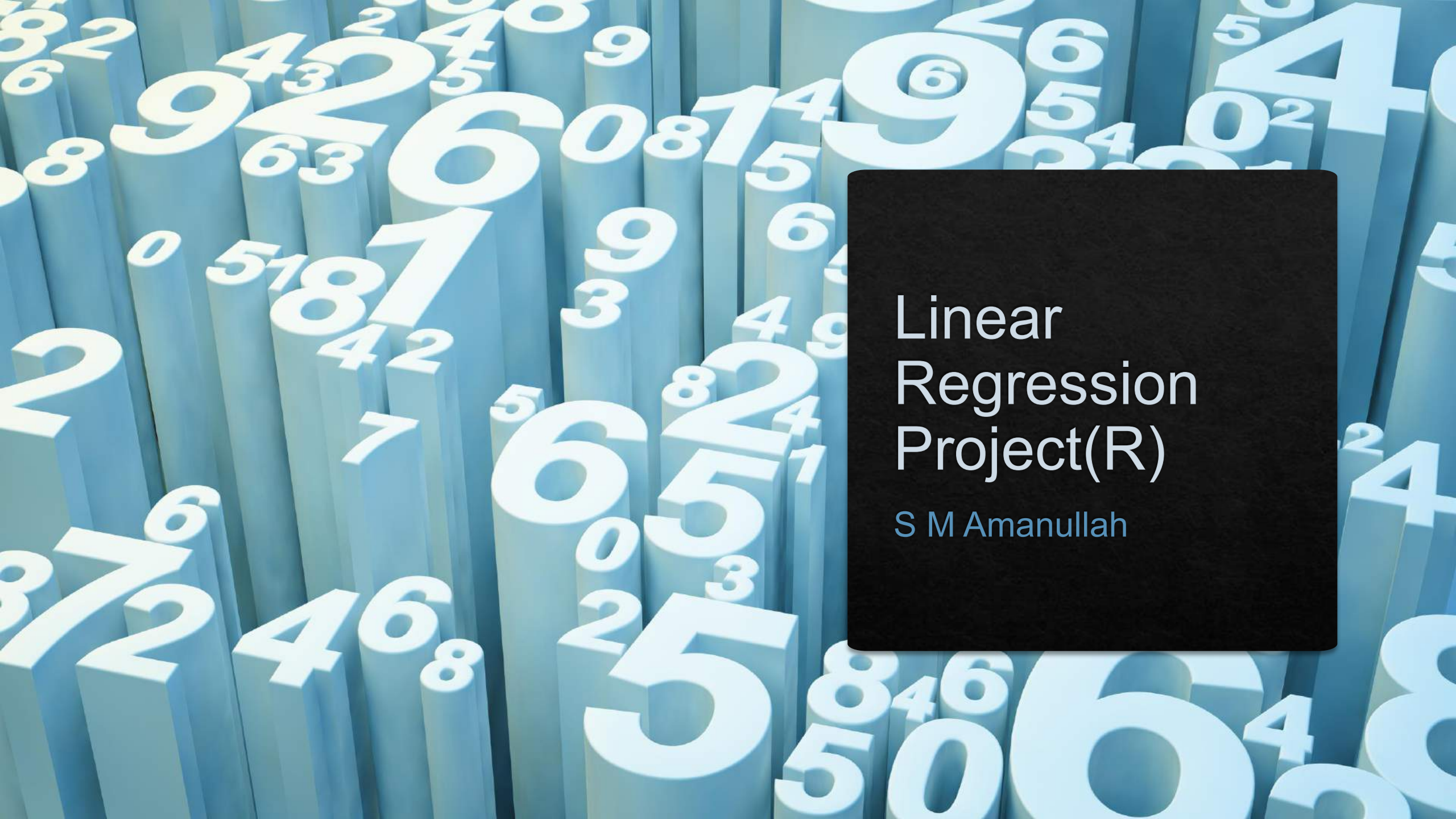


Linear Regression Project(R)

S M Amanullah

Objectives

- The dataset is related to life expectancy with many factors in the 4 significant categories being health, social, economic, mortality, and immunization. We have to predict next year value of average life expectancy using linear regression.

- Considering the objective, using correlation function, we can accurately predict life expectancy.

- By knowing which factors have the most significant role in a lower life expectancy, a country can decide to spend more money & resource on those certain things.

Count of Variables & Observations

- Total variables: 23
- Target variable is Life Expectancy.
- Variables which were correlated with the target variable:
  - Status
  - Adult Mortality
  - Percentage Expenditure
  - BMI
  - Polio
  - Diphtheria
  - Per Capita GDP
  - Income Composition of Resources
  - Schooling
- Left out variables were not correlated or had very less relation with the Target Variable.

Significant Variables

- Status
- Percentage Expenditure
- Polio
- Per Capita GDP
- Schooling
- Adult Mortality
- BMI
- Diphtheria
- Income Composition of Resources

Insignificant Variables

- Country
- Year
- Life Expectancy
- Infant Deaths
- Alcohol
- Hepatitis-B
- Measles
- Under five Deaths
- Total Expenditure
- HIV/AIDS
- GDP
- Population
- Thinnes_1-19_Years
- Thinness_5-7_Years

Explanation of Significant Variables-1

- Status (Developed or Developing status)
- Percentage Expenditure (Expenditure on health as a percentage of Gross Domestic Product per capita(%).)

  Affects because it has direct relationship with life expectancy.
- Polio (Pol3) immunization coverage among 1-year-olds (%))

  Between 5% and 10% of people who develop paralytic polio die.
- Per Capita GDP (GDP per capita is gross domestic product divided by midyear population.)

  Higher values of GDP per capita leads to higher life expectancy at birth

  suggesting that longevity of people.

Explanation of Significant Variables-2

- Schooling :Number of years of Schooling in years.

  People with more education have significantly longer life expectancy.

- Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population).

  Directly relates to the life expectancy.

- BMI: Average Body Mass Index of entire population

  Those who are extremely obese, life expectancy may be reduced by an estimated five to 20 years.

- Diphtheria: Diphtheria tetanus toxoid and pertussis (DTP3) immunization coverage among 1-year-olds (%)

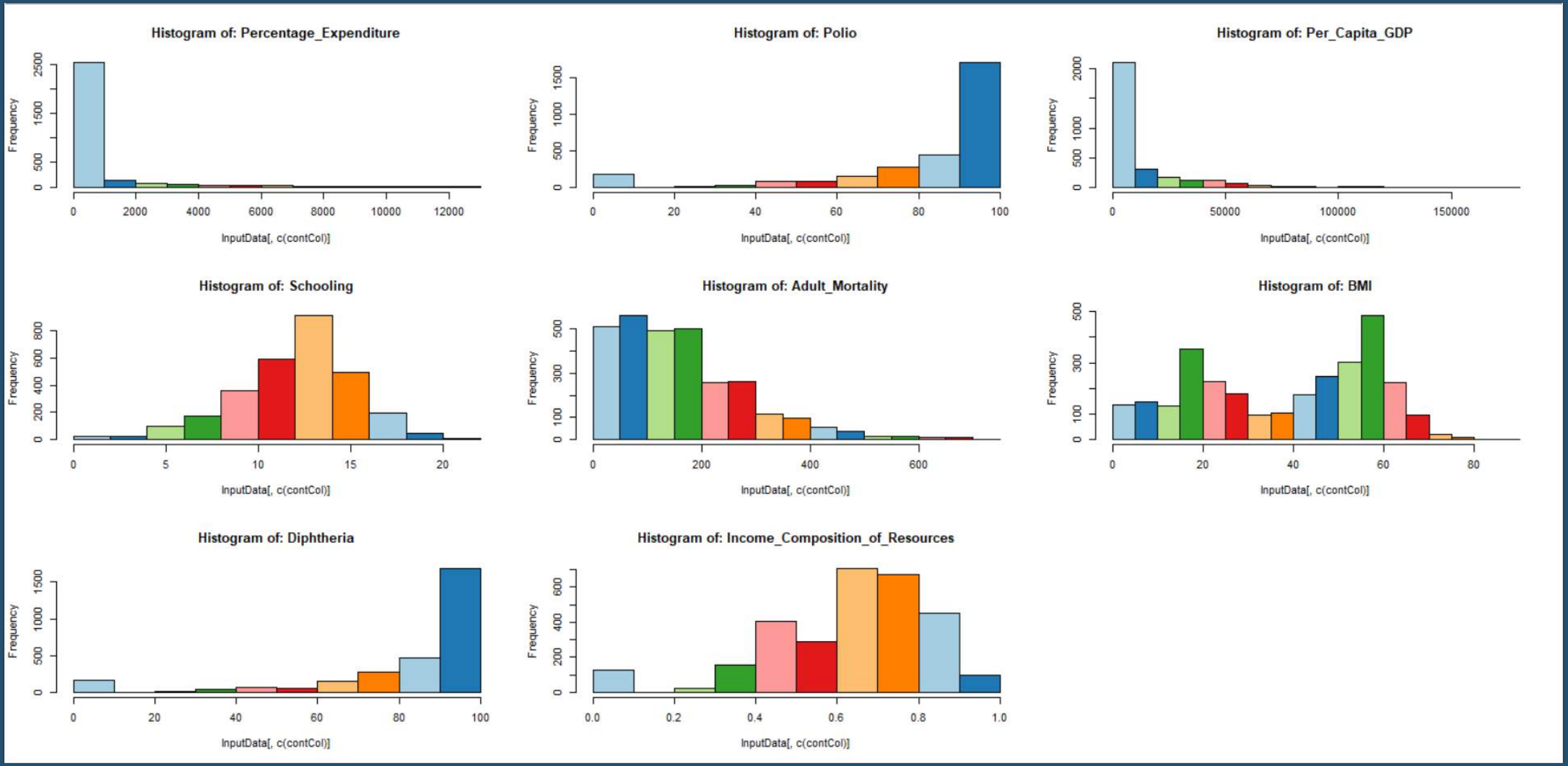
  It is an immunity booster so affects Life expectancy.

- Income Composition of Resources: Human Development Index in terms of income composition of resources (index ranging from 0 to 1)

  Have the highest correlation coefficient of 0.91 which means that if a country

  utilizes its resources productively, it is more likely to see its citizens live longer than expected.

Graphical display of Significant Variables



Results Obtained

- We can predict a country's life expectancy as infrastructure, society, and resources change over the years and it depends on the status, adult mortality, percentage expenditure, BMI, polio, per capita GDP, diphtheria, income, schooling. By focusing on these factor we can improve life expectancy of that country and a country can decide to spend more money & resource on those certain things

- Mean accuracy of the Linear Model is 94.57.
- Median Accuracy of Linear Model is: 96.33.
- Multiple R-Square error: 0.7649
- Adjusted R-Square error: 0.7639
- Durbin Watson Test

  lag Autocorrelation D-W Statistic p-value

  1 -0.01324788 2.025856 0.552

Thank You!