Statistical data analysis via R and PHP: A case study of the relationship between GDP and foreign direct investments for the Republic of Moldova

Ştefan Cristian CIUCU

PhD Candidate, Cybernetics and Statistics Doctoral School, The Bucharest University of Economics Studies, Romania (stefanciucu@yahoo.com).

2014
Introduction

- R is a language and environment for statistical computing and graphics. It is a GNU project (free software);
- PHP is a popular general-purpose scripting language that is especially suited to web development;
- The idea of integrating R with PHP has been around for a few years => R-PHP software developed within the Department of Statistical and Mathematical Sciences of the University of Palermo (Italy) – (Mineo, Pontillo, 2006).
The use of PHP in statistics

- Mathematical extensions that include numerous functions:
  - “Math” (with basic Mathematical Functions),
  - “Statistics” – a statistics extension that contains functions for statistical computations
  - “Trader” – Technical Analysis for Traders which contains some functions for linear regressions

Examples of statistical functions of PHP:
- stats_absolute_deviation; stats_cdf_f; stats_cdf_t; stats_standard_deviation; stats_variance
An example of PHP for statistics implementation

```php
function correlation($x, $y)
{
    $PPMCC = stats_stat_correlation($x, $y);
    echo "Pearson product-moment correlation coefficient is " .
         $PPMCC;
}
$array_x = array(5,3,6,7,4,2,9,5);
$array_y = array(4,3,4,8,3,2,10,5);
correlation($array_x, $array_y);
?>
```
In this article we will take a look at a way to access R via PHP already implemented, with a software called R–PHP.

R–PHP implements **two modules**:
- The first module allows the simple insertion of the R code and it prints its output (analyses and plots) in another page.
- The second module makes some statistical analysis by using a GUI.
R-PHP module no. 1

File upload

Command line box

Select a file from your computer:
Browse... No file selected.

Header:
- yes
- no

Send  Reset

Send

R-PHP module no. 2
Data analysis using R–PHP

- This section has two major goals:
  - to show basic way of using R–PHP for data analysis;
  - to interpret data and draw some conclusions about the economic situation in the Republic of Moldova.
```r
# Load data from a tab-delimited text file
date.analizate <- read.table("a.tab.txt", header=TRUE)

# Attach data from file
attach(date.analizate)

# Print column names
names(date.analizate)

date.analizate
```

```
> date.analizate <- read.table("a.tab.txt", header=TRUE)
> attach(date.analizate)
> names(date.analizate)
[1] "Year"  "GDP"  "FDI"
> date.analizate
     Year  GDP     FDI
  1 1992 2319243407 17000000
  2 1993 2371812924 14000000
  3 1994 1702314353  11568000
  4 1995 1752995314  25910000
  5 1996 1695130484  23740000
  6 1997 1930071445   78740000
  7 1998 1639497207   75500000
  8 1999 1170785048   37890000
  9 2000 1288420223  127540000
 10 2001 1480656894   54540000
 11 2002 1661818168   84050000
 12 2003 1980901554   73750000
 13 2004 2598231467   87690000
 14 2005 2988172424  190700000
 15 2006 3408451498  259680000
 16 2007 4402495921  536020000
 17 2008 6054806101  726610000
 18 2009 5439422031  135150000
 19 2010 5811622394  201500000
 20 2011 7015201446  276420000
 21 2012 7252769934  184940000
> 
```
R-PHP – first module – code insertion

date.analizate<-read.table("a.tab.txt", header=TRUE)
summary(date.analizate)

> date.analizate<-read.table("a.tab.txt", header=TRUE)
> summary(date.analizate)

A tabular summary of the data is shown below:

<table>
<thead>
<tr>
<th></th>
<th>GDP</th>
<th>FDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1.171e+09</td>
<td>11568000</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1.695e+09</td>
<td>37890000</td>
</tr>
<tr>
<td>Median</td>
<td>2.319e+09</td>
<td>84050000</td>
</tr>
<tr>
<td>Mean</td>
<td>3.141e+09</td>
<td>153425619</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>4.402e+09</td>
<td>190700000</td>
</tr>
<tr>
<td>Max.</td>
<td>7.253e+09</td>
<td>726610000</td>
</tr>
</tbody>
</table>

>
R-PHP – first module – code insertion

date.analizate<-read.table("a.tab.txt", header=TRUE)
summary(lm(date.analizate$GDP~date.analizate$FDI))
```
> date.analizate <- read.table("a.tab.txt", header=TRUE)
> anova(lm(date.analizate$GDP ~ date.analizate$FDI))
```

Analysis of Variance Table

Response: date.analizate$GDP

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>date.analizate$FDI</td>
<td>1</td>
<td>3.1702e+19</td>
<td>3.1702e+19</td>
<td>12.607</td>
<td>0.002135 **</td>
</tr>
<tr>
<td>Residuals</td>
<td>19</td>
<td>4.7779e+19</td>
<td>2.5147e+18</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
The second module is quite interesting because no programming skills in R are required.

All of the commands that were ran in section 4.1. can be run in this module using the appropriate command buttons in the upper part of the Web page.

The only step that might cause some problems is the data preparation step. After uploading the file with data, the application becomes quite self-explanatory.
Conclusions

- In the transition Republic of Moldova, as seen from the data presented in this article, the GDP over the 1992 and 2012 time period has an increasing trend. But an increasing GDP doesn’t always mean a good life for everyone. The Republic of Moldova’s biggest problem is that income is distributed unevenly.

- The foreign direct investments are very important for a country economy because it can create jobs (reduce unemployment) and also increase productivity. In the Republic of Moldova’s case, the FDI are not stable. There are some years with very high FDI and other years with low FDI.

- From the linear regression model briefly presented, it can be observed that, in the Republic of Moldova, between GDP and FDI there is a positive relationship.

- Using R via PHP, means that a user can access a Web page and run commands on data, without installing R on their computer.
References