**Program Exercise:**

We have been given an abridged version of the Pennsylvania Hospital Discharge Dataset (PHC4). The data is at the discharge level—that is, there is one observation per patient-discharge. These data contain all hospital discharges within the state of Pennsylvania over the time period of the data. Discharges are clustered within attending physicians. Because most physicians take care of multiple patients, each physician in this data generally has multiple discharges.

From this dataset we would like you to use SAS or STATA to create an analytical dataset. The analytical dataset is a physician-level dataset, where the unit of observation is the attending physician who delivers babies. All attending physicians who deliver babies in Pennsylvania should be in this file and physicians who do not deliver babies should be excluded. The analytical dataset should include the following variables: the id# of the doctor, the number of babies delivered, the number of c-sections performed, and, for each physician, the percentage of patients who are black (race=”B”).

We know that the procedure codes indicating a delivery are:

720, 721, 724, 726, 728, 729, 731, 733, 736, 738, 740, 741, 742, 744

The procedure codes for deliveries by C-section are:

740, 741, 742, 744

Once you have created this dataset, please summarize and include the following:

1. Number of doctors in the state delivering babies - 177
2. Average number of deliveries and c-sections per doctor – 15-16 deliveries,6-7 c-sections per doctor
3. An easy way to understand the distribution of deliveries per doctor. Is this a skewed distribution? What are other important characteristics of this distribution – right skewed, box plot, one outliner

![Chart, histogram

Description automatically generated]()

1. The percentage of mothers delivering babies who are black – 4.52% of them

Table

Description automatically generated

1. Are the patients of high-volume doctors more likely to be black than the patients of low-volume doctors? – Patients by high-volume doctors are more likely to be black than those of low-volume doctors

Table

Description automatically generated