**IE266: ENGINEERING STATISTICS**

**Case Study 1**

*“Academic integrity is expected of all students of METU at all times, whether in the presence or absence of members of the faculty. Understanding this, I declare that I shall not give, use, or receive unauthorized aid in this study.”*

Group 22

Işıl Çiçek - 2629509

Ekin Akçay - 2577419

Müjde Derya Ocaktan – 2629822

21.04.2024

**Part 1 – Descriptive Statistics**

For this part, descriptive measures and descriptive graphs have been used to display and summarize the requested data in each question. The utilized data set is an Excel file about historical data on global agricultural activities and countries’ environmental impacts collected by FAO, named “Agricultural\_Impact\_Data.xlsx”. It consists of meat, rice, and cereal production data of 158 countries from 1995 to 2020. The file includes information about the following features: **Country, Region, Subregion, Development, Year, Population, Agricultural Land, Item, Emissions,** **Production,** and **Freshwater withdrawals.**

**a)** In the first question, we are requested to show the relative frequency distribution of each item between the years 1995-2020 over the subregions.

A graph with different colored bars

Description automatically generated with medium confidence

*Figure 1: Relative frequency distribution of each item between the years 1995-2020 over the subregions*

We can observe that rice production is more prominent in subregions of Asia whereas in other subregions, cereal production is much more common. However, in both cases, the relative frequency of meat production is quite low.

**b)** To determine which countries accounted for 80% of the GHG emissions due to rice production in 2020, we have used a Pareto chart as it displays the cumulative percentage and each country’s percentage simultaneously.

A graph of global emissions

Description automatically generated

*Figure 2: Countries’ GHG Emissions due to rice production in 2020*

Figure 2 above shows that China, India, Indonesia, Thailand, Philippines, Vietnam, Bangladesh, and Myanmar account for 80% of total GHG emissions due to rice production in 2020. Their individual percentages are displayed as a pie chart in Figure 3 below:

A pie chart with different colored circles

Description automatically generated

*Figure 3: GHG Emission Percentages of top countries contributing to 80%*

**c)**  For the years 1999, 2009, and 2019, box-and-whisker plots of the cereal production amount per capita of the developed countries have been shown in the following figure:

**A graph showing a number of blue rectangular objects

Description automatically generated with medium confidence**

*Figure 4: Cereal Production Amount per Capita in years 1999, 2009, and 2019 respectively*

Among 158 countries, there were only three outliers in 1999 and one in 2009. Thus, we can say that the data mostly falls between upper and lower limits. However, the upper limit on cereal production showed an increasing trend starting from lower than 1.25 in 1999 to over 1.5 in 2019. From the increase in the length of the box, it can be interpreted that the interquartile range has also increased throughout the years. Therefore, the cereal production amount per capita has increased for most of the countries as years go by.

**d)** For the first part of this question, the top 12 countries in the world for total rice production between 2011 and 2020 and their respective total GHG emission amounts have been listed below in Table 1:

|  |  |  |
| --- | --- | --- |
| Country | Total Rice Production | Total GHG Emission |
| **China** | 2,084,871,288 | 1,985,628,941 |
| **India** | 1,660,998,625 | 1,491,409,798 |
| **Indonesia** | 575,885,769 | 799,122,026 |
| **Bangladesh** | 524,779,158 | 412,632,638 |
| **Vietnam** | 436,422,239 | 466,504,296 |
| **Thailand** | 329,240,527 | 568,742,453 |
| **Myanmar** | 265,819,735 | 337,831,081 |
| **Philippines** | 184,353,040 | 468,128,321 |
| **Brazil** | 117,641,244 | 64,093,172 |
| **Japan** | 108,516,100 | 103,400,028 |
| **Pakistan** | 101,118,387 | 129,253,747 |
| **Cambodia** | 99,614,640 | 144,899,789 |

*Table 1: Top 12 countries on rice production between years 2011-2020 and their total GHG emissions*

The table mostly consists of Asian countries as expected since rice production is easier due to the climate conditions of Asia and countries with more rice production account more for the emissions, however, Brazil has a relatively lower emission amount.

In the second part, we have calculated each country’s average GHG emission amount from rice between the years 1995-2020. After removing the top 20 countries with the highest and lowest emissions, the plotted histogram of the emission amounts and their frequencies is the following:

A graph of the number of emissions

Description automatically generated

*Figure 5: Average GHG Emissions of Countries from Rice Between Years 1995-2020*

After several trials, we have determined the number of classes as nine as it gives a better understanding of the data distribution. The shape of the histogram shows that the distribution of data is close to a right-skewed distribution.

**e)** The value of emission efficiency has been defined as the production amount per GHG emission amount. We drew separate line charts to observe fluctuations in each item’s emission efficiency, as in Figure 6. Cereal’s emission efficiency is significantly higher compared to rice and meat for all years. Meat also has extremely low emission efficiency, close to zero. Therefore, meat accounts for more GHG emissions per product. Although the emission efficiency of cereal has always been the highest, the value of the efficiency also increased throughout the years. This may be a result of more eco-friendly production techniques. However, the efficiency values for rice and meat have followed a more stable trend.

A graph showing the growth of a number of people

Description automatically generated with medium confidence

*Figure 6: Emission efficiency of items throughout years 1995-2020*

We have also examined if there is a relation between the development status of a country and its emission efficiency. For this purpose, the dataset of the top 100 countries in rice production in 2020 and their emission efficiencies in rice has been used.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **0-250** | **251-500** | **501-750** | **751-1000** | **1001-1250** | **1251- 1500** | **1501-1750** | **1751-2000** | **2001-2250** | **2251-2500** | **2501-5000** |
| **Developed** | 4 | 4 | 1 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| **Developing** | 6 | 10 | 14 | 6 | 6 | 3 | 1 | 1 | 0 | 0 | 3 |
| **Underdeveloped** | 4 | 6 | 10 | 5 | 2 | 6 | 1 | 0 | 2 | 1 | 1 |

*Table 2: Relation between development status of countries and emission efficiency*

As observed from the contingency table in Table 2, there is not a considerable relation between these two variables. Yet, it can be said that developed countries generally have a lower emission efficiency probably because the advanced production techniques these countries access cause a lot more GHG emissions.

**f)** To have a better understanding of how rice production in 2020 and freshwater withdrawals affect GHG emissions, we have plotted a scatter diagram and a best-fit line to show whether there is a correlation between GHG emissions/freshwater withdrawals and production amount, as seen in Figure 8 and 9 below.

A graph with a red line

Description automatically generated

*Figure 7: Relation between production amounts and GHG emissions*

A graph showing a curve

Description automatically generated

*Figure 8: Relation between Freshwater Withdrawals and GHG Emissions*

These diagrams indicate that GHG emissions have a positive correlation with both freshwater withdrawals and production amounts. The correlation coefficient for production and emissions is calculated as ρ1 = 0.9792479 whereas for freshwater withdrawals and emissions it is ρ2 = 0.9734407. Thus, production has a higher correlation with emission amounts.