Abstract:
With a purpose to evaluate the possible role of statistical societies in digitalisation and globalisation era, selected “digital skills – related” and “globalisation” variables are put into the focus of the desk research, hoping that conclusions might be stimulating for statistical societies’ enhanced support to “digital skills – related” statistical education, knowledge, literacy and skills. Regarding this, possibilities of national statistical societies’ and Federation’s of National Statistical Societies (FENStatS, 2019) a more intensive standardized communication is suggested. The goal of the desk research was to explain one digital skills related variable, with three others. In this research, data for Percentage of individuals aged 16 to 74, who use the internet for ordering goods or services (Y2017_IntOrderGoods), related to three development indicators, as independent variables, for 31 European countries for 2017, has been explored. Among several indicators influencing the main variable under study positively, Percentage of individuals aged 16-74, who have basic or above basic overall digital skills (X3_DigitalSkill), showed to have the strongest positive correlation with it. Descriptive summary statistics, linear trend modelling, correlation, regression and hierarchical cluster analysis were used. Analysis included 27 countries of the EU-28, plus four EU candidates, Montenegro, FYR of Macedonia, Serbia and Turkey. The estimated linear trend for the EU-28 countries shown that Percentage of individuals aged 16 to 74, who use the internet for ordering goods or services increased with the slope of 2.73 percentage points yearly. After discovering quite strong positive correlations between the dependent and each of independent variables, a two multiple linear regression models were fully developed, both including X3_DigitalSkill, as the independent variable. Both models show the average increase in the regression value to be statistically highly significant for each of the regressors considered. After hierarchical clustering, the Scandinavian countries with Germany and United Kingdom, all clustered within the separate cluster, as the absolute European leaders regarding the Digital Society indicator and Percentage of persons who order goods and services online, while the South-East European (SEE) countries, being similar, were clustered, as those at the end of the queue. Here recognized differently developed clusters of countries, with different percentages of individuals for Y2017_IntOrderGoods, and X3_DigitalSkill, variables taken as examples for digitalisation and globalisation, have also different communication habits of their NSAs. Activities of the NSAs should be studied with the goal to standardize their communication, to be better fitted into the global digitalised World, by communicating over the common platform, fulfilling their activities’ goals.

Keywords: Globalisation; Digital Skills; Hierarchical clustering; Statistical societies; FENStatS

1. Introduction:
Purpose of this study is to recognize statistical societies’ role and acting opportunities in the era of digitalisation and globalisation, based on digital related variables research findings. Since the overall
present digitalization initiates changes in lives of people in all the societies and creates a growth of economies in countries all over the World, it is considered as the natural component of our reality. In this research, selected indicators, which influence the main variable under study, Percentage of individuals who use the internet for ordering goods or services (Y_{2017\text{IntOrderGoods}}), which increased for EU-28 from 30% in 2007 to 60% in 2018, were analysed. Among several considered, one Digital Society related indicator, called Percentage of individuals who have basic or above basic overall digital skills (X_3\text{DigitalSkill}), showed to be positively correlated, with the strongest positive influence on the main variable under study. Three indicators, with fully available data, are included for correlation and regression analysis: GDP per capita in PPS, Index, EU-28=100; Percentage of households who have internet access at home, for the population aged 16 to 74; and Percentage of individuals aged 16-74, who have basic or above basic overall digital skills. Since the descriptive exploration found that the Luxembourg’s Gross Domestic Product (GDP) per capita in PPS, X_1\text{GDPpcPPS}, has been an extremely high outlier, the further analyses with 31 countries for 2017 were performed: the 27 countries of EU-28, Montenegro, FYR of Macedonia, Serbia and Turkey. The authors studied mutual interaction of the variables with correlation, regression and cluster analysis in 2017. Dumičić, Žmuk & Novkovska (2017) performed regression analysis of e-commerce, focusing the selected EU candidates and the EU countries. Further, Dumičić, Žmuk & Mihajlović (2017) executed profile analysis of clustered European countries constructed using a number of indicators impacting the e-commerce for individuals. Dumičić, Skoko Bonić & Žmuk (2018) conducted statistical analysis of nine development indicators’ impacting the Internet purchases by individuals, defined as a share of individuals who made a purchase by using the Internet in the last 12 months in the total population of a country for 2013. Further, variety of development indicators influences special forms of internet purchases were analysed, too. So, online banking analysis, given in Dumičić, Čeh Časni & Palić (2015), includes variety of multivariate analysis, while development indicators impacts on online booking for travel and accommodation is presented in Žmuk, Dumičić & Mihajlović (2014), and Dumičić, Žmuk & Čeh Časni (2015), showing separate clusters of European countries. Žmuk & Mihajlović (2018), paid attention to position of the Western Balkan countries within Europe regarding online booking influenced by economic and digital development level indicators. Nagy (2016) explored e-commerce on Hungarian market, and Nagy (2017) compared Hungarian and Ukrainian digital economy and society. A number of visualised digital economy impacts of various aspects of customer behaviour is given in Consumer Barometer, as given at CB (2017).

In this paper, a trend analysis for Y_{2017\text{IntOrderGoods}}, over the period 2007 to 2018, follows. The added indicators have been as follows: X_1\text{GDPpcPPS}, Level of internet access for households, given as Percentage of households who have internet access at home, for the population aged 16 to 74, and (X_3\text{DigitalSkill}). The research hypotheses state, firstly, that a positive correlation exists between Y_{2017\text{IntOrderGoods}} and each of three included explanatory indicators, so that clear clusters of countries could be created; and there are increased motivation and opportunities for statistical societies to act in digitalised and globalised World by improved communication when promoting statistical and digital related improvements in education, knowledge, literacy and skills. Performed statistical data description, performed correlation, regression and cluster analysis, helped testing these hypothesis.

2. Methodology:
After investigation on the recent official data availability, Eurostat (2019), the following indicators have been analysed:

- Y_{2017\text{IntOrderGoods}} - Percentage of individuals aged 16-74, who use the internet for ordering goods or services, being Digital Society related indicator and the main variable of the interest in this study;
- X_1\text{GDPpcPPS} - GDP per capita in PPS, Index, EU-28 = 100;
- X_2\text{AccHome} - Percentage of households who have internet access at home, for the population aged 16 to 74, also called Level of internet access for households; and
- X_3\text{DigitalSkill}, Percentage of individuals aged 16-74, who have basic or above basic overall digital skills.

Firstly, an Ordinary Least Squares (OLS) liner trend model for the variable Y_{2017\text{IntOrderGoods}} for the EU-28 area over the period 2007 to 2018, and short term forecasts for 2019 and 2020 were estimated.
Further, all four considered variables are described using Descriptive Statistics for 31 countries, as shown in Table 1. Since, in exploratory data analysis the Luxembourg’s data for \( X_{1, \text{GDPpcPPS}} \) appeared to be a serious outlier, with \( z = 3.84 \), it must be removed from further analysis. The EU-28 countries, without Luxembourg, with Montenegro, FYR of Macedonia, Serbia and Turkey, in 2017 proceeded for the further analysis.

Multiple linear regression (MLR) analysis, using the software MegaStat and EViews, was used, Gujaraty & Porter (2010). The population model is given in (1):

\[
y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{ij} + \epsilon_i, \quad i = 1, 2, \ldots, n
\]

Regression parameters are estimated using the OLS method. The model with estimated parameters is given in (2):

\[
\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j x_{ij}, \quad i = 1, 2, \ldots, n
\]

And, finally, for \( Y_{2017\text{IntOrderGoods}} \), and three explanatory variables, \( X_{1, \text{GDPpcPPS}} \), \( X_{2, \text{AccHome}} \) and \( X_{3, \text{DigitalSkill}} \), a hierarchical cluster analysis was performed.

3. Result:

OLS linear trend model for the variable \( Y_{2017\text{IntOrderGoods}} \) for the EU-28 area over the period 2007 to 2018, and short term forecasts for 2019 and 2020 in Figure 1 are illustrated.

![Figure 1. Y_{2017\text{IntOrderGoods}}. EU-28, period 2007-2018 and forecasts for 2019 and 2020](image)

Source: Eurostat, Authors’ creation.

| Table 1. Descriptive statistics, for n=31 selected European countries in 2017 |
|-------------------|-------------------|-------------------|-------------------|
| \( Y_{2017\text{IntOrderGoods}} \) (percentages) | \( X_{1, \text{GDPpcPPS}} \) (Index, EU-28 = 100) | \( X_{2, \text{AccHome}} \) (percentages) | \( X_{3, \text{DigitalSkill}} \) (percentages) |
| Count | 31 | 31 | 31 |
| Mean | 47.48 | 87.10 | 82.61 | 53.61 |
| Standard Error | 3.78 | 5.90 | 1.50 | 2.46 |
| Standard Deviation | 21.05 | 32.84 | 8.38 | 13.70 |
| Coeff. of variation= | 44.33% | 37.71% | 10.14% | 25.55% |
| Range | 69 | 145 | 31 | 50 |
| Minimum | 13 | 36 | 67 | 29 |
| Maximum | 82 | 181 | 98 | 79 |
| 1st quartile | 32 | 67 | 77 | 46 |
| median | 46 | 82 | 82 | 54 |
| 3rd quartile | 61 | 108 | 88 | 61 |
| interquartile range | 29 | 42 | 12 | 15 |
| Mode | 32 | 75 | 81 | 50 |
| Kurtosis | -1.03 | 0.70 | -0.62 | -0.53 |
| Skewness | 0.068 | 0.684 | 0.076 | 0.035 |
| zmin | ME= -1.64 | MK and RS = -1.56 | BG= -1.86 | RO= -1.80 |
| zmax | UK= 1.64 | IE= 2.86 | NL= 1.84 | NL= 1.85 |

Source: Eurostat, Authors’ creation

For EU-28 countries, an increase of the main variable studied, with the linear trend slope of 2.73 percentage points yearly over 2007 to 2018, and 99.50% of total variation explained by the model is shown in Figure 1. It is visible that the percentage of individuals ordering goods and services increased from 30% in 2017 to 60% in 2018. According to e-commerce statistics for individuals, Eurostat (2019), data for the EU-28 countries, Montenegro, FYR of Macedonia and Serbia, sorted downward by the
percentage of those who purchased online, and this was compared to the percentage of those who used the Internet in the same period in 2018. Denmark leads with 98% for those who purchased online within the last 12 months, and 84% for those who used the Internet within the last 12 months with. Montenegro is with 12% at the bottom for those who purchased online within, while for the percentage of those who used the Internet Bulgaria appeared with 67% as being at the bottom.

All four variables, Y2017 IntOrderGoods, X1 GDPpcPPS, X2 AccHome and X3 DigitalSkill, were explored and described statistically, Table 1. The highest variability of data over 31 countries is calculated for Y2017 IntOrderGoods, with Coefficient of variation of CV=44.33%, and the lowest variability was noticed for X2 AccHome (CV=10.14%). Regardless the high outlier for the variable X1 GDPpcPPS for Ireland, with standardized value of Z=2.86, it was remained for the further analysis. The same moderate low extremes were found for FYR of Macedonia and Serbia (Z=1.56). The correlation matrix for 31 countries in 2017, Table 2, shows correlations between all the pairs of variables under consideration, being all positive and strong.

### Table 2. Correlation matrix, for n = 31 countries and 2017.

<table>
<thead>
<tr>
<th></th>
<th>Y2017 IntOrderGoods</th>
<th>X1 GDPpcPPS</th>
<th>X2 AccHome</th>
<th>X3 DigitalSkill</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y2017 IntOrderGoods</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X1 GDPpcPPS</td>
<td>0.7654</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>X2 AccHome</td>
<td>0.9027</td>
<td>0.8129</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>X3 DigitalSkill</td>
<td>0.9166</td>
<td>0.6852</td>
<td>0.8471</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Source: Eurostat data, Authors’ creation.

The OLS MLR modelling was performed for 31 countries in 2017. Among several models built and tested for validity regarding explanation the dependent variable, Y2017 IntOrderGoods, the following two multiple regression models were fully developed and interpreted: In MLR Model 1, for explanation the dependent variable, Y2017 IntOrderGoods, two independent variables were used: X1 GDPpcPPS and X3 DigitalSkill. In MLR Model 2, for explanation the dependent variable, Y2017 IntOrderGoods, two independent variables were used: X2 AccHome and X3 DigitalSkill.

The estimated Model 1 looks as given in (3):

\[
\hat{Y}_{2017 \text{IntOrderGoods}} = -27.86 + 0.17 \cdot X_{1 \text{GDPpcPPS}} + 1.14 \cdot X_{3 \text{DigitalSkill}},
\]

\[
\hat{Y}_{2017 \text{IntOrderGoods}} = 7.683; R = 0.936; R^2 = 0.876; \hat{R}^2 = 0.867; F = 98.58; n=31.
\]

For one index point increase in the variable GDP per capita in PPS, Index, EU28 = 100, X1 GDPpcPPS, with the other independent variable unchanged, the regression value of Y2017 IntOrderGoods, would increase by 0.17 percentage points. For one percentage point increase in the variable X3 DigitalSkill, having the remaining independent variable fixed, the regression value of Y2017 IntOrderGoods, would increase by 1.14 percentage points.

The F-Test for overall regression, with p-value=2.12E-13, shows statistical significance of the whole Model 1 at even 1% significance level. Coefficient of determination R^2 tells that 87.6% of the total sum of squares is explained. The individual two-sided t-Tests, shows that variable X1 GDPpcPPS is statistically significant at 1% significance level (p-value=0.009). The variable X3 DigitalSkill is statistically significant at 1% significance level (p-value=8.52E-09). No model assumptions violation is present.

The estimated Model 2 is given in (4):

\[
\hat{Y}_{2017 \text{IntOrderGoods}} = -89.61 + 0.12 \cdot X_{2 \text{AccHome}} + 0.83 \cdot X_{3 \text{DigitalSkill}}.
\]

\[
\hat{Y}_{2017 \text{IntOrderGoods}} = 7.008; R = 0.947; R^2 = 0.897; \hat{R}^2 = 0.889; F = 121.30; n=31.
\]

For one index point increase in the variable X2 AccHome, with the other independent variable unchanged, the regression value of Y2017 IntOrderGoods, would increase by 0.17 percentage points. For one percentage point increase in the variable X3 DigitalSkill, having the remaining independent variable fixed, the regression value of Y2017 IntOrderGoods, would increase by 0.83 percentage points.

Coefficient of determination R^2 shows that 99.7% of the total sum of squares is explained by Model 2. Since the overall F-Test has p-value = 1.61E-14, the whole regression Model 1 is statistically significant at 1% significance level. Using two-sided t-Test, the variable X2 AccHome is statistically significant, with t-statistic = 3.908 and p-value = 0.0005, at 1% significance level. The variable X3 DigitalSkill Is statistically significant, with p-value = 0.0001, at 1% significance level, too. Regression diagnostics’ tests for residuals were performed, showing no assumptions violation is present.
Cluster analysis: In the next step, for 31 countries in 2017, based on all four variables examined in the regression modelling, \( Y_{2017\text{InternetGoods}}, X1_{\text{GDPpPPS}}, X2_{\text{AccHome}}, \) and \( X3_{\text{DigitalSkills}} \), cluster analysis using Ward linkage and Squared Euclidean distances, according to Hair et al. (2008), and Field (2011), was performed, Table3.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>No. of countries; ( n = 31 )</th>
<th>Countries grouped into the clusters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>10</td>
<td>Belgium, France, Austria, Czech R., Slovakia, Estonia, Spain, Malta, Slovenia, Ireland</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>6</td>
<td>Denmark, Germany, Finland, United Kingdom, Netherlands, Sweden</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>8</td>
<td>Bulgaria, Serbia, Greece, Croatia, Montenegro, Romania, FYR of Macedonia, Turkey</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>7</td>
<td>Italy, Cyprus, Portugal, Lithuania, Latvia, Hungary, Poland</td>
</tr>
</tbody>
</table>

4. Discussion and Conclusion:
Individuals using the internet for ordering goods or services, as the percentage of individuals aged 16 to 74, \( Y_{2017\text{InternetGoods}} \), which doubled for the EU-28 countries from 30% in 2007 to 60% in 2018, resulted with the highly representative estimated linear trend, with the yearly slope of 2.73%. For the period from 2007 until 2015, such a trend slope was a little bit higher, 2.82%. Because of the highest correlation, \( r = 0.9166 \), both here analysed MLR models, built for explanation of the main variable under study, included the Digital Society related indicator named Percentage of individuals aged 16-74, with basic or above basic overall digital skills. Additionally, Model 1 included GDP per capita in PPS, while Model 2 included additionally Level of internet access for households, performing the second regressor. Both models, shown that an increase in each of the independent variables would result with a statistically significant increase in the regression value of the main variable under study at 1% significance level. Clustering method resulted with four clusters. Regarding all four variables studied here, as the result of clustering, highly developed countries, Belgium, France, Austria, Czech R., Slovakia, Estonia, Spain, Malta, Slovenia, Ireland, gathered in Cluster 1. The countries, being economically and digitally the most developed ones, Denmark, Germany, Finland, United Kingdom, Netherlands and Sweden, clustered in Cluster 2, separately. These countries are leaders regarding the Digital Society indicators. Eight SEE countries, the EU-28 members plus four EU candidate countries (Bulgaria, Serbia, Greece, Croatia, Montenegro, Romania, FYR of Macedonia, Turkey, with exception of Cyprus), clustered in the same, Cluster 3. Cluster 4 collected Italy, Cyprus, Portugal, Lithuania, Latvia, Hungary and Poland.

More efforts for improved education regarding the prerequisite for better performance of Digital Society over European countries, which includes internet usage and digital literacy, based on improved feeling for financial transactions security, should come from various sides, from governments, educators, employers, employees, citizens, and from national statistical societies. So, governments should offer more adequate legal frame and support for it; educators might should develop and adopt adjusted educational programs; employers might show more understanding by giving more opportunities for additional education on digital and related types of literacy, such as statistical, media and financial literacies; and employees’, facing some skill gaps, should pose request for continuous adults’ education not only for digital, but also numerical, statistical and financial literacy. Regarding digitisation and globalisation, for what here studied main variable studied (Percentage of individuals who ordered goods or services via Internet), is just an example, the national statistical societies (NSAs) might have an extra challenge and could be of higher influence. NSAs’ role at national, European and international level might be of higher importance and more visible, if they could find the way to support “digital skills – based” statistical literacy improving and spreading. Especially those NSAs, which, either research and analyse statistical literacy indicators and respective education process development, trying to influence them, or directly encourage and support up-to-date “digital skills – based” education programs in statistics, might contribute to an increase of Digital Society related indicators (Internet Ordering, Purchases or other similar), through rising voice pro more adjusted programs, not solely regarding statistical, but also digital, media and financial literacies, truly needed for better adoption of Digital Society benefits in modern globalized World per se, as well.

In the future desk research, the limitations of this study may be overcome if more variables, e.g. gender and age groups, as well as different educational levels, should be analysed for better explanation of digital society related indicators, additionally. Further, the impact of the NSA’s activities in the European countries may be analysed and measured, by analysing the existence and the efficiency of
their attempts to contribute to the improvement of statistical and related digital literacy, through promotion of educational programs, which should nowadays be more often visible or offered at the NSA websites. Recently, since the beginning of 2019, FENStatS (2019) has been announcing (parallel to the announcing the courses organized by European Courses on Advanced Statistics, ECAS), the upcoming NSAs’ Statistical and Data Science conferences more often, which is based on the improved overall digital skills adoption. This activity is encouraging and supportive to the NSAs’ acting goals and efforts in the same direction to reach better informed, “digital based” statistical literacy improved, for better citizenship and society.

Here recognized clusters of countries, with significantly different development levels and percentages of individuals for Y_2017IntOrderGoods, and X_3_DigitalSkill, which have been taken in this research as the symbols for digitalisation and globalisation, have also more or less active communication habits of their NSAs. Because of the research time constrains, their activity and communication style were not explored with full potential for this paper, but generally speaking, each NSA’s activity alone and towards the FENStatS could be in many countries intensified through regular standardized communication procedure and praxis, being regularly current (e.g. on quarterly or similar basis), efficient and formalized. Proposal for a common society, such as FENStatS, is directed to an idea of intensifying communication between NSAs by using the existing FENStatS’ web site, based on standardized information collection and current information refreshment, which would enable information announcement, flow and exchange. Since such an approach should be simple, a short standardized format could be suggested for use. Current news of interest might include the categories: Statistical and Data Science education and training programs; including statistical and digital knowledge and literacy improvement possibilities; Conferences organised by NSAs Call for Papers announcements; and Other news. This would improve networking among the CSAs’ members and the overall importance of the FENStatS’ activity, based on practical use of available ICT with an increasing potential for improvements in literacy and knowledge regarding Statistics and Data Science in the World of digitalisation and globalisation.

References:


