**Courseware**     **Course Info**     **Discussion**     **Syllabus**     **Download R and RStudio**     **R Tutorials**     **Readings**     **Contact Us**

**Progress**     **Office Hours**     **Community**

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? Does this model hold through 2012?

**Help**

(3/3 points)

## Check the Data

Let's begin by examining our data in R.

1. Open RStudio. Make sure you've installed 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

**Help**

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

1987

1987

Show Answer      *You have used 1 of 2 submissions*

(4/4 points)

# 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

**Help**

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

Show Answer      *You have used 2 of 2 submissions*

(1/2 points)

# 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

**Help**

**CORRECT. R-SQUARED CAN BE CALCULATED FOR A LINEAR, EXPONENTIAL, OR LOGISTIC MODEL AND IS A GOOD MEASURE OF HOW THE WELL A PARTICULAR MODEL FITS THE DATA POINTS IN A SAMPLE--THAT IS, HOW WELL THE MODEL EXPLAINS THE PATTERN IN THE DATASET.**

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.

**INCORRECT**

The distance of a data point from the line of best fit is a single numerical value. Therefore, it does not make sense to find an "average" distance.

**CORRECT. A RESIDUAL IS THE ACTUAL VALUE MINUS THE PREDICTED VALUE OF A MODEL. IN OTHER WORDS, IT IS A MEASURE OF THE DISTANCE ALONG THE Y-AXIS BETWEEN A DATA POINT AND THE VALUE OF THE MODEL AT A PARTICULAR VALUE OF THE**

**X-VARIABLE.**

Hide Answer          *You have used 2 of 2 submissions*

**Help**

**Help**

EdX offers interactive online classes and MOOCs from the world's best universities. Online courses from MITx, HarvardX, BerkeleyX, UTx and many other universities. Topics include biology, business, chemistry, computer science, economics, finance, electronics, engineering, food and nutrition, history, humanities, law, literature, math, medicine, music, philosophy, physics, science, statistics and more. EdX is a non-profit online initiative created by founding partners Harvard and MIT.

© 2015 edX Inc.

EdX, Open edX, and the edX and Open edX logos are registered trademarks or trademarks of edX Inc.

Terms of Service and Honor Code

Privacy Policy (Revised 10/22/2014)

## About edX

About

News

Contact

FAQ

edX Blog

Donate to edX

Jobs at edX

## Follow Us

Twitter

Facebook

Meetup

LinkedIn

Google+