Examine the Data

Reflect on the Question | Analyze the Data | Draw Conclusions

Lab 6: Worldwide Trends in Internet Usage

The World Bank is a data collection of information on all the world's countries. Data is collected by country, and include items such as total population, CO^2 emissions, and the number of mobile device subscriptions. We will examine some of the trends in this dataset and interpret the parameters of the fitted models to best describe the change over time.

Primary Research Question

What model best describes the first decade of internet usage (1990-1999) in the United States? Which model is a better long-term fit?

---

problem

3/3 points (graded)

Check the Data

Let's begin by examining our data in R.

1. Open RStudio. Make sure you've installed the **current version** of the SDSFoundations package.
2. Type **`library(SDSFoundations)`** This will automatically load the data for the labs.
3.Type **`world <- WorldBankData`** This will assign the data to your Workspace.
4. Look at the spreadsheet view of the data to answer the following questions.

Alternatively, you can use follow the steps in the "Importing a Data Frame" R tutorial video, and use the WorldBankData.csv file. (Right-click and "Save As.") Make sure to **name** the dataframe "world" when importing.

1. Open RStudio.
2. Click on "Import Dataset" button at the top of the workspace window. Choose *"from text file."*
3. Click on the location of the WorldBankData.csv file you just downloaded.
4. Click on the WorldBankData.csv file. Then, click Upload.
5. Look at the spreadsheet view of the data to answer the following questions.

1a) What is the first “Low Income” country in the dataset?

Afghanistan ✔

1b) What was the rural population of Aruba in 1970? *(Report without commas)*

29164 ✔

29164

1c) When was the first year Australia had data on the number of mobile device subscriptions? (Subscriptions more than 0)

1987 ✔

1987

Submit You have used 1 of 1 attempt

✔ Correct (3/3 points)

problem

4/4 points (graded)

Check the Variables of Interest

Let's find the variables we need to answer the question.

2a) Which variable tells us the *number* of internet users in a specific year? The variable name in the dataset is:

internet.users ✔

2b) What type of variable is this?

quantitative ✔

2c) Which variable tells us *when* the number of internet users was recorded? The variable name in the dataset is:

year ✔

2d) What type of variable is this?

quantitative ✔

Submit You have used 1 of 1 attempt

✔ Correct (4/4 points)

problem

2/2 points (graded)

Reflect on the Method

Which method should we be using for this analysis and why?

3a) What **statistic** helps us determine how well a particular model fits the data?

- Variance
- R-squared ✔

- Mean

3b) In this lab, we will calculate **residuals** after fitting both an exponential and a logistic model to a set of data. What is a **residual**?

- A residual is the difference between a predicted value and the actual observed value. ✔
- A residual is the average distance of a data point to the line of best fit.
- A residual is a statistic that tells you how much variation is explained by the model.

Submit You have used 1 of 1 attempt

✔ Correct (2/2 points)