

## Challenges and Innovations in Statistics Education

Multiplier Conference of PROCIVICSTAT  
Szeged (Hungary) September 7-9 2017



Promoting Civic Engagement via Exploration of Evidence:  
Challenges for Statistics Education

Strategic Partnership of the Universities of Durham, Haifa, Ludwigsburg, Paderborn, Porto and Szeged  
<https://www.procivicstat.org>

# Lifelong learning development level in selected European countries: A perspective for improving statistical literacy

Ksenija Dumičić

Berislav Žmuk

University of Zagreb, Faculty of Economics and Business, Croatia

# Outline

1. Introduction
2. Data and methods
3. Descriptive statistics analysis of lifelong learning variable (LLL development level indicators) in selected European countries in recent years
4. K-means clustering of countries with 10 LLL variables, 33 countries
5. Regression analysis of GDP per capita in PPP regarding: LLL variable and unemployment level
6. Education for improving Statistical Literacy and Statistical Knowledge deserves improved opportunities
7. Conclusions

# 1. Introduction

## The research hypothesis

- The higher the lifelong learning development level (LLL) or the participation rate in education and training of adults in a country the more persons are competitive on the labour market and the greater the chance for a national economy to be more competitive.
- All these lead to higher development level in a country and to higher well-being level of citizens.

# Introduction

## Lifelong learning (LLL) definition (newly: Adults participation in LLL)

- Here the **lifelong learning (LLL)** is **defined** as a participation rate in education and training of people aged from 25 to 64 years (Eurostat). (After 2016 LLL was changed to **Adults participation in education**)
- **LLL encompasses all purposeful learning activities** whether formal non-formal or informal undertaken on an ongoing basis with the **aim of improving knowledge skills and competence**. *Source: Eurostat*
- **Useful definitions for understanding LLL:**
  - Formal education
  - Non-formal education
  - Informal learning
  - Continuing vocational training
  - Adult participation in lifelong learning

# Introduction:

## Lifelong learning (LLL) and Lifelong learning statistics

- According to Eurostat Glossary **LLL** is the lifelong voluntary and self-motivated pursuit of knowledge for personal or professional reasons.
- Within the domain of **LLL statistics** **formal education** covers education and training in the regular system of schools universities and colleges. **Non-formal education and training** includes all taught learning activities which are not part of a formal education programme. The information collected relates to all education or training regardless of whether it is relevant to the respondent's current or possible future job.
- In contrast to **LLL** as a concept **LLL statistics** do **not** cover **informal** learning which corresponds to self-learning (through the use of printed material computer-based learning/training online Internet-based web education visiting libraries etc.).
- The target population of Eurostat's **LLL statistics** is all members of private households aged between 25 and 64. Data are collected through the EU Labour force survey (LFS).
- There is also: Survey of Adult Skills (PIAAC)

## 2. Data and methods

For Europe countries recent data were used

For cluster analysis: 10 LLL variables (LLL development level indicators) were used

- LLL by gender (male female)
- LLL by employment status (employed unemployed)
- LLL by educational attainment level (primary secondary tertiary) and
- LLL by degree of urbanisation (city rural areas).

For the regression analysis

- Dependent: GDP per capita in PPP
- Independent: LLL & unemployment rate

# Table 1 List of selected lifelong learning variables for cluster analysis

No	Code	Variable brief description
1.	LLTotal	Participation rate in education and training % of total population aged 25-64
2.	LLMale	Participation rate in education and training % of males aged 25-64.
3.	LLFemale	Participation rate in education and training % of females aged 25-64.
4.	LLEmp	Participation rate in education and training % of employed persons aged 25-64.
5.	LLUnemp	Participation rate in education and training % of unemployed persons aged 25-64.
6.	LLEduc1	Participation rate in education and training % of persons with less than primary primary and lower secondary education (levels 0-2) aged 25-64.
7.	LLEduc2	Participation rate in education and training % of persons with upper secondary and post-secondary non-tertiary education (levels 3 and 4) aged 25-64.
8.	LLEduc3	Participation rate in education and training % of persons with tertiary education (levels 5-8) aged 25-64.
9.	LLCity	Participation rate in education and training % of persons living in cities aged 25-64.
10.	LLRural	Participation rate in education and training % of persons living in rural areas aged 25-64.

# Data and methods

## Data source:

- Eurostat; for descriptive and cluster analysis
- World Bank and Eurostat for Regression analysis

## Time:

- mostly 2005 to 2014; and a selected year (2011 for details; or 2015 etc.)

## Geography:

- **33** European countries: all **EU-28** member states **plus the FYR of Macedonia (FYROM)**  
**Iceland, Norway, Switzerland** and **Turkey**.

## Methods:

- Descriptive statistics analysis (exploration of outliers etc.);
- Cluster analysis: non-hierarchical cluster analysis (K-means)
- Correlation and OLS linear regression analysis; Trend Forecasting

# Methods

## Research steps

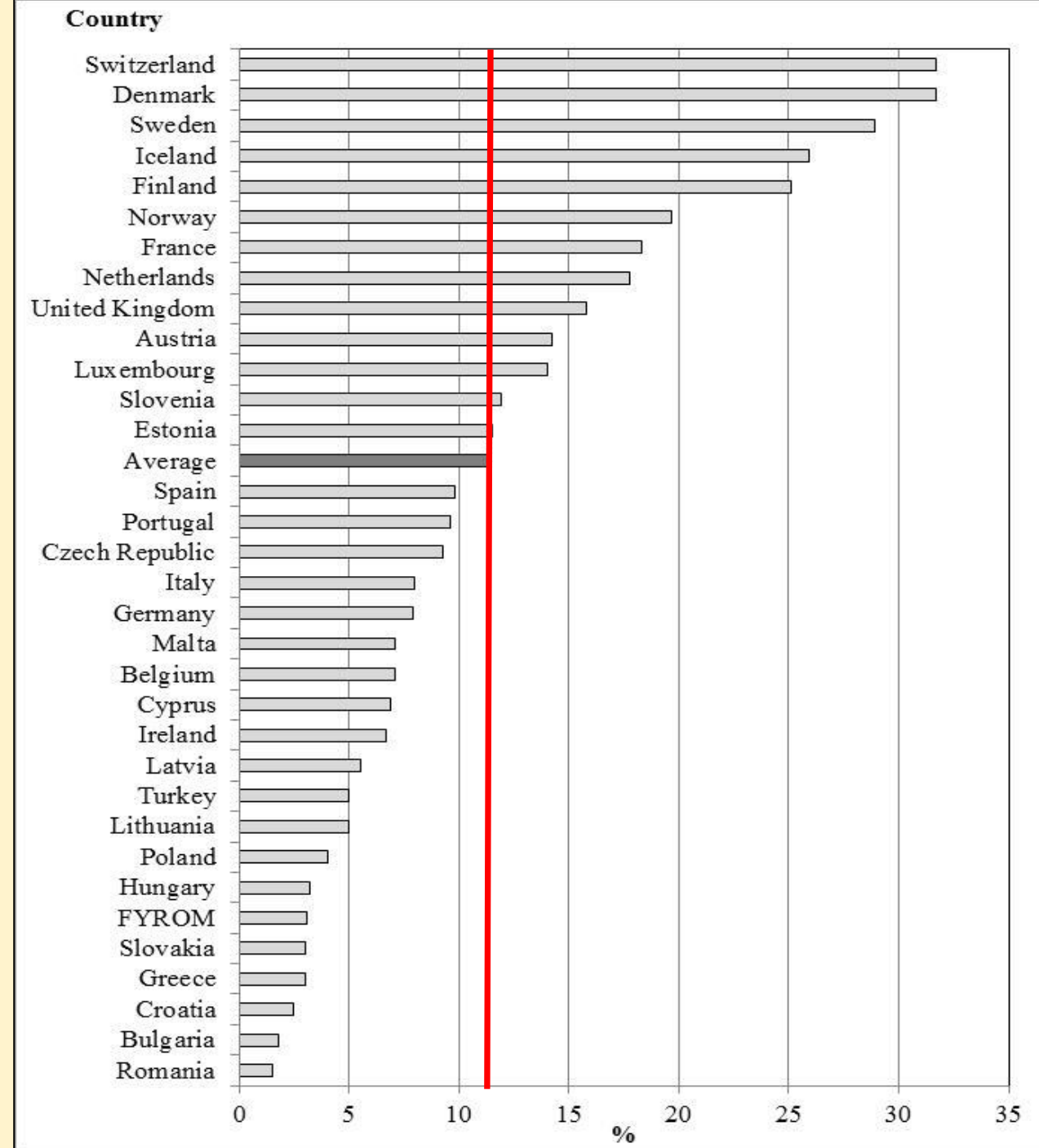
- The **descriptive exploratory data analysis** and **outlier detection**
- **Missing data problem:** Unfortunately data for all countries for the whole observed period are not available.
- In the **cluster analysis** the non-hierarchical approach is used.
- In order to observe the strength of impacts of **LLL development level change on the economic development** the **correlation** and **regression modelling** is conducted.
- In the analysis of **Gross domestic product (GDP) per capita** in purchasing power parity (PPP) (World Bank 2015)

# 3.Descriptive statistics analysis

**Figure 1**  
Lifelong learning as participation rate in education and training percentage of **total** population aged from 25 to 64 years (**LLTotal**) in 33 European countries in 2014

**Conclusion:**

**13 countries above and  
20 countries below average of 11.4%**



# Interpretation:

- There is a great disproportion in the **participation rates in education and training between the observed European countries.**
- In 33 European countries together in average **11.4%** of **total** population aged from 25 to 64 years (**LLTotal**) participated in education and training in **2014**.
- However there are 20 countries which had value of variable **LLTotal** lower than 10% in 2014.
- On the other hand only in **Switzerland (31.7%)** and **Denmark (31.7%)** more than 30% of total population aged from 25 to 64 years participated in education and training in 2014.
- It has to be emphasized that countries with the lowest percentage of total population aged from 25 to 64 years which participated in education and training in 2014 are **Romania (1.5%) Bulgaria (1.8%)** and **Croatia (2.5%).**

Table 2 Basic descriptive statistics of LLL as participation rate in education and training % of total population aged 25 -64 in 33 European countries 2005 to 2014

Statistics	Year									
	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
N	31	33	33	33	33	33	33	33	33	33
Mean	10.48	9.95	9.92	10.22	10.20	10.55	10.87	11.08	11.45	11.41
Std. dev.	8.23	8.27	8.07	8.21	8.05	8.75	8.57	8.68	8.74	8.93
Coef. var.	78.55	83.06	81.38	80.34	78.85	82.92	78.81	78.35	76.28	78.27
Skewness	1.03	1.12	1.11	1.06	1.01	1.11	1.06	1.06	0.98	1.05
Kurtosis	-0.16	0.09	0.13	0.00	0.03	0.28	0.28	0.15	-0.06	0.10
Minimum	1.3	1.3	1.3	1.4	1.4	1.2	1.3	1.3	1.7	1.5
1st quartile	5.1	4.2	4.4	4.7	4.4	3.9	4.4	4.5	4.3	5.0
Median	7.4	6.8	7.0	6.9	6.8	7.2	7.5	7.4	7.8	8.0
3rd quartile	15.6	15.0	14.8	13.9	14.6	16.2	15.7	14.1	16.1	15.8
Maximum	27.6	29.2	29.0	29.9	31.2	32.5	32.3	31.6	31.4	31.7

Conclusions:

High variability of data in all the years median is smaller than the mean always positive skewness

**Table 3 Basic descriptive statistics for nine LLL variables percentages in 33 European countries in 2014**

Stat.	Variable								
	LLMale	LLFemale	LLEmp	LLUnemp	LLEdu1	LLEdu2	LLEdu3	LLCity	LLRural
N	33	33	33	33	30	33	33	31	31
Mean	10.30	< 12.53	11.99	11.29	5.72	10.01	18.00	14.09	> 9.94
Std. dev.	7.79	10.28	9.28	9.82	5.97	7.74	11.23	9.74	8.27
Coef. var.	75.62	82.05	77.39	86.99	104.32	77.37	62.40	69.12	83.28
Skewness	1.10	1.14	0.96	1.40	1.56	0.98	0.72	0.99	0.94
Kurtosis	0.62	0.33	-0.02	2.11	1.86	0.02	-0.24	0.01	-0.06
Min	1.6	1.3	1.2	0.8	0.3	1.6	3.0	2.1	0.6
Q1	4.5	4.9	5.5	3.7	2.1	3.3	9.4	7.1	2.9
Median	8.0	8.3	8.7	7.9	3.2	8.1	17.5	10.7	7.0
Q3	14.2	17.4	17.4	16.2	7.8	13.1	24.6	19.5	15.3
Max	32.2	37.5	34.1	42.9	23.0	28.3	44.3	36.8	29.7

**Conclusions:**

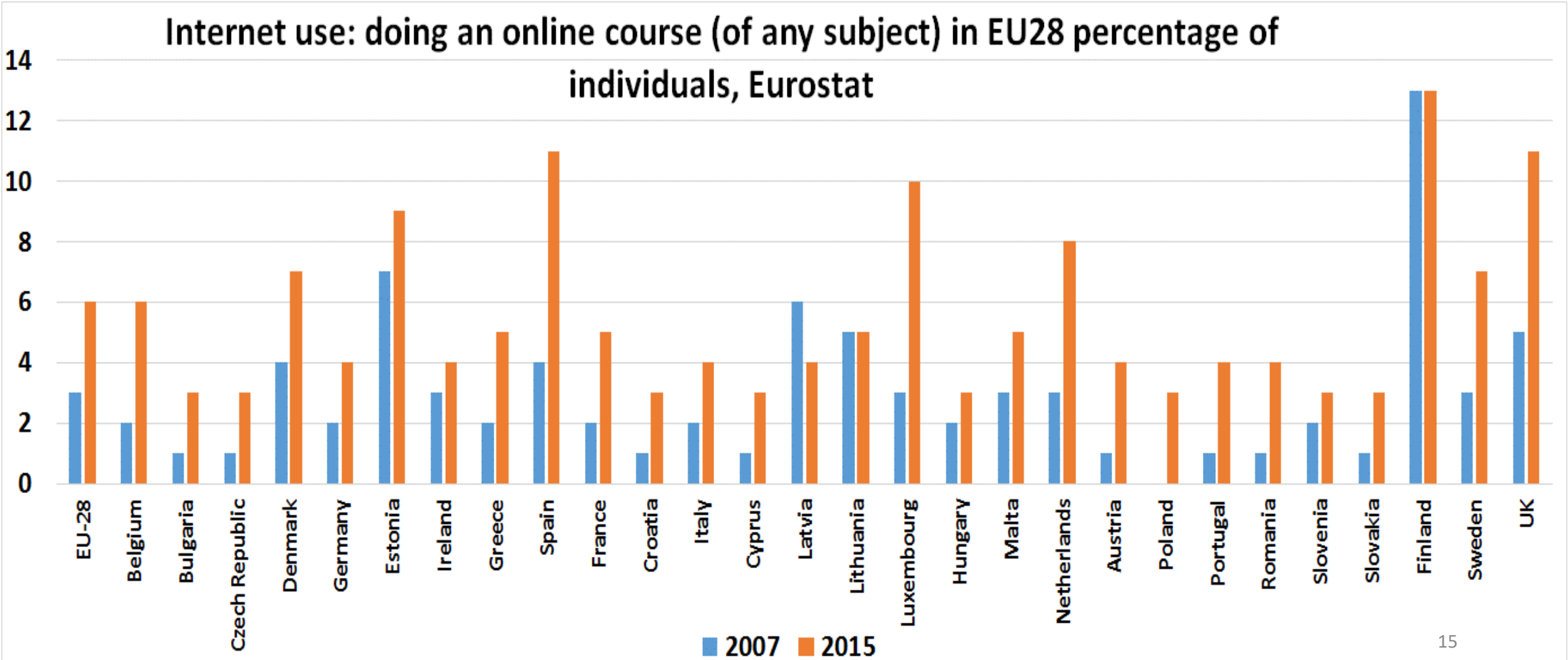
LLL % is higher for Females and in the Cities; the highest for LLEdu3 and the lowest for LLEdu1

High positive skewness; the medial always smaller than the mean

# Interpretation of Table 3

- Basic descriptive statistics results of the other **9 selected LLL** variables when all selected European countries are observed together in **2014**.
- Unfortunately there are some **missing data**. So there are no data available for the variables **LLCity** and **LLRural** for the **FYROM and Turkey**.
- Considering variable **LLEduc1** there are no data available for Bulgaria, Lithuania and Slovakia, whereas for Croatia the most recent available data from 2009 was used
- At the variable **LLUnemp** for **Lithuania** data from 2013 was used as an estimate for 2014.
- The **coefficients of variation** noticeably **higher than 30%** it can be concluded that data variation level at each variable is high.
- If the **median** values between the observed variables are compared it can be concluded that the **highest share of persons who participate in the LLL processes is when persons with tertiary education are observed (variable LLEduc3)**.

# Recent dynamics: increase of Internet use for doing an online course (of any subject) in EU-28 percentage of individuals, Eurostat



# 4. K-means clustering of countries with 10 LLL variables, 33 countries

**Two non-hierarchical K-means cluster analysis with three clusters of countries in 2014 were performed:**

## First:

The following variables were omitted because of **missing data**:

- **LLEduc1, LLCity and LLRural.**
- **Finally: 7 variables and 33 countries**

## Second:

The following countries were omitted because of the **outliers**:

- **Bulgaria, FYROM, Lithuania, Slovakia and Turkey.**
- **Finally: 28 countries and all 10 selected lifelong learning variables**

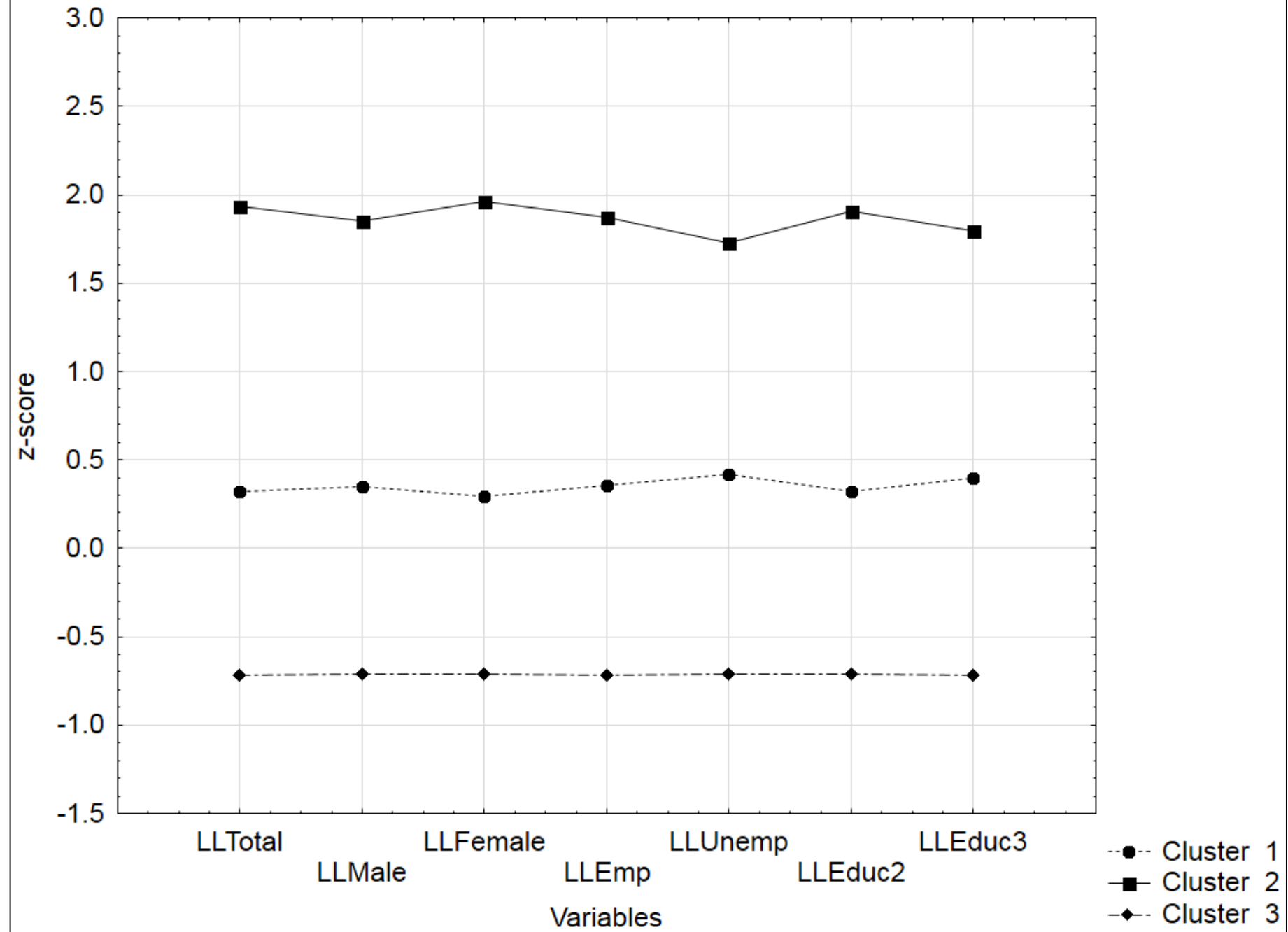
**Table 4** Analysis of variance table non-hierarchical K-means clustering method **7 selected lifelong learning variables in 33 European countries** data from **2014**

Variable	Between SS	df	Within SS	df	F	Signif. p
LLTotal	28.86	2	3.14	30	137.75	0.0000
LLMale	27.40	2	4.60	30	89.46	0.0000
LLFemale	29.15	2	2.85	30	153.43	0.0000
LLEmp	28.10	2	3.90	30	107.93	0.0000
LLUnemp	25.86	2	6.14	30	63.12	0.0000
LLEduc2	28.26	2	3.74	30	113.22	0.0000
LLEduc3	27.03	2	4.97	30	81.53	0.0000

**Figure 3**

Plot of means for each cluster non-hierarchical cluster analysis **K-means** clustering method

**7** selected LLL variables in **33** European countries data from 2014



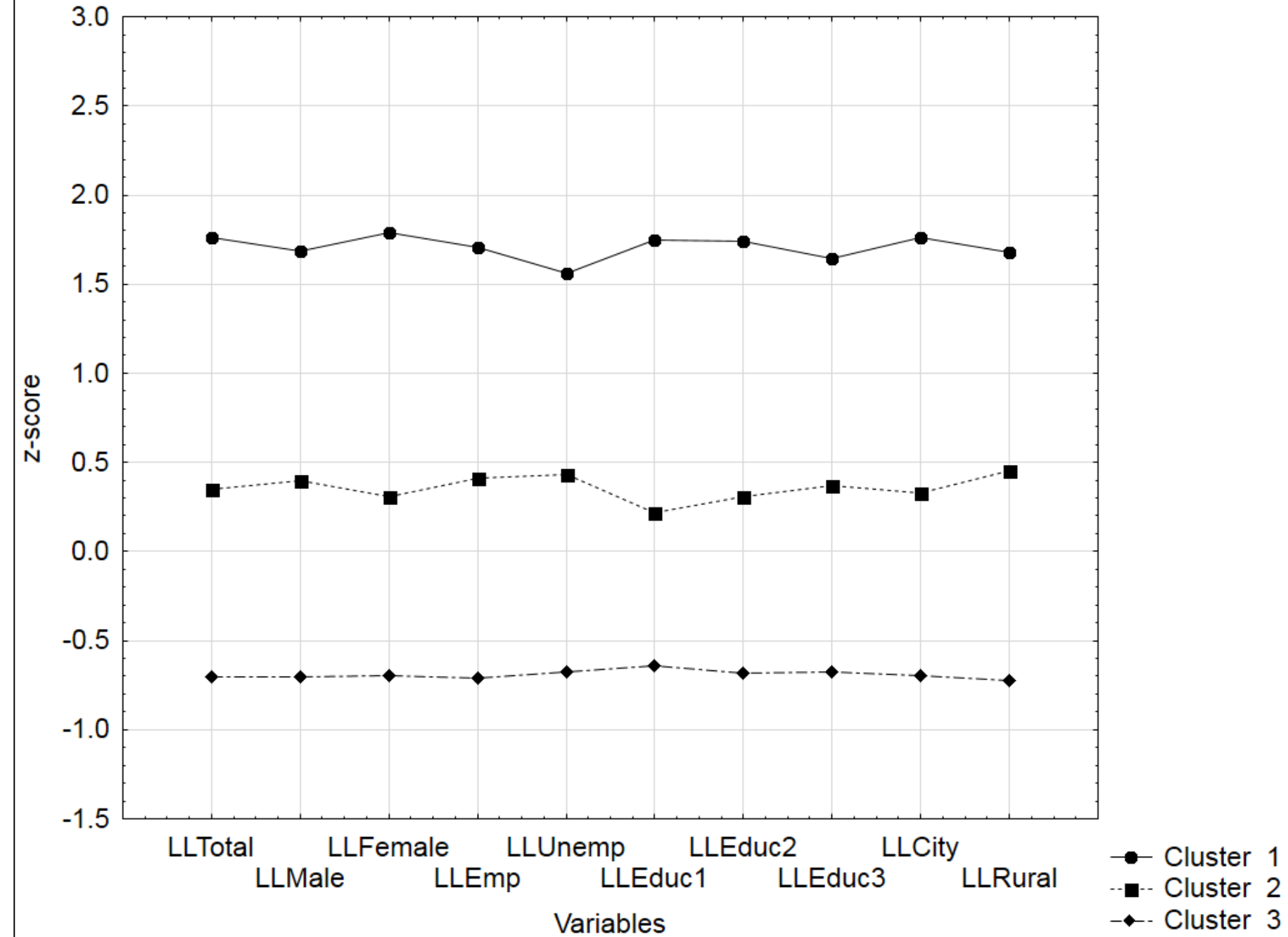
**Table 5** Classification of countries in the clusters non-hierarchical cluster analysis **K-means clustering** method **7 selected LLL variables in 33 European countries** data from **2014**

Cluster 1 10 countries	Cluster 2 5 countries	Cluster 3 18 countries
Austria Estonia France Luxembourg Netherlands Norway Portugal Slovenia Spain <b>United Kingdom</b>	<b>Denmark</b> Finland Iceland Sweden Switzerland	Belgium Bulgaria <b>Croatia</b> Cyprus Czech Republic FYROM Germany Greece <b>Hungary</b> Ireland Italy Latvia Lithuania Malta Poland Romania Slovakia Turkey
<b>Cluster 1</b> in average have the <b>medium</b> LLL development level.	<b>Cluster 2</b> in average has the <b>highest</b> LLL%	<b>Cluster 3</b> in average has the <b>lowest</b> LLL%

**Table 6** Analysis of variance table non-hierarchical K-means clustering method **10 selected LLL** variables in **28 European countries** in **2014**

Variable	Between SS	df	Within SS	df	F	Signif. p
LLTotal	24.34	2	2.66	25	114.40	0.0000
LLMale	23.19	2	3.81	25	76.07	0.0000
LLFemale	24.44	2	2.56	25	119.08	0.0000
LLEmp	23.83	2	3.16	25	94.17	0.0000
LLUnemp	20.92	2	6.08	25	42.99	0.0000
LLEduc1	22.16	2	4.84	25	57.29	0.0000
LLEduc2	23.20	2	3.80	25	76.38	0.0000
LLEduc3	21.74	2	5.26	25	51.68	0.0000
LLCity	24.06	2	2.94	25	102.18	0.0000
LLRural	23.94	2	3.06	25	97.75	0.0000

**Figure 4** Plot of means for each cluster non-hierarchical cluster analysis **K-means clustering** method 10 selected LLL variables in **28** European countries data from **2014**



**Table 7** Classification of countries in the clusters non-hierarchical cluster analysis **K-means** clustering method **10** selected LLL variables in **28** European countries data from **2014**

Cluster 1 5 countries	Cluster 2 7 countries	Cluster 3 16 countries
Denmark Finland Iceland Sweden Switzerland	Austria France Luxembourg Netherlands Norway Slovenia <b>United Kingdom</b>	Belgium <b>Croatia</b> Cyprus Czech Republic <b>Estonia</b> Germany Greece <b>Hungary</b> Ireland Italy Latvia Malta Poland <b>Portugal</b> Romania <b>Spain</b>
<b>Cluster 1</b> in average has the <b>highest</b>	<b>Cluster 2</b> in average have the <b>medium</b> lifelong learning development level.	<b>Cluster 3</b> in average has the <b>lowest LLL%</b> (those green moved from the <b>medium</b> to <b>the lowest</b> LLL% Cluster)

## 5. Regression analysis of GDP per capita in PPP regarding: LLL variable and unemployment level

- How does LLL variable impact the GDPpc in PPP indicator?
- How does unemployment rate impact the GDPpc in PPP?
- European countries (32)
- 2014

# Correlation and regression analysis of GDPpc and lifelong learning and long-term unemployment in selected European countries

- **Dependent:**  $Y_{GDPpcPPP}$  - The economic development level measured by the variable *GDPpcPPP* (in current international US\$).
- **Regressors:**
- $X_{LLTotal}$  - lifelong learning development level *LLTotal* and
- $X_{LTTotal}$  - long-term unemployment level *LTTotal*
- **33 European countries were initially observed**
- **Luxembourg data was omitted from the further analysis since at the variable *GDPpcPPP* it deviates for 3.56 standard deviations from the average of this variable.**
- **Consequently the observed number of countries is reduced to 32.**
- **It has to be emphasized that there are some missing data. So for Luxembourg Malta and Switzerland data for variable *GDPpcPPP* are used from 2013. For all other countries and at all other variables data from **2014 are used.****

# Correlation matrix

Variable	Variable		
	$Y_{GDPpcPPP}$	$X_{LLTotal}$	$X_{LTTotat}$
$Y_{GDPpcPPP}$	1.0000		
$X_{LLTotal}$	<b>0.7349*</b>	1.0000	
$X_{LTTotat}$	<b>-0.5252*</b>	<b>-0.6851*</b>	1.0000
*Correlation is significant at the 0.01 level (1-tailed)			

# Simple linear regression *Model R1*

- In the regression analysis firstly simple linear regression models are developed. Afterwards OLS estimated multiple linear regression modelling is performed.
- In all regression models the dependent variable is variable *GDPpcPPP* whereas independent variables are variables *LLTotal* and *LTTotal*. In the regression modelling data for 32 observed European countries from 2014 are used.
- The first simple linear regression model ***Model R1*** with parameters estimated using the ordinary least squares method is:

$$\begin{array}{lll} \hat{Y}_{GDPpcPPP} = 22,870.97 + 983.44 \cdot X_{LLTotal} & n = 32 & R^2 = 0.5401 \\ (2,388.49) \quad (165.67) & \hat{\sigma} = 8,357.23 & \hat{V} = 24.57\% \end{array}$$

# Interpretation

- **The regressor  $X_{LLTotal}$  is statistically significant in the model at the significance level of 1% (p-value=0.0000).**
- The linear regression model diagnostics results confirmed that in the regression model given above at significance level of 1% ( $\alpha=0.01$ ) there is neither heteroskedasticity (White test statistics=1.173 p-value=0.556) nor non-normality of residuals (Jarque-Bera test statistics=7.269 p-value=0.026) problems.
- **The coefficient of determination indicates that the variable  $LLTotal$  explains 54.01% of the total variation.**
- **The regression coefficient of variation is 24.57%.**

- If LLL (participation rate in education and training defined as a percentage of total population aged from 25 to 64 years) increases for 1 percentage point the  $Y_{GDPpcPPP}$  will in average increase by 983.44 international US\$ in the observed European countries.
- If the standardized simple linear regression model is observed it can be concluded that **one standard deviation increase in variable *LLTotal* will result in an average increase of variable  $Y_{GDPpcPPP}$  of 0.73 standard deviations.**

## Conclusion:

- Obviously an increase of population share which takes part in LLL has positive impact on the economic development level in a country.

# Simple linear regression *Model R2*

- The second simple linear regression model *Model R2* where impact of long-term unemployment level on the economic development level is following:

$$\begin{array}{lll} \hat{Y}_{GDPpcPPP} = 51,223.79 - 384.39 \cdot X_{LTTotal} & n = 32 & R^2 = 0.2758 \\ (5,418.99) \quad (113.72) & \hat{\sigma} = 10,487.40 & \hat{V} = 30.83\% \end{array}$$

- The multiple regression model was investigated but there the regressor  $X_{LTTotal}$  was not statistically significant.

- So if long-term unemployment share increases for 1 percentage point the  $Y_{GDPpcPPP}$  will in average decrease by 384.39 international US\$ in the observed European countries.
- Also the standardized simple linear regression model has shown that one standard deviation increase in  $Y_{GDPpcPPP}$  of 0.53 standard deviations.

## Conclusion:

- As expected, and opposite to the LLL development level, **the long-term unemployment level** has **negative impact** on the economic development level in a country.

# 6. Education for improving Statistical Literacy and Statistical Knowledge deserves improved opportunities

**SL Perspective:**

***To be a Statistician is an Excellent Career Choice (and Data Scientists, too)!***

**The World is becoming more quantitative and data-focused**

**Increasing job opportunities in statistics**

**Demand for statisticians and data analysts - 4.4 million new jobs expected worldwide in the years ahead ([www.worldofstatistics.org](http://www.worldofstatistics.org), 2016).**

# Statistics as a key competence for lifelong learning

**M. Vichi (CESS-2016) interpreted Statistics as a key competence for lifelong learning since „*The workers in 21<sup>st</sup> century must have a stock of information-processing skills*”**

(Pellizzari M. and A. Fichen (2013) “A New Measure of Skills Mismatch: Theory and Evidence from the Survey of Adult Skills (Programme for the International Assessment of Adult Competencies (PIAAC)” OECD Social Employment and Migration Working Papers No. 153 OECD Publishing. <http://dx.doi.org/10.1787/5k3tpt04lcnt-en>)

**„Statistics is a new key competence for lifelong learning enhancing employability and the ability to remain employable throughout life.”**

The *European Parliament* and the *Council of the European Union* in 2006 has recommended a list of **eight key competences for lifelong learning** which regard: the ability to communicate in mother and foreign languages, **mathematical and digital competences** learning to learn social and civic competences sense of initiative and cultural awareness. ...

In the Information Society also the **ability to read and interpret reality and make decisions with data by means of statistics is a new fundamental competence for the 21<sup>st</sup> century citizens.**

# SL should be promoted to be implemented into the new jobs:

## New jobs are appearing related to:

- Actuarial science; Astrostatistics; Business analytics; Chemometrics; Computational Statistics; Data Science; Demography; Econometrics; Environment statistics; Epidemiology; Geostatistics; Information Engineering; Machine Learning; Operations research; Population ecology; Psychometrics; Quality control; Quantitative psychology; Reliability Engineering; Statistical finance; Statistical mechanics; Statistical physics; Statistical Signal Processing; Statistical thermodynamics; Social Statistics; ....

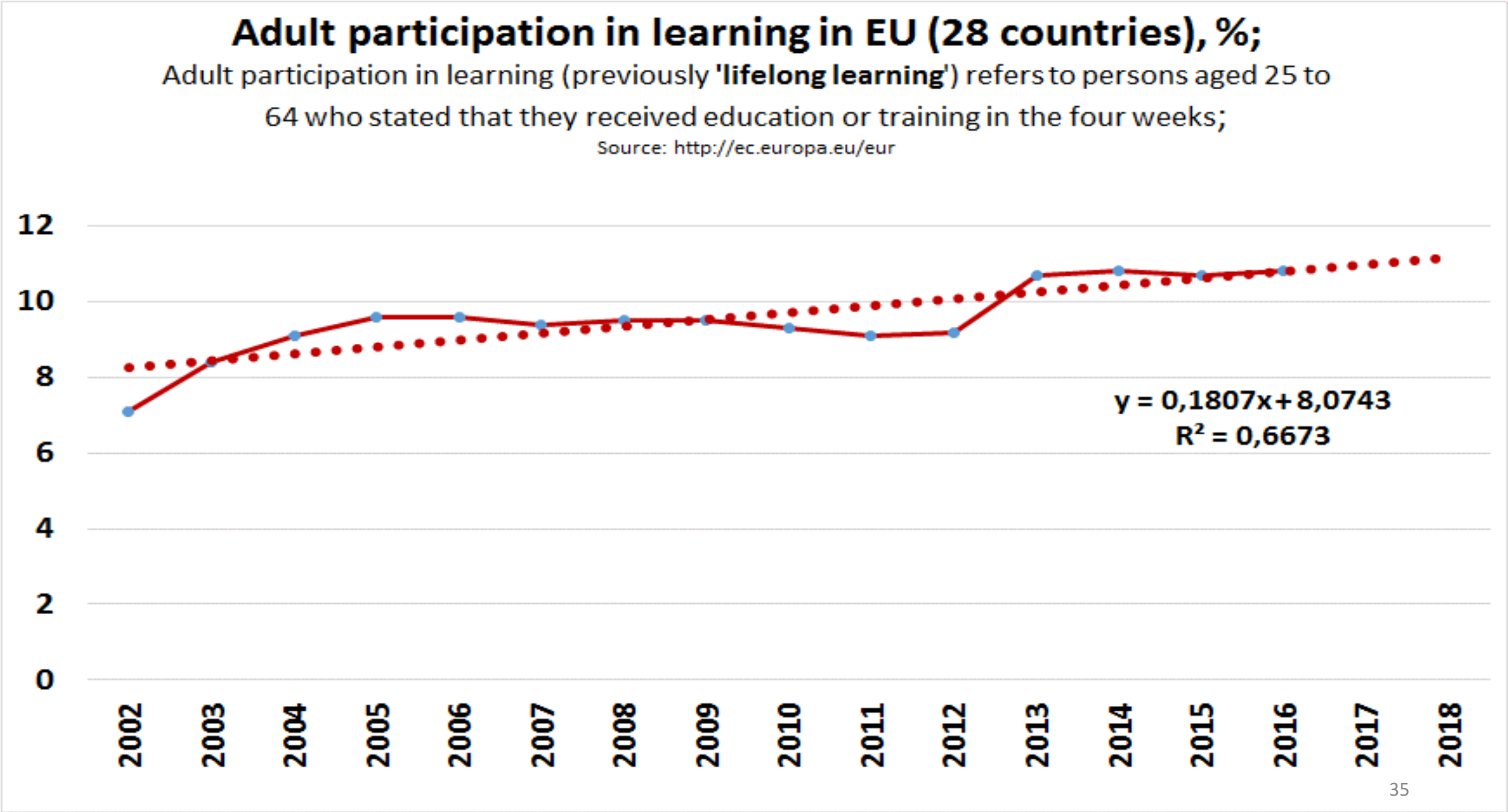
# SL Perspective: The strategy for the future

**The goal set by the governments should be: to create material and staff-equipped basis for continuous improvement of SL of the society through education and training programs!**

## **Steps:**

- 1. learning** elements of SL in primary schools;
- 2. learning** at secondary schools;
- 3. attending statistical courses** during the high school and at the university up to doctoral degree;
- 4. permanent statistical education** during the working life - **SL as a lifelong (learning) activity** (Professor *M. Vichi (FenStats; CESS-2016, Budapest)*- the role of data producers and of all kinds of educational institutions universities etc.)

# Good perspective: Increasing trend of *Adult Participation in learning*



# 7. Conclusions

- **LLL development level** in the **33** observed European countries increased in last 10 years from 10% to 11.5% (in 2014: 13 countries above and 20 countries below the average of 11.4%)
- **The higher the LLL development level or the higher the participation rate of adults in education and training in a country,**
- **the higher the chance that the employment rate will be higher, since the more persons are competitive on the labour market and**
- **the greater the chance for a national economy to be more competitive.**
- All these lead to **higher development level in a country** and to **higher well-being level of citizens.**

# 7. Conclusions

- **Permanent education for improving Statistical Literacy/ Knowledge deserves improved opportunities helping the unemployment rate to be reduced**
- The statistical education for adults (already employed or unemployed) for improving Statistical Literacy and Knowledge **should be improved in (all) the countries (respective indicators should be increase)**
- **Statistics is a new key competence for LLL enhancing employability and the ability to remain employable throughout life.**
- **The role of statistics educators** (related to government support, ministry of education & science, etc.) becomes very responsible, especially for those that design and organize educational programs in statistics for all society segments.
- **Future research: LLL perspective for improving Statistical Knowledge should be investigated and implemented more**

# Thank you for your attention!

Köszönöm a figyelmet!

תודה על תשומת הלב שלך

Vielen Dank für Ihre Aufmerksamkeit!

Obrigado pela sua atenção!

Hvala Vam na pozornosti!