New Challenges for Statistical Software - The Use of R in Official Statistics

INFLUENCE FACTORS OF THE ECONOMIC DEVELOPMENT LEVEL ACROSS EUROPEAN COUNTRIES

April 7-8, 2016, Bucharest

Diana Ioana POPA

National Institute of Statistics Romania
diana.popa@insse.ro
Overview

- GDP per capita - a measure of a country’s economic development level
- Factors of influence: various, from social and economical to environmental and government policies
- Aim of research: to discover through statistical tools available in R potential factors
- Data from 31 European countries was used in the interest years 2008 and 2013
Methods

- Multiple Linear Regression models were applied
- Goal: to explain the relationship between GDP per capita and certain independent variables
- 11 independent variables were chosen for testing
The relationship between GDP per capita vs. the independent variables, in 2008 and 2013

Source data: Eurostat
Variables of the model

- Dependent variable: GDP per capita

- The best fitted model contains three independent variables:
  - Unemployment rate (15-64 years)
  - Share of the working age population with tertiary education
  - Life expectancy at birth
Analysys

- The multiple regression equation of the best fitted model:

\[ gdp\_cap_i = \beta_0 + \beta_1 \times unempl\_rate_i + \beta_2 \times sh\_tert_i + \beta_3 \times life\_ex_i + \varepsilon_i \]

- Two equations were computed initially, one for each year of interest

- Luxembourg – outlier; two more equations were computed after eliminating it
# Coefficient estimates for Model 3 (2008) and Model 4 (2013)

<table>
<thead>
<tr>
<th>Variables / Year</th>
<th>Coeff. / Std. error</th>
<th>P-value</th>
<th>Confidence interval</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>-178170.6 (38001.3)</td>
<td>7.65e-05</td>
<td>-256283.3340 -100057.7691</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>-231050.8 (50556.4)</td>
<td>0.000105</td>
<td>-334970.9131 -127130.6638</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployment rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>-1690.8 (733.8)</td>
<td>0.02945</td>
<td>-3199.0499 -182.5346</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>-1294.6 (307.4)</td>
<td>0.000268</td>
<td>-1926.3737 -662.8098</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share of tertiary</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>736.7 (232.1)</td>
<td>0.00384</td>
<td>259.6531 1213.7082</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>812.8 (266.3)</td>
<td>0.005188</td>
<td>265.3323 1360.2085</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Life expectancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2008</td>
<td>2523.5 (485.9)</td>
<td>2.02e-05</td>
<td>1524.8130 3522.2138</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2013</td>
<td>3141.0 (656.8)</td>
<td>5.97e-05</td>
<td>1791.0197 4490.9592</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Significance level

- $R^2$
  - 2008: 0.7495 (p value - 5.606e-08)
  - 2013: 0.7276 (p value - 1.645e-07)

- The significance level and $R^2$ of both models are similar in both years studied.
Conclusions

- As level, the coefficients estimates are different in the years studied.

- The relationship type between the dependent and independent variables - the same in both years of interest.
References


THANK YOU!