Week 1

R Output.R

Student Name: Zihan Ma

Class NameALY6010: Probability Theory and Introductory Statistics

Instructor: Tom Breur

02/26/2023

Report part 3:

For the Dataset, I’m using the free Dataset that I found from GitHub (<https://github.com/nytimes/covid-19-data>); the chart below shows the first five observations of the data.

Table

Description automatically generated

Since It’s a .csv file that has already been cleaned, so I don’t have to apply “gsub” or “ifelse()” to process the data. But the Dataset contains no categorical variables, so I create some ordinal variables by making classifying them to existing numerical variables; in this case. I cut the cases and death rate(cases/deaths) into ten ordinal variables on the equal amount that I can work with, and removed the dips row, which I will not be using in this report. I also make a sub-data frame summarising every state’s case together.

Report part 4:

Chart

Description automatically generated

Above is a bar blot showing each state’s total COVID-19 cases and the death rate. The plot shows that COVID-19’s death rate is visible but not in all states. There are four states. Taxes/New York/Florida and California have red bars, as seen above, and all four states have relatively high COVID cases. To find out whether there is a relationship between total cases and the death rate. I made a few tables below.

Graphical user interface, application

Description automatically generated

Alaska is a good presentation of the low death rate example, and there are only level 1 and 2 death rates for all countries’ daily reports. Which means in January 1st and February 26th, and there are 1471 times that a country’s daily report shows that local death is the lowest level and 125 times for the level 2 death rate in USwide.

Table

Description automatically generated

Massachusetts is an example of the death rate distribution at just below the average.

Table

Description automatically generated

Florida represents the states that have a high death rate distribution.

Table

Description automatically generated

Table

Description automatically generated

Here is the death rate level probity for five states that chose using systematic randomization.

From table 4.2, We can observe the rank of death rate distribution and estimate the rank of the danger level for five states.

Which are Alaska<Massachusetts<Indiana<Florida<Georgia.

Table

Description automatically generated

Table

Description automatically generated

Table 5 shows the daily case level for different countries by state.

If we do the same, estimate the danger level rank for five states by their case level distribution.

We could rank them: Alaska <Georgia<Indiana <Florida <Massachusetts

Compared to the result from table 4, Alaska<Massachusetts<Indiana<Florida<Georgia.

We could raise a presumption that the low COVID cases dose limited the death rate, but as the cases amont increases, other elements become more dominant reasons that affect the death rates.

A sheet of music

Description automatically generated with medium confidence

Table

Description automatically generated

The last two tables show how the death rates distribute at the same case level.

Without the limit of states, we found something interesting: the death rate is relatively high in some low-case reports. There are two possibilities for this fact.

It could relate to different COVID strains, some of which have lower infectivity but higher lethality. (Zimmer et al., 2021) That could explain why in all low cases amount reports (1~5), there are huge gaps between a low death rate and a high death rate.

Or it might be caused by the sample table I chose not being big enough, and the flawed random method caused some bias.

Other than those, if we ignored rows 9 and 10, we could get the similar result we obtained by comparing tables 4 and 5.

References:

Nytimes. (n.d.). *Nytimes/COVID-19-data: A repository of data on coronavirus cases and deaths in the U.S.* GitHub. Retrieved February 27, 2023, from https://github.com/nytimes/covid-19-data

Zimmer, C., Anthes, E., & Jacobs, A. (2021, November 30). *Omicron: What we know about the dominant coronavirus variant*. The New York Times. Retrieved February 27, 2023, from https://www.nytimes.com/article/omicron-coronavirus-variant.html