48	164	0,484	164	0,4849	-0,0009	0,489	-0,005	0,4898	-0,0058	0,4883	0,4885	0,4939
		BEN	0,473	165	0,476	165	0,4772	-0,0012	0,4846	-0,0086	0,4916	-0,0156	0,4805	0,4855	0,4982
		SDN	0,472	166	0,473	166	0,4769	-0,0039	0,4797	-0,0067	0,494	-0,021	0,4764	0,4765	0,4947
		TGO	0,47	167	0,473	167	0,4749	-0,0019	0,4764	-0,0034	0,4724	0,0006	0,4765	0,4773	0,4762
		HTI	0,469	168	0,471	168	0,4732	-0,0022	0,4728	-0,0018	0,4743	-0,0033	0,4744	0,4739	0,476
		AFG	0,466	169	0,468	169	0,474	-0,006	0,4698	-0,0018	0,4853	-0,0173	0,4741	0,4833	0,4902
		DJI	0,465	170	0,467	170	0,4704	-0,0034	0,4689	-0,0019	0,4716	-0,0046	0,4708	0,4715	0,4731
		CIV	0,448	171	0,452	171	0,4521	-0,0001	0,455	-0,003	0,4538	-0,0018	0,4561	0,4558	0,459
		GMB	0,438	172	0,441	172	0,4396	0,0014	0,4486	-0,0076	0,4582	-0,0172	0,4446	0,4478	0,4553
		ETH	0,429	173	0,435	173	0,4375	-0,0025	0,4415	-0,0065	0,4469	-0,0119	0,4415	0,443	0,4523
		MWI	0,411	174	0,414	174	0,4145	-0,0005	0,4219	-0,0079	0,4199	-0,0059	0,4171	0,4248	0,425
		LBR	0,407	175	0,412	175	0,4128	-0,0008	0,4116	0,0004	0,4152	-0,0032	0,4165	0,4182	0,4201
		MLI	0,406	176	0,407	176	0,4105	-0,0035	0,4156	-0,0086	0,4231	-0,0161	0,4105	0,4186	0,422
		GNB	0,396	177	0,396	177	0,3941	0,0019	0,4027	-0,0067	0,394	0,002	0,3951	0,3985	0,3968
		MOZ	0,389	179	0,393	178	0,3936	-0,0006	0,395	-0,002	0,4066	-0,0136	0,3978	0,3957	0,4013
GIN	0,391	178	0,392	179	0,3954	-0,0034	0,3939	-0,0019	0,3938	-0,0018	0,3947	0,3958	0,3965		
BDI	0,386	180	0,389	180	0,3899	-0,0009	0,3894	-0,0004	0,4035	-0,0145	0,3929	0,3967	0,4018		
BFA	0,385	181	0,388	181	0,3931	-0,0051	0,3935	-0,0055	0,3949	-0,0069	0,3937	0,3949	0,3956		
ERI	0,38	182	0,381	182	0,383	-0,002	0,3829	-0,0019	0,3832	-0,0022	0,3832	0,3826	0,3826		
SLE	0,368	184	0,374	183	0,3749	-0,0009	0,3766	-0,0026	0,3769	-0,0029	0,3791	0,3803	0,3825		
TCD	0,37	183	0,372	184	0,3778	-0,0058	0,3742	-0,0022	0,3784	-0,0064	0,3756	0,3742	0,3801		
CAF	0,365	185	0,341	185	0,3692	-0,0282	0,3685	-0,0275	0,3754	-0,0344	0,331	0,3335	0,3668		
COD	0,333	187	0,338	186	0,3386	-0,0006	0,3504	-0,0124	0,3414	-0,0034	0,3419	0,3359	0,3468		
NER	0,335	186	0,337	187	0,3413	-0,0043	0,3449	-0,0079	0,3425	-0,0055	0,3419	0,351	0,3482		

\* Diferença entre o valor real e o predito. P - previsões.



**APÊNDICE 4 – Correlogramas e Séries Temporais Observadas para a Modelagem dos Modelos ARIMA para Previsão do IDH 2013 dos Países da América Latina**

Figura 37 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 da Argentina (ARG) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

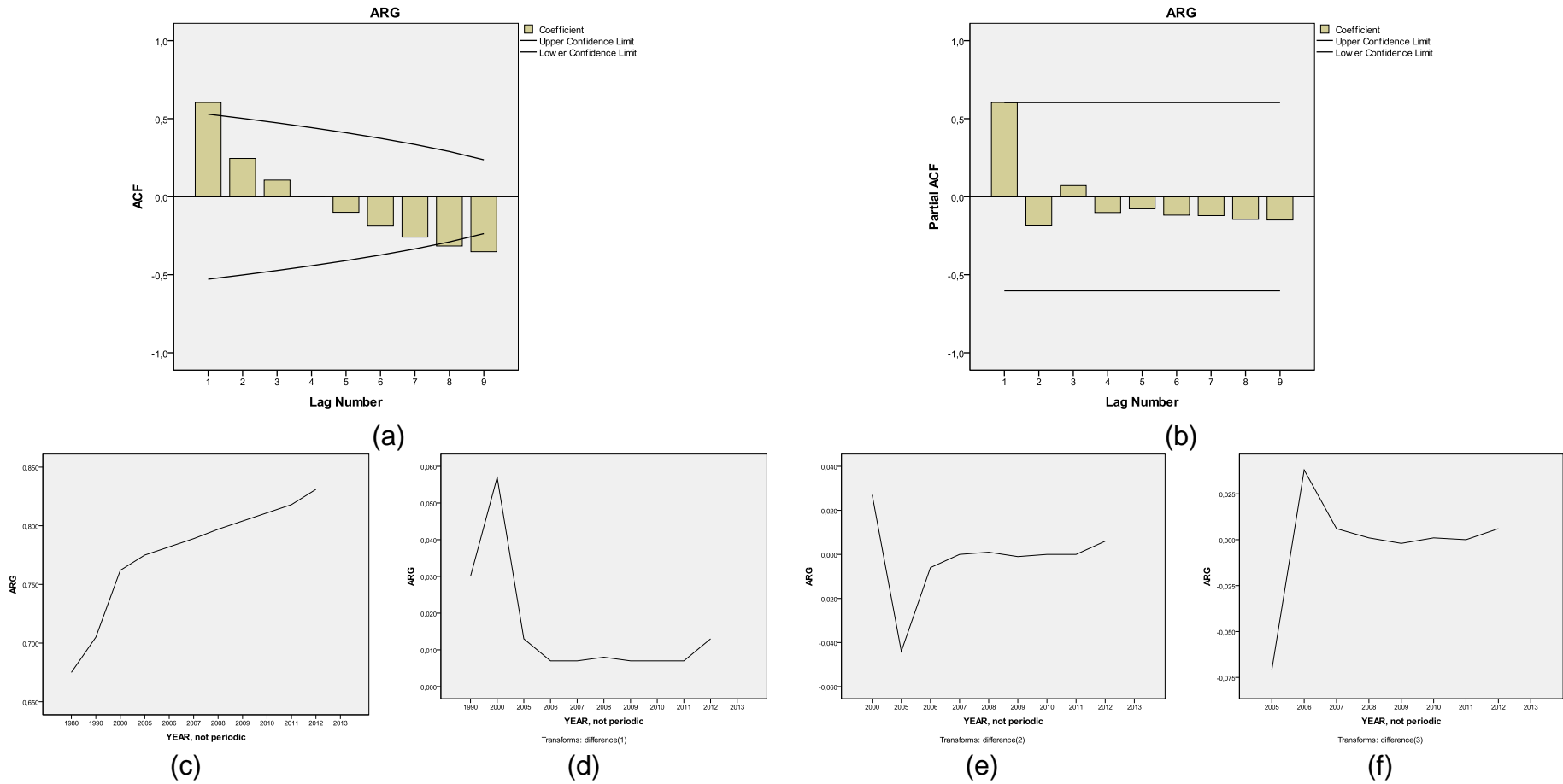


Figura 38 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Chile (CHL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

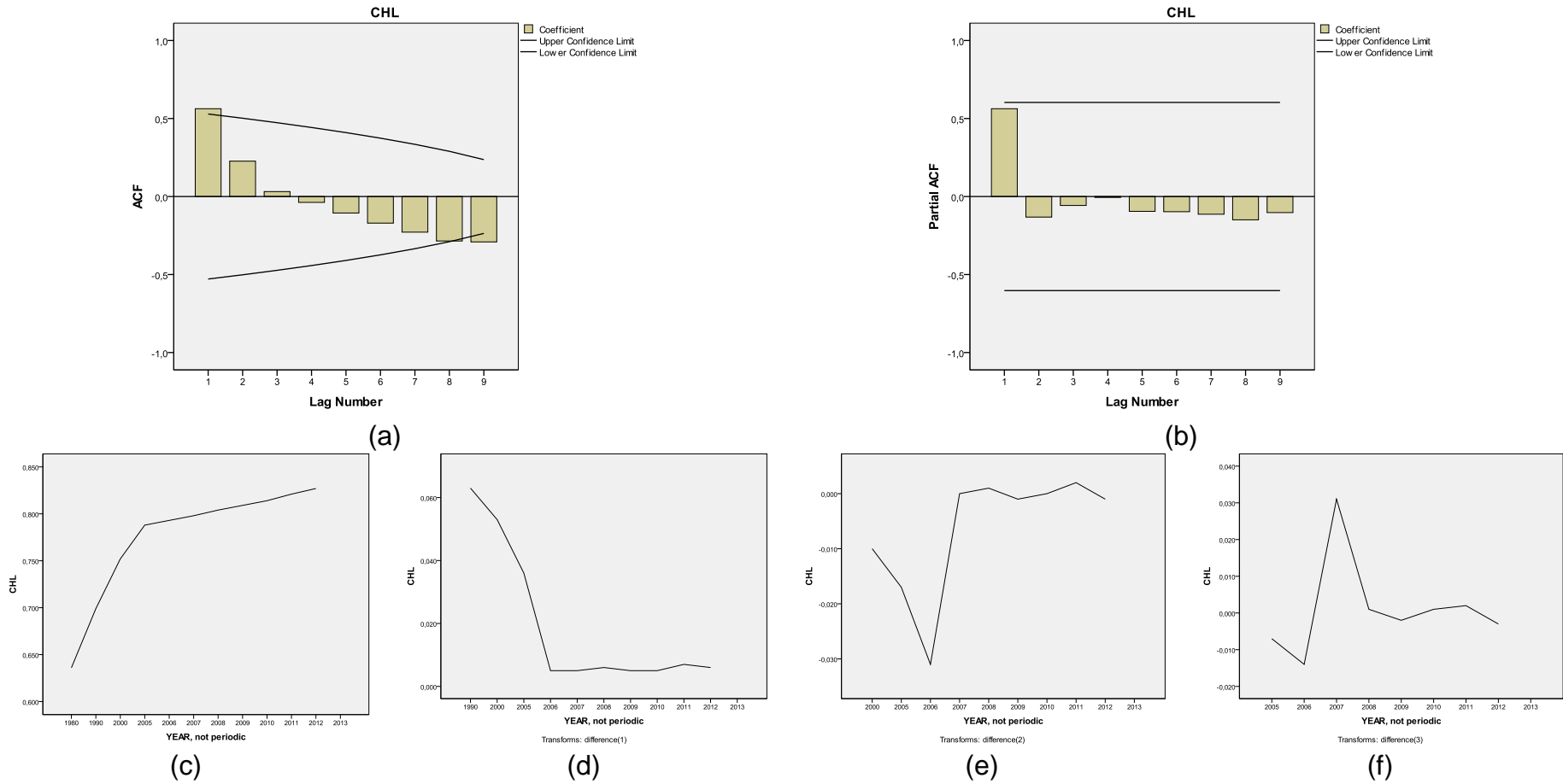


Figura 39 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Uruguai (URY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

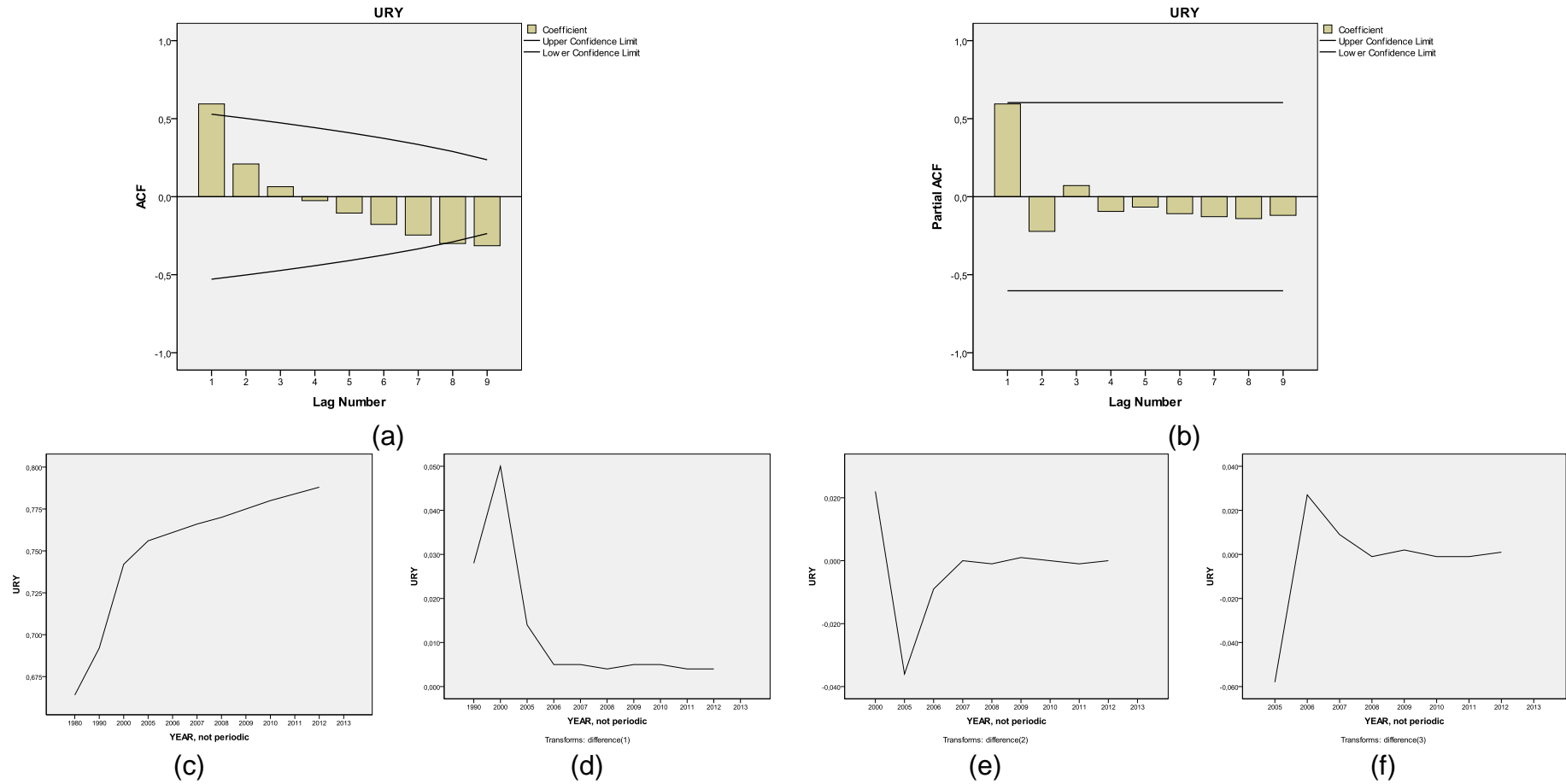


Figura 40 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Panamá (PAN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

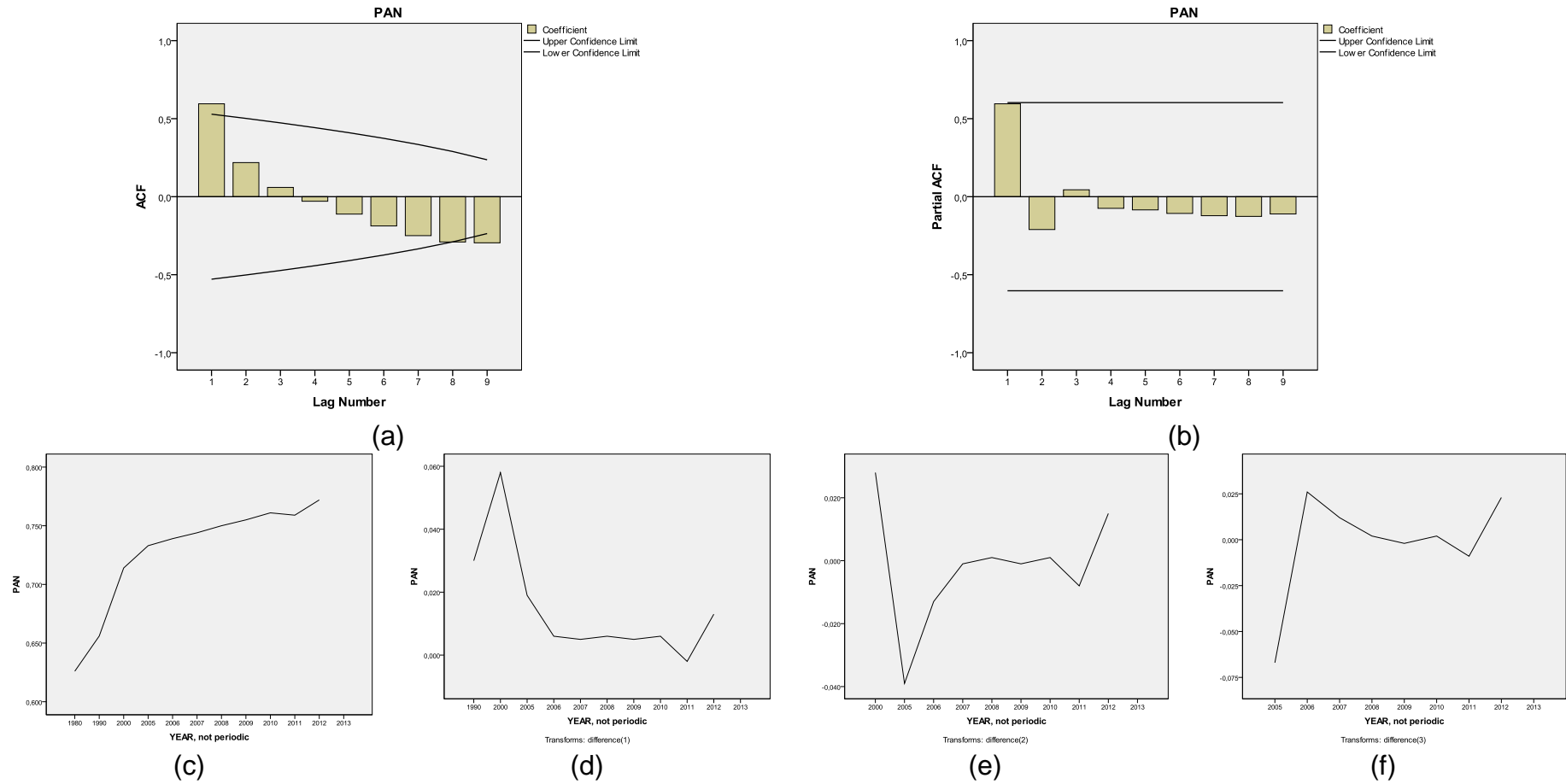


Figura 41 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Cuba (CUB) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

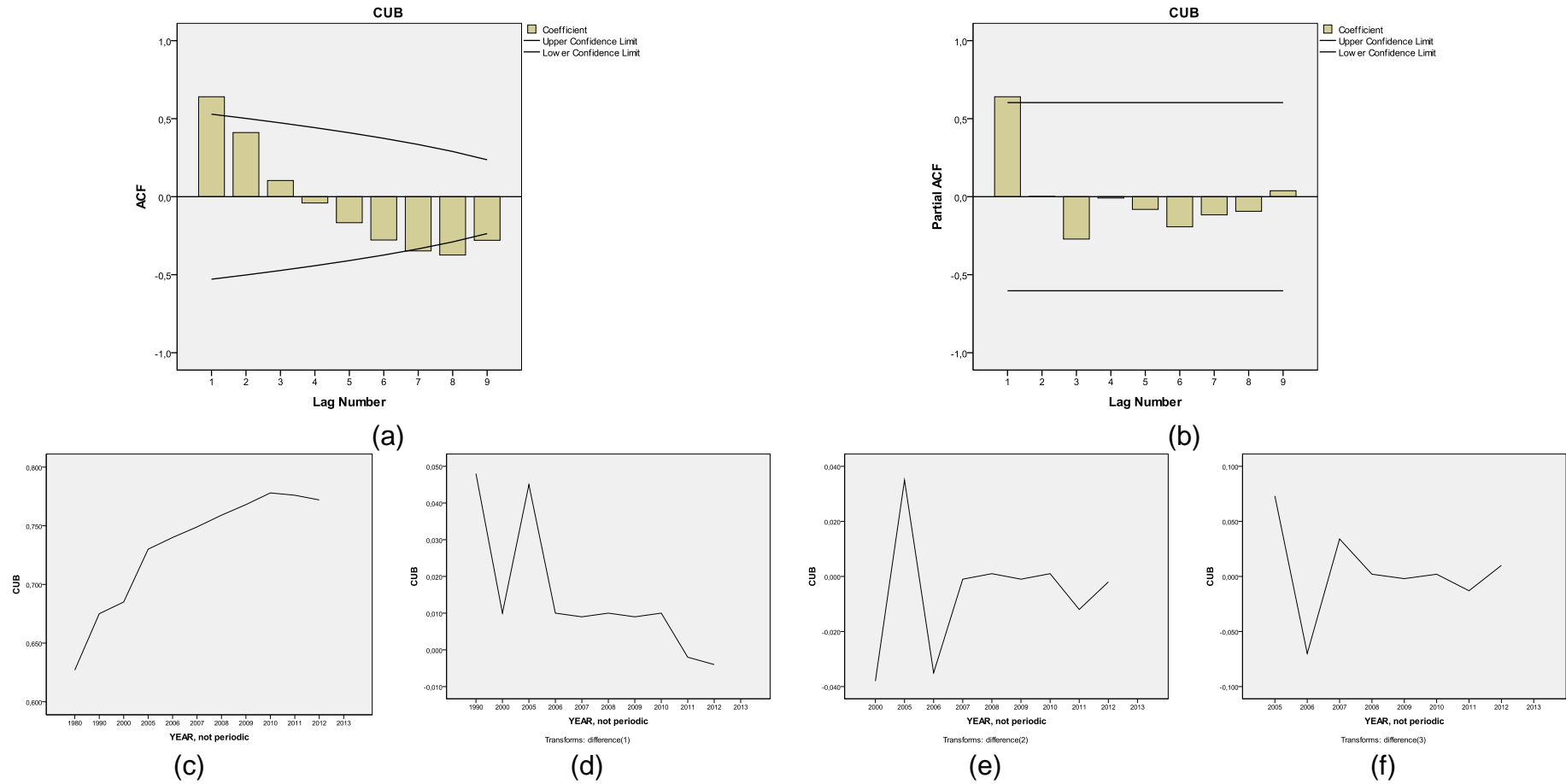
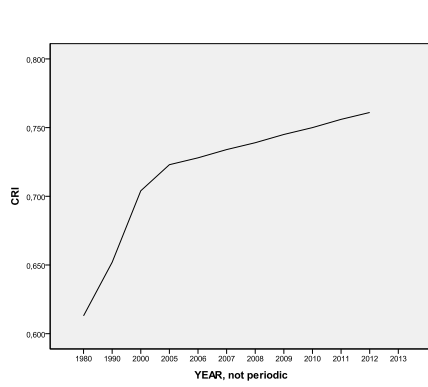
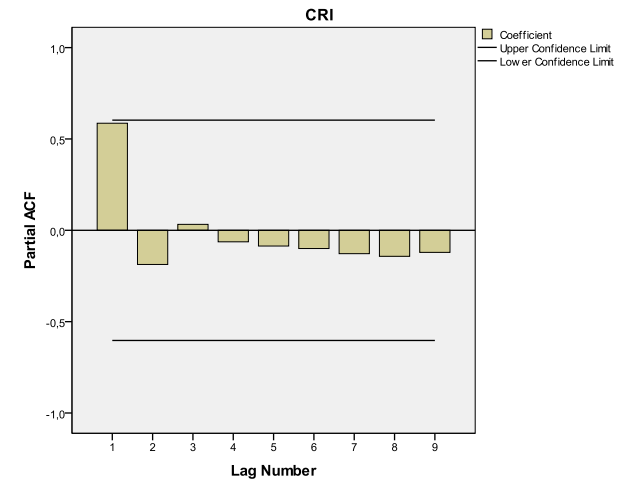
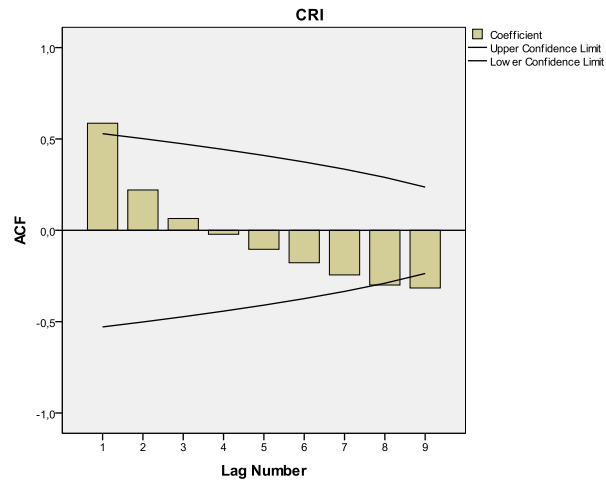
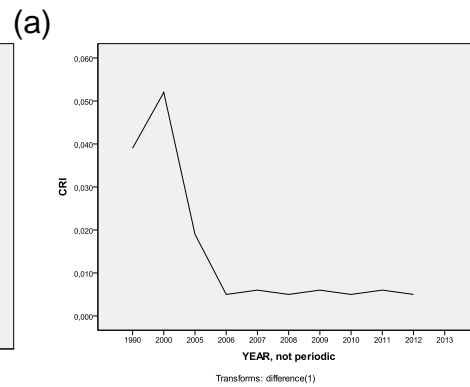


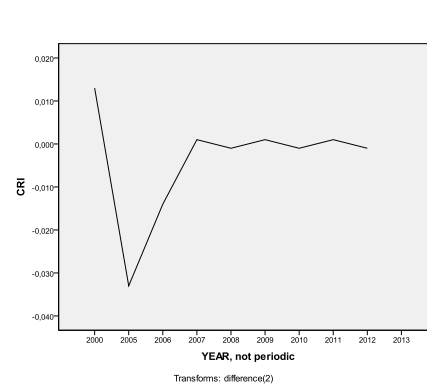
Figura 42 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Costa Rica (CRI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)



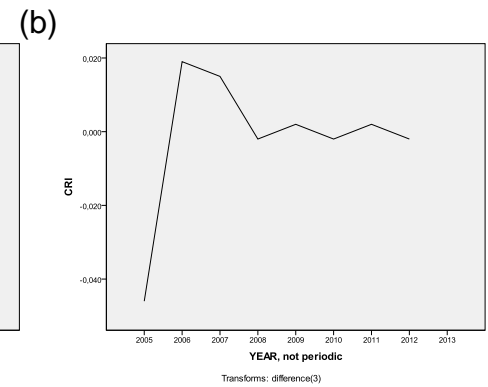
(c)



(d)



(e)



(f)

)

Figura 43 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Venezuela (VEN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

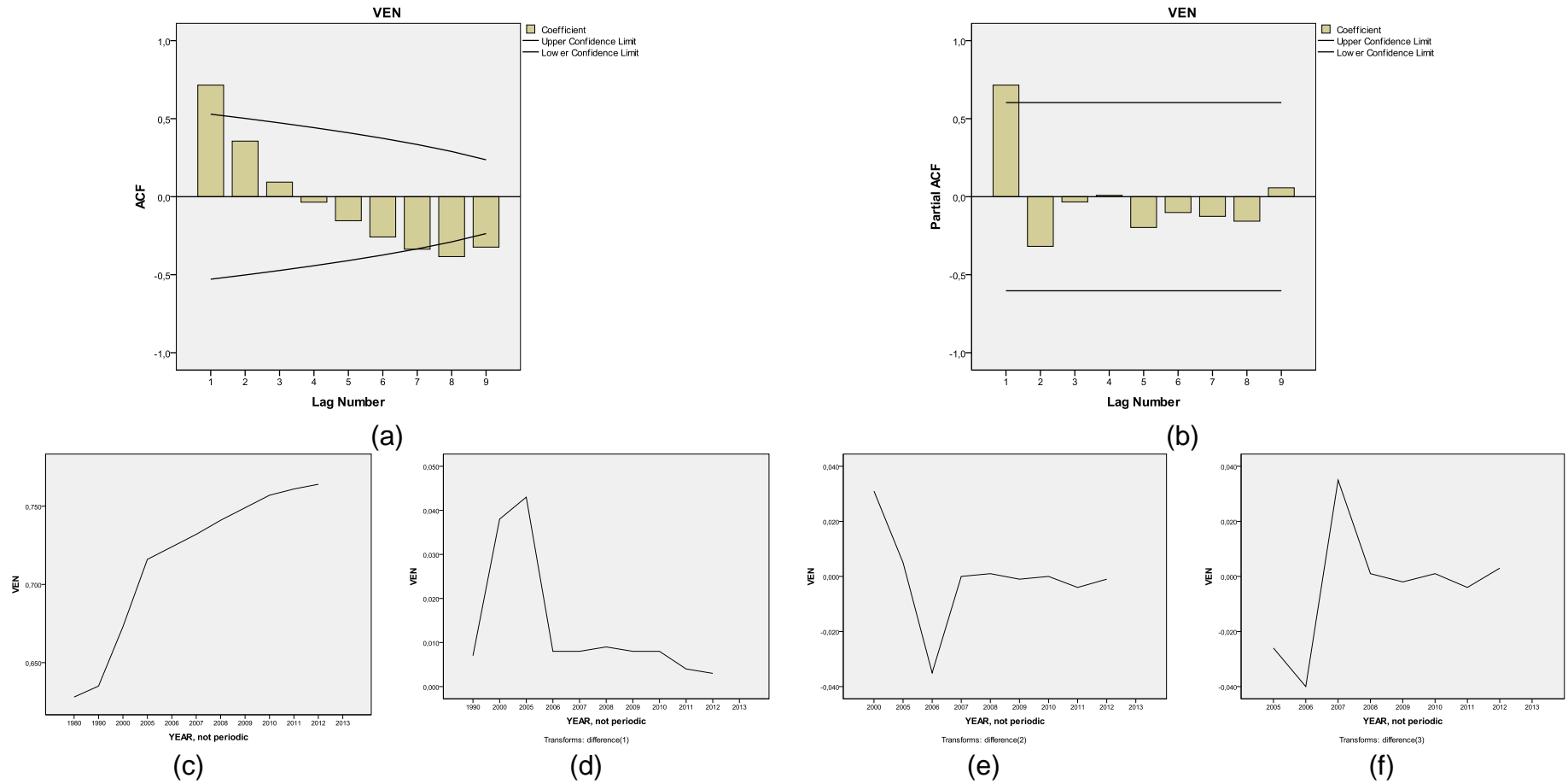




Figura 44 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do México (MEX) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

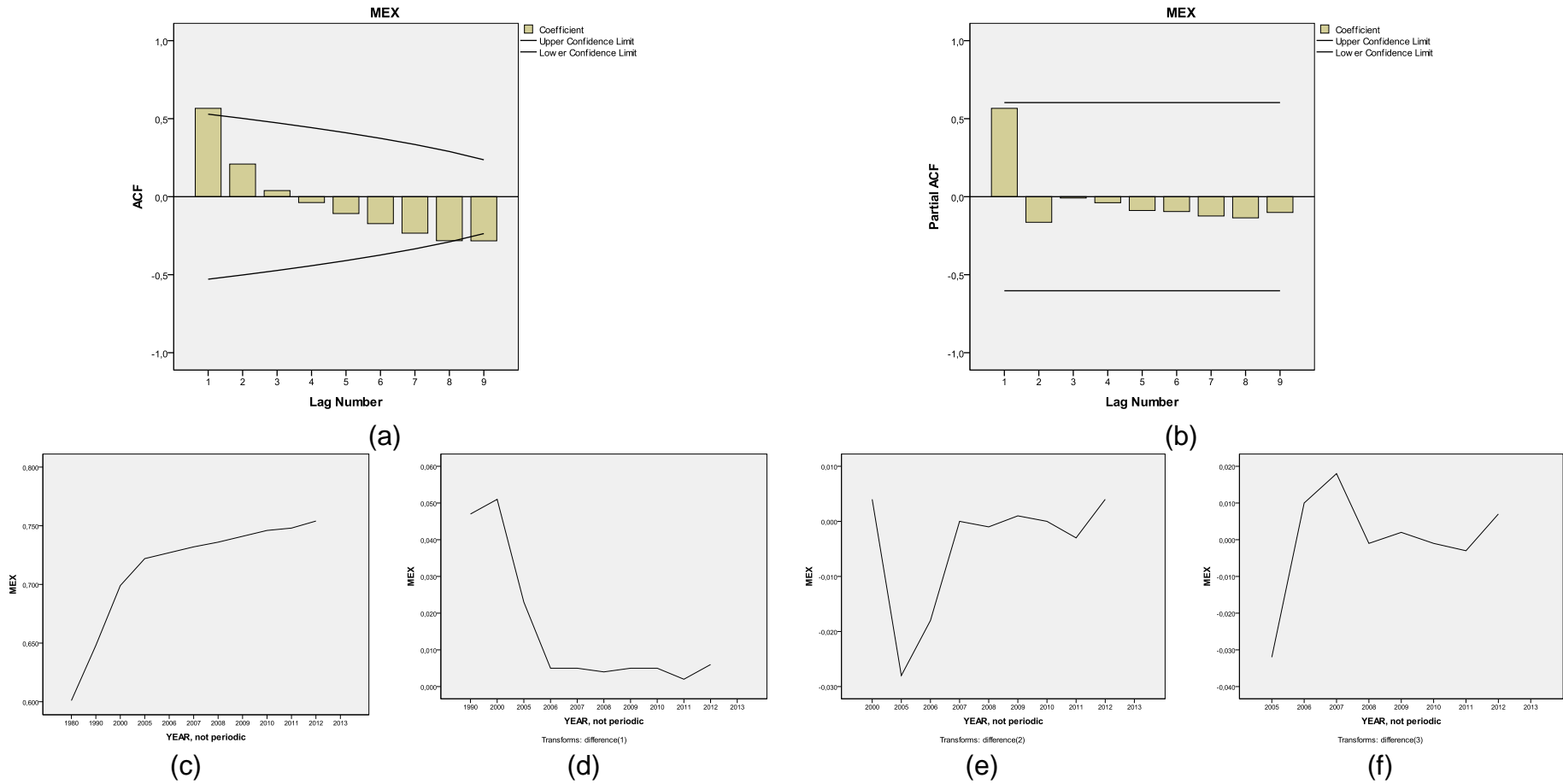


Figura 45 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Brasil (BRA) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

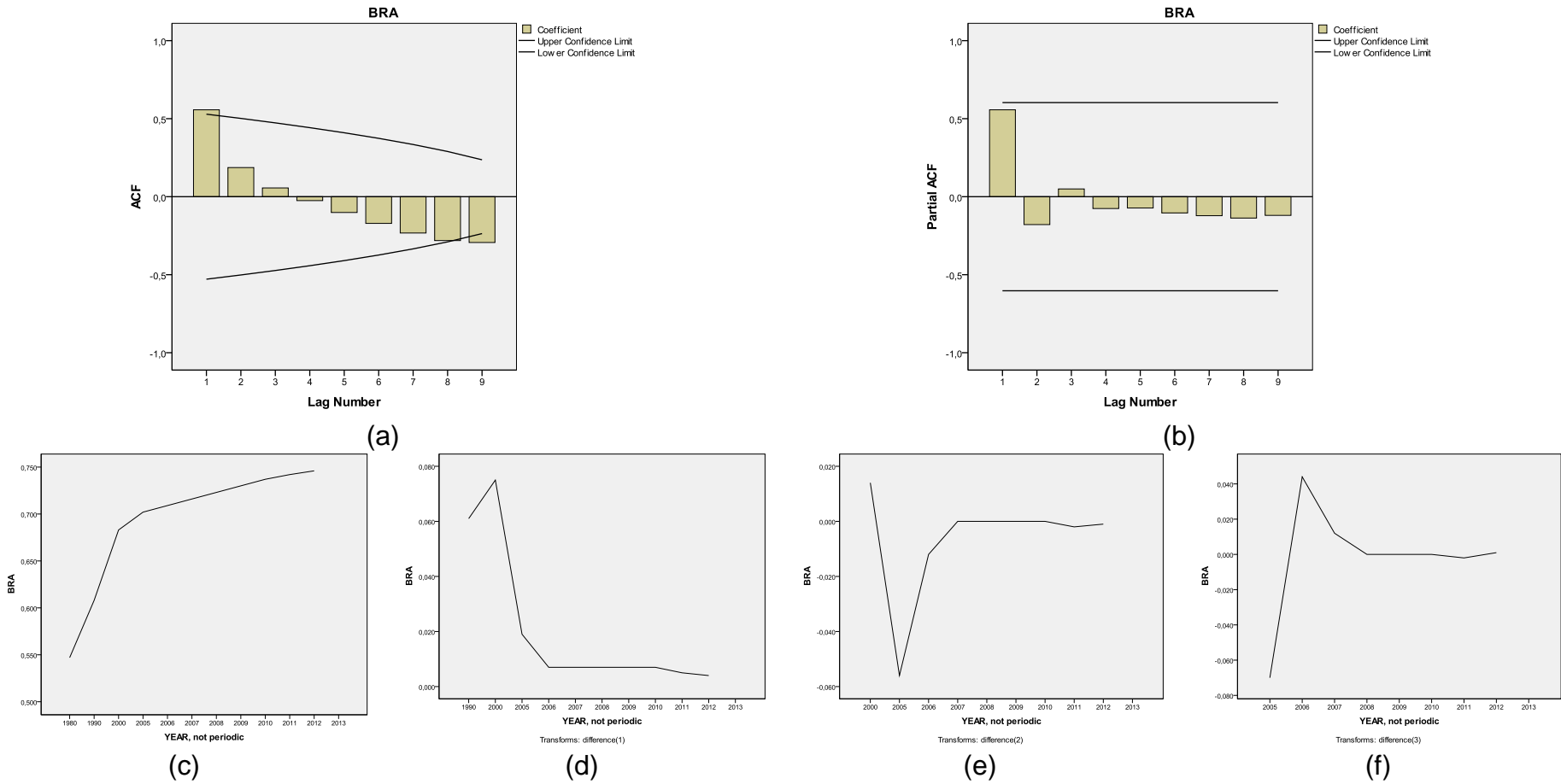


Figura 46 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Peru (PER) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

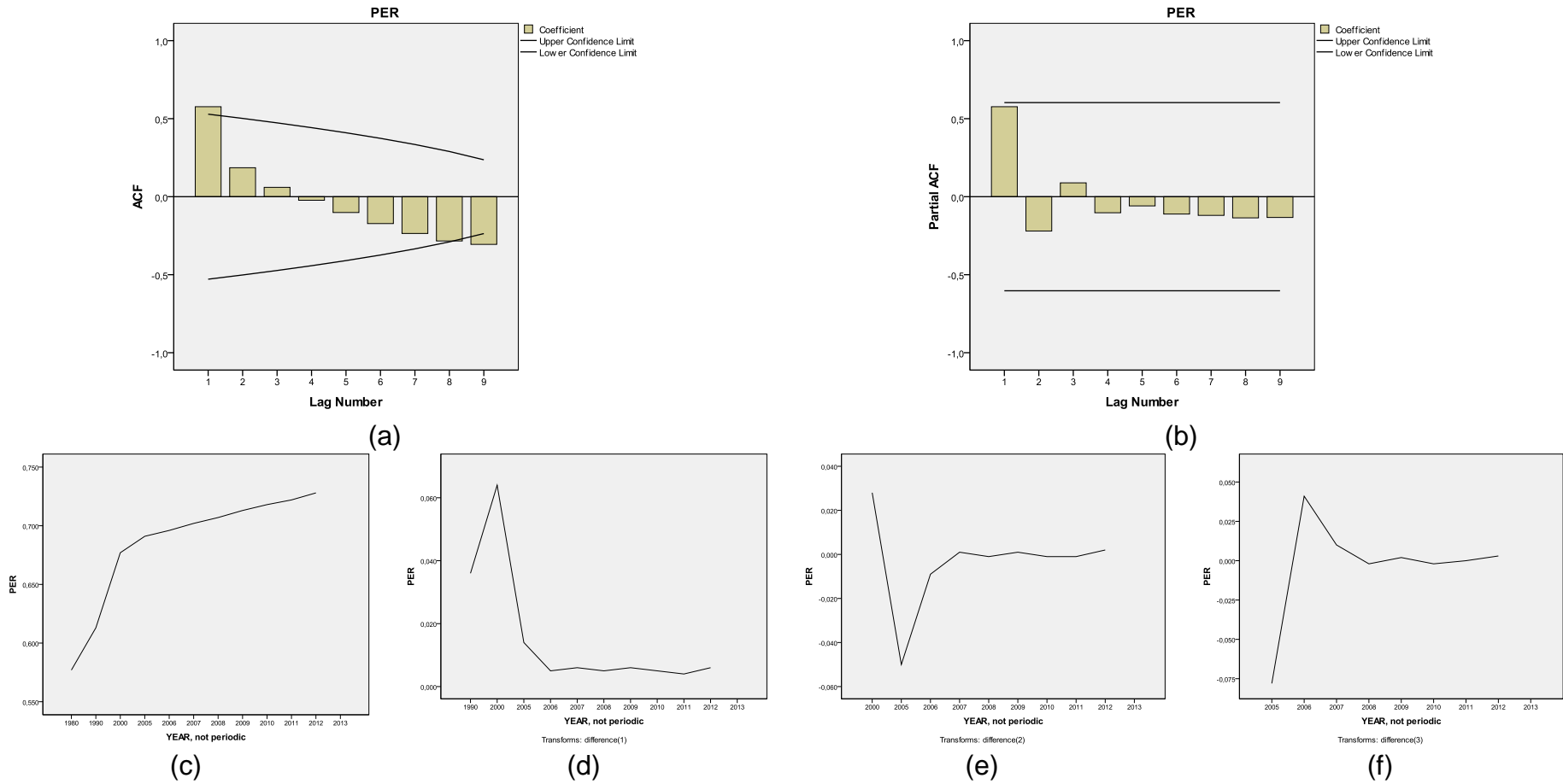


Figura 47 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Equador (ECU) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

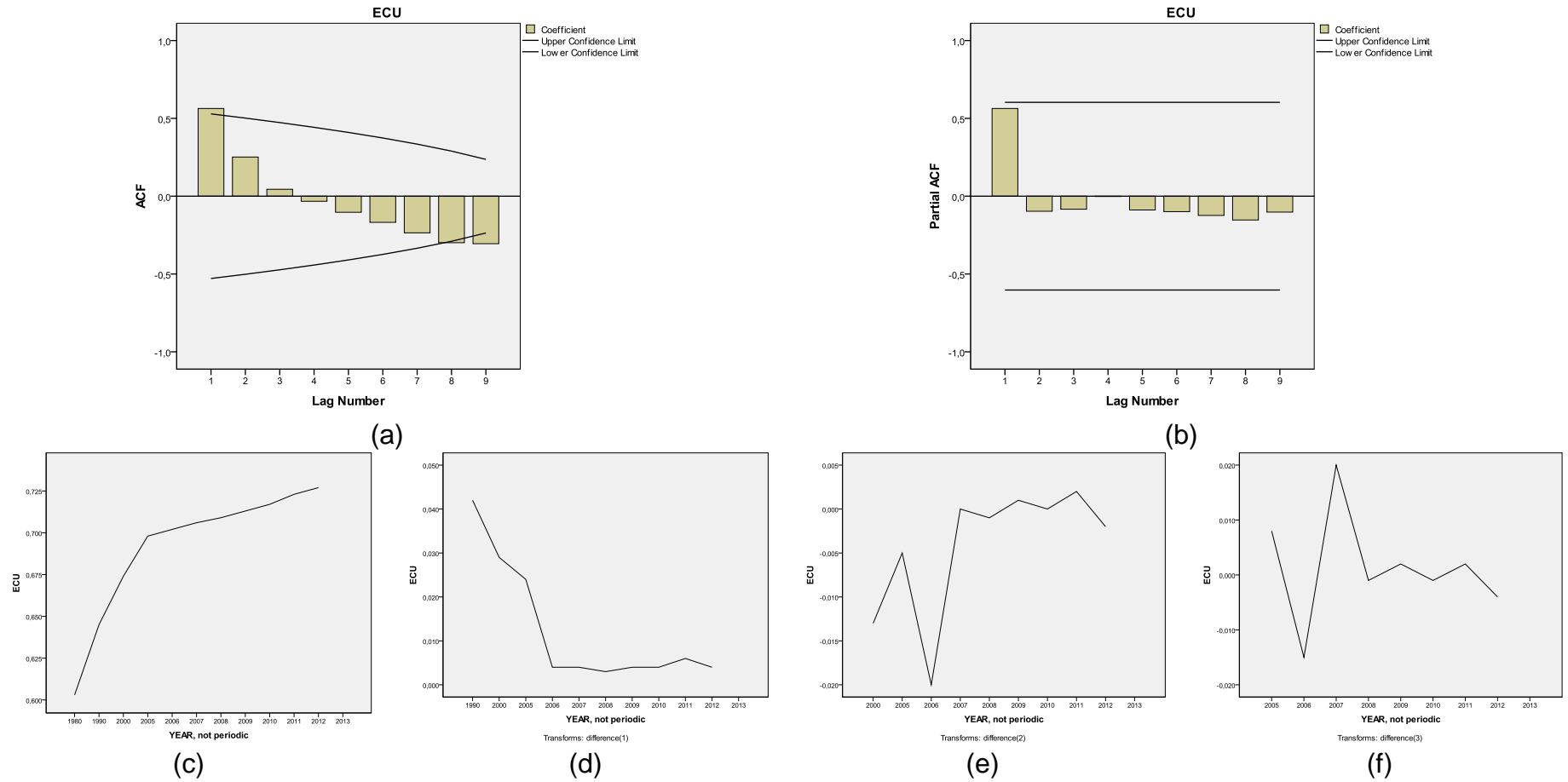


Figura 48 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Colômbia (COL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

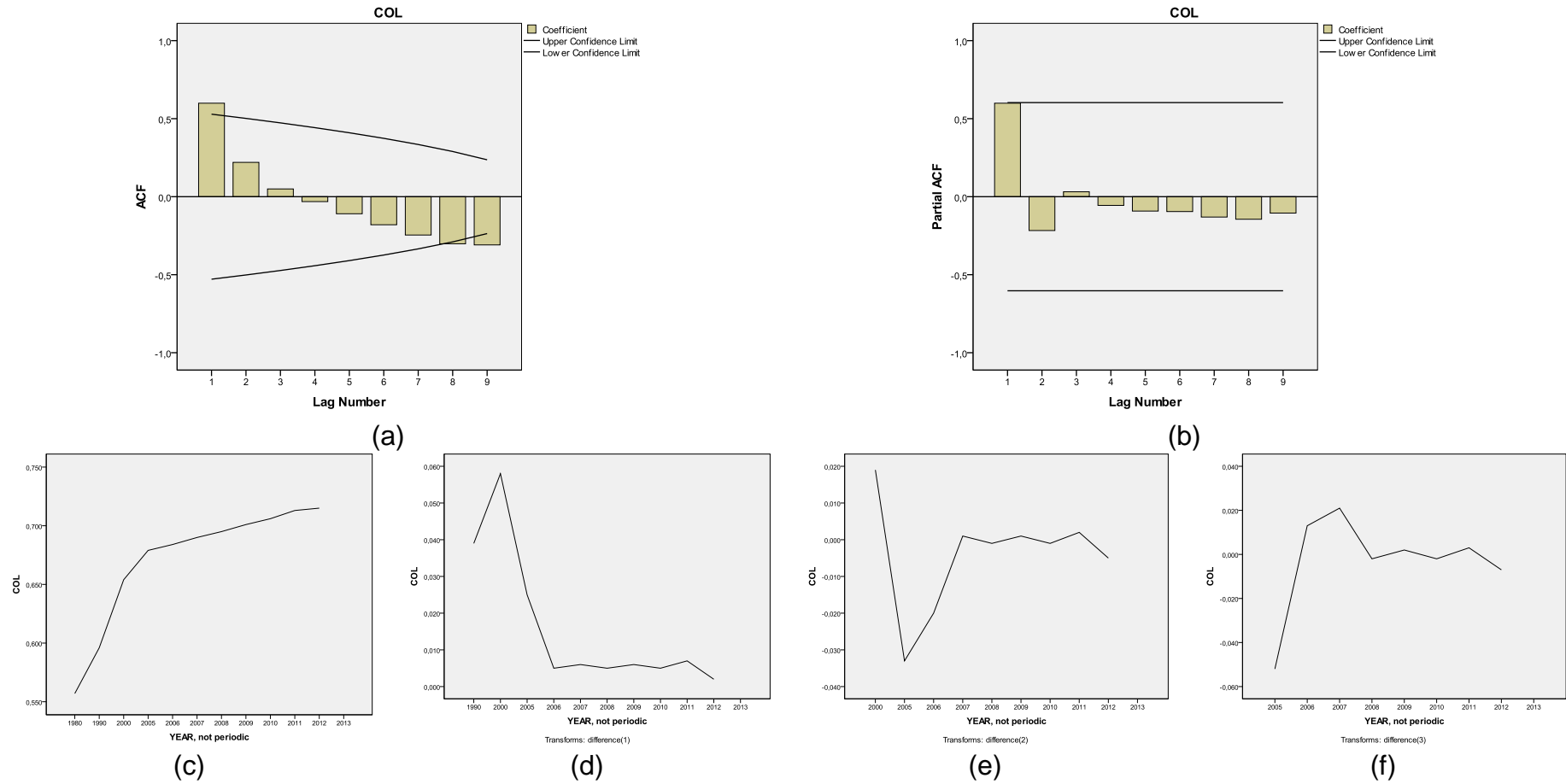


Figura 49 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Belize (BLZ) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

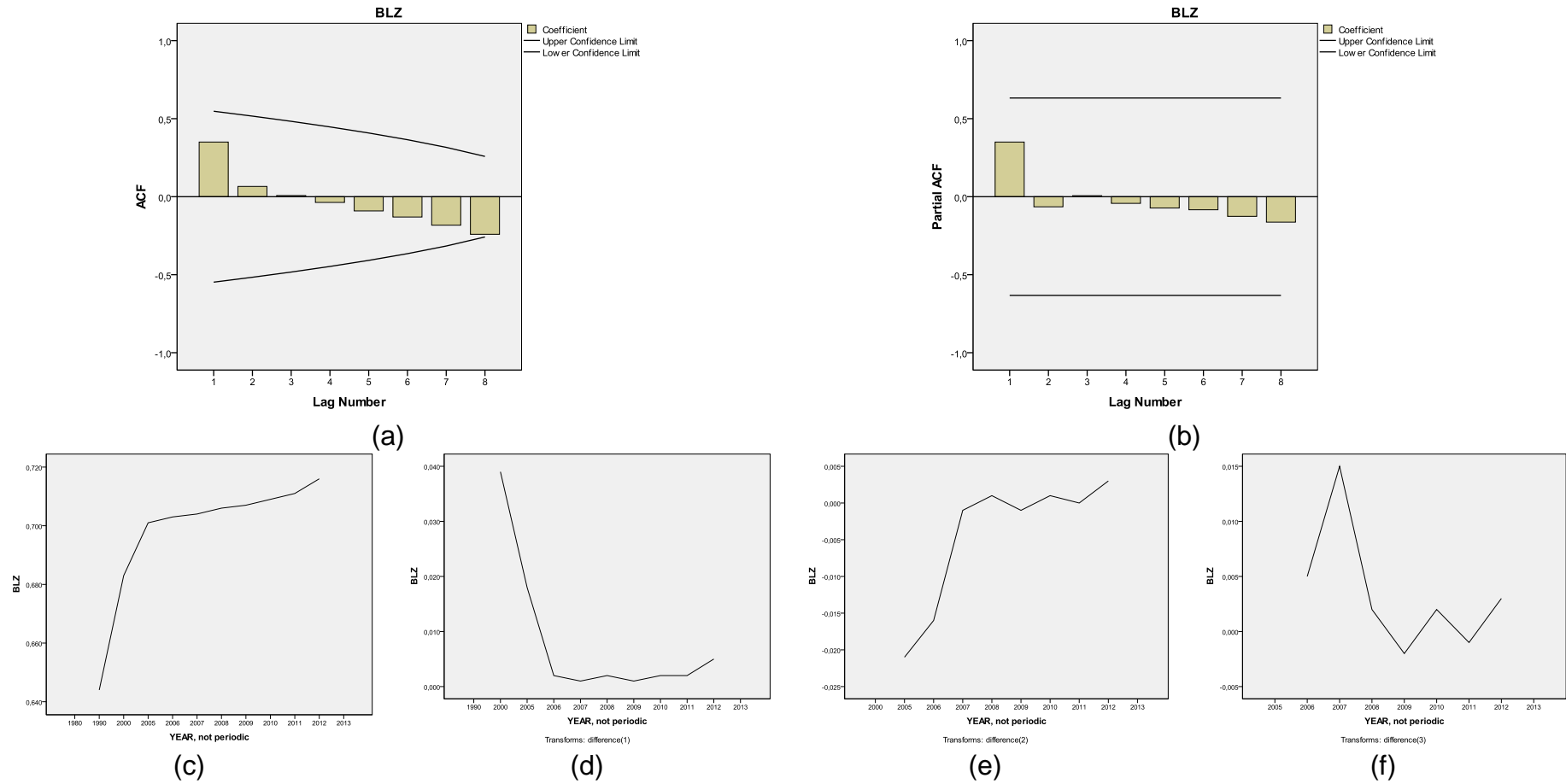


Figura 50 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do República Dominicana (DOM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

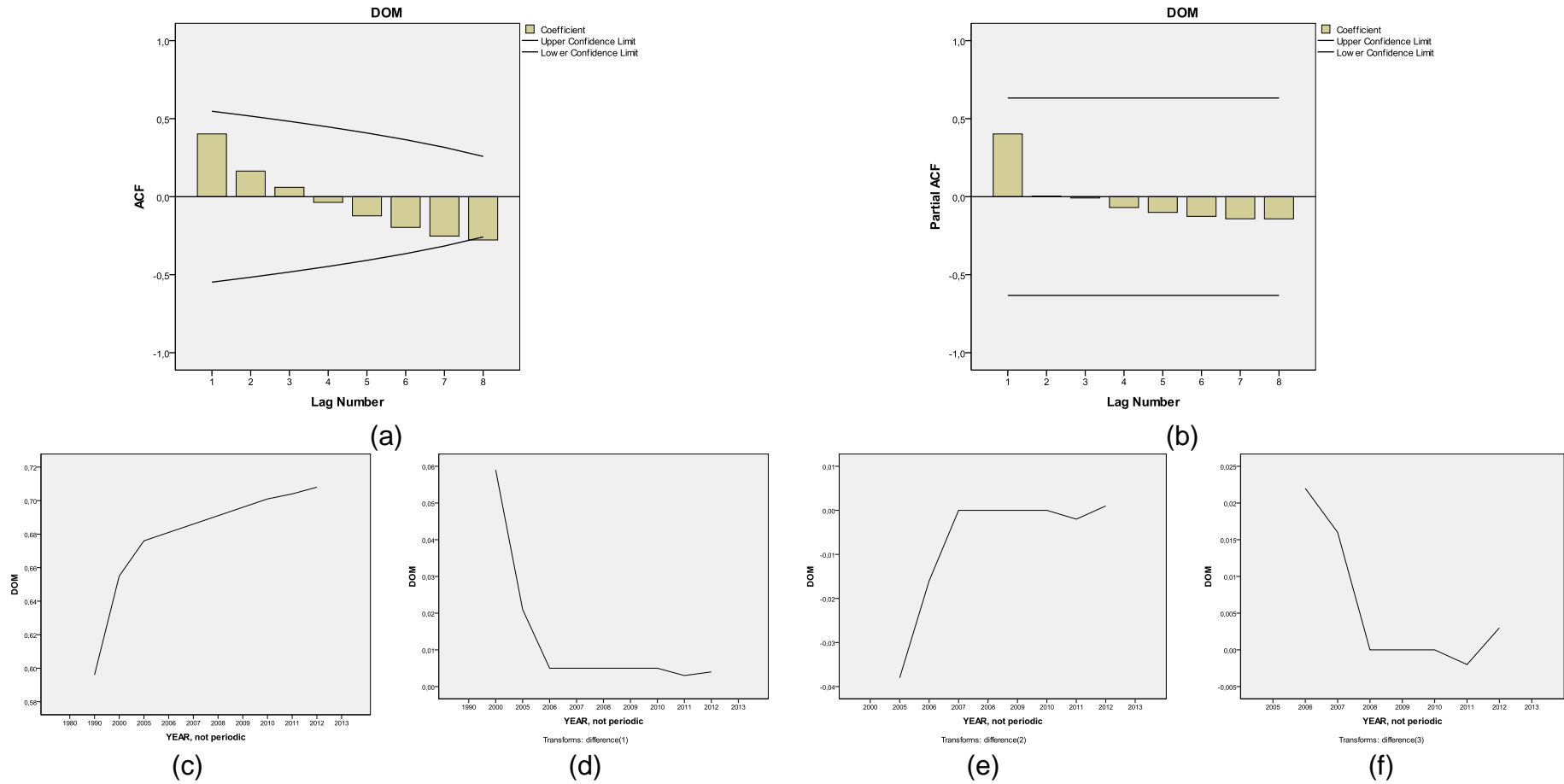


Figura 51 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Paraguai (PRY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

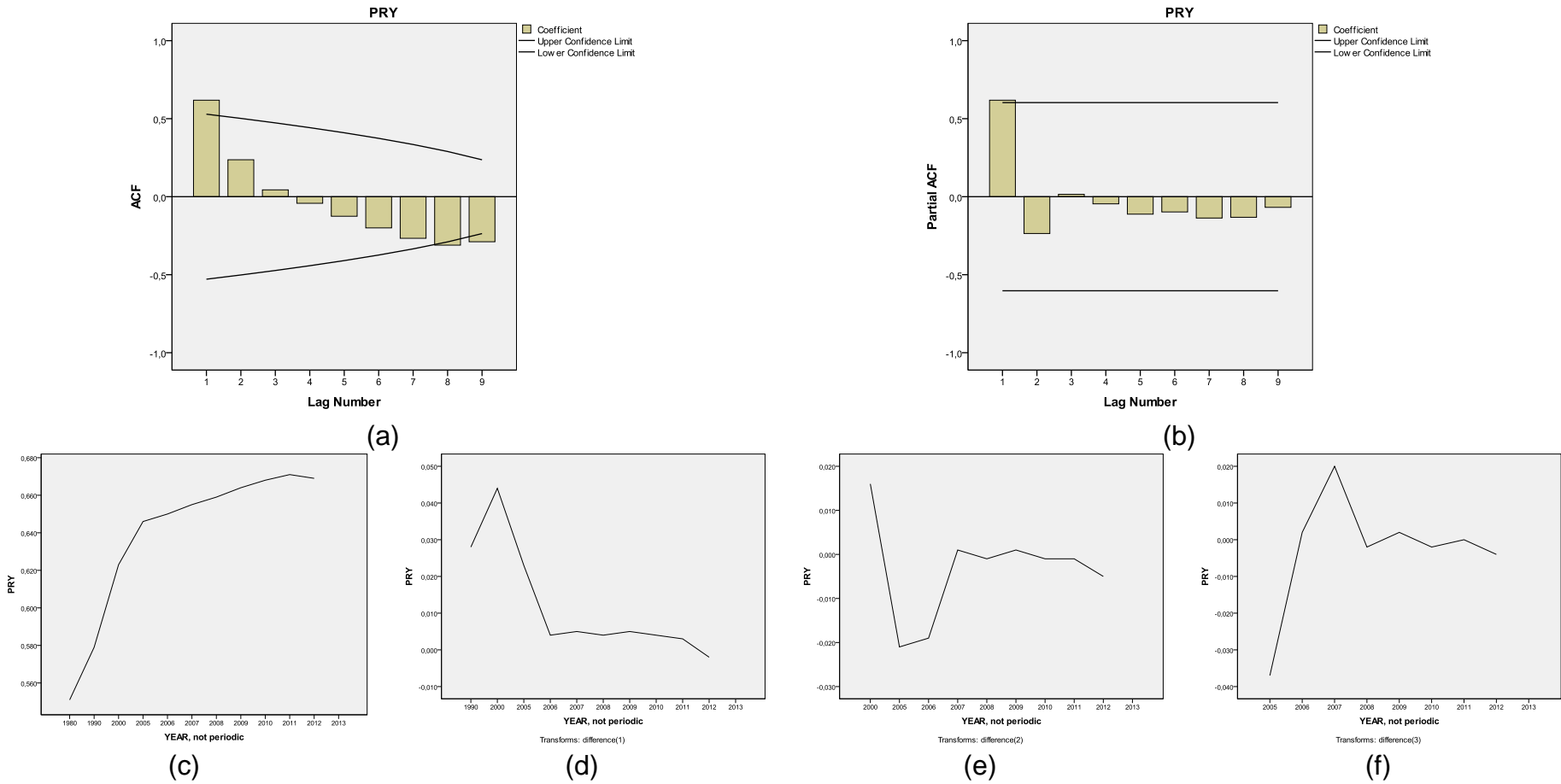




Figura 52 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do El Salvador (SLV) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

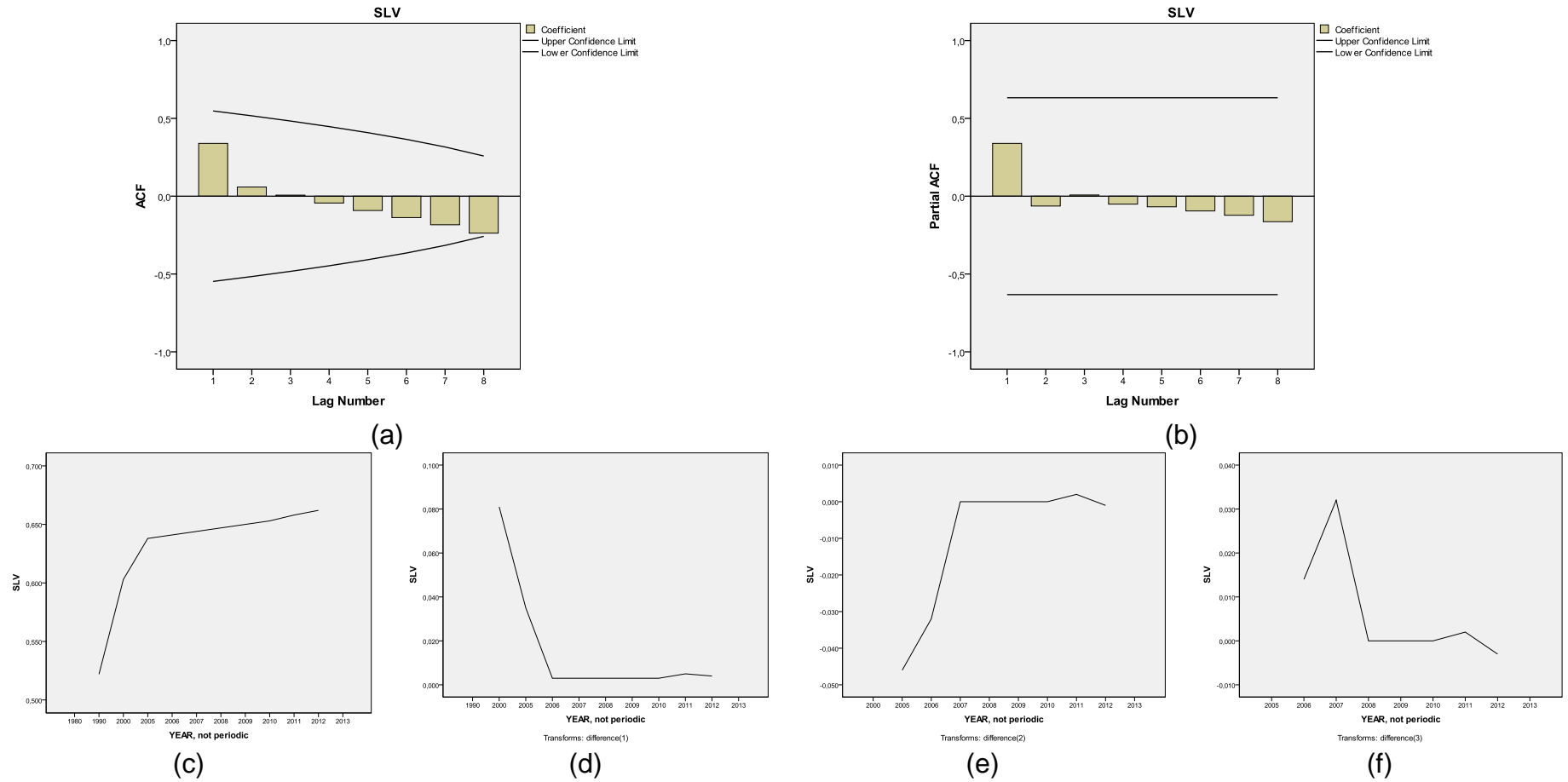


Figura 53 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Bolívia (BOL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

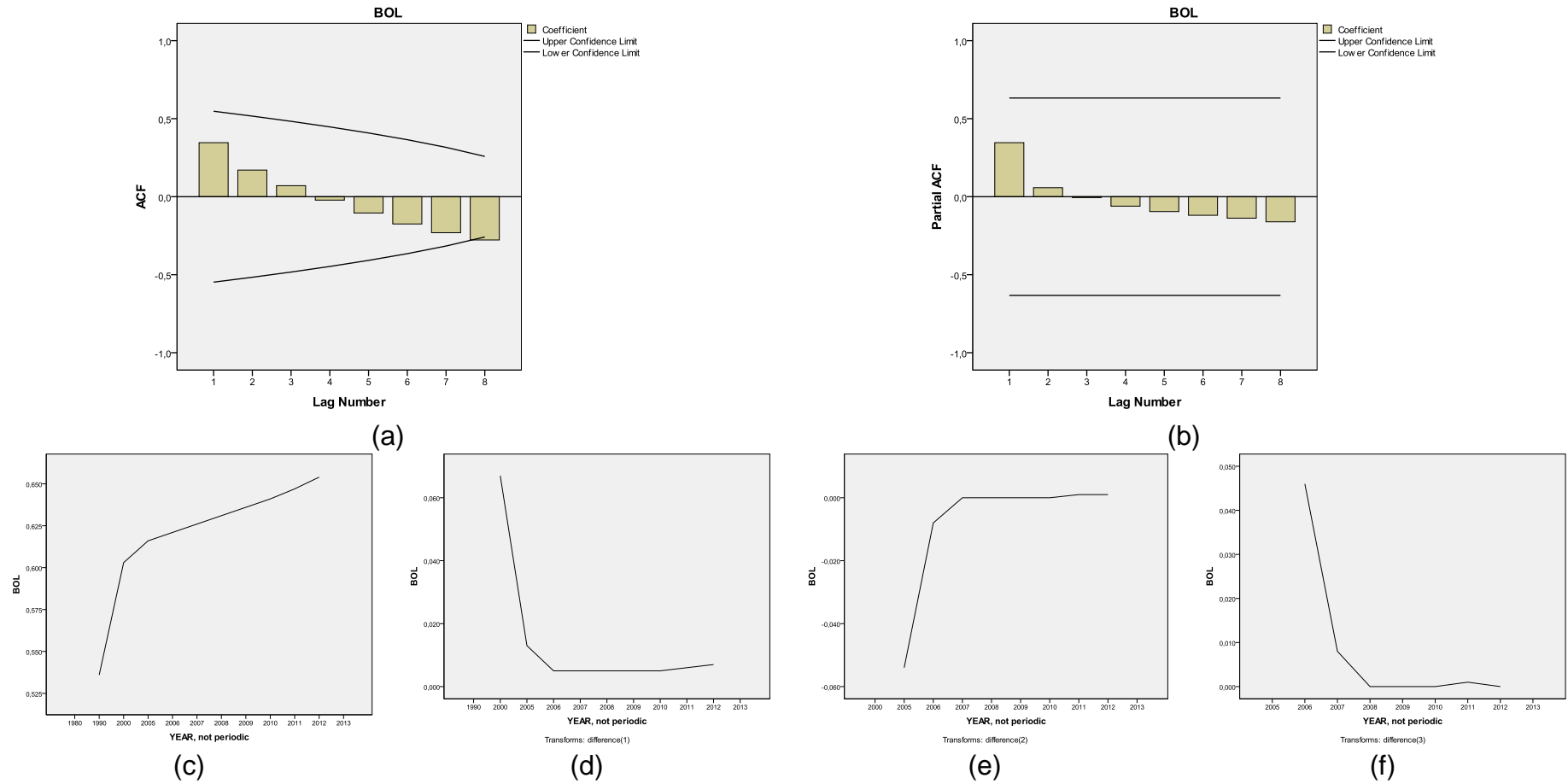


Figura 54 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2013 do Guiana (GUY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

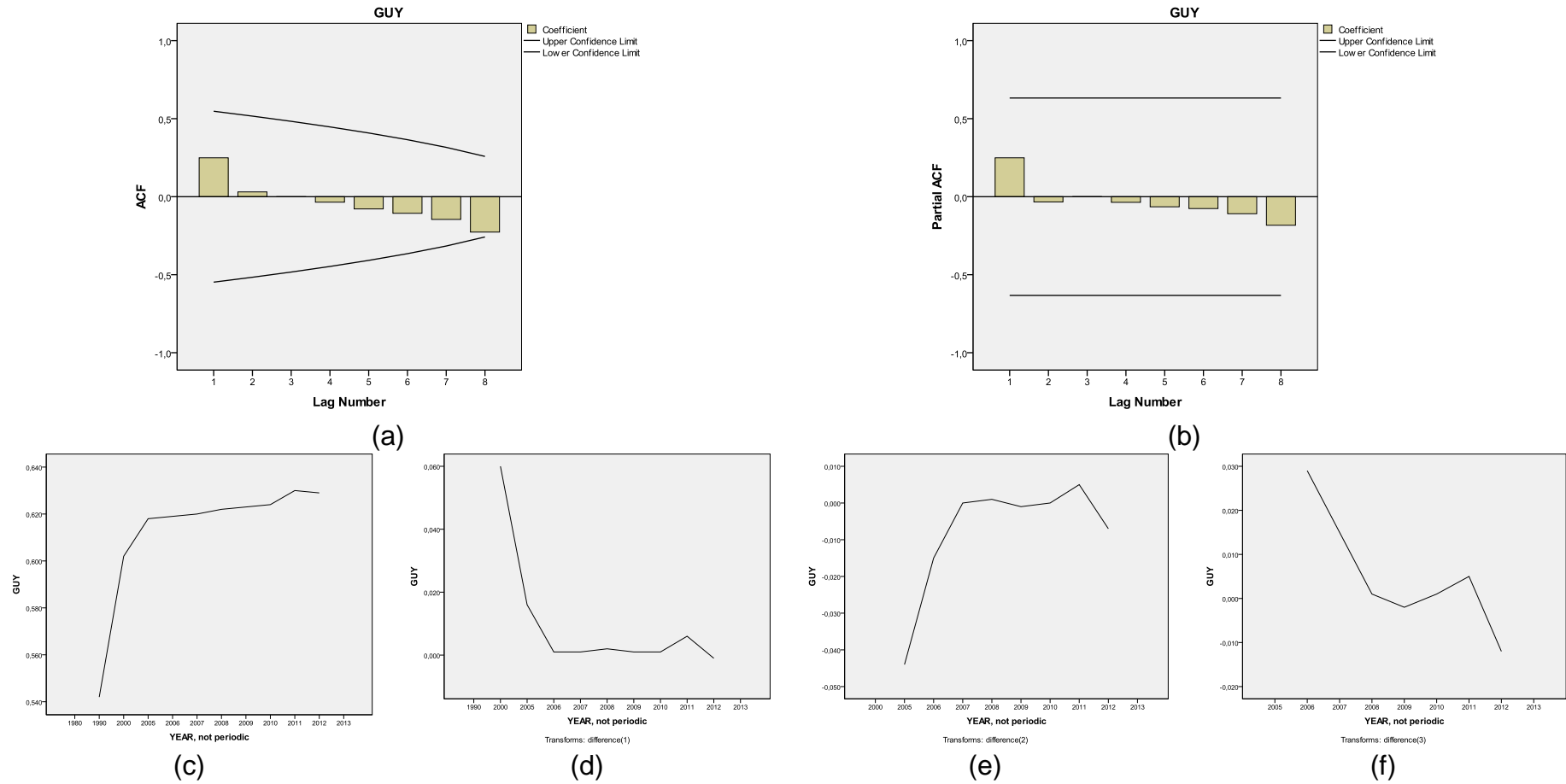


Figura 55 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Nicarágua (NIC) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

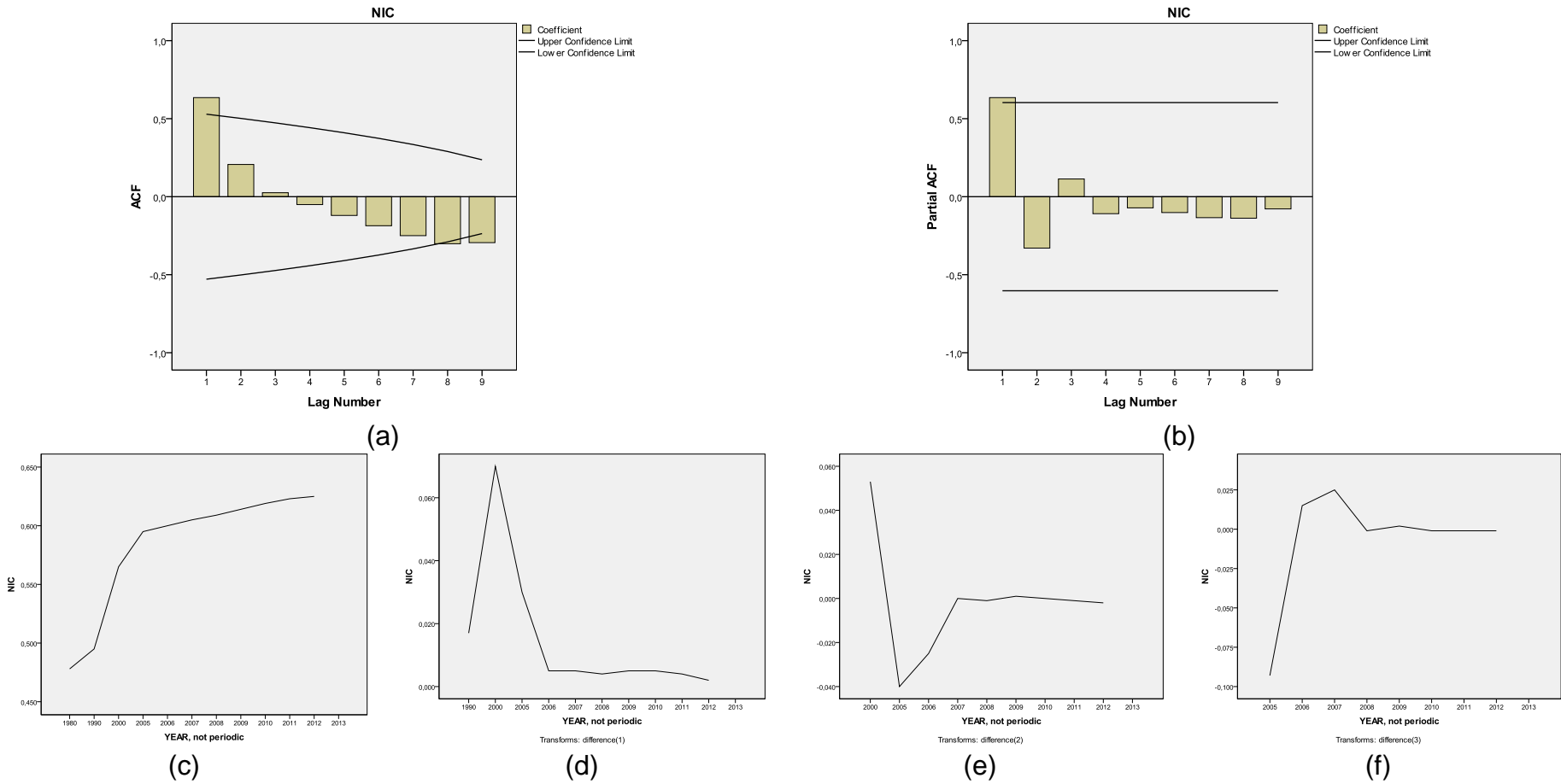


Figura 56 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2013 do Guatemala (GTM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

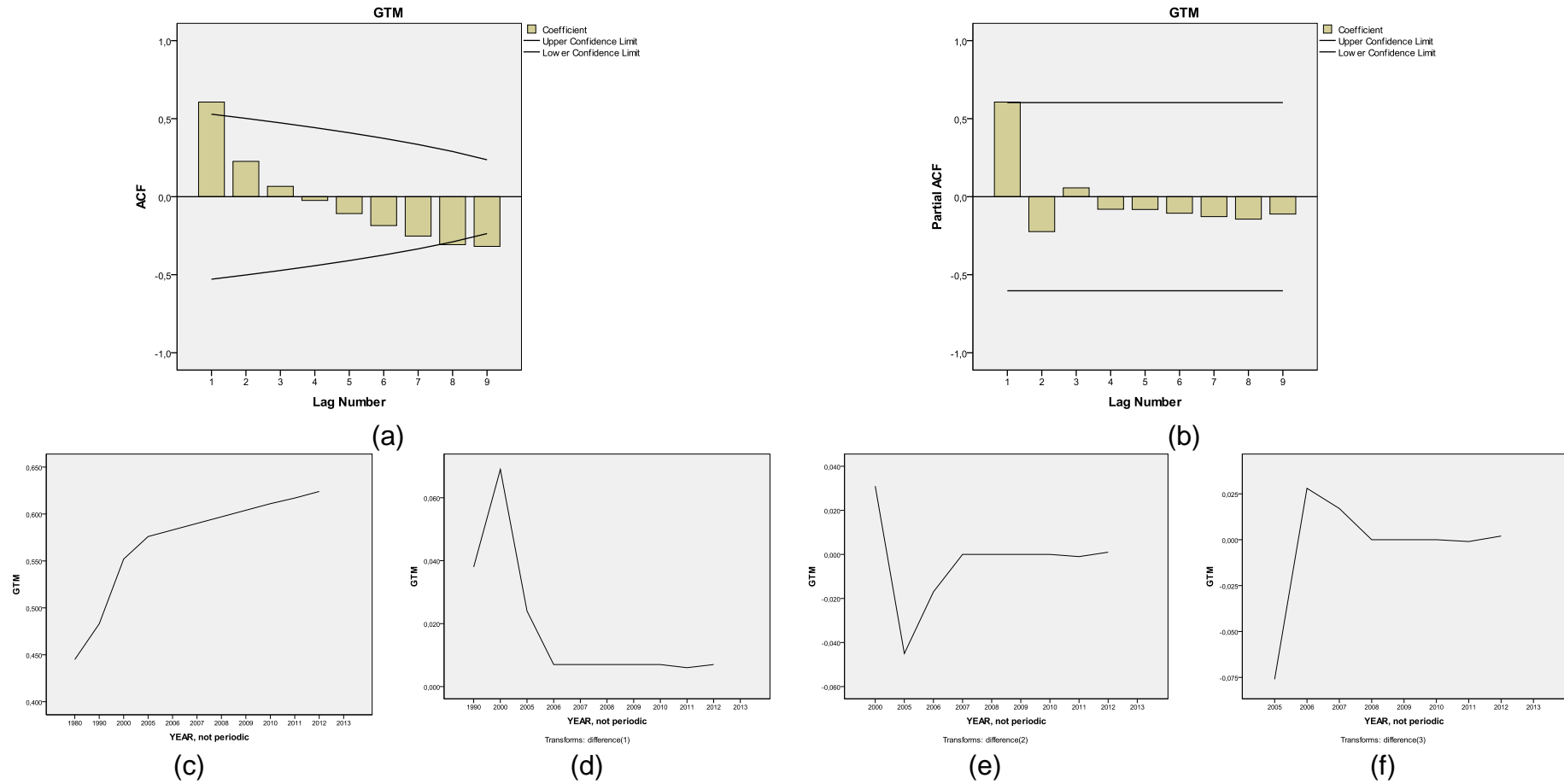


Figura 57 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,0) para Previsão do IDH 2013 do Honduras (HND) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

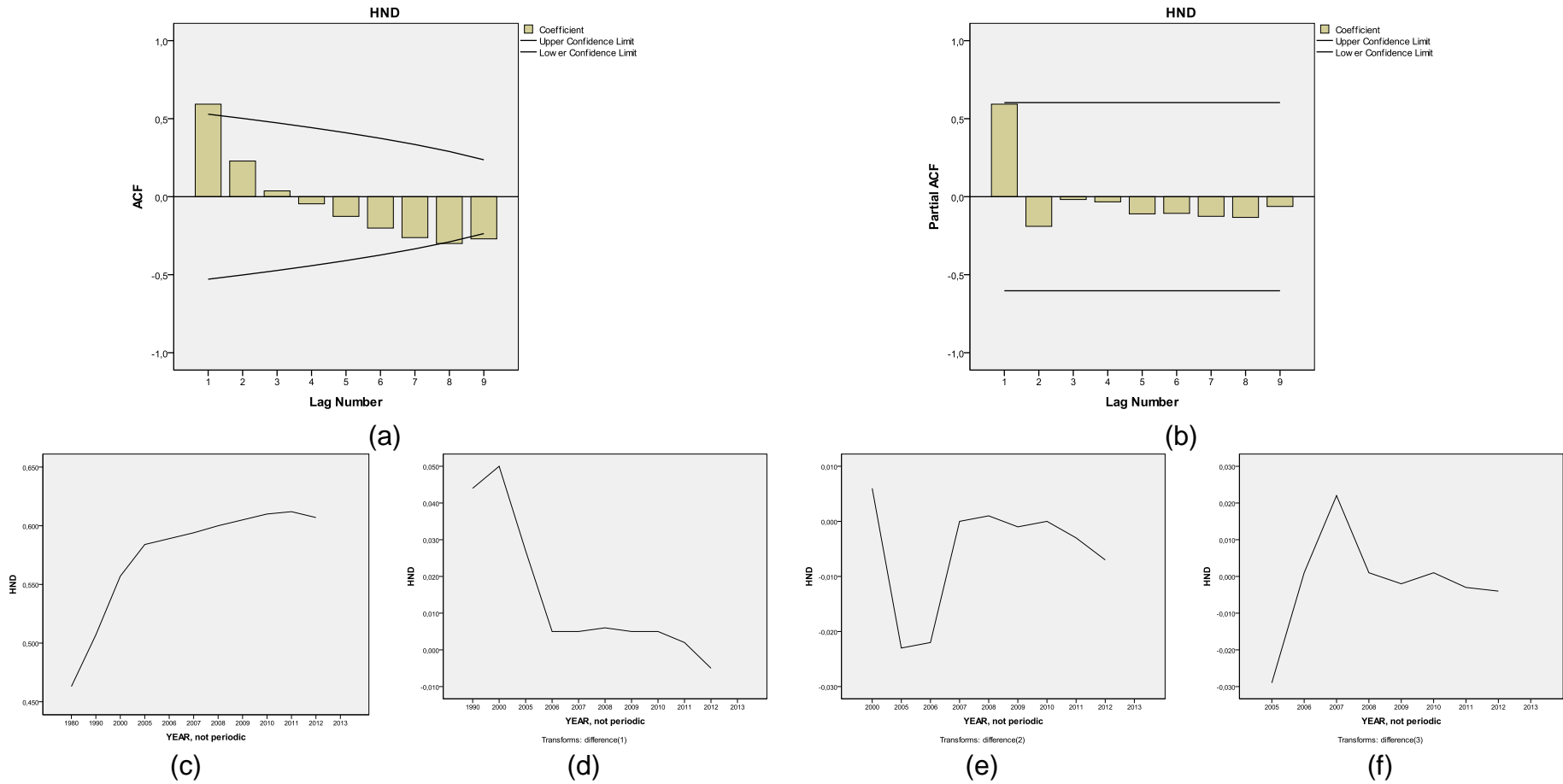
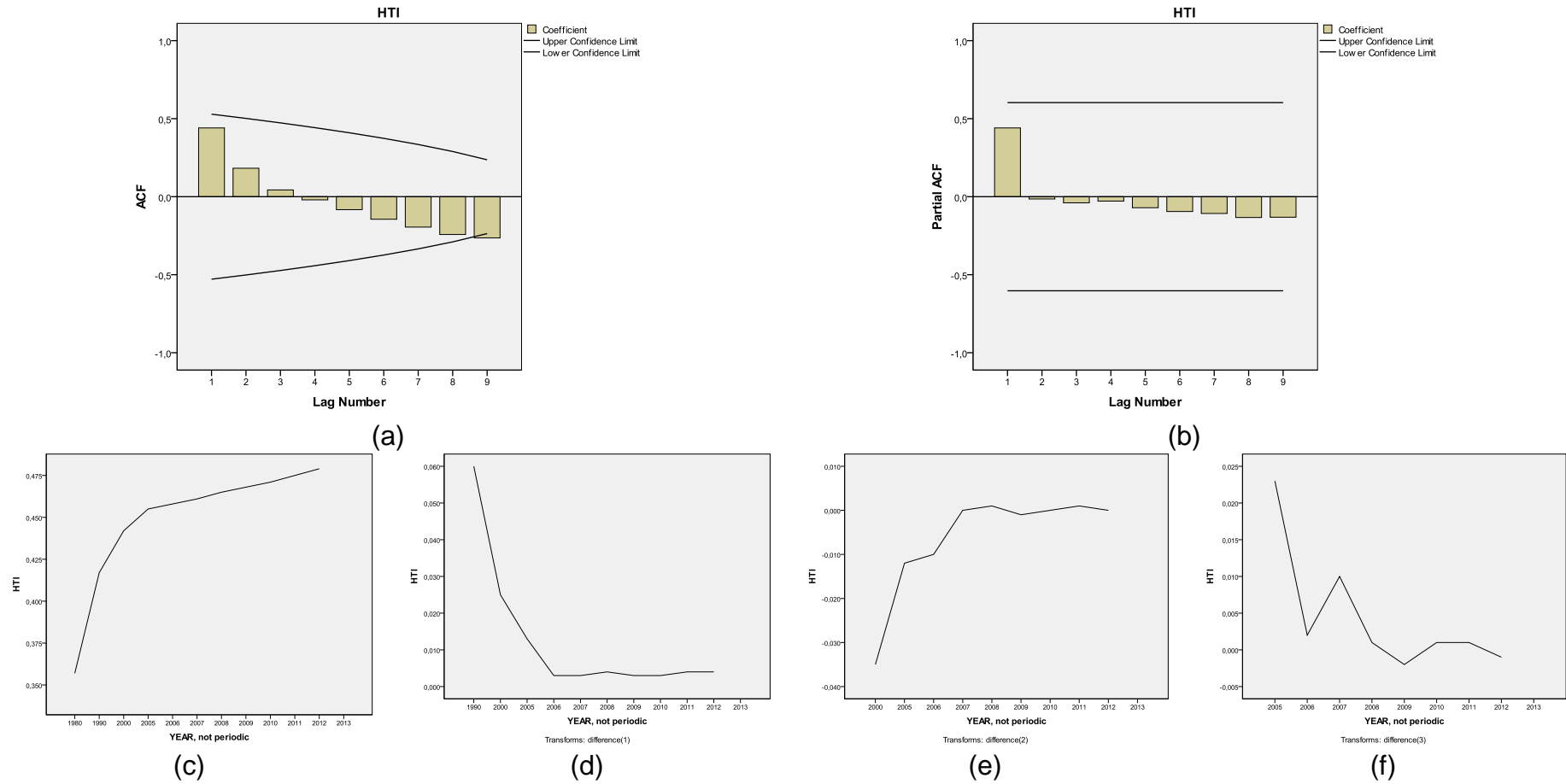


Figura 58 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,1,0) para Previsão do IDH 2014 do Haiti (HTI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)



**APÊNDICE 5 – Correlogramas e Séries Temporais Observadas para a  
Modelagem dos Modelos ARIMA para Previsão do IDH 2014 dos Países  
da América Latina**



Figura 59 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 da Argentina (ARG) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

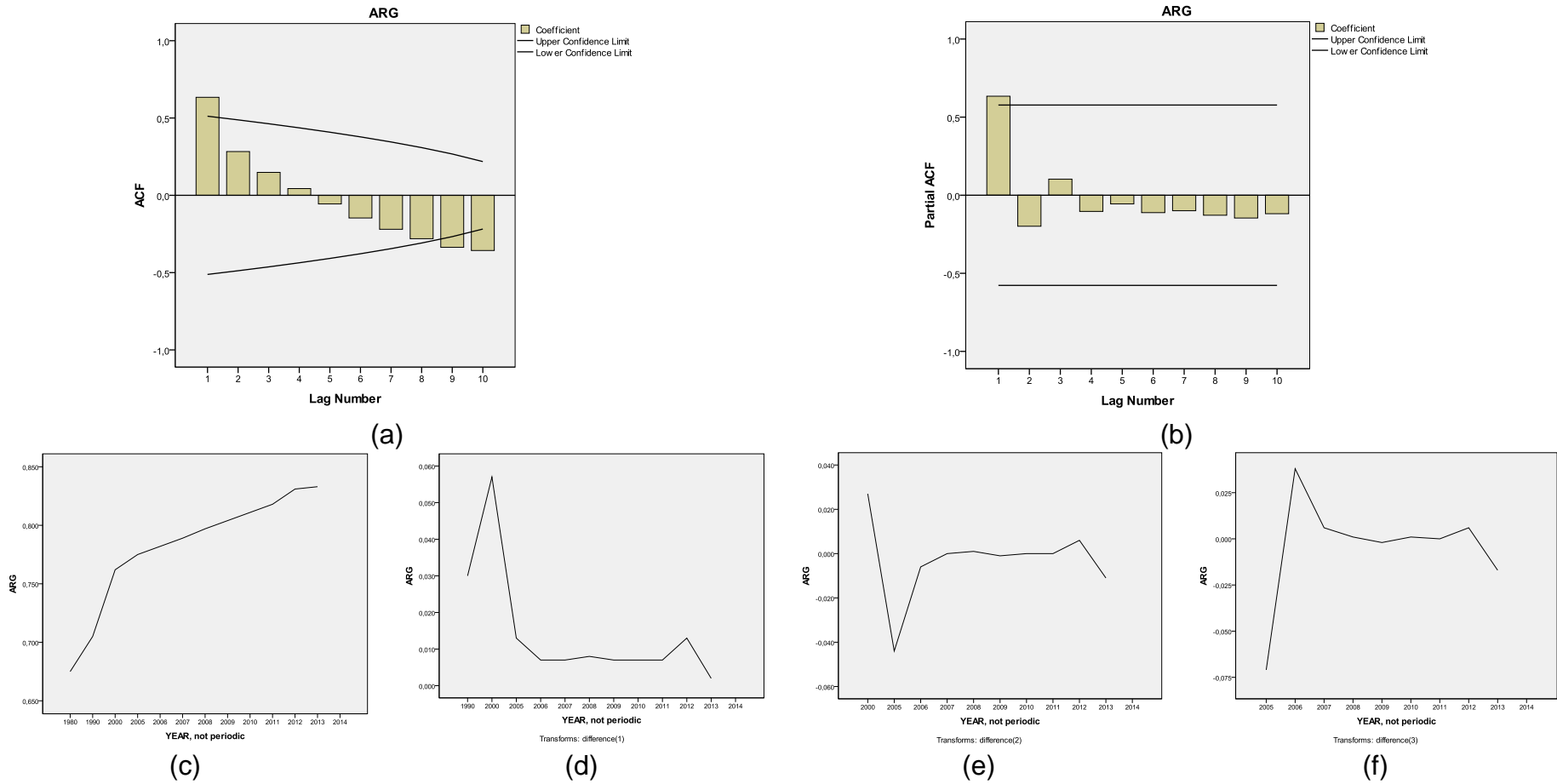


Figura 60 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Chile (CHL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

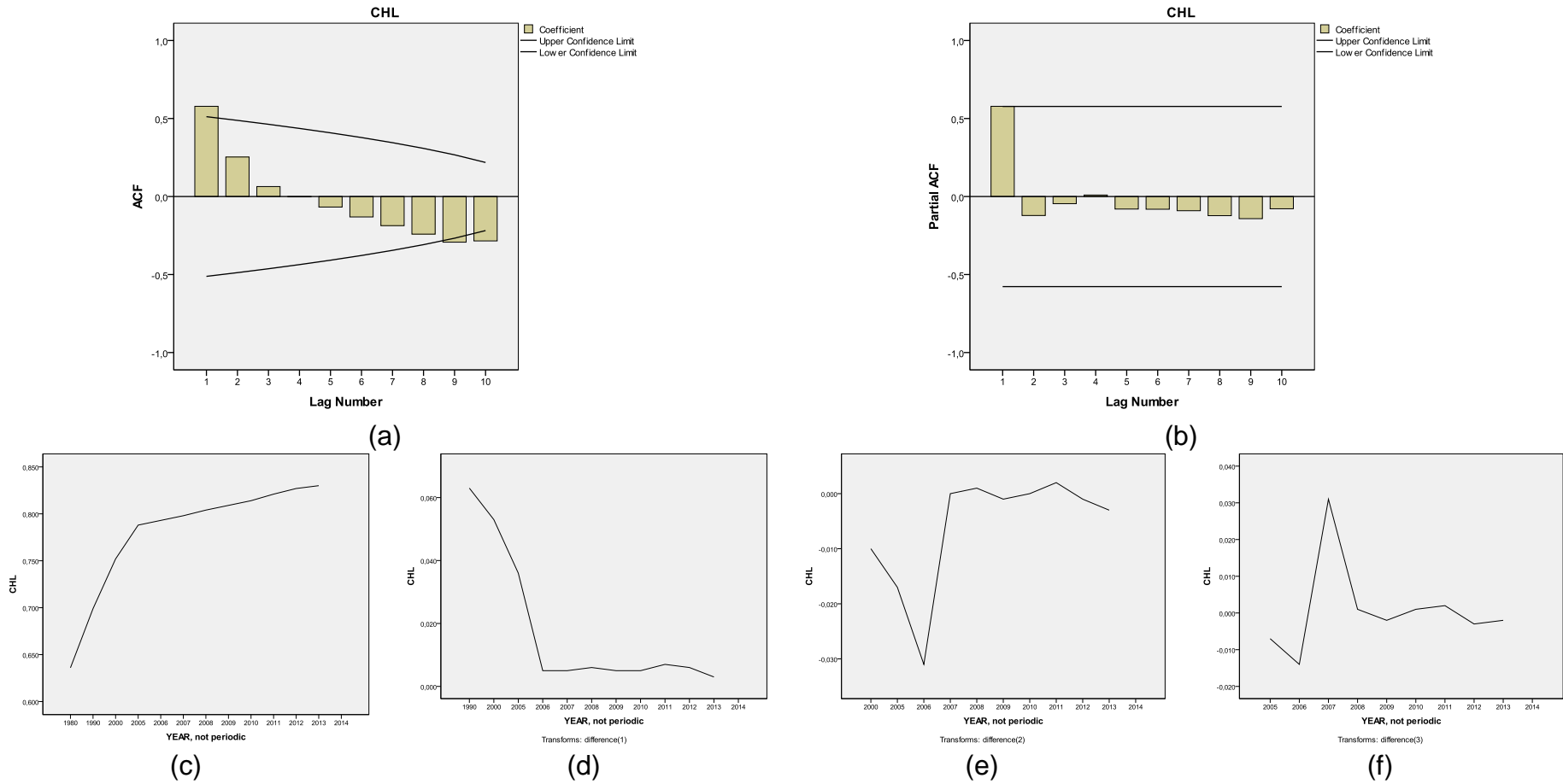


Figura 61 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Uruguai (URY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

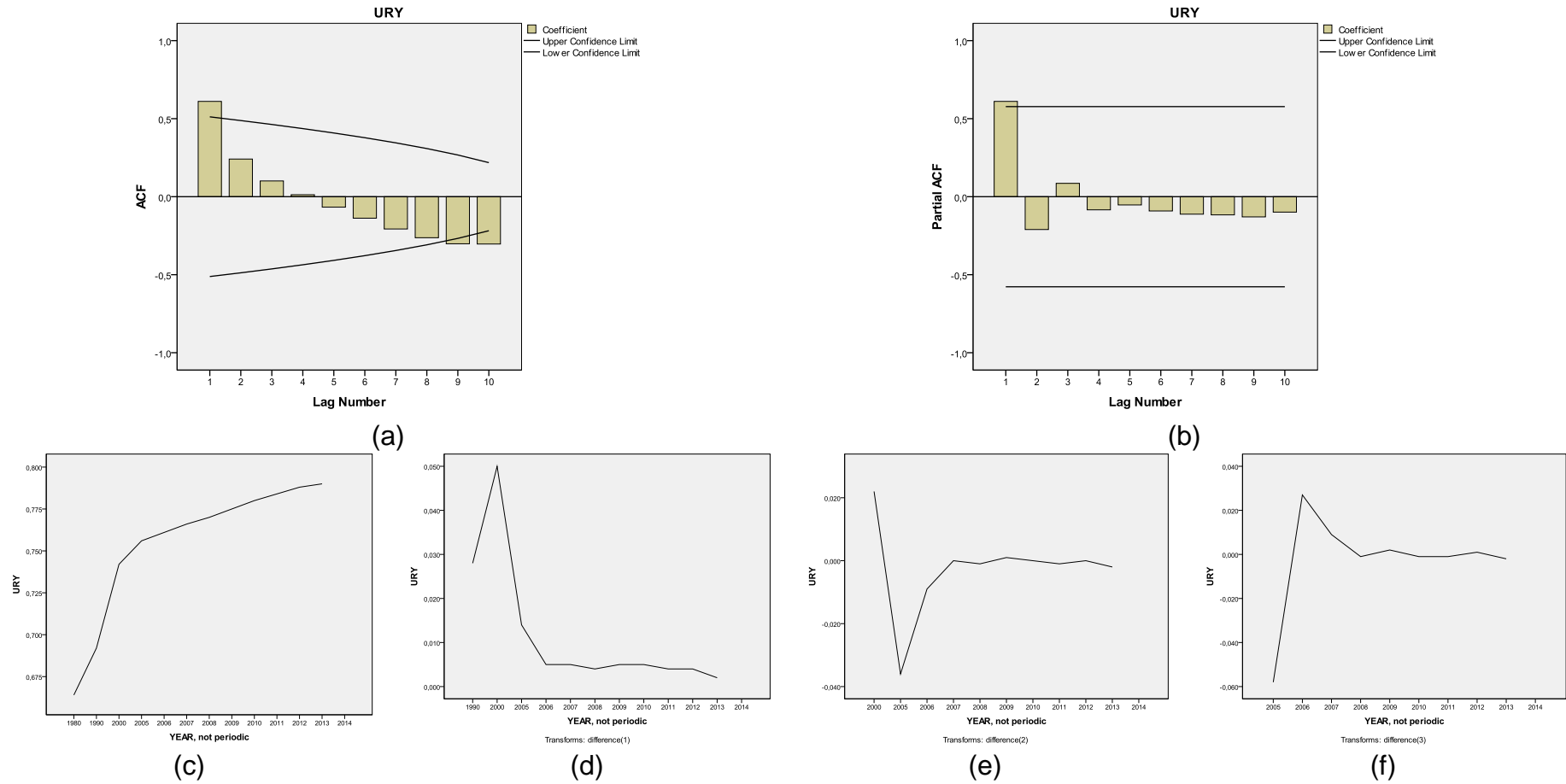


Figura 62 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Panamá (PAN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

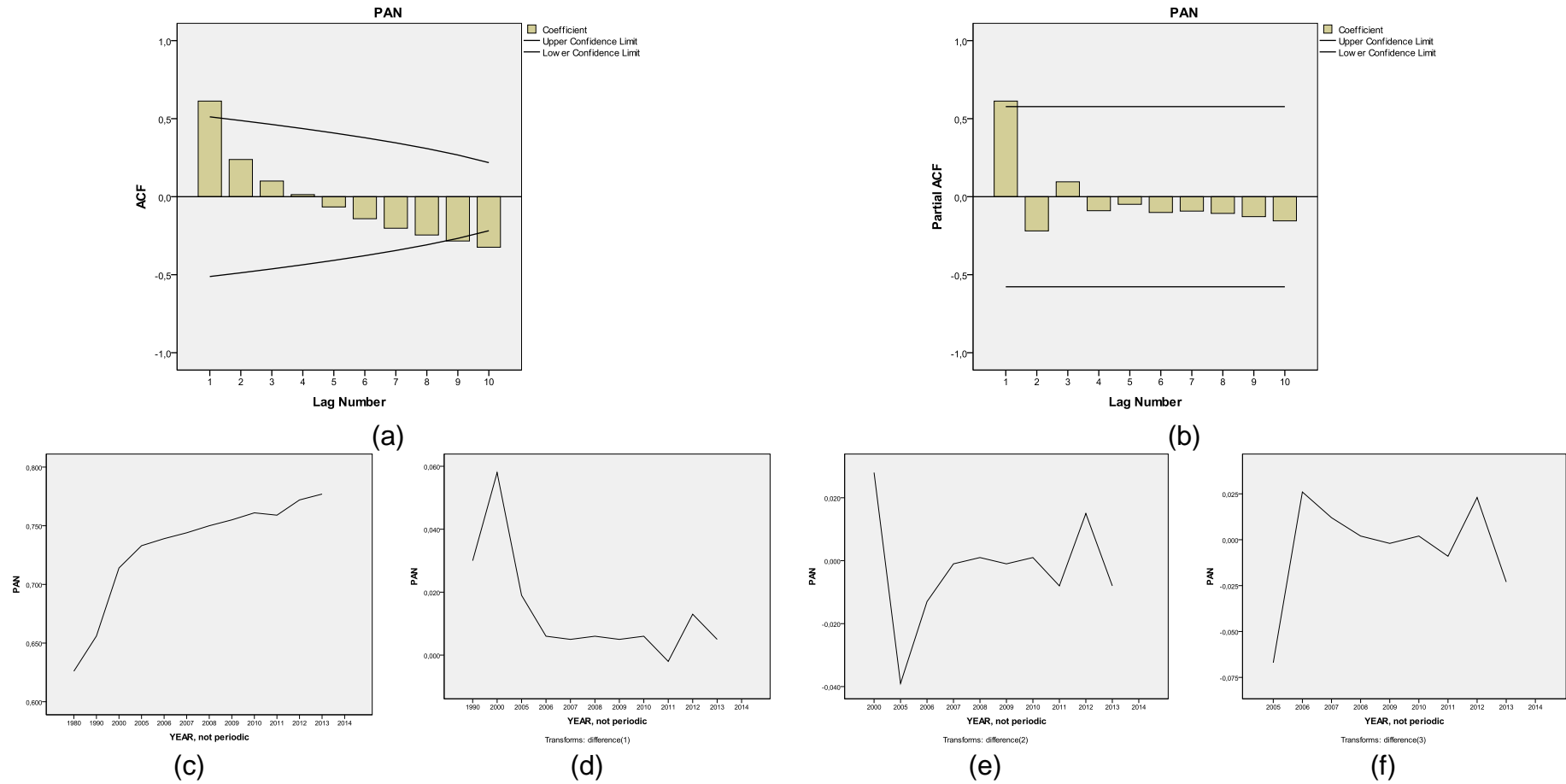


Figura 63 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Cuba (CUB) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

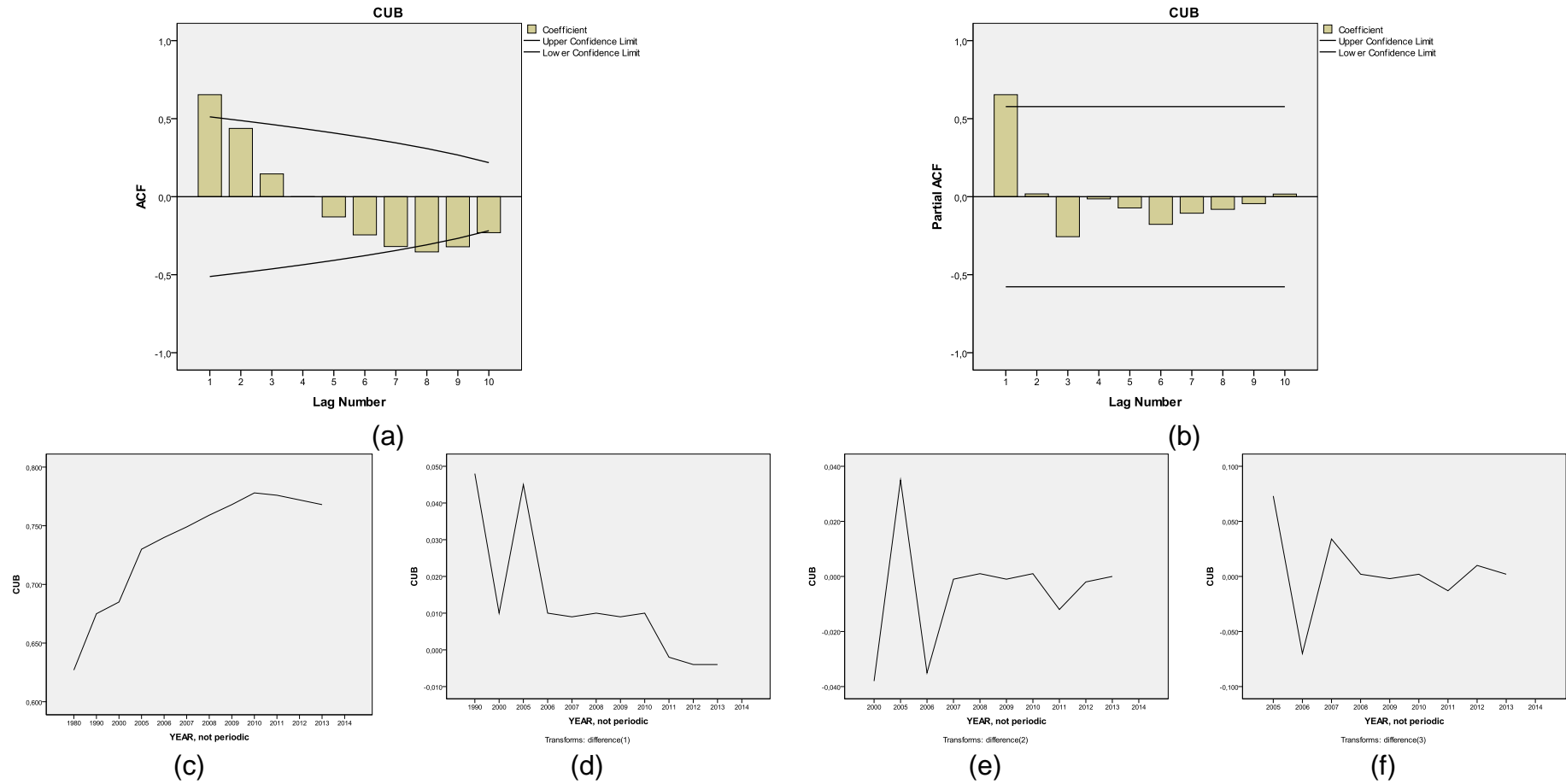


Figura 64 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Costa Rica (CRI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

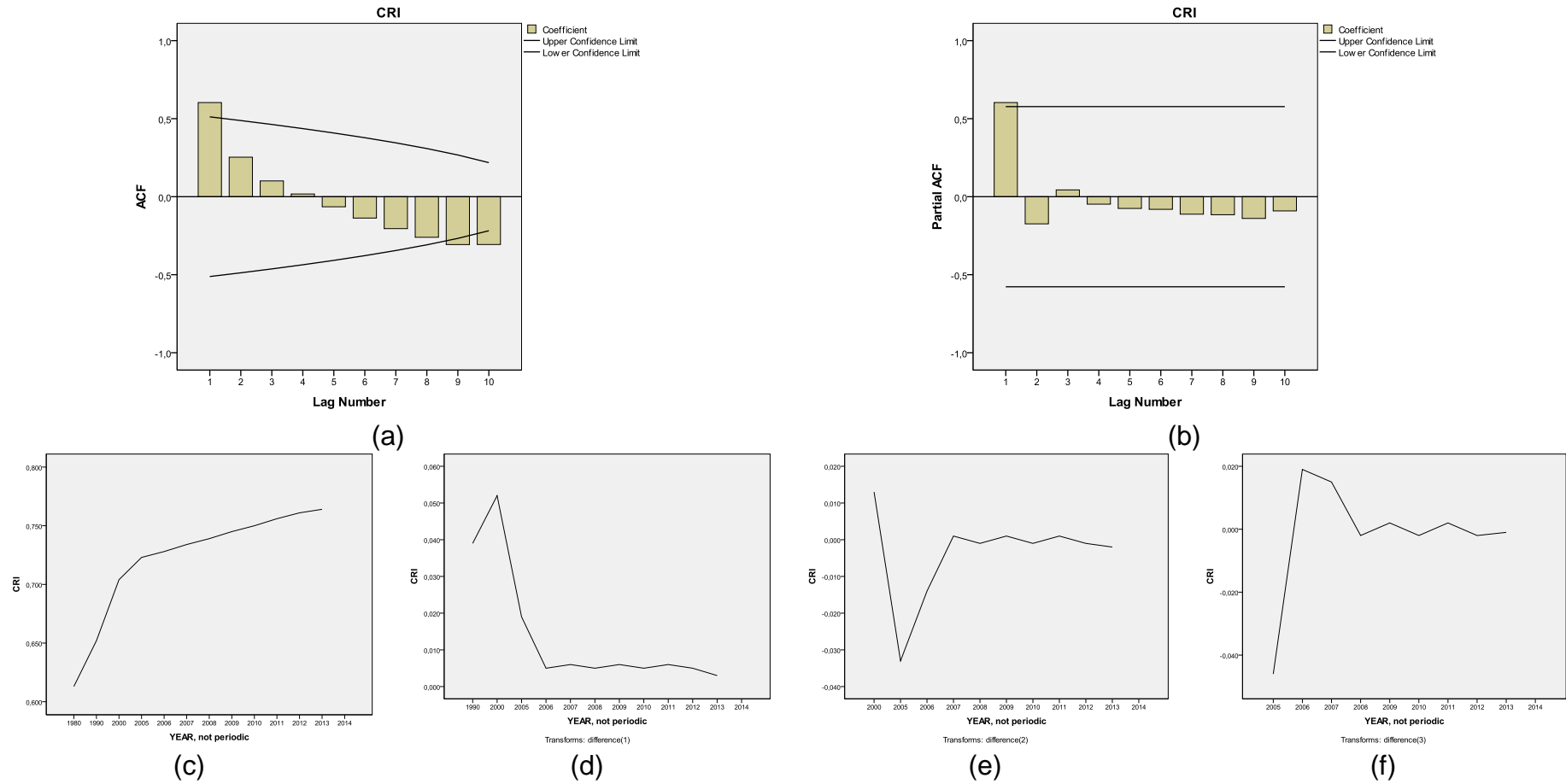


Figura 65 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Venezuela (VEN) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

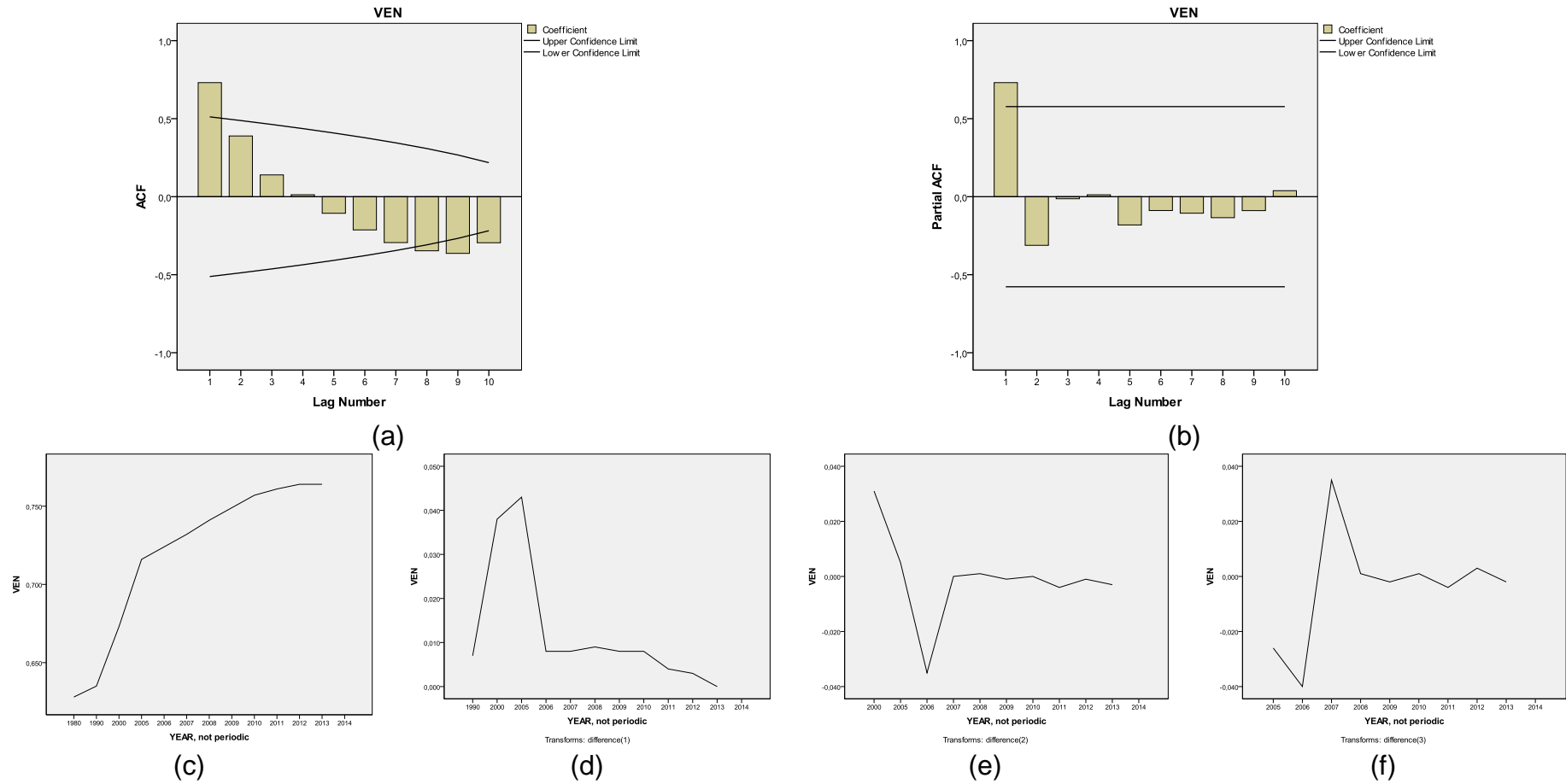


Figura 66 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do México (MEX) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

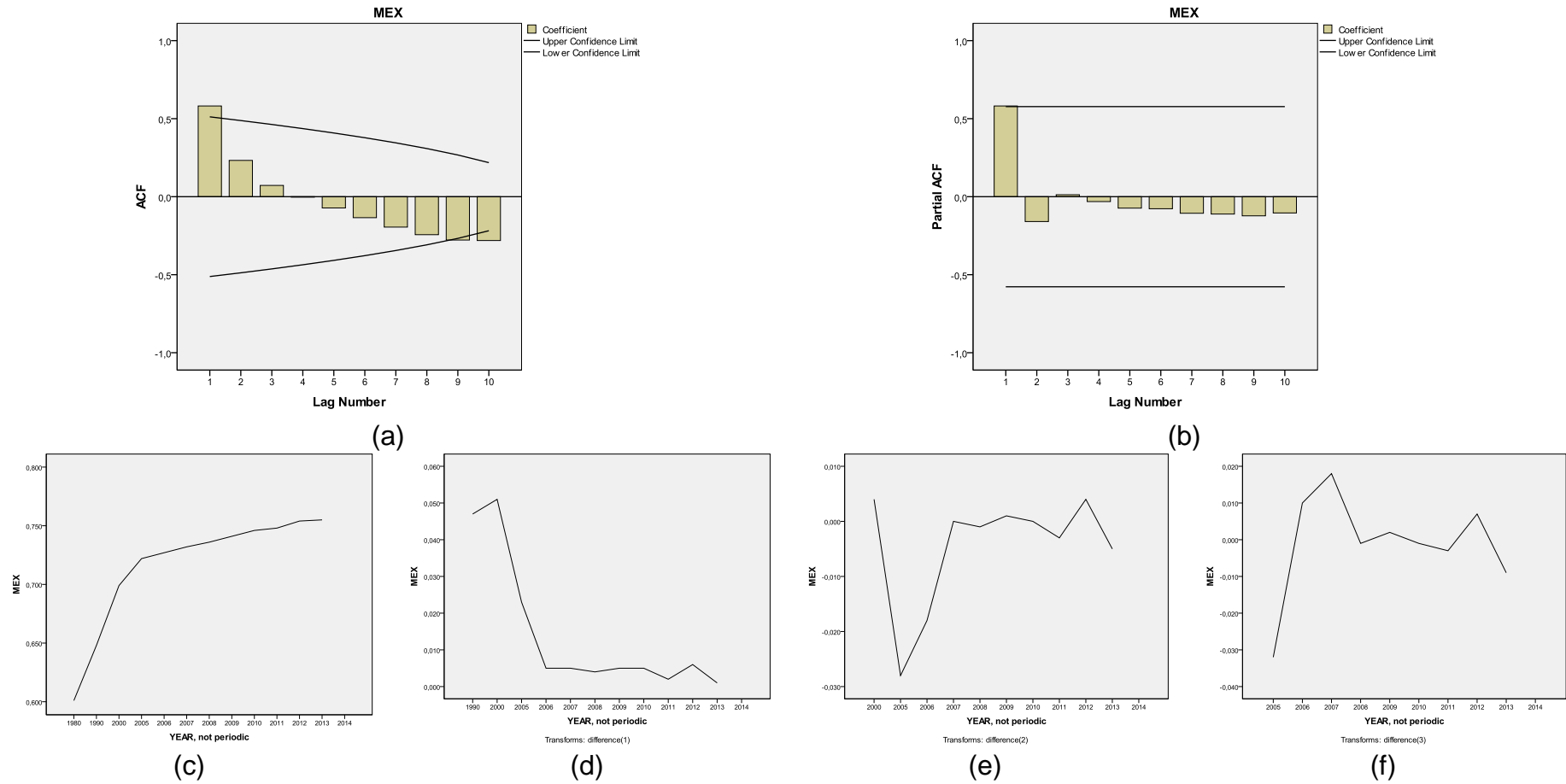




Figura 67 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Brasil (BRA) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

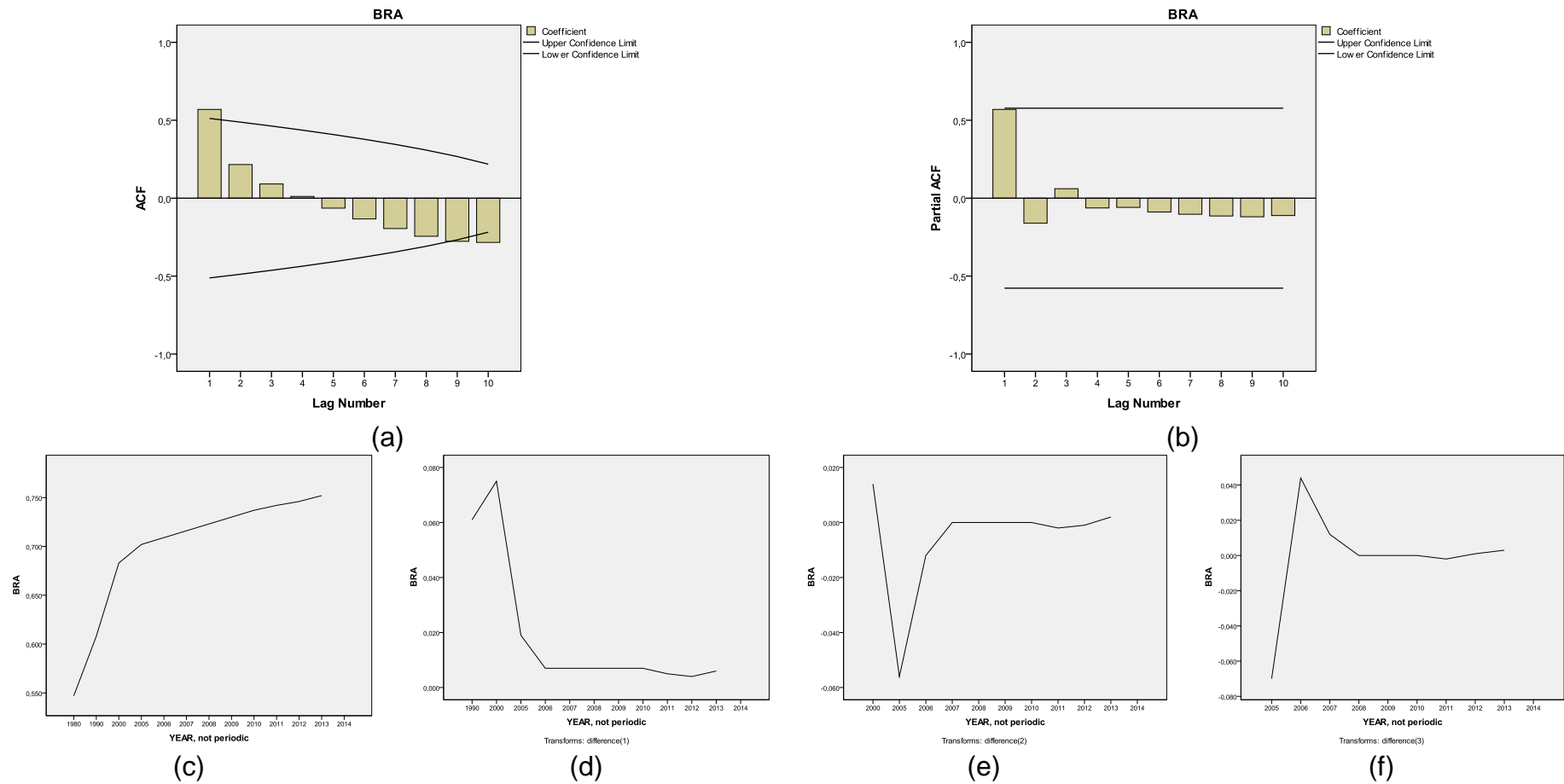


Figura 68 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Peru (PER) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

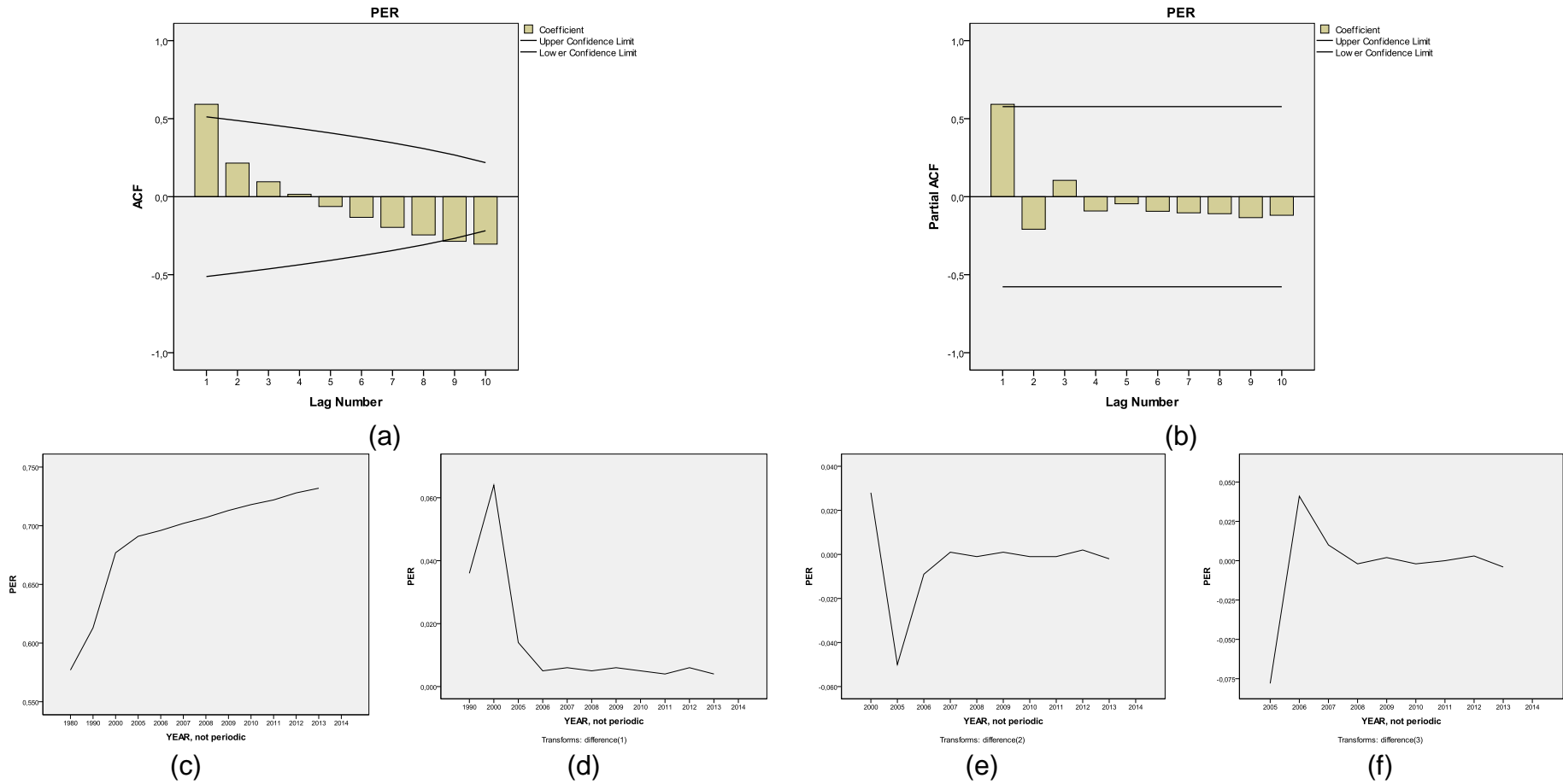


Figura 69 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Equador (ECU) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

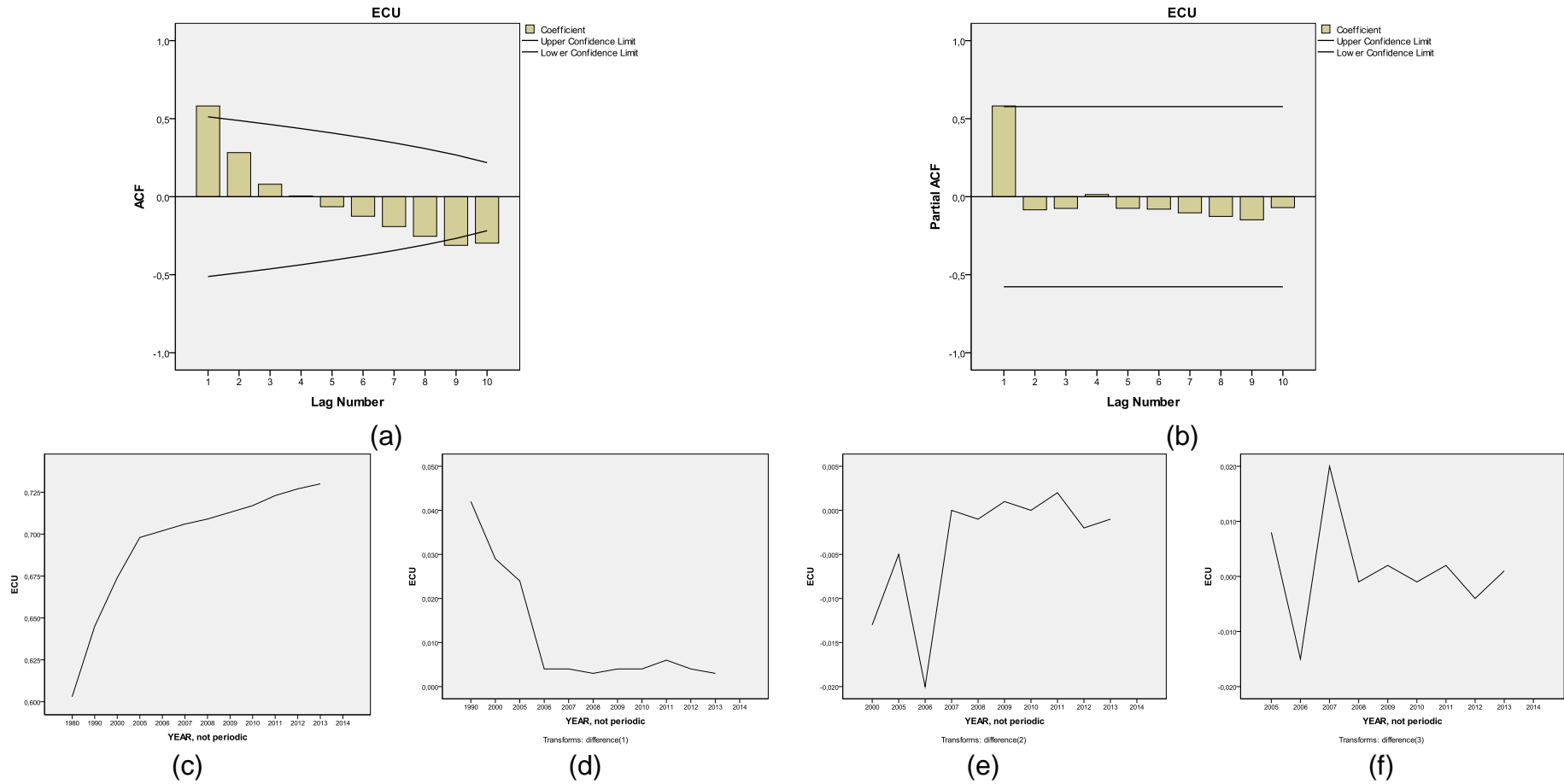


Figura 70 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Colômbia (COL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difencições(e) e três diferenciações (f)

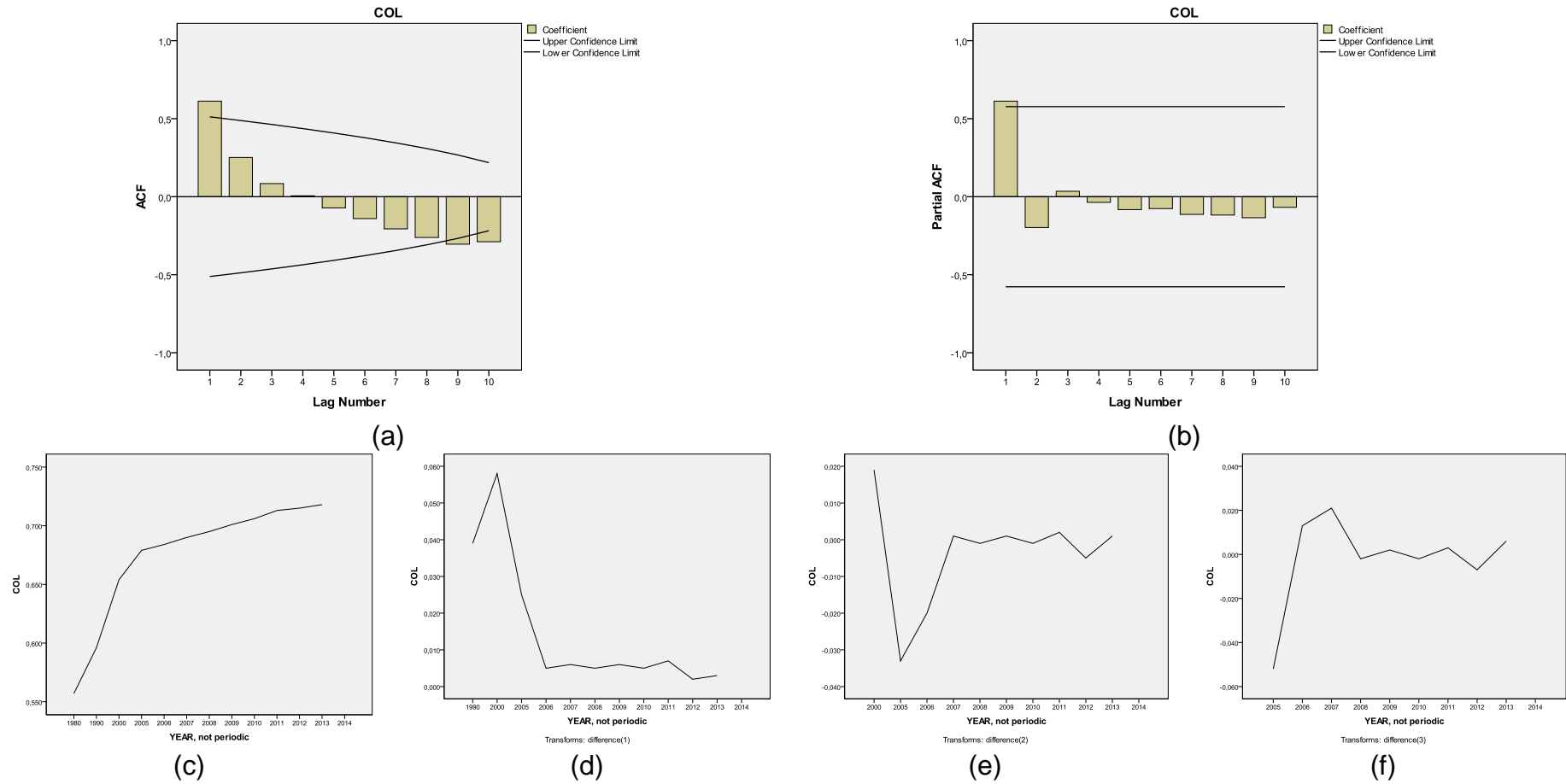


Figura 71 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Belize (BLZ) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

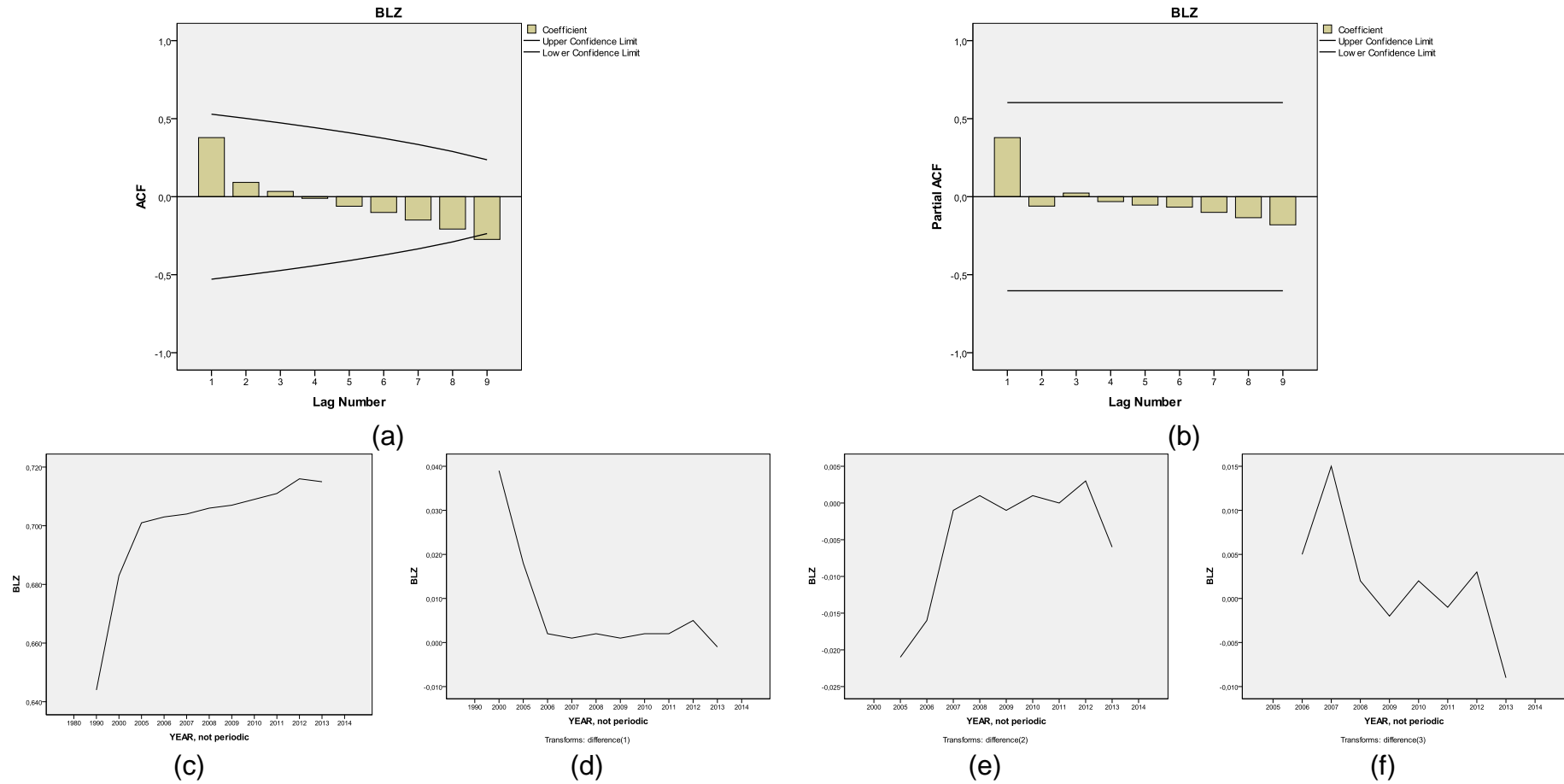


Figura 72 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do República Dominicana (DOM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difincações(e) e três diferenciações (f)

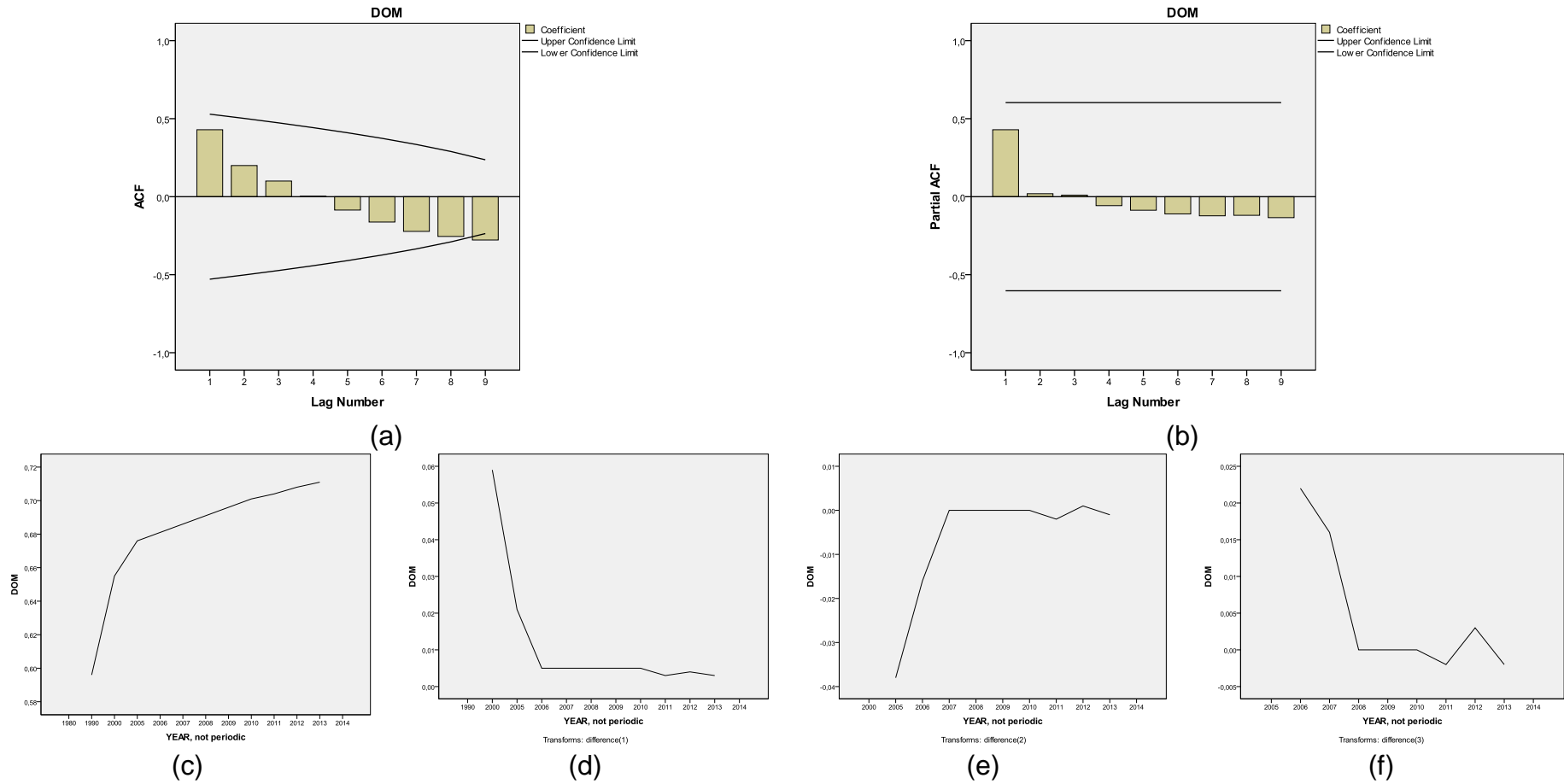


Figura 73 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Paraguai (PRY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

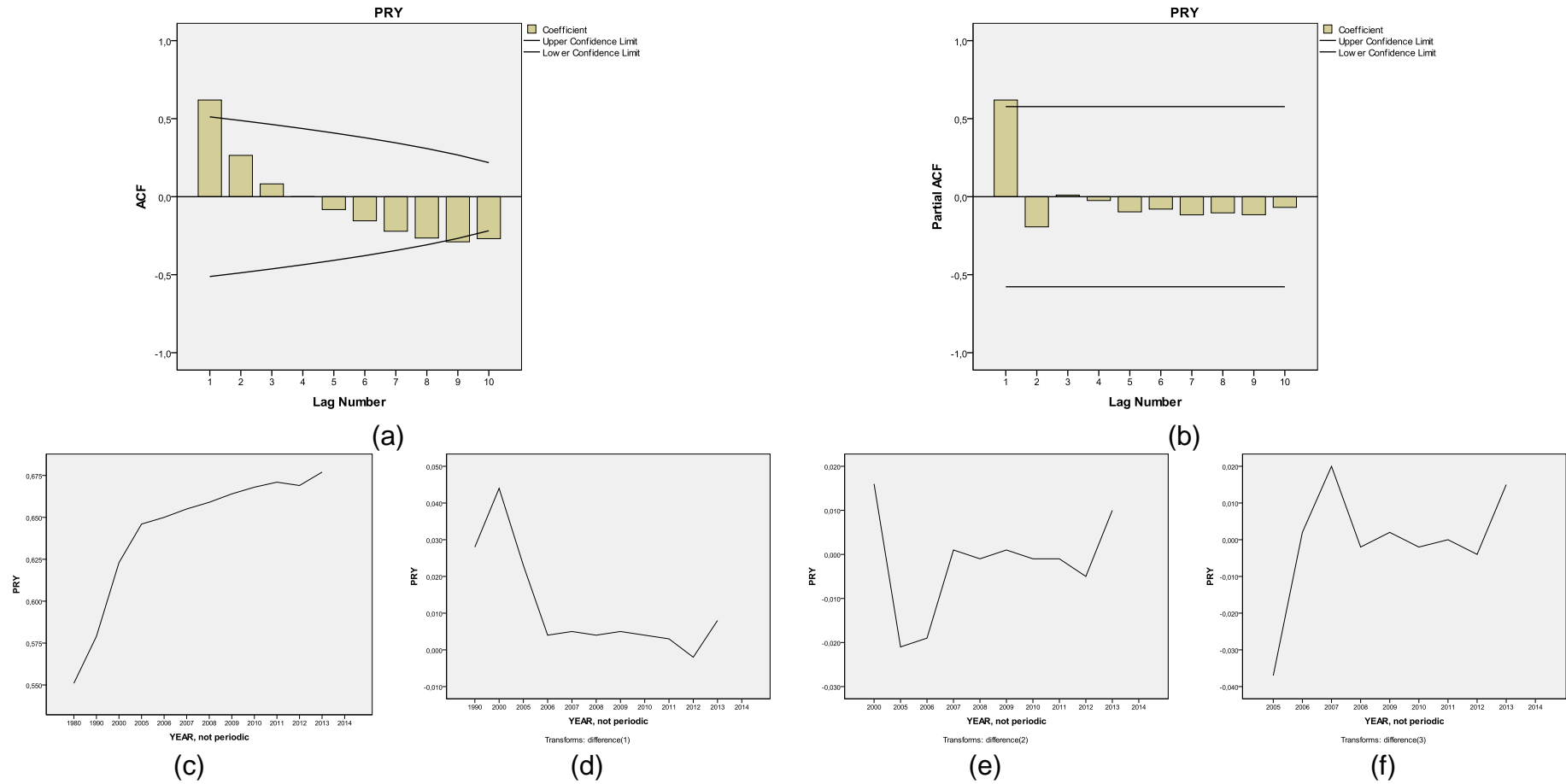


Figura 74 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do El Salvador (SLV) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

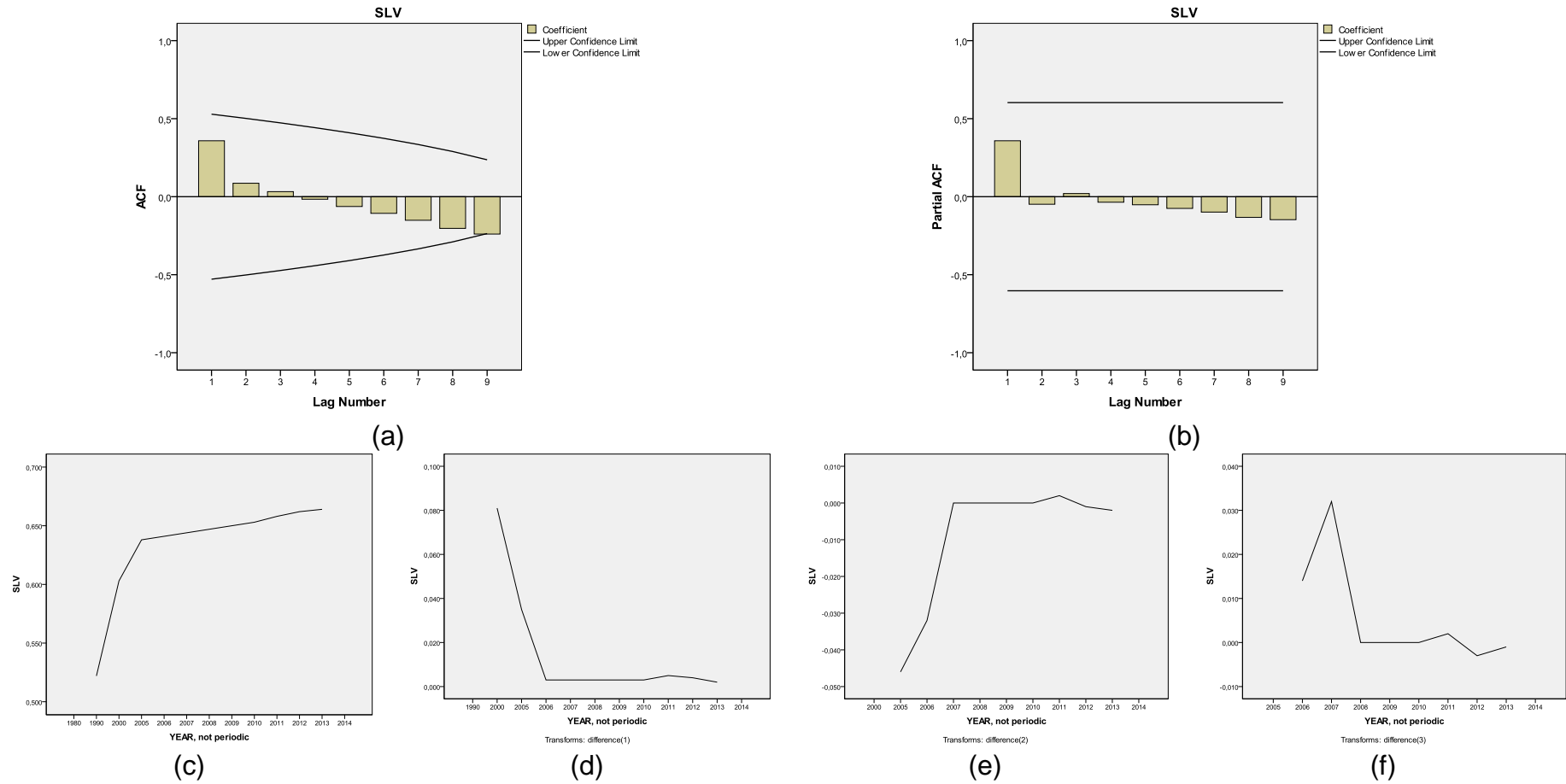




Figura 75 - Correlogramas ACF(a) e PACF(b), do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Bolívia (BOL) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas diferenciações(e) e três diferenciações (f)

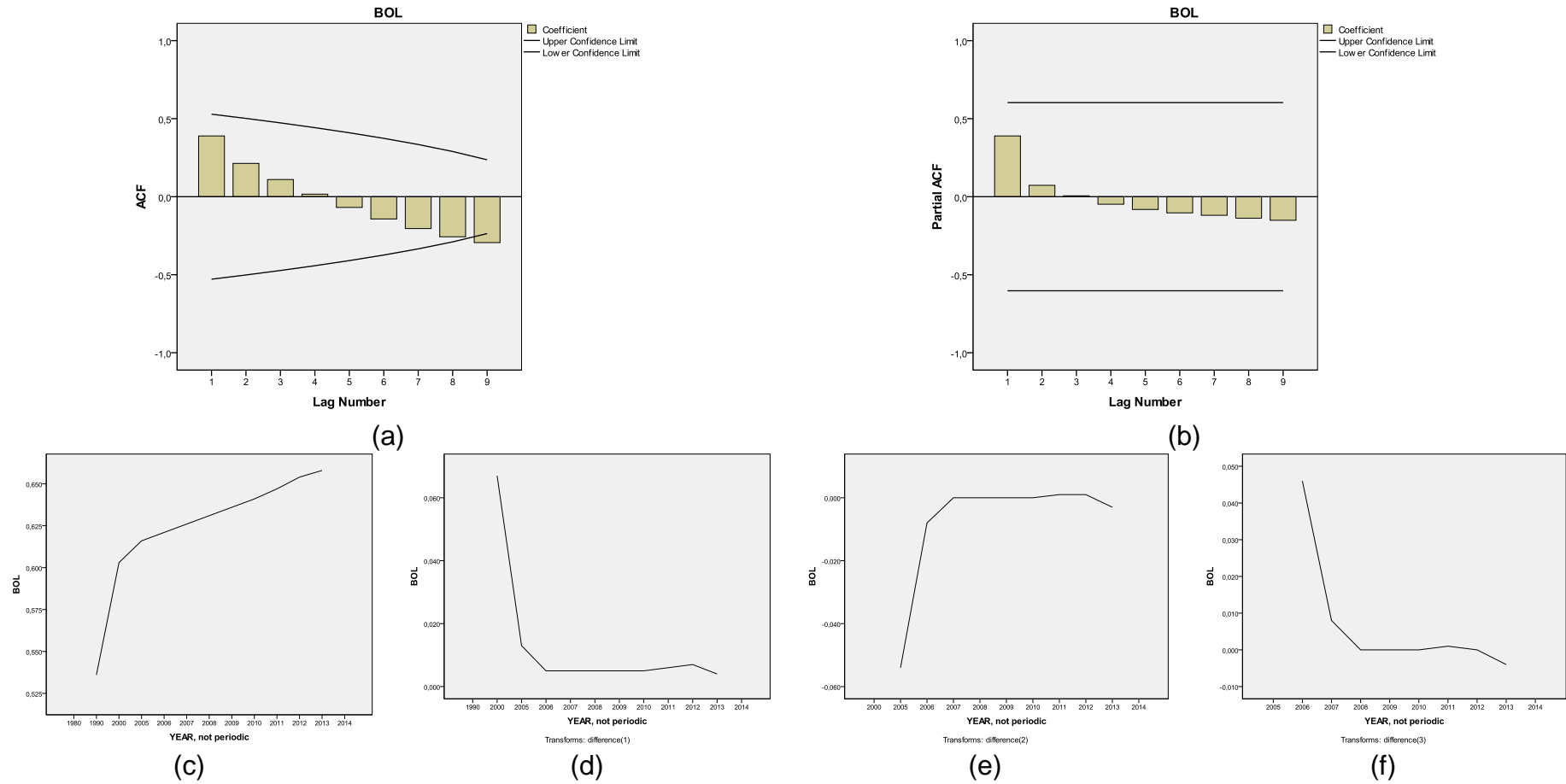


Figura 76 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (0,2,0) para Previsão do IDH 2014 do Guiana (GUY) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

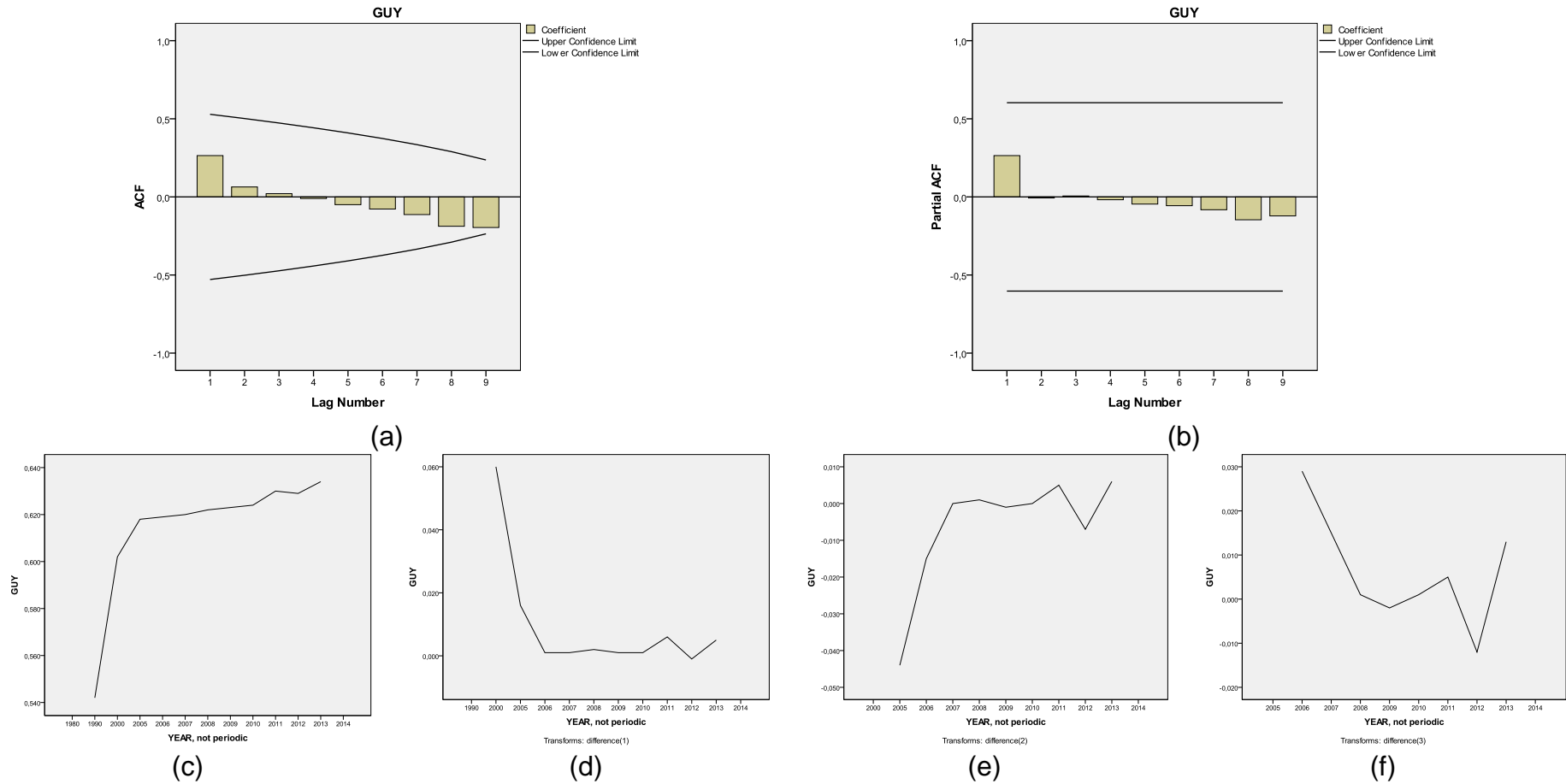


Figura 77 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Nicarágua (NIC) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

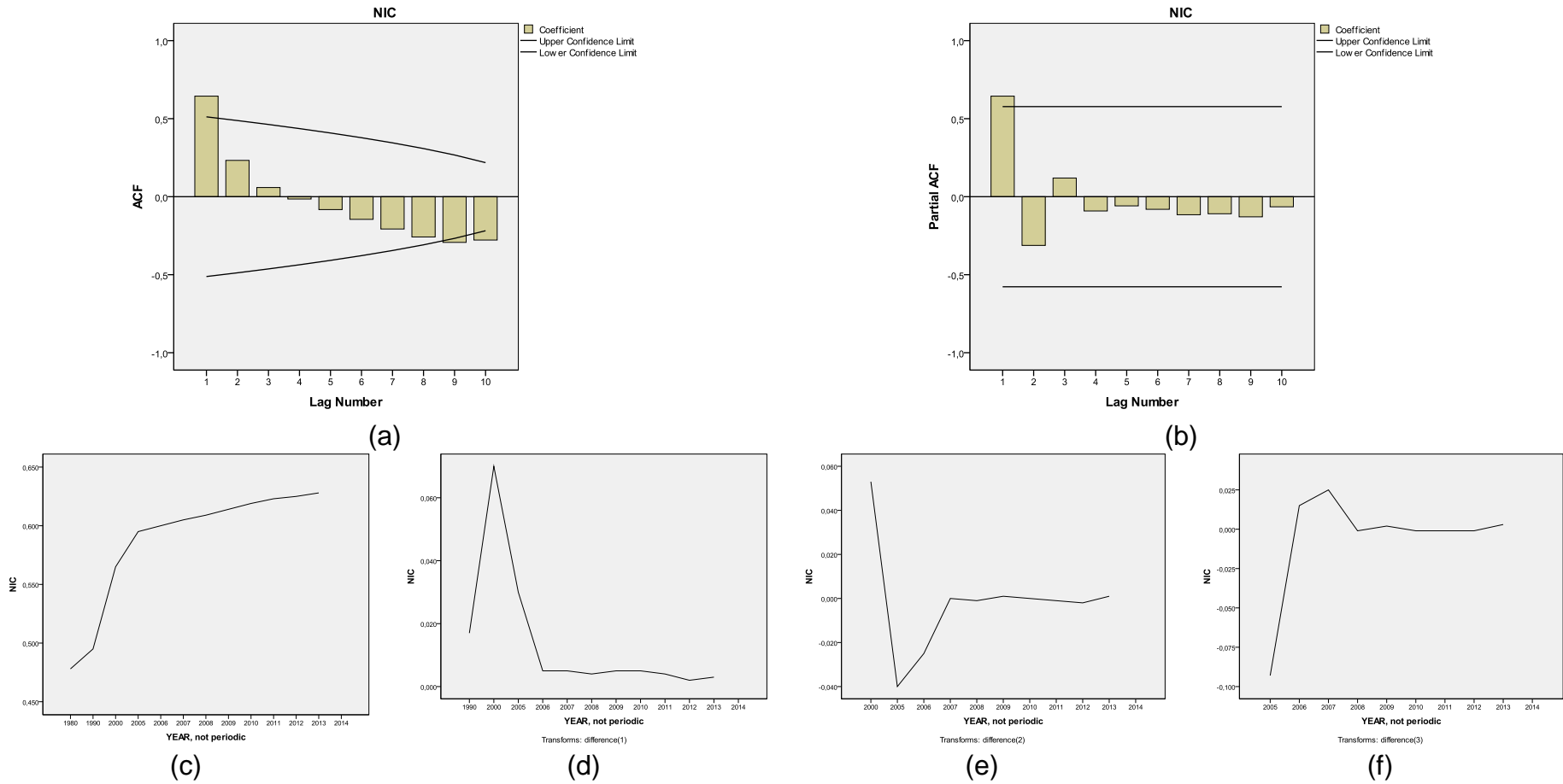


Figura 78 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Guatemala (GTM) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

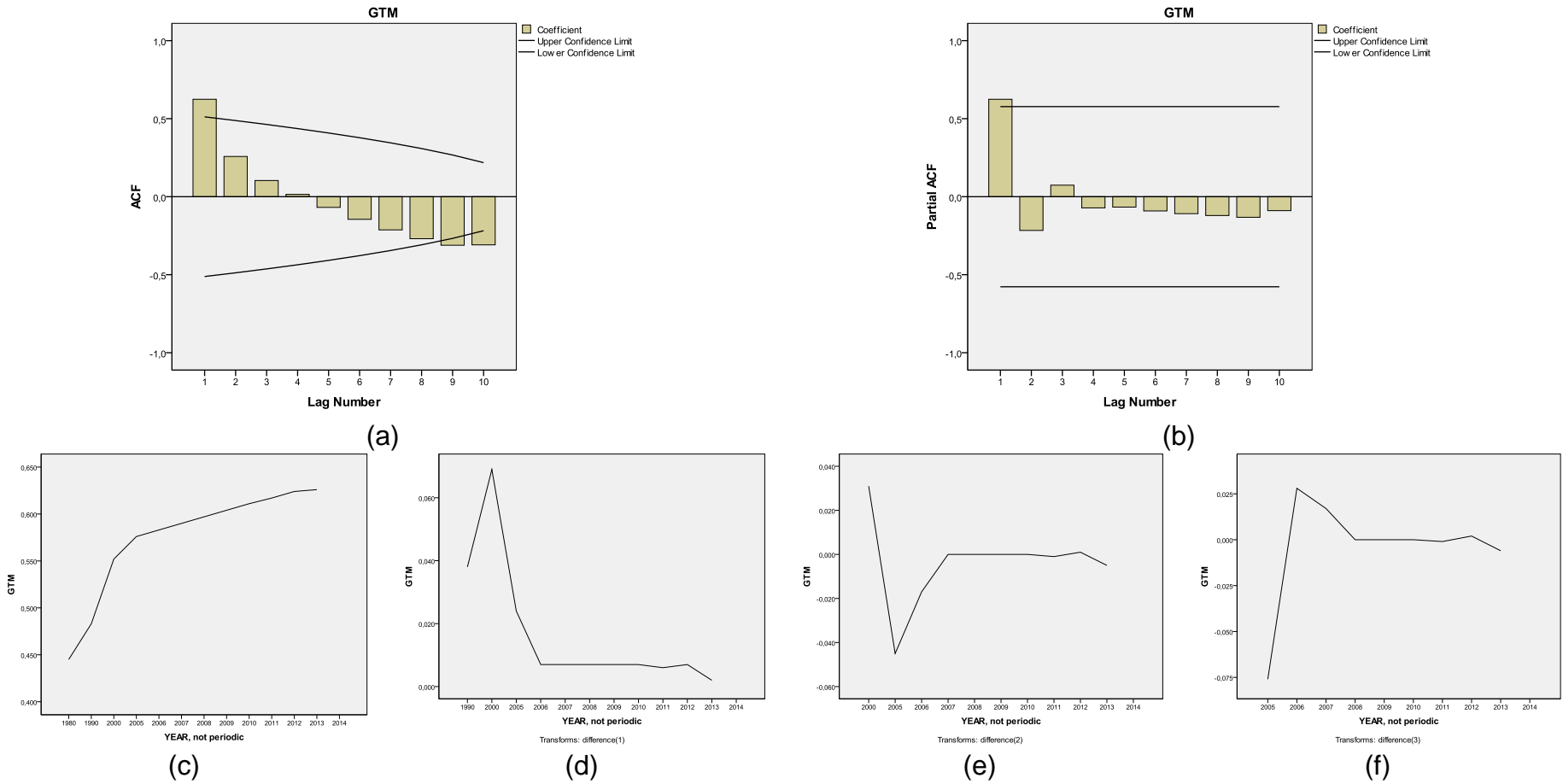


Figura 79 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA (1,1,1) para Previsão do IDH 2014 do Honduras (HND) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difenciações(e) e três diferenciações (f)

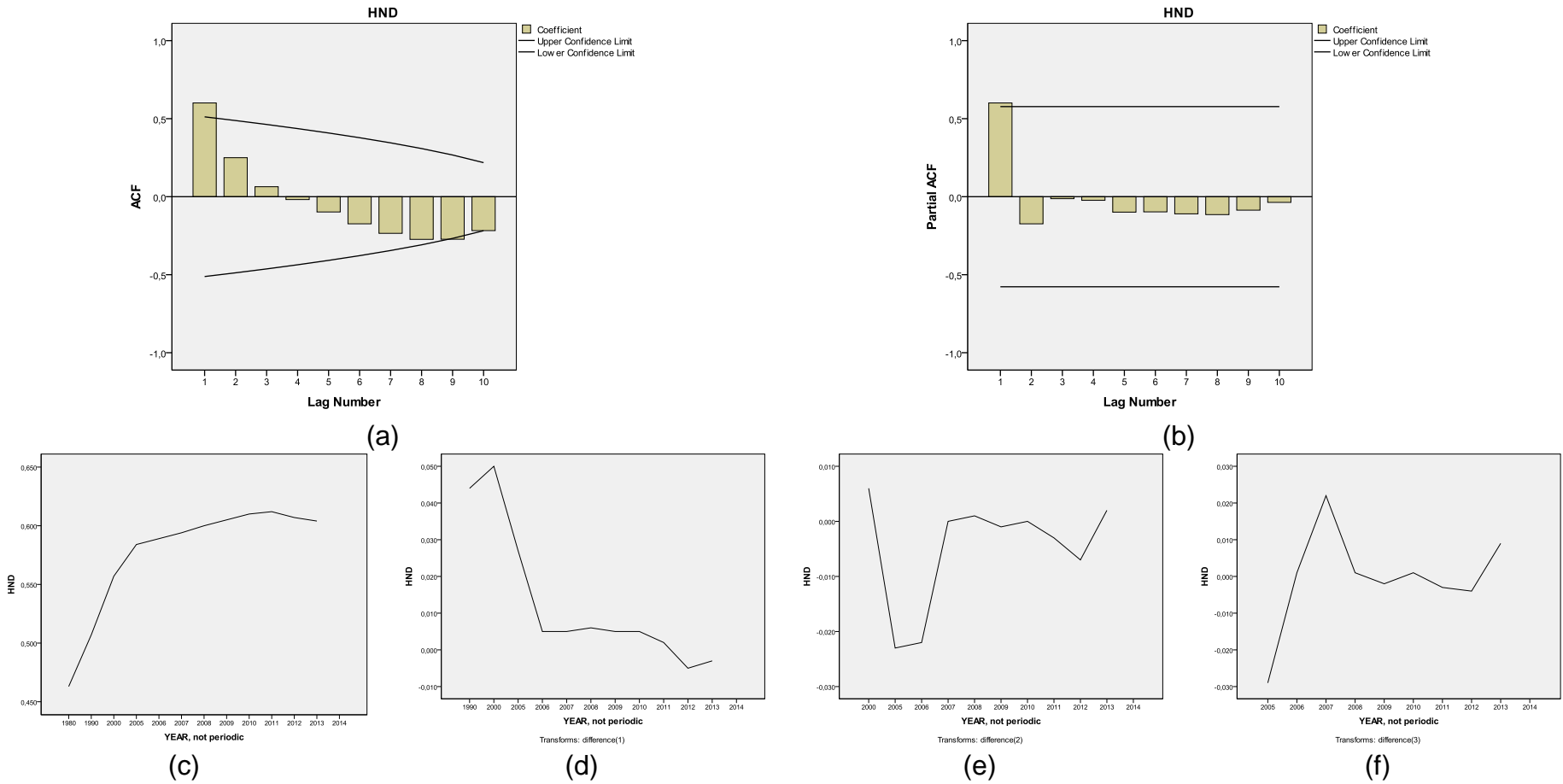
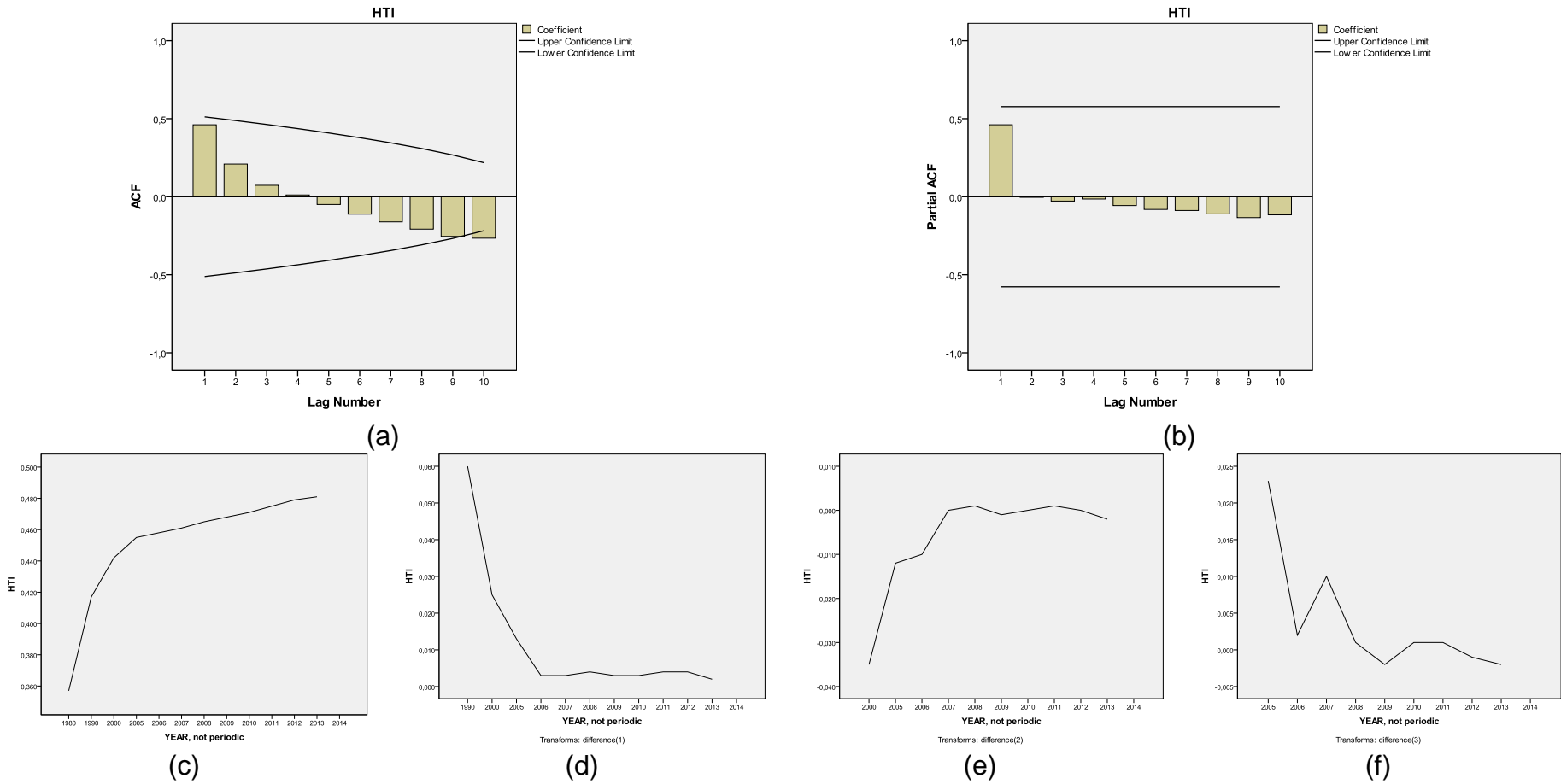


Figura 80 - Correlogramas ACF(a) e PACF(b),do Modelo ARIMA(0,1,0) para Previsão do IDH 2014 do Haiti (HTI) e suas séries temporais: original(c) e transformada com: uma diferenciação(d), duas difincações(e) e três diferenciações (f)



**APÊNDICE 6** – Comparação das Previsões Realizadas com Metodos de  
Mineração de Dados e ARIMA

**Tabela 10 - Previsões do IHD 2013 E 2014 dos países da América Latina por modelos e suas tendências divulgadas pela UNDP**

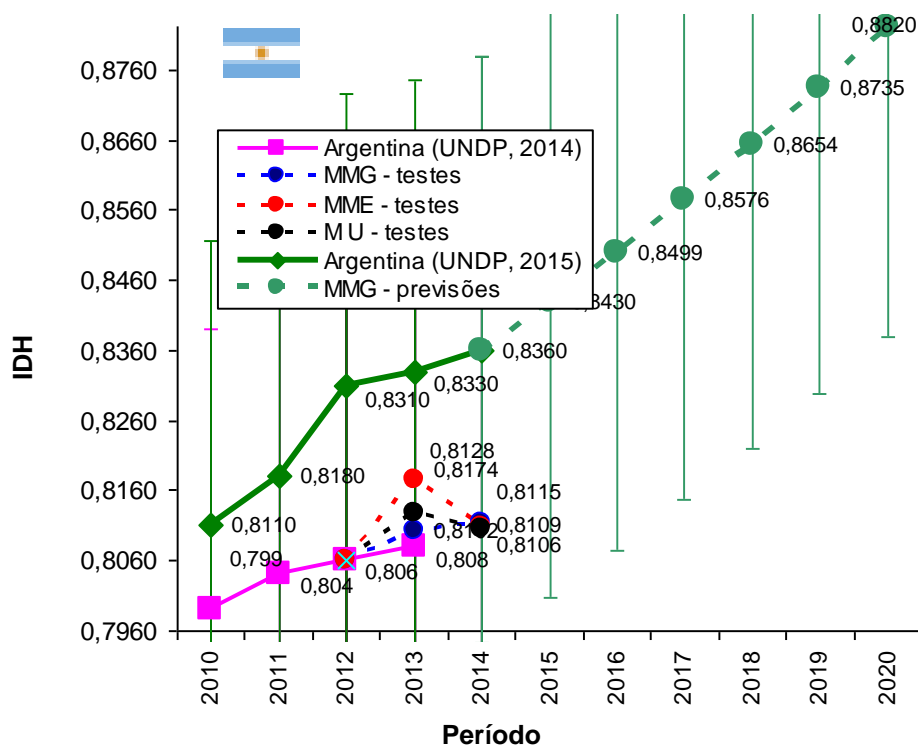
País	IDH								Tendências divulgadas (UNDP)		
	Previsões							ARIMA	**	***	***
	2013			2014							
	Modelos de MD			Modelo de MD							
MMG	MME	MU	ARIMA	MMG	MME	MU					
ARG	0.8102	0.8174	0.8128	0.819	0.8115	0.8109	0.8106	0.821	0.808	0.833	0.836
CHL	0.8249	0.8253	0.8213	0.824	0.8265	0.8272	0.8276	0.834	0.822	0.83	0.832
URY	0.791	0.7931	0.7891	0.801	0.7949	0.7934	0.7932	0.8	0.79	0.79	0.793
PAN	0.764	0.7663	0.7651	0.789	0.7693	0.7708	0.7691	0.796	0.765	0.777	0.78
CUB	0.8097	0.8104	0.8097	0.789	0.8157	0.8228	0.8197	0.787	0.815	0.768	0.769
CRI	0.7671	0.7671	0.7611	0.781	0.7668	0.7687	0.7667	0.791	0.763	0.764	0.766
VEN	0.765	0.7672	0.7742	0.76	0.7656	0.7689	0.7749	0.753	0.764	0.764	0.762
MEX	0.7594	0.76	0.7631	0.784	0.759	0.7578	0.7652	0.789	0.756	0.755	0.756
BRA	0.7455	0.7507	0.7486	0.743	0.7476	0.7501	0.7488	0.743	0.744	0.752	0.755
PER	0.7402	0.7385	0.7404	0.751	0.7421	0.7413	0.7436	0.757	0.737	0.732	0.734
ECU	0.712	0.7107	0.7125	0.733	0.7144	0.7144	0.7161	0.732	0.711	0.73	0.732
COL	0.7098	0.7157	0.7174	0.727	0.714	0.7132	0.7188	0.727	0.711	0.718	0.72
BLZ	0.7409	0.7371	0.7402	0.705	0.7378	0.7404	0.7369	0.709	0.732	0.715	0.715
DOM	0.7026	0.7017	0.7041	0.709	0.7043	0.7036	0.7051	0.711	0.7	0.711	0.715
PRY	0.6706	0.6703	0.6797	0.675	0.68	0.6824	0.6824	0.677	0.676	0.677	0.679
SLV	0.6646	0.6623	0.6629	0.691	0.6656	0.666	0.6637	0.694	0.662	0.664	0.666
BOL	0.6675	0.6658	0.6653	0.684	0.672	0.6691	0.6697	0.692	0.667	0.658	0.662
GUI	0.637	0.6388	0.6392	0.648	0.6422	0.6431	0.6444	0.657	0.638	0.634	0.636
NIC	0.6148	0.6153	0.6139	0.605	0.6177	0.6187	0.6164	0.618	0.614	0.628	0.631
GTM	0.6327	0.6332	0.6322	0.59	0.6335	0.6392	0.6374	0.588	0.628	0.626	0.627
HND	0.6189	0.6185	0.6273	0.641	0.6197	0.6223	0.6313	0.636	0.617	0.604	0.606
HTI *	0.4732	0.4728	0.4743	0.466	0.4744	0.4739	0.476	0.468	0.471	0.481	0.483

\* outliers, \*\* UNDP, 2014 (UNDP, 2014) e \*\*\* UNDP, 2015 (UNDP, 2015)

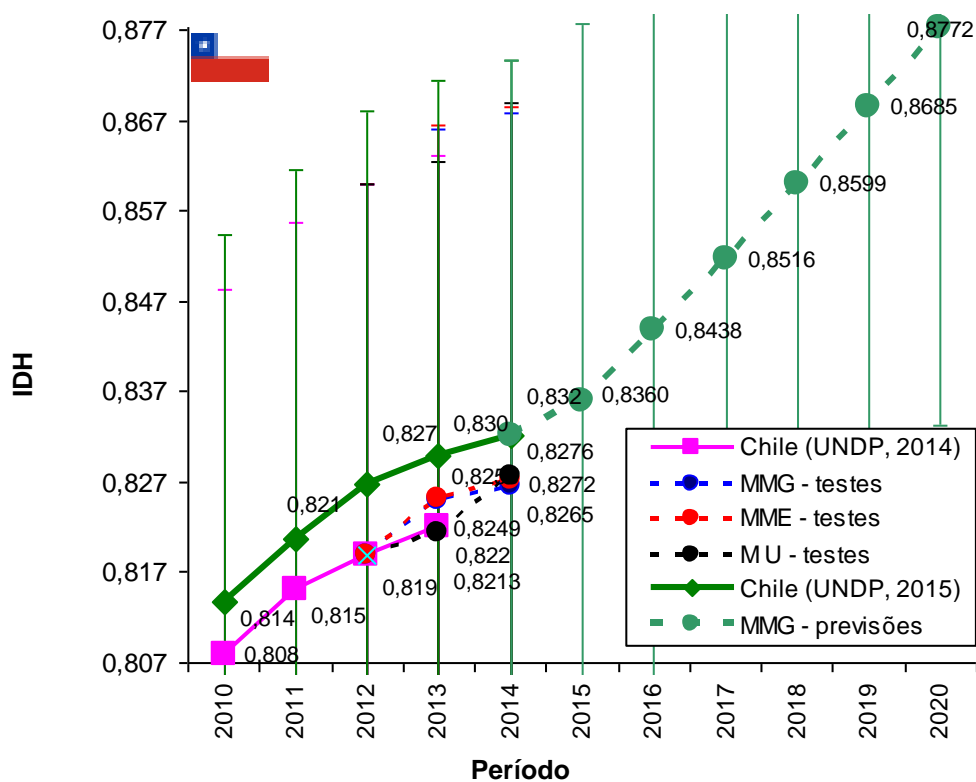


**APÊNDICE 7 – IDH dos Países da América Latina: Testes de Previsões**  
por Modelos (MMG, MME E MU) para 2013 e 2014, Tendência por  
Relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a  
2020

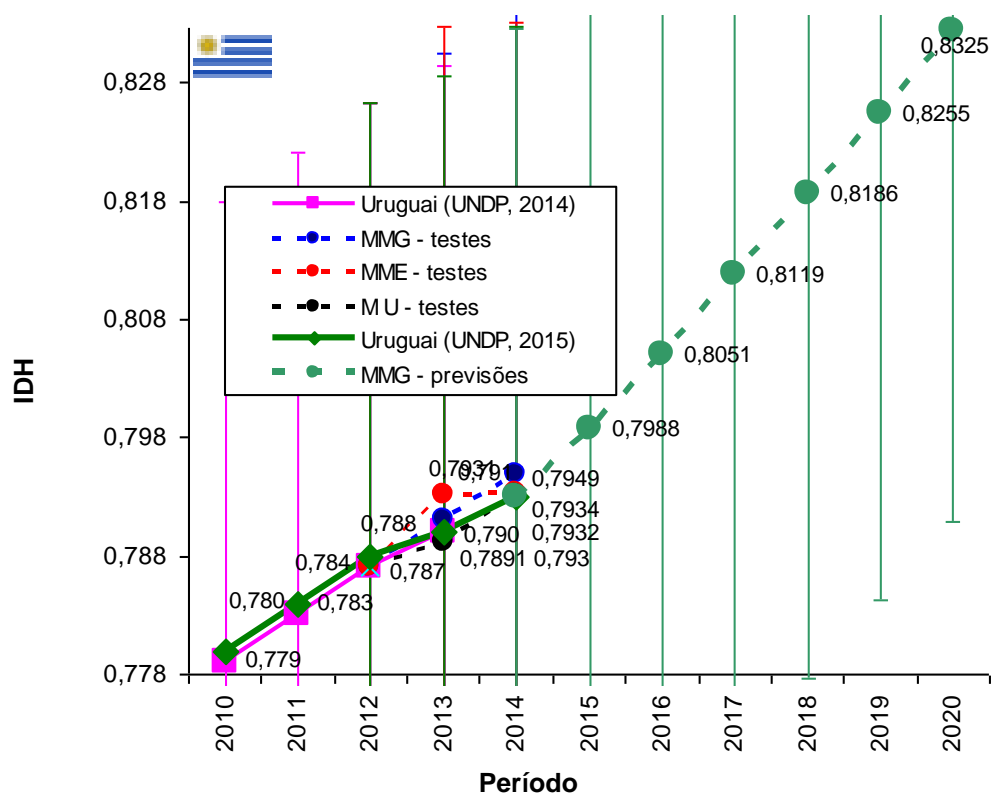
**Figura 81 - IDH da Argentina: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



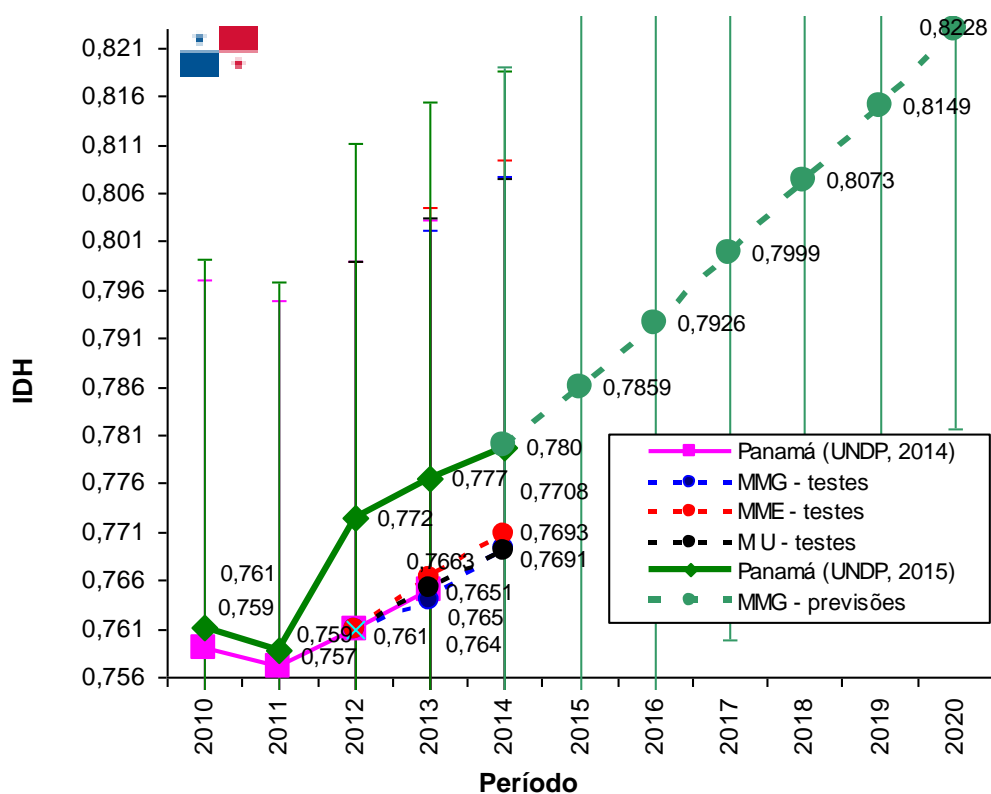
**Figura 82 - IDH do Chile: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



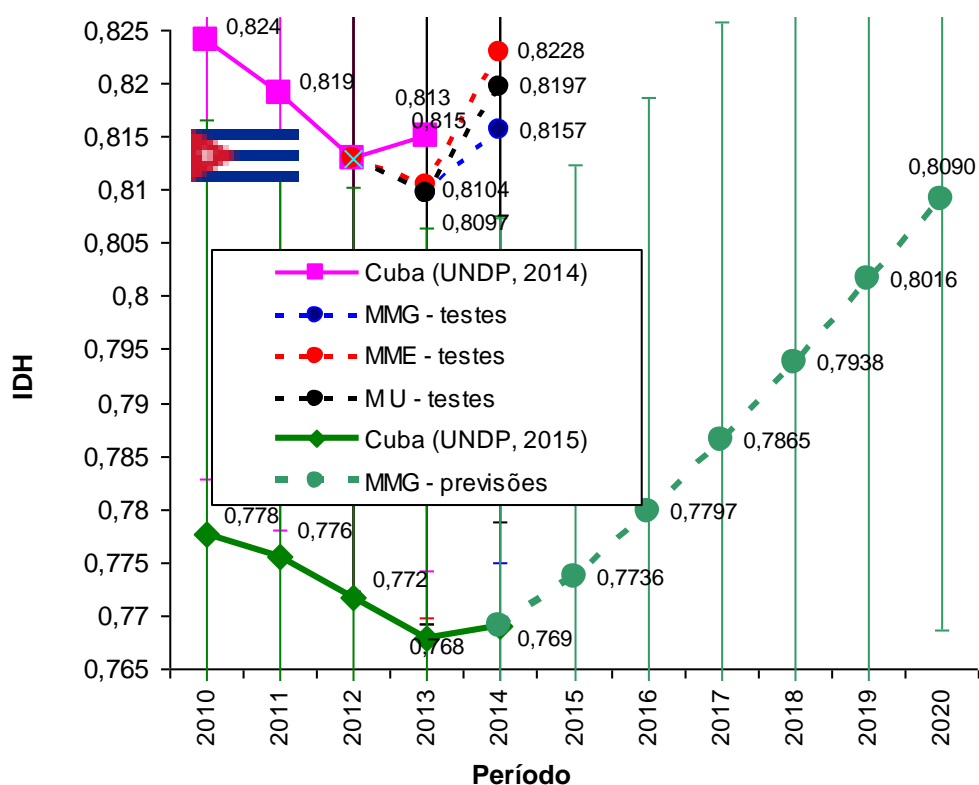
**Figura 83 - IDH do Uruguai: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



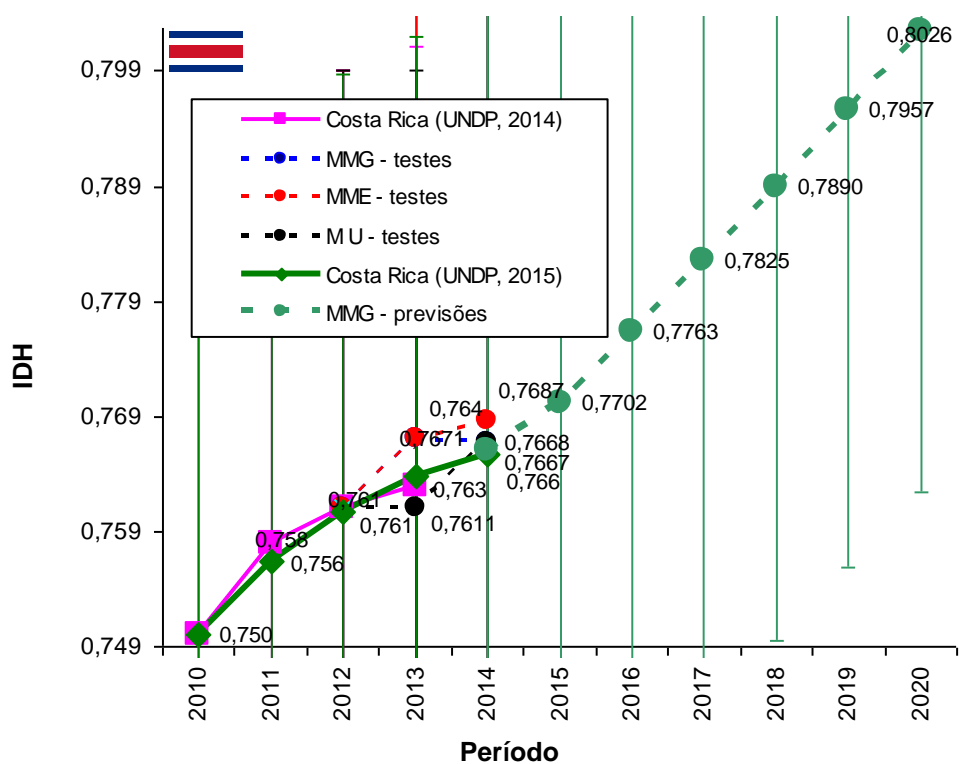
**Figura 84 - IDH do Panamá: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



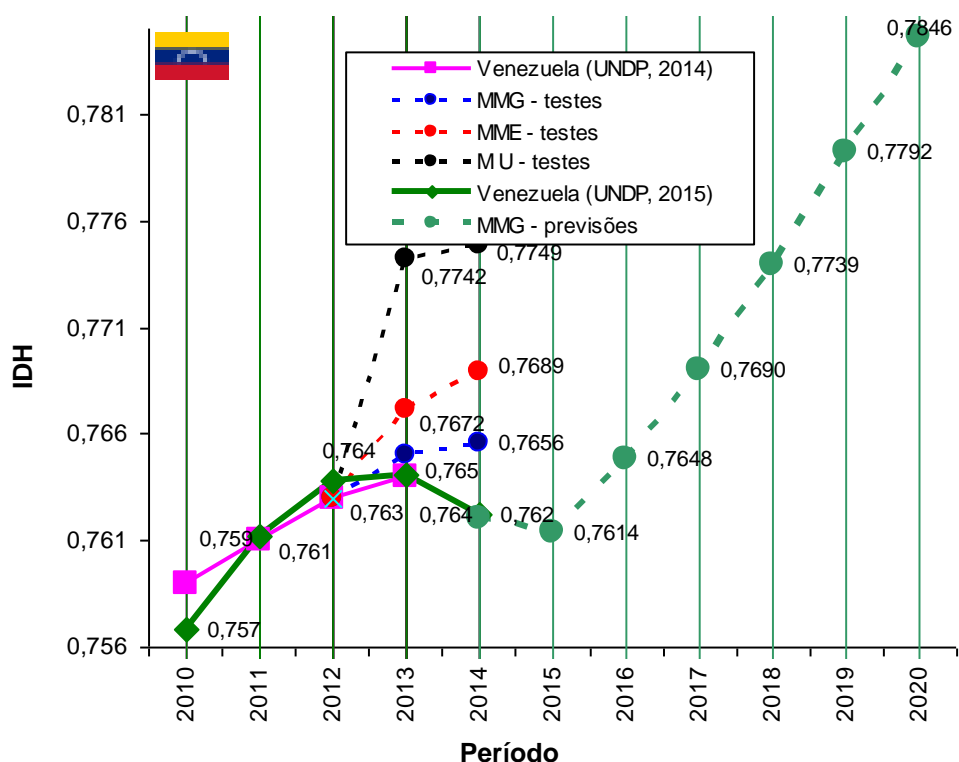
**Figura 85 - IDH de Cuba: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



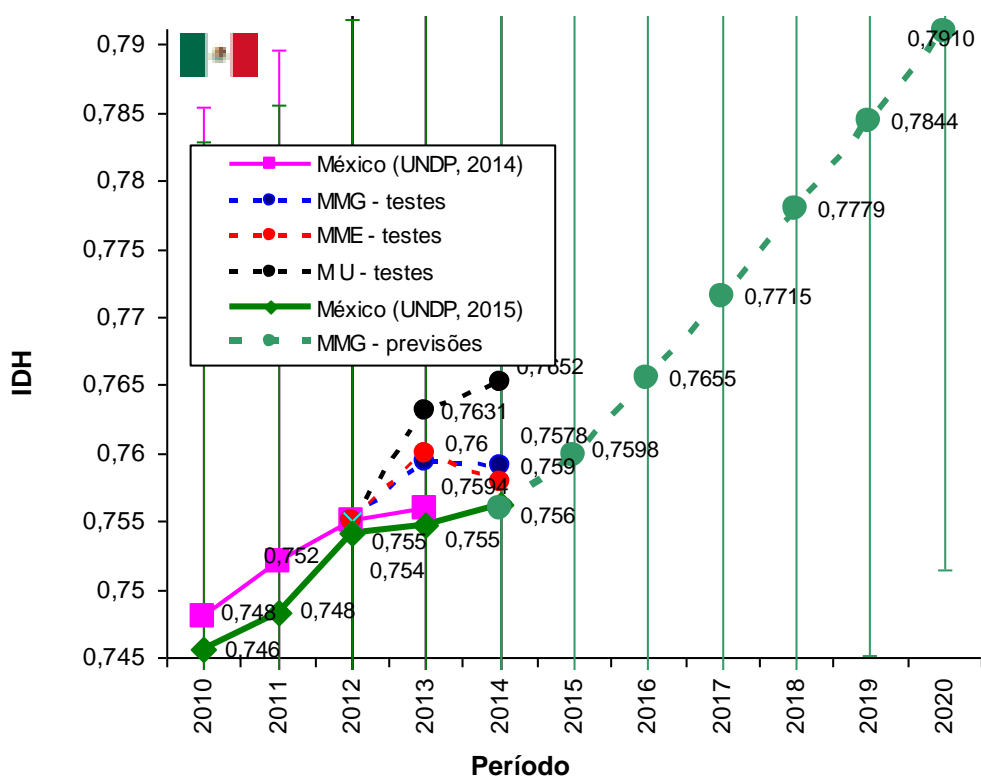
**Figura 86 - IDH de Costa Rica: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



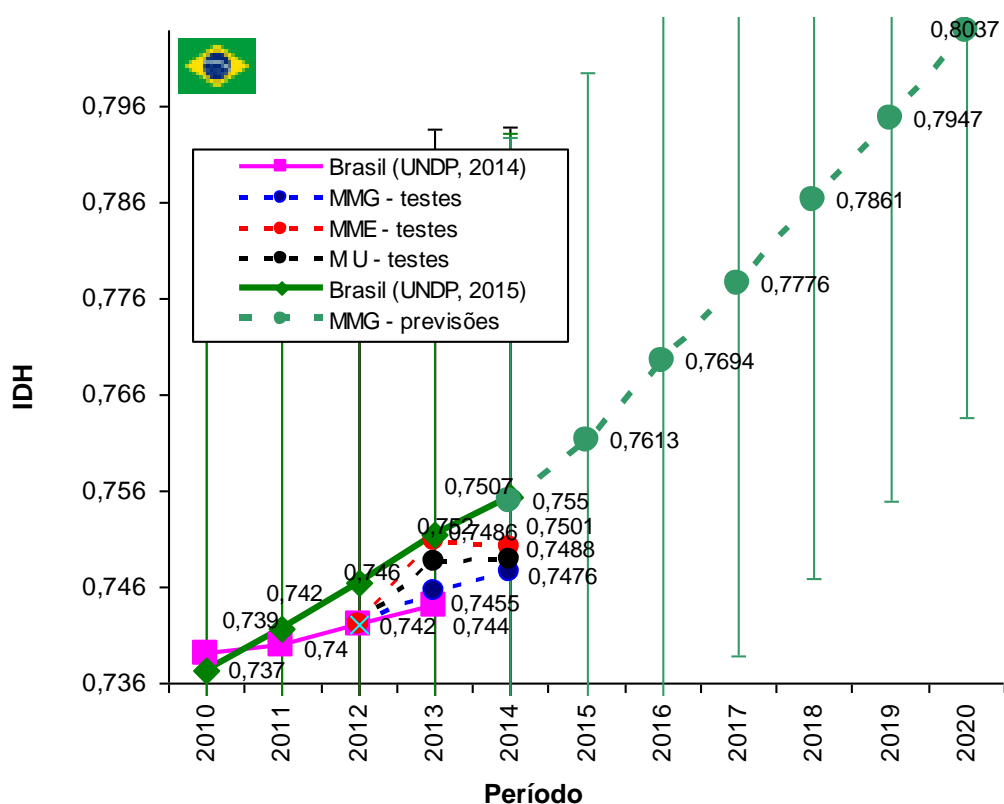
**Figura 87 - IDH da Venezuela: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



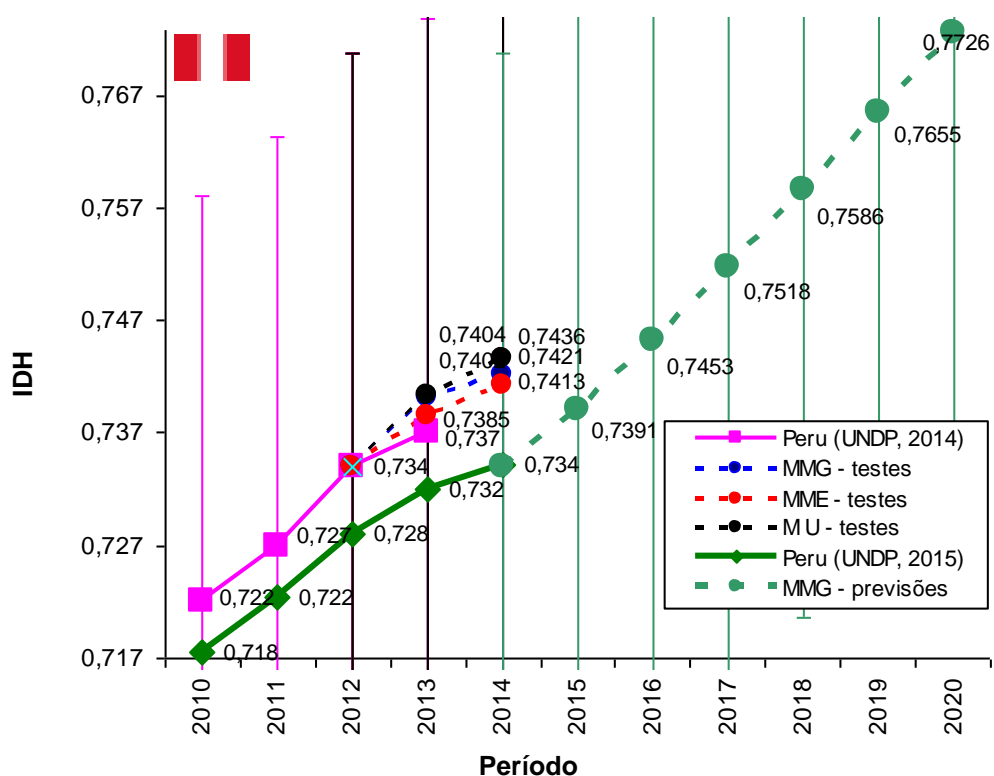
**Figura 88 - IDH do México: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



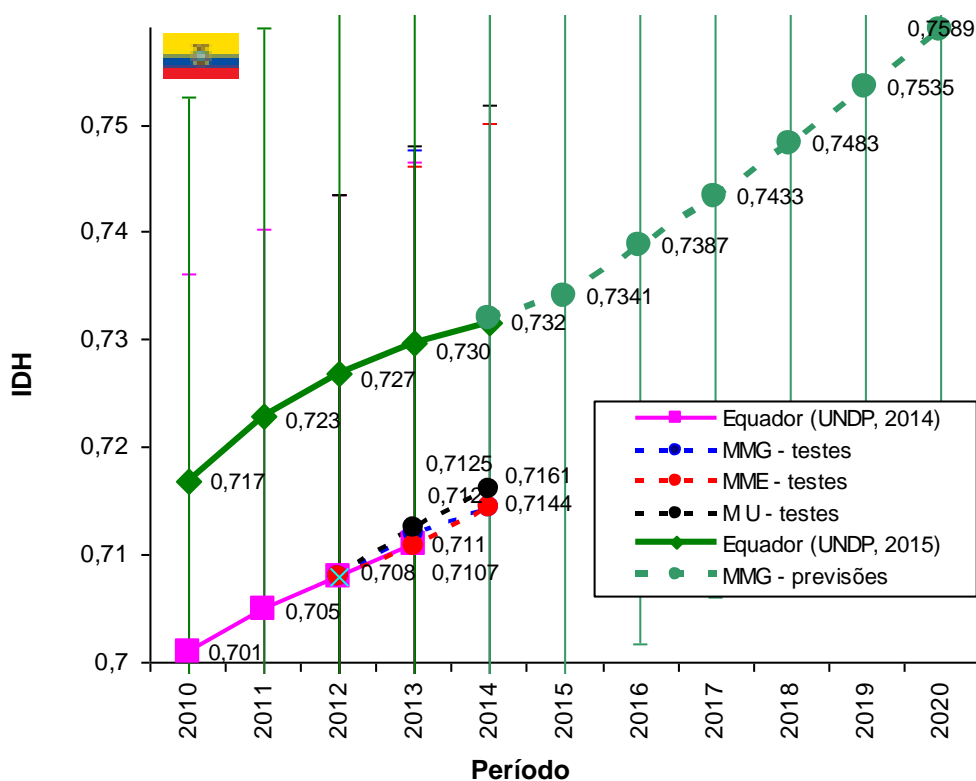
**Figura 89 - IDH do Brasil: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



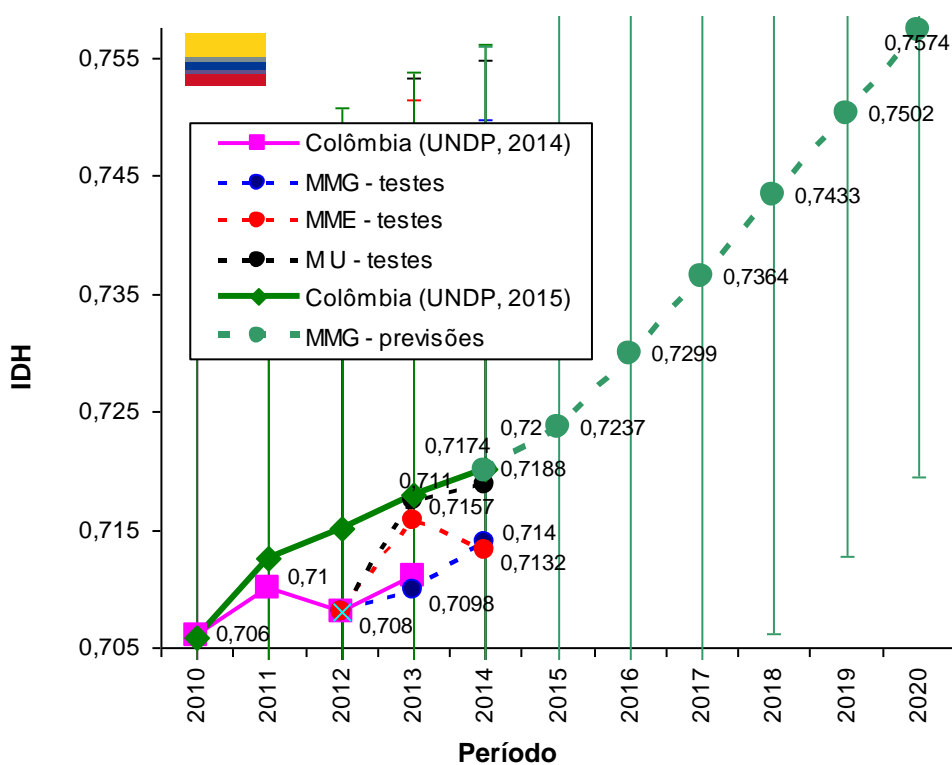
**Figura 90 - IDH do Peru: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



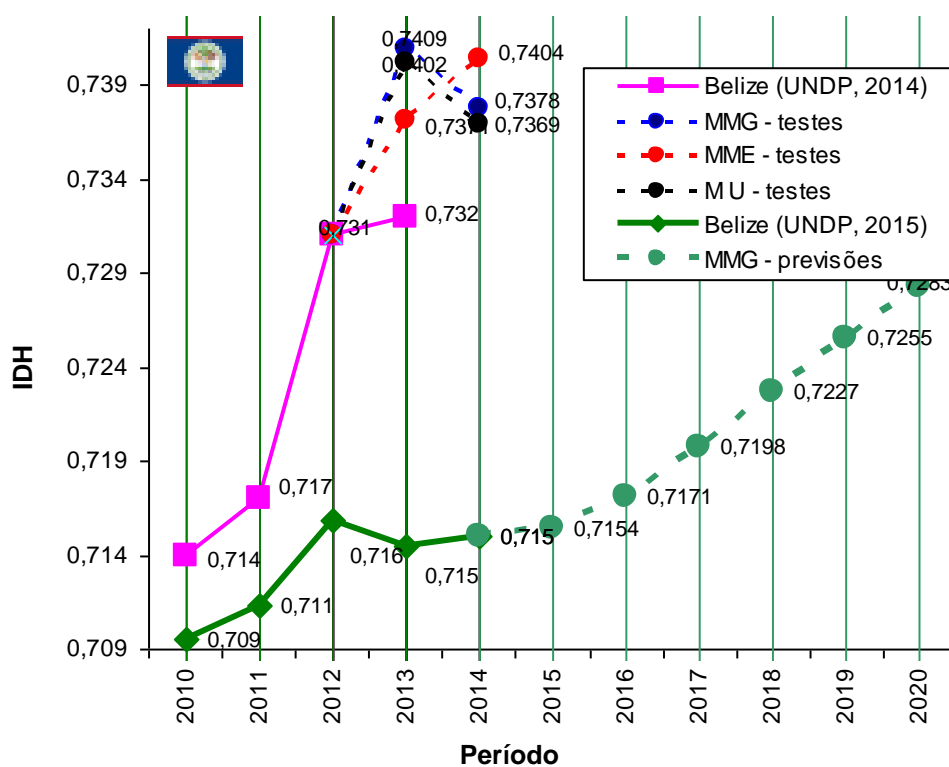
**Figura 91 - IDH do Equador: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



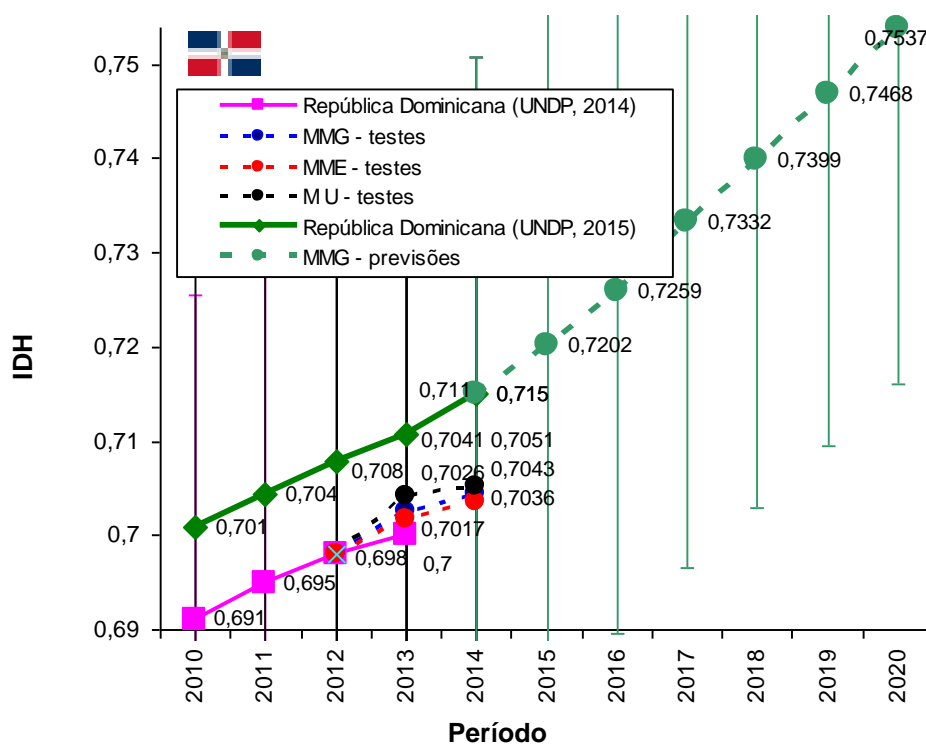
**Figura 92 - IDH da Colômbia: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



**Figura 93 - IDH de Belize: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**

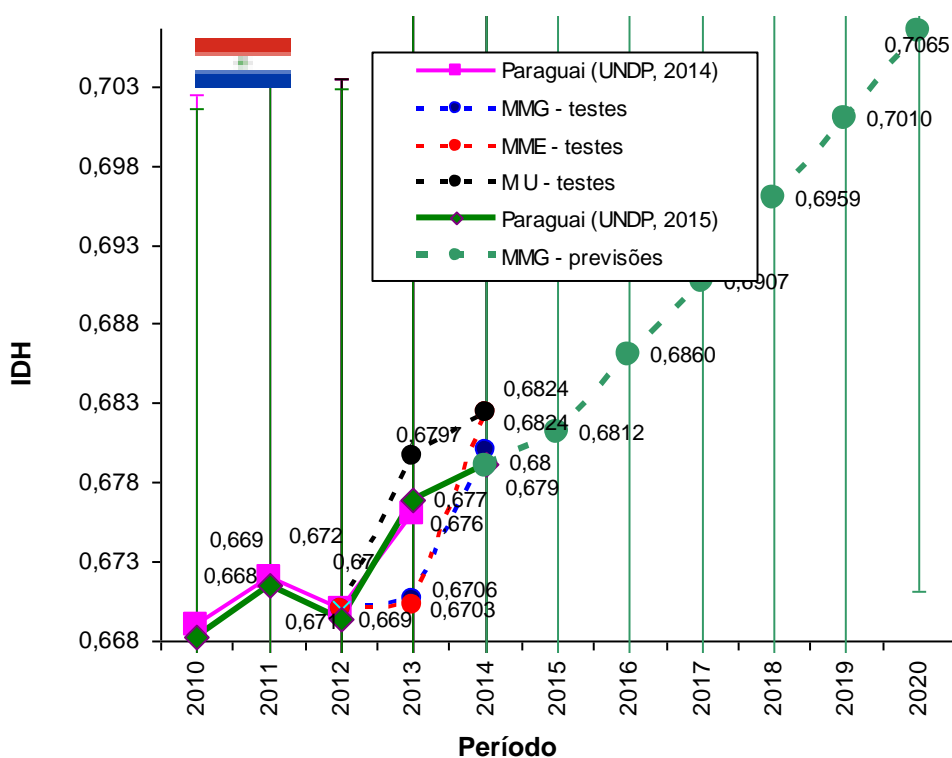


**Figura 94 - IDH da República Dominicana: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**

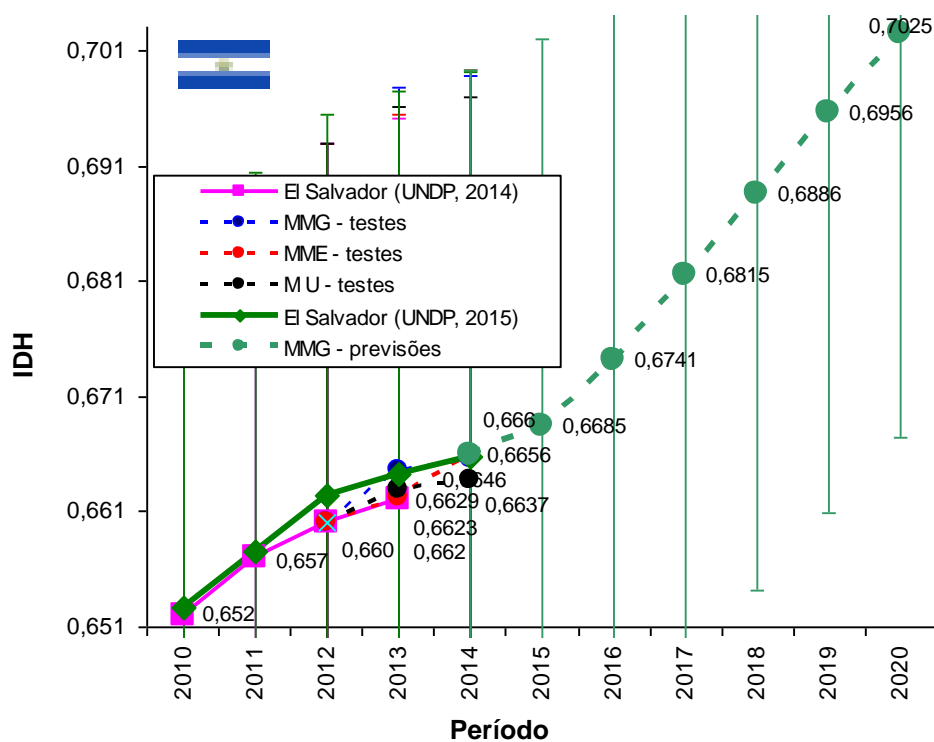




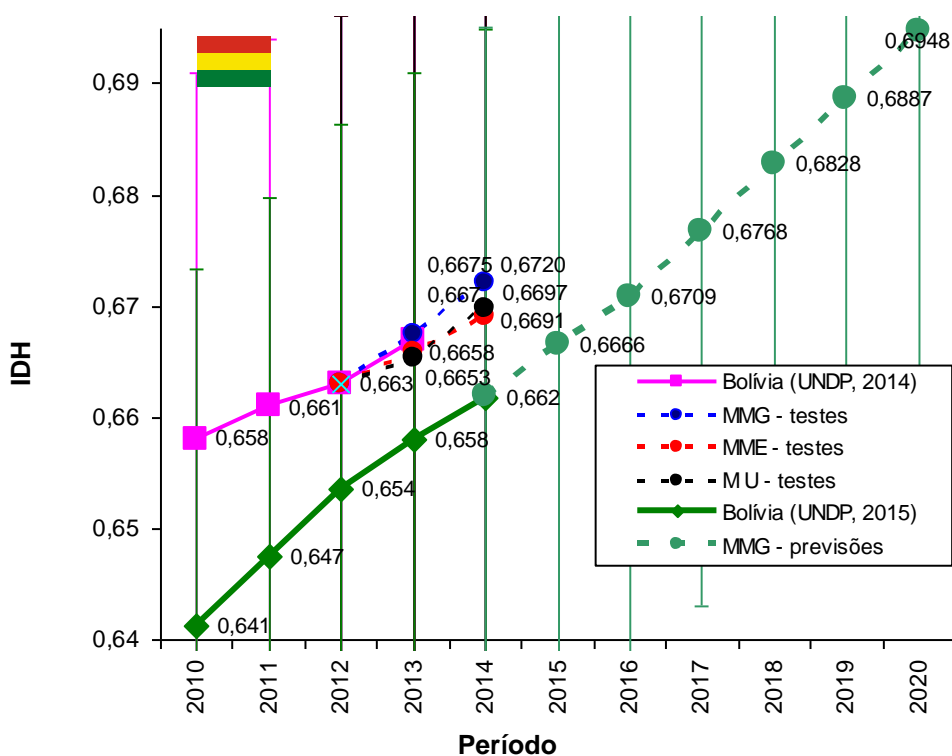
**Figura 95 - IDH do Paraguai: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



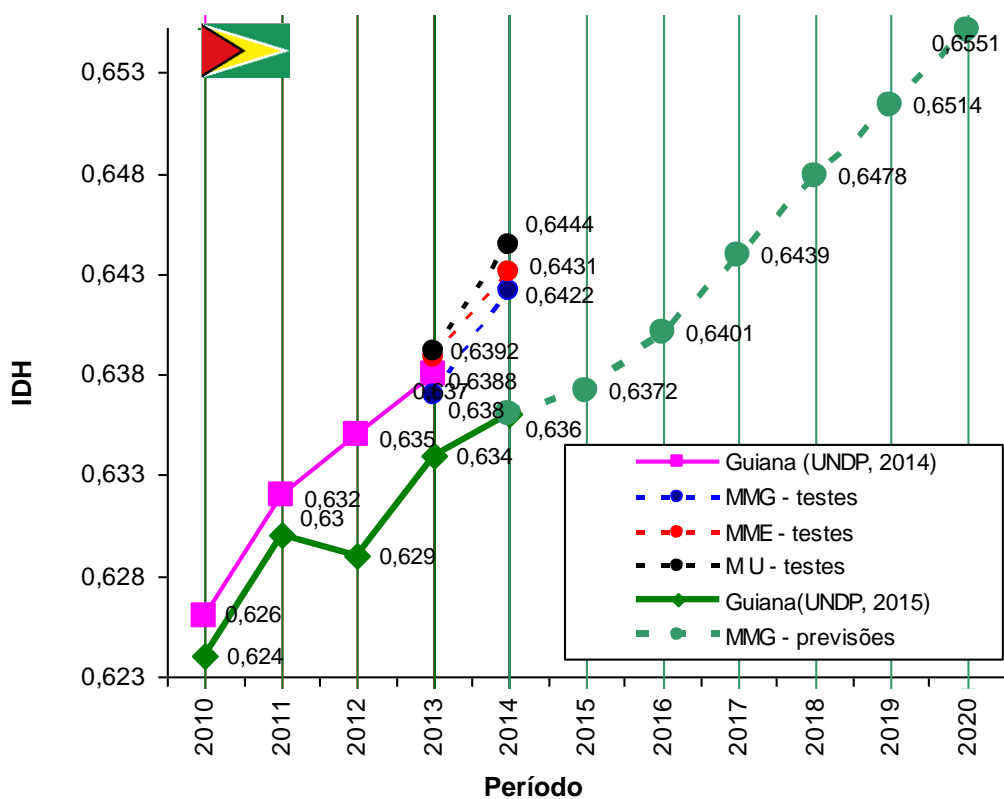
**Figura 96 - IDH de El Salvador: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



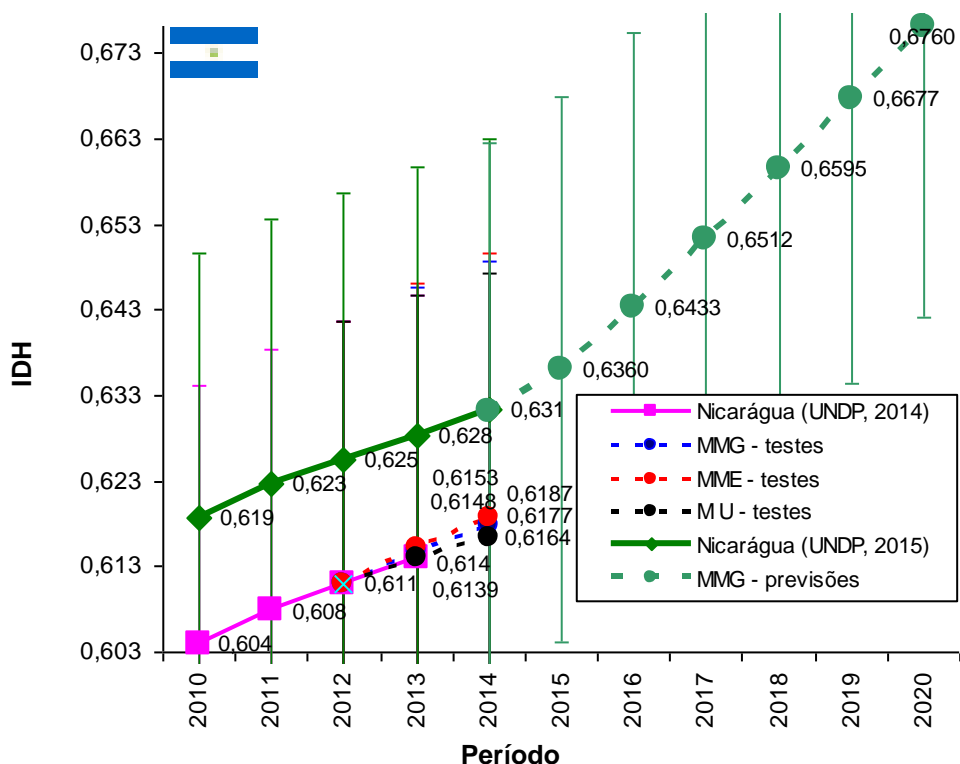
**Figura 97 - IDH da Bolívia: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



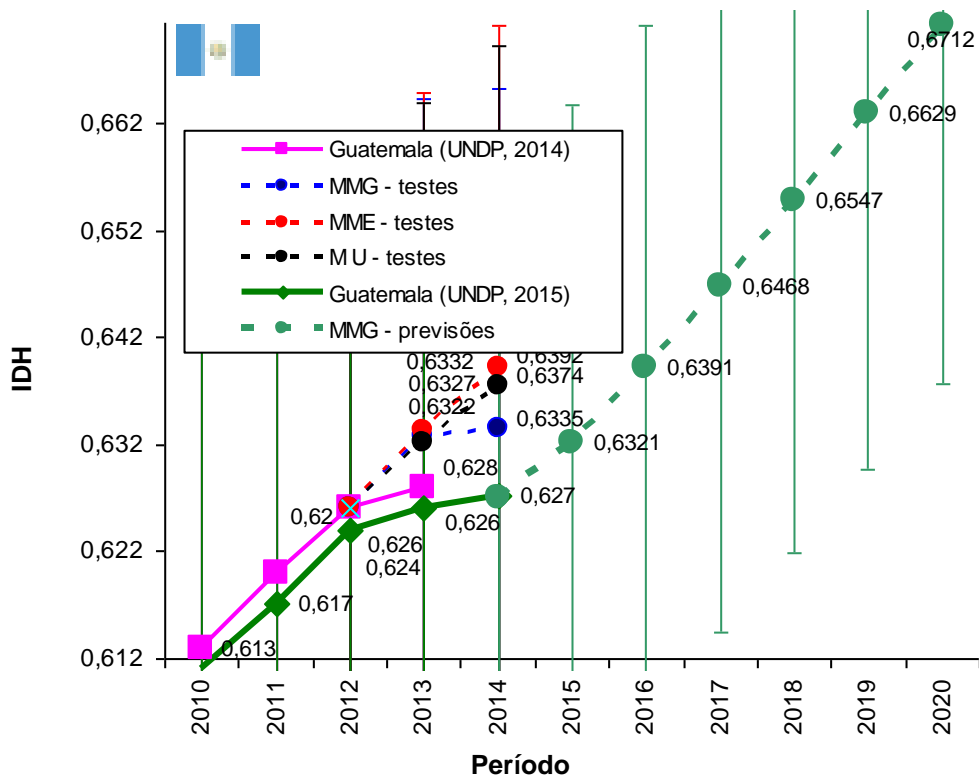
**Figura 98 - IDH da Guiana: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



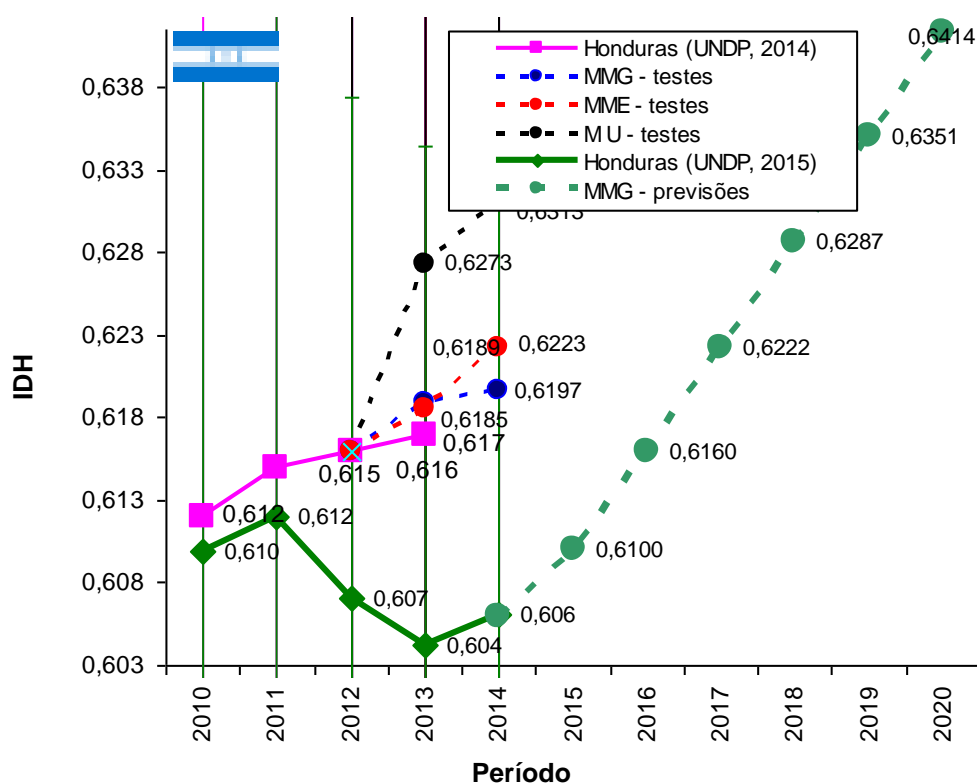
**Figura 99 - IDH da Nicarágua: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



**Figura 100 - IDH de Guatemala: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



**Figura 101 - IDH de Honduras: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**



**Figura 102 - IDH do Haiti: Testes de previsões por modelos (MMG, MME e MU) para 2013 e 2014, tendência por relatório do UNDP (2013 e 2014) e Previsões do MMG para 2015 a 2020**

