Lab – Basic Data Analysis

1. Objectives

Use very simple methods to describe existing data, fill-in missing data values and to make simple predictions.

Part 1: Learn how to Use Data as Information

Part 2: Plot data and predict values

1. Background/Scenario

Data is meaningless in and of itself. Information is meaningful and useful. Data only becomes information when it used in context to answer specific questions. In this lab, you will use graphs of existing data to create missing values and to predict values based on trends.

1. Required Resources

* PC or mobile device with Internet access
* Browser capable of playing a video from the Internet.
* Audio capability to listen to video narration.

1. Learn how to Use Data as Information

Data analysis can occur in many different ways. The ultimate goal is to discover something in the data that gives insight into what has happened or to predict what may happen in the future. Descriptive statistics summarizes what happened and provides the data in a numeric or graphical way. Predictive analytics answers the question of what may happen in the future based on past data.

* 1. Describe the data.
     1. Review the data chart shown on Worksheet 1.
     2. Over the 110-year period for the data set, what is the range (lower and upper limits) of median ages for males and females at first marriage?

Male median age: \_\_\_\_\_ to \_\_\_\_\_\_

Female median age: \_\_\_\_\_ to \_\_\_\_\_\_

* + 1. The range is a type of descriptive statistic that summarizes the data. It has been presented above in a numerical format. But if you need to look at the trend of the data, it may be better to graph or plot the data.
  1. Plot the data.
     1. Plot the data on the chart that is found on Worksheet 1 at the end of this lab. Use different colors for the male and female data. **Do not connect the data points until you are told to do so in the video in Step 3**.
  2. Perform simple data prediction.
     1. Khan Academy is an excellent source for lessons on a wide range of topics including statistics and statistical concepts. You will watch a short video and follow along with the narrator using the data table and chart provided at the end of this lab. Navigate to <https://youtu.be/aVDiAGZmcPo> and follow the Predicting with Linear Models video provided by Khan Academy. Pause the video and complete the activities along with the video instructor. Use the plot you created.

Watch the entire video. You will only work with the first dataset that the instructor discusses in this lab. In the video, the instructor demonstrates how to use data points as information to create new estimated data points.

What is a linear model?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The instructor in the video teaches the processes of interpolation and extrapolation as tools for estimating or predicting data in a linear model. Define each term.

interpolation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

extrapolation \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

What are two interesting observations that the instructor in the video makes regarding the trends in the median age of marriage and the ages of the males and females who marry?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. Plot data and Predict Values

Other social trends can make use of the simple linear model that was demonstrated in the video. In this section of the lab, you will interpolate and extrapolate values from a new data set.

* + 1. Plot the data on Worksheet 2 at the end of the lab. Use different colors for the two different variables.
    2. Note that the data was collected at different intervals. Before 2000, the data was collected every ten years, however, after 2000 it was collected every 5 years. Interpolate values for the three missing years. Plot your values.

|  |  |  |
| --- | --- | --- |
| Missing Year | Woman hours | Men Hours |
| 1970 |  |  |
| 1980 |  |  |
| 1990 |  |  |

* + 1. Extrapolate values for the year 2020 by creating a line that best summarizes the values for the previous five periods.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* + 1. Another kind of information that can be derived from this data is about the gap between the number of hours of housework for men versus the number of hours of housework for women. This will display another trend regarding the equality between men and woman over this period. Complete the table below by filling it in with the amount of time that women do house work subtracted from the amount of time that men do housework.

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Men housework hours/week** | **Woman housework hours/week** | **Women hours - Men hours** |
| 1965 | 4.4 | 31.9 |  |
| 1975 | 6 | 23.6 |  |
| 1985 | 10.2 | 20.7 |  |
| 1995 | 10.2 | 18.9 |  |
| 2000 | 10 | 18.6 |  |
| 2005 | 9.2 | 19.1 |  |
| 2010 | 10 | 17.4 |  |
| 2015 | 9.8 | 17.8 |  |

* + 1. Graph the calculated values on the chart provided on Worksheet 3 at the end of this lab.

What has been the trend for equality between men and women doing housework?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

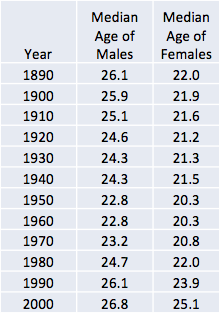
If men and woman were completely equivalent in the amount of housework they do per week, in 2020, where would the next data point be plotted?

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

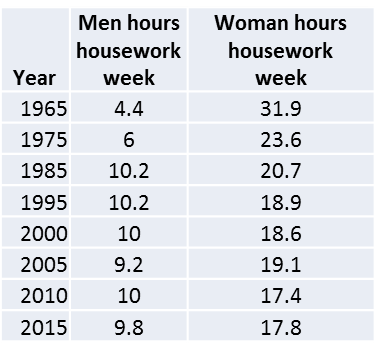
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

In the IoT, Big Data comes from many sources. Sometimes values are missing because a sensor temporarily lost connectivity or data points were lost in transmission. Interpolation can serve as one strategy for replacing missing data. Extrapolation is used to predict values for events that have not yet occurred. Because the IoT yields so much data, predictive analytic models can be built that reliably see into the future by extrapolating trends from historical data.

Worksheet 1



Worksheet 2



Worksheet 3

