

Modelagem do número de patentes nos países americanos via Regressão múltipla

Diallison Teixeira de Carvalho¹

Luiz Alberto Beijo²

Eduardo Gomes Salgado³

1 – Introdução

A análise de regressão linear múltipla é uma técnica estatística que possibilita avaliar a relação entre vários fatores sobre uma variável principal, descrevendo tal relação por meio de um modelo. Este permite realizar previsões da variável de interesse utilizando valores dos fatores em estudo (TRIOLA, 2005).

A patente é uma forma de proteção dos direitos autorais sobre um produto. O número de patentes pode ser utilizado como um indicador de desenvolvimento dos países. Nos últimos anos houve um aumento significativo no número de patentes em vários países do continente americano, porém, não se sabe quais fatores influenciam nesse aumento.

Busca ajustar-se um modelo de regressão linear múltipla que possibilite avaliar se os seguintes fatores influenciam significativamente no número de patentes (NP) em alguns países do continente americano: número de artigos publicados (NART), Produto Interno Bruto (PIB), população (POP), variação do número de patentes em relação ao ano anterior (NPT₁) e variação do número de patentes em relação aos 2 anos anteriores (NPT₂).

2 – Metodologia

Os dados são referentes ao período de 1997 a 2012 (16 observações). O número de patentes depositadas pelos países do continente americano nesse período foi obtido a partir da Organização Mundial de Propriedade Intelectual (WIPO). Os dados sobre o Produto Interno Bruto (PIB) e a população foram obtidos junto ao Banco Mundial. O número de publicações científicas dos países no período mencionado foram obtidos na página do *SCImago Journal & Country Rank*.

1- Graduando em biotecnologia. UNIFAL. Correio eletrônico: diallison-carvalho@hotmail.com

2- Estatístico. Pesq. UNIFAL. Correio eletrônico: luizbeijo@yahoo.com.br

3- Engenheiro de produção. Pesq. UNIFAL. Correio eletrônico: egsalgado@yahoo.com.br

Os dados obtidos para o número total de patentes e o número de artigos são analisados em milhares, a população em milhões e o PIB em bilhões.

Estabeleceu-se como critério de inclusão para análise os países que apresentassem o número de patentes >1000 em pelo menos um dos últimos três anos do período analisado.

Para a pré-seleção dos fatores que iriam compor o ajuste do modelo, foi realizada a análise de correlação de Pearson (r), e posteriormente ajustado um modelo com todos os fatores. Os fatores que apresentaram efeito significativo ao nível de 5% pelo teste t foram investigados mais detalhadamente.

O modelo final foi selecionado, entre os que satisfizeram as pressuposições básicas do modelo de regressão, conforme o maior número de covariáveis significativas presentes no mesmo (de acordo com um teste t) e com base no maior coeficiente de determinação (R^2).

Para avaliar as pressuposições de normalidade e independência dos resíduos foi aplicado, respectivamente, o teste de Shapiro-Wilk e o teste de Box-Pearce.

Todas as análises foram realizadas considerando o nível de 5% de significância, utilizando o Sistema Computacional Estatístico R (R DEVELOPMENT CORE TEAM, 2013).

3 – Resultados e discussão

3.1 – Ajustes dos modelos de regressão múltipla

Ajustou-se um modelo contendo todos os cinco fatores em estudo e as informações sobre a significância (valor p) de cada um dos fatores nesse modelo para cada um dos países são dadas a seguir (Tabela 1).

Tabela 1: Avaliação dos fatores do modelo.

| Covariáveis | Estados Unidos | Brasil | Argentina | Canadá | México |
|---------------------|----------------|---------|-----------|---------|--------|
| Intercepto | 0,0667 | <0,0001 | 0,1012 | 0,2974 | 0,2920 |
| NART | 0,0027 | 0,1520 | 0,2095 | 0,0450 | 0,1843 |
| PIB | 0,5441 | 0,0750 | 0,9944 | 0,0138 | 0,1910 |
| POP | 0,0580 | <0,0001 | 0,2417 | 0,3629 | 0,9528 |
| NPt ₁ | 0,2792 | 0,1950 | 0,3763 | 0,1617 | 0,1859 |
| NPt ₂ | 0,2542 | <0,0001 | 0,0826 | 0,7085 | 0,0844 |
| p – valor do modelo | <0,0001 | <0,0001 | 0,0380 | <0,0001 | 0,0009 |

Baseado nos fatores que apresentaram efeito significativo ao nível de 5%, ajustaram-se vários modelos de regressão múltipla para cada país, sendo escolhido o modelo com maior coeficiente de determinação (R^2) entre os que satisfizeram as pressuposições básicas do modelo de regressão.

Os modelos selecionados são dados a seguir (Tabela 2), bem como a validação de cada modelo (Tabela 3).

Tabela 2: Estimativas dos modelos ajustados para cada país.

| Covariáveis | Estimativas de β | Erro padrão | p - valor |
|-----------------------|------------------------|-------------|-----------|
| Estados Unidos | | | |
| Intercepto | 898,715 | 330,504 | 0,0175 |
| NART | 0,054 | 0,007 | <0,0001 |
| POP | -4,214 | 1,444 | 0,0120 |
| Canadá | | | |
| Intercepto | 3,875 | 1,012 | 0,0028 |
| NART | 0,006 | 0,002 | 0,0349 |
| PIB | 0,139 | 0,056 | 0,0298 |
| NPt ₁ | 0,372 | 0,177 | 0,0599 |
| Brasil | | | |
| Intercepto | 2,648 | 0,121 | <0,0001 |
| NART | -0,001 | 0,0003 | 0,0019 |
| PIB | 0,108 | 0,011 | <0,0001 |
| NPt ₁ | 1,001 | 0,265 | 0,0030 |
| Argentina | | | |
| Intercepto | 1,162 | 0,028 | <0,0001 |
| NPt ₂ | 0,506 | 0,130 | 0,0047 |
| México | | | |
| Intercepto | -0,058 | 0,165 | 0,731 |
| PIB | 0,114 | 0,015 | <0,0001 |
| NPt ₂ | 0,0002 | 0,0002 | 0,234 |

Estimativas do modelo para os Estados Unidos.

$$NP = 0,054 (NART) - 4,214(POP) + 898,715 \quad (1)$$

De acordo com o modelo ajustado (1), nos Estados Unidos o número de patentes é influenciado positivamente pelo número de artigos publicados e negativamente pelo número de habitantes (POP), ou seja, o número de patentes aumenta com o aumento do número de artigos publicados e diminui com o aumento da população.

Estimativas do modelo para o Canadá.

$$NP = 0,006(NART) + 0,139(PIB) + 0,372 (NPt_1) + 3,875 \quad (2)$$

De acordo com o modelo ajustado (2), no Canadá o número de patentes é influenciado positivamente pelo aumento do número de artigos publicados e pelo aumento do PIB.

Embora presente no modelo, o fator NPt₁ não influencia no número de patentes (NP) (p= 0,0599).

Estimativas do modelo para o Brasil.

$$NP = -0,001(NART) + 0,108(PIB) + 1,001(NP_{t-1}) + 2,648 \quad (3)$$

De acordo com o modelo ajustado (3), no Brasil o número de patentes é influenciado positivamente pelo PIB e pela variação do número de patentes em relação ao ano anterior e negativamente pelo número de artigos publicados.

Estimativas do modelo para a Argentina.

$$NP = 0,506(NPt_{.2}) + 1,162 \quad (4)$$

De acordo com o modelo ajustado (4), na Argentina o número de patentes é influenciado positivamente pela variação do número de patentes em relação aos 2 anos anteriores, indicando que algum fator (s) não estudado aqui interfere positivamente no número de artigos.

Estimativas do modelo para o México.

$$NP = 0,114(PIB) + 0,0002(NPt_{.2}) - 0,058 \quad (5)$$

De acordo com o modelo ajustado (5), no México o número de patentes é influenciado positivamente pelo PIB, ou seja, há um aumento do número de patentes com o aumento do PIB.

Embora presente no modelo, o fator $NPt_{.2}$ não influencia no número de patentes ($p=0,234$).

Para validação dos modelos de regressão múltipla ajustados, analisaram-se as pressuposições básicas do modelo de regressão múltipla, como a normalidade (Teste de Shapiro-Wilk) e independência dos resíduos (Teste de Box-Pearce). São também apresentados os valores p de cada modelo e seu R^2 (coeficiente de regressão múltipla, que indica quanto o modelo pode explicar sobre a variação do número de patentes) (Tabela 3).

Tabela 3: Validação dos modelos ajustados.

| Países | Teste de Shapiro | Teste de Box-Pearce | R^2 ajustado | p - valor |
|----------------|------------------|---------------------|----------------|-----------|
| Estados Unidos | 0,4489 | 0,8792 | 0,9861 | <0,0001 |
| Canadá | 0,7695 | 0,1331 | 0,9775 | <0,0001 |
| Brasil | 0,8734 | 0,9702 | 0,9733 | <0,0001 |
| Argentina | 0,6214 | 0,6333 | 0,6091 | 0,0047 |
| México | 0,7734 | 0,0721 | 0,8186 | <0,0001 |

Ao que se pode observar (Tabela 3), as pressuposições da análise de regressão múltipla não foram violadas, assim as conclusões obtidas a partir dos modelos são válidas.

4 – Conclusões

Baseado nas estimativas obtidas notou-se que o número de patentes é influenciado diferentemente conforme o país em estudo. Um fator que influencia positivamente no número de patentes em um país pode ser negativamente em outro, bem como não ter influência.

O número de patentes nos Estados Unidos é influenciado positivamente pelo aumento do número de artigos publicados e negativamente com o aumento da população.

O aumento do número de artigos publicados e do PIB influencia positivamente no número de patentes no Canadá.

No Brasil, o número de patentes é influenciado positivamente pelo aumento do PIB e pela variação do número de patentes em relação ao ano anterior e negativamente pelo número de artigos publicados.

Na Argentina sabe-se que o número de patentes é influenciado positivamente pela variação do número de patentes em relação aos dois anos anteriores, o que indica que outro(s) fator(s) não estudado(s) neste trabalho pode influenciar positivamente no número de patentes nesse país.

No México, um aumento do PIB influencia positivamente no número de patentes.

REFERÊNCIAS BIBLIOGRÁFICAS

BANCO MUNDIAL. Disponível em: <<http://data.worldbank.org>>. Acesso em: 30 jan. 2014.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. Disponível em <<http://www.r-project.org>>. Acessado em 30 de janeiro de 2014.

SCIMAGO JOURNAL & COUNTRY RANK. Disponível em: <<http://www.scimagojr.com>>. Acesso em: 30 de janeiro de 2014.

TRIOLA, M. F. **Introdução à estatística**. 9 ed. Rio de Janeiro: LTC, 2005

WORLD INTELLECTUAL PROPERTY ORGANIZATION (WIPO). Disponível em: <<http://www.wipo.org>>. Acesso em: 30 de janeiro de 2014.