Prepare for the Analysis

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?

Breakdown Your Analysis

Let's break this analysis into its required steps:

1. Create a subset of the dataset that contains only the information for the United States.
2. Create a subset of the US data that contains only the years 1990 to 1999.
3. Use a function to fit both an exponential and a logistic model to the data.
4. Using each model, predict the number of internet users in 2006.
5. Compare the size of the residuals for 2006 to determine which model appears to have better long-term fit.

Note: In this lab, you will be creating three new subsets of data by constraining the original world dataset. The names of these three datasets are shown in the table below. A brief description of each is also provided to help you navigate the R code.

name	description of dataset
us	contains only the USA data from the world dataset
us_select	contains only the USA data for the years 1990 and beyond
us_select_10	contains only the USA data for the years 1990 - 1999

Here is the code you will use:

```
# Subset data for just the United States and name the new data frame "us"
us <- world[world$Country.Code == "USA",]

# Select the years from 1990 and name the new data frame "us_select"
us_select <- us[us$year >= 1990, ]

# Create a new variable in our datset called internet.mil to make the number of users
more interpretable (into millions)
us_select$internet.mil <- us_select$internet.users / 1000000

# Create a new variable in our dataset called time that represents "years since 1990"
us_select$time <- us_select$year - 1990

# Select the first 10 years (from 1990 to 1999) and name the new data frame "us_select_10"
us_select_10 <- us_select[us_select$time < 10,]

# Use a function to fit an exponential and logistic model for 1990-1999
expFit(us_select_10$time, us_select_10$internet.mil)
logisticFit(us_select_10$time, us_select_10$internet.mil)

# Based on the prior model parameters, predict the number of internet users in 2006
e <- expFitPred(us_select_10$time, us_select_10$internet.mil, 16)
l <- logisticFitPred(us_select_10$time, us_select_10$internet.mil, 16)
```

problem

1/1 point (graded)

1) In both of these models, what does **Year 0** correspond to?

- ◉ 1990 ✔
- ○ 2012
- ○ 1995
- ○ 2000

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

problem

1/1 point (graded)

2) Which **functions** will you use to fit exponential and logistic models to the data?

- ○ internet.mil()
- ○ us_select
- ◉ expFit() and logisticFit() ✔

- exp(model) and log(model)

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

problem

1/1 point (graded)

3) What will be the value of "**us_select$time**" for 2006?

- 6
- 16 ✔
- 90
- 2006

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

problem

1/1 point (graded)

4) In 1990, the number of internet users in the US was 1,958,863. What will be the value of "**us_select$internet.mil**" for 1990 after this line of code is run?

```
# Create a new variable in our dataset called
internet.mil to make the number of users more
interpretable (into millions)
us_select$internet.mil <- us_select$internet.users /
1000000
```

- 19.58863
- 1.958863 ✔
- 1958,863
- 195.8863

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)

problem

1/1 point (graded)

5) Suppose we wanted to create a subset of data that contained only Belgium entries that were from 1990 and onward. What error was made in the code below? **(Notice that R does not report some types of errors, making it very important to check your data as you go.)**

```
world <- WorldBankData
belgium <- world[world$Country == 'Belgium',]
recent_belgium <- world[world$year >= 1990,]
```

- The "Country" variable has only 3 letter entries.
- We pulled our data from the wrong dataset in the second line.
- We pulled our data from the wrong dataset in the third line. ✔
- There is not a "Country" variable in the dataset to use.

Submit You have used 1 of 1 attempt

✔ Correct (1/1 point)