**Capstone Project: Youth Smoking**

**Introduction:   
The Capstone Project is a semester long, critical thinking project utilizing a single, large data set. You will be examining this data from the point of view of the concepts in each chapter of the statistics course. The goal is to give you experience in manipulating and analyzing data and to help you understand how decisions are made and conclusions are reached using data.**

**Intermediate assessments will have deadlines scattered throughout the semester. The assignments range from answering questions and making graphs to running hypothesis tests and drawing conclusions regarding the meaning of the data set. For all written responses you are expected to use proper English grammar and to write in complete sentences and, where needed, paragraphs. Statistical graphs should be properly labeled and in a format large and clear enough to be easily understood.**

**Links to the original raw data will be provided as needed.**

***Adapted from data and article in “Journal of Statistics Education” by Michael Kahn, Wheaton College.***

**Part I: Chapter 1: Sampling**

**The data set that you will be using is a study of the forced expiratory volume (FEV) of 654 youths, aged 3 to 19, in East Boston. The data was gathered by researchers in the mid to late 1970’s and measures five variables for each participant:**

|  |  |
| --- | --- |
| **Variable Label** | **Measured In** |
| **Age** | **Years** |
| **FEV (forced expiratory volume)** | **Liters** |
| **Ht (height)** | **Inches** |
| **Sex** | **Female coded 0, Male coded 1** |
| **Smoke (Smoking status)** | **Nonsmoker coded 0, Smoker coded 1** |

**Provide link to data set here.**

**Based on this description of the data and research setup, answer the following questions.**

**Task 1:**

* **Would this be considered an observational or experimental research project? Explain your choice.**
* **Could you expect the conclusions you reach in analyzing this data to apply to all youth, aged 3-19, in the United States? Explain why or why not.**

**Task 2: Determine the level of measurement for each of the 5 variables.**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Level of Measurement** | **Explanation** |
| **Age** |  |  |
| **FEV** |  |  |
| **Height** |  |  |
| **Sex** |  |  |
| **Smoking Status** |  |  |

**Task 3:**

* **What kind of sampling technique appears to be used?**
* **What problems can you identify with the sampling technique used?**