Bootstrapping

Caelan Bryan, Jenna Dufresne, Jamielee Jimenez Perez

Fall 2022

Data Description

The dataset we will be using to explore Bootstrapping is the "Causes of Death - Our World in Data" dataset from kaggle. The causes of death dataset was also expanded with population data from the world bank.

The raw data contains a multitude of death statistics broken down by the continent, region, country, and territory. The statistics for cause of death are given in number of deaths and split up by the cause of death. Data is presented over the span of multiple years. The population dataset from the world bank is a list of countries, regions, and territories but contains the population for each year from 1960 to 2021 where available.

To get more meaningful numbers, the population for each year 1990 through 2019 was populated into the causes of death dataset. In order to perform a more predictable experiment, the causes of death dataset was cleaned by first removing all non-country entries. The number of executions and terrorism deaths columns were also removed due to a lack of data for the majority of countries. As a final note, the Vatican and Liechtenstein are the two countries that did not have cause of death statistics available in the dataset.

The columns in the cleaned dataset are as follows:

Name	Description	Type
Entity	Name of Country	Nominal
Population	Population of Country at Specific Year	Discrete
Code	Three Letter Country Code	Nominal
Year	Year for Causes of Deaths	Nominal
Deaths_Meningitis	Number of Deaths Caused by Meningitis	Discrete
Deaths_Neoplasms	Number of Deaths Caused by Neoplasms	Discrete
Deaths_FireHeatHotSubstances	Number of Deaths Caused by Fire, Heat, or Hot Substances	Discrete
Deaths_Malaria	Number of Deaths Caused by Malaria	Discrete
Deaths_Drowning	Number of Deaths Caused by Drowning	Discrete
Deaths_InterpersonalViolence	Number of Deaths Caused by Interpersonal Violence	Discrete
Deaths_HIVAIDS	Number of Deaths Caused by HIV/AIDS	Discrete
Deaths_DrugUseDisorders	Number of Deaths Caused by Drug Use Disorders	Discrete
Deaths_Tuberculosis	Number of Deaths Caused by Tuberculosis	Discrete
Deaths_RoadInjuries	Number of Deaths Caused by Road Injuries	Discrete
Deaths_MaternalDisorders	Number of Deaths Caused by Maternal Disorders	Discrete
Deaths_LowerRespiratoryInfections	Number of Deaths Caused by Lower Respiratory Infections	Discrete
Deaths_NeonatalDisorders	Number of Deaths Caused by Neonatal Disorders	Discrete
Deaths_AlcoholUseDisorders	Number of Deaths Caused by Alcohol Use Disorders	Discrete
Deaths_ExposureToForcesOfNature	Number of Deaths Caused by Exposure to Forces of Nature	Discrete
Deaths_DiarrhealDiseases	Number of Deaths Caused by Diarrheal Diseases	Discrete
Deaths_EnvironmentalHeatAndColdExposure	Number of Deaths Caused by Environmental Heat and Cold Exposure	Discrete
Deaths_NutritionalDeficiencies	Number of Deaths Caused by Nutritional Deficiencies	Discrete
Deaths_Selfharm	Number of Deaths Caused by Self-Harm	Discrete
Deaths_ConflictAndTerrorism	Number of Deaths Caused by Conflict and Terrorism	Discrete

Name	Description	Type
Deaths_DiabetesMellitus	Number of Deaths Caused by Diabetes Mellitus	Discrete
Deaths_Poisonings	Number of Deaths Caused by Poisoning	Discrete
Deaths_ProteinEnergyMalnutrition	Number of Deaths Caused by Protein Energy Malnutrition	Discrete
Deaths_CardiovascularDiseases	Number of Deaths Caused by Cardiovascular Diseases	Discrete
Deaths_ChronicKidneyDisease	Number of Deaths Caused by Chronic Kidney Disease	Discrete
Deaths_ChronicRespiratoryDiseases	Number of Deaths Caused by Chronic Respiratory Diseases	Discrete
Deaths_CirrhosisOtherChronicLiverDiseases	Number of Deaths Caused by Cirrhosis or Other Chronic Liver Diseases	Discrete
Deaths_DigestiveDiseases	Number of Deaths Caused by Digestive Diseases	Discrete
Deaths_AcuteHepatitis	Number of Deaths Caused by Acute Hepatitis	Discrete
Deaths_AlzheimersDiseaseOtherDementias	Number of Deaths Caused by Alzheimers Disease or Other Dementias	Discrete
Deaths_ParkinsonsDisease	Number of Deaths Caused by Parkinsons Disease	Discrete

Introduction

Methods

Statistical Modeling

Conlusion

References