Basic Econometrics

ECON 1313

Semester B

SGS Campus – Group 11

Mr. Bob Baulch

Tran Huong Ly

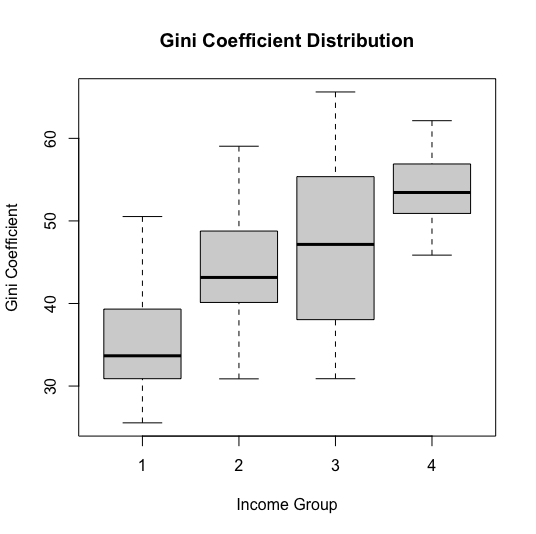
S4000175

Assignment 1

**Question 1.**

**A graph of a graph with a number of squares

Description automatically generatedA graph of a diagram

Description automatically generatedA graph with a number of squares

Description automatically generatedi)** The dataset assigned to this analysis consists of low and low-middle income countries in the year 2017 (referred to as countries in income groups 3 and 4 in the dataset). By World Bank’s definition (2022), these are the countries with the Gross National Income per capita of US$ 4095 or less. Also often cited as “developing countries”, these countries typically share characteristics of having a lower Human Development Index (HDI) and a higher Gross Domestic Product (GDP) growth and higher inflation rate than countries in higher income groups, as visualised in the graphs below. Lower income countries also exhibit relatively higher Gini coefficients of the income groups 3 and 4 when compared to the other income groups, which also seemingly support Kuznets’ (1955) theory of the Gini coefficient increasing before peaking economic growth then declining after said peak.

**ii)** The data was taken from the World Development Indicator database by the World Bank, which contains 1400 time series indicators relating to all aspects of development of a country, including Poverty and Inequality, People, Environment, Economy, States and Markets, and Global Links (World Bank n.d.) The database currently holds comparable statistics for 217 countries and more than 40 country groups, making it one of the most comprehensive databases relating to global statistics.

The selected dataset for this analysis consists of 54 countries with 11 variables, including highly relevant statistics such as GDP, inflation, Gini coefficient, Human Development Index (HDI), GDP per capita, GDP growth.

The mean, median, and standard deviation for GDP, GDP per capita, and inflation are presented as following:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Mean | Median | Standard deviation |
| GDP | 99,891,780,128 | 17,117,786,362 | 364,429,442,673 |
| GDP per capita | 1,647.6 | 1,405.2 | 1,046.449 |
| Inflation rate | 6.198 | 4.398 | 6.456538 |

It is notable that India as an abnormally high GDP comparing to the rest of the countries in this income group, which may have influenced the mean and the standard deviation to be abnormally large. The abnormally large standard deviation of GDP indicates a wide spread of GDP values, which is the same for the also relatively large GDP per capita and inflation rate. The large difference between the mean and the median of the GDP also suggests that the data is not normally distributed. The median for GDP per capita is US$ 1405.2, which is only US$ 360.2 higher than the World Bank’s upper threshold for the categorisation of the Low Income group (World Bank 2022), thereby suggesting that nearly 50% of the data set belong in income group 4.

**iii)** India has been removed as an outlier in the second graph illustrating GDP distribution for the purpose of better visualisation of other data points. However, it is worth noting that there were still some data points that can be considered as outliers in this figure, due to the nature of being extremely positively skewed. Overall, these histograms show that all three indicators are not normally distributed but rather have a positive skew, meaning most of the data is more concentrated near the 0 value. This suggests that most countries in this data set belong to income group 4, with only a few categorised as in income group 3.

A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generatedA graph of a graph

Description automatically generatedA graph of a number of people

Description automatically generated

As there is no given population standard deviation for this dataset, the calculations below assume the sample mean follows a student’s (t) distribution and therefore the sample standard deviation is used for questions 2 and 3. All calculated results beyond this point are rounded to the nearest 3 significant figures.

**Question 2.**

**i)** The formula for estimating the confidence interval is:

where and is the sample mean and standard deviation respectively, is the critical value of the t distribution with degrees of freedom, n being the sample size.

With the margins of error () for a 95% confidence interval with 53 degrees of freedom being known as 0.03 for the HDI and 2.47 for the Gini coefficient, the population mean of the HDI is estimated to be 0.542 ≤ μ ≤ 0.597 and the Gini coefficient to be 46.5 ≤ μ ≤ 51.4, meaning the population mean of the HDI is likely to be between 0.542 and 0.597, and the population mean of the Gini coefficient is likely to be between 46.5 and 51.4. Both estimations have a relatively narrow interval width, which indicates a high precision of the data.

**ii)** The population of HDI and Gini coefficient can be assumed to be normally distributed due to the sample size being larger than 30, according to the Central Limit Theorem which specifies the larger the sample size, the more the distribution of the sample mean resembles the normal distribution.

**iii)** Using the above formula but with a 99% confidence interval, the margins of error for the HDI and Gini coefficient are calculated to be 0.04 and 3.29 respectively, which are notably larger comparing to the margins of error for the 95% confidence interval estimates. The estimations for the population mean of the HDI and the Gini coefficient are therefore calculated as 0.533 ≤ μ ≤ 0.606 and 45.6 ≤ μ ≤ 52.2 respectively, which both have a wider width of interval compared to the 95% confidence interval estimates.

**Question 3.**

**i)** The HDI is calculated by finding the geometric mean of adjusted Gross National Income per capita (Income Index), Life Expectancy Index and Education Index, which comprises of Mean Years of Schooling Index and Expected Years of Schooling Index. The generic formula for the HDI calculation is as below:

**ii)** The mean and standard deviation of the HDI is calculated as 0.569, while the standard deviation is calculated as 0.101.

The alternative hypothesis (H1) is determined to be μ > 0.5, which indicates that this is an upper-tail hypothesis testing. The null hypothesis is therefore μ ≤ 0.5. The formula used for the test statistic are as following:

where and are the sample mean and standard deviation respectively, represents the population mean, which equals 0.5 in this hypothesis, and n is the sample size.

Because the significance level is 0.05 and the critical value is known to be 1.67, the test statistic is 5.07. The null hypothesis is rejected because the test statistic is greater than the critical value, indicating that there is sufficient evidence that the mean of HDI in the population is larger than 0.5.

**iii)** Changing the significance level to 0.01 would give a critical value of 2.40. This would not change the conclusion as the test statistics, being independent from the significance level, would still be greater than the critical value.

**References:**

Kuznets S (1955) ‘Economic Growth and Income Inequality’, *The American Economic Review*, 45(1):1–28, https://www.jstor.org/stable/1811581.

World Bank (2022) *New World Bank country classifications by income level: 2022-2023*, World Bank, https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2022-2023.