

Student Version

Title: Where one lives matters when it comes to Covid-19

Summary of the project: In this project teams of students will attempt to determine the best predictor(s) for the number of Covid-19 infections. The project involves identifying sources of data and regression analysis where we build a model for the number of Covid cases per capita for a region (country or state) as a function of various predictors such as population density, Gross National Income (GNI), average age of the population, geography of the region, etc.

Instructions to students:

- **Background information:** A quick look at the number of Covid-19 infections around the world shows us that the pandemic has affected countries differently. The number of infections in some countries like New Zealand, Taiwan and Vietnam was small and the pandemic was controlled fairly quickly. In contrast, China, Russia, western European and North American countries had huge number of infections. Furthermore, later on some of these countries, like Germany and Italy managed to control the pandemic and reduced the number of new infections drastically. However, countries such as Russia, Sweden and United States are still reporting significant number of infections. What is/are best predictors of the number of infections in a region or country?
- Our goal is to build a simple linear regression model, where the dependent variable is the number of infections per one million and the independent variables (predictors) are population density, GNI, average age of the population, type of the government (democratic or not) and geography of the country (island or not). There are other predictors, such as rating of the health care system, annual number of visiting tourists to the country, etc., that for a more detailed study can be included in the model.
- **Individual or group project:** This project will be assigned to teams of students
- **Data:**

For data on the daily and total number of Covid-19 cases and number of death visit:

<https://ourworldindata.org/coronavirus>

For GNI per capita of a country refer to:

[https://en.wikipedia.org/wiki/List_of_countries_by_GNI_\(nominal\)_per_capita](https://en.wikipedia.org/wiki/List_of_countries_by_GNI_(nominal)_per_capita)

For median age of countries see:

<https://www.cia.gov/library/publications/the-world-factbook/fields/343.html>

Other needed information such as population density (population/area of the country), geography of a country and democratic or totalitarian nature of the government are readily available online

- **Procedure/plan of action to complete the project:**

Pre-project Assignment for students:

Think about these questions, we'll discuss in class:

- 1) What do you think determines the number of Covid-19 cases in a country?
- 2) Why should we normalize the number of Covid cases by dividing the total number by the population in millions? Why not just consider the total number of cases?
- 3) Why the geography of a country and population density matter?
- 4) Do you think pandemic can be controlled better in a country run by democratic system or non-democratic?
- 5) Do a search of GNI per capita, see what it means and why it matters.

After the preliminary discussion on these questions, the instructor prepares 2-3 list of 25 countries with reliable data and assigns them to different teams of students. Step by step instructions on using a software (such as Minitab) or an online computational site such as WolframAlpha will be given to students in a separate document. After the models are created, an in-class discussion on the meaning and significance of the coefficients of the regression model should take place. Specifically, student should consider the following questions:

- 1) What does the sign of a coefficient for each predictor mean? For example, GNI per capita is (+) whereas the sign for the coefficient of average age is (-). Interpret the signs.
- 2) For a better model, what other predictors might you include in this model?

- **Example(s):**

We selected 25 countries from Europe, North America, South America Asia, and New Zealand and Australia and created a linear regression model. For categorical variables we assign Island = 1, not island = 0, Democratic = 1, not democratic = 0, Early shelter-in-place = 1, no/late shelter-in-place = 0

Using data from 25 countries, we found the following:

Regression Equation

$$\begin{aligned} \text{Total} &= 12103 + 0.248 \text{ Popdnsty} + 0.0886 \text{ GNI} - 271 \text{ Av. Age} - 3561 \text{ Shelter} \\ \text{Case} &+ 2792 \text{ Dem} - 2644 \text{ Island} \end{aligned}$$

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	12103	5446	2.22	0.040	
Popdnsty	0.248	0.199	1.25	0.228	2.36
GNI	0.0886	0.0405	2.19	0.043	1.66
Av. Age	-271	165	-1.64	0.118	1.56
Shelter	-3561	1765	-2.02	0.060	1.31

Dem	2792	2640	1.06	0.305	2.57
Island	-2644	1591	-1.66	0.115	1.38

Information on the specific countries used in this example maybe provided to students in a separate document

- **Duration:** This is a one to two-week project depending on the scope of the project and details expected from the students
- **Deliverables and evaluation:**
 - Students are expected to be prepared for the pre-project questions.
 - Each team will present their results in a five-minute presentation.
 - Students are expected to be prepared for the post-project discussions.