**Data Analysis**

**Task 2**

The file Worldbank.xlsx contains World Bank data on various variables related to export activities by countries for the year 2018.

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Variable Label** | **Values** |
| Year | Year |  |
| CountryName | Country Name |  |
| CountryCode | Country Code |  |
| Agriculturalrawmaterialsexpor | Agricultural raw materials exports (% of merchandise exports) |  |
| ArmsexportsSIPRItrendindica | Arms exports (SIPRI trend indicator values) |  |
| Commercialserviceexportscurr | Commercial service exports (current US$) |  |
| Communicationscomputeretc | Communications, computer, etc. (% of service exports, BoP) |  |
| Exportsofgoodsandservices | Exports of goods and services (% of GDP) |  |
| Exportsofgoodsandservicesa | Exports of goods and services (annual % growth) |  |
| ExportsofgoodsandservicesB | Exports of goods and services (BoP, current US$) |  |
| Exportsofgoodsandservicesc | Exports of goods and services (constant 2010 US$) |  |
| Foodexportsofmerchandisee | Food exports (% of merchandise exports) |  |
| GoodsexportsBoPcurrentUS | Goods exports (BoP, current US$) |  |
| Fuelexportsofmerchandisee | Fuel exports (% of merchandise exports) |  |
| Hightechnologyexportsofma | High-technology exports (% of manufactured exports) |  |
| Hightechnologyexportscurrent | High-technology exports (current US$) |  |
| TaxesonexportscurrentLCU | Taxes on exports (current LCU) |  |
| Transportservicesofservice | Transport services (% of service exports, BoP) |  |
| Travelservicesofcommercial | Travel services (% of commercial service exports) |  |
| Travelservicesofserviceex | Travel services (% of service exports, BoP) |  |
| high\_income | High income countries | Takes 2 values:  1 - belongs to the high income group of countries  0 - no |
| income | income | Takes 3 values:  - high income  - middle income  - low income |
| development | Developed or developing country | Takes 2 values:  - developing  - developed |

Please consider the following tasks

1. For interval or ratio scale variables run suitable normality tests to check whether their distribution is close to the normal distribution. Formulate hypothesis. Make conclusions. Create suitable graphs to compare the variables’ distributions with the normal distribution.

2. Run a chi-square test to analyze the relationship between categorical variables. You can take the existing variables or create new categorical variables based on interval or ratio scale variables. Explain why the chi-square test is applicable to analyze the relationship between the selected pairs of variables. Create a contingency table based on these variables and describe the tendencies that you can observe in frequency distribution. Formulate hypotheses for the chi-square test, interpret the results of analysis and make conclusions. Create a suitable graph to demonstrate the relationship between the selected variables.

3. Calculate appropriate correlation coefficients between any three pairs of variables. Explain the selection of the correlation coefficient. Fill in the table below. Interpret the results. Create suitable graphs to visualize the analyzed relationships.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Variables | Appropriate correlation coefficient | Hypotheses | Strength of the relationship | Direction of the relationship | Significance of the relationship |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

4. Calculate a paired correlation coefficient between any variables. Now calculate the partial correlation coefficient between the same pair of variables controlling for any other third variable. Interpret the results of analysis. Create suitable graphs to visualize the analyzed relationships.

Please send the Jupyter Notebook with comments and answers to amelikyan@hse.ru from your **personal** e-mail. The task can be done individually or in a group of 2-3 students. If the task is done in a group, then one member of the group can send the task indicating the names of the group members in the subject of the email.