**Analytics Assignment**

**Format: Powerpoint to answer questions accompanied by STATA do file. Use one slide per question (e.g. part a, slide 1, part b, slide 2…). Please answer each question part on a single slide and format appropriately as if this were a presentation. You must submit a “log file” and a “do file” as well via email.**

1. **Download the 2016 SPARCS Dataset available from Health.Data.NY.gov (Hospital Inpatient Discharges (SPARCS De-Identified))**

*(I recommend using the “insheet” command to import the data, keeping only the data that you need and then saving the data in STATA format (.dta))*

1. Where did patients get discharged (e.g. home, hospice, etc) to when they were discharged from the hospital? Where did the men go and where did the women go? You must use the data dictionary and codebook to determine this. Show your answer in a table with percentages.
2. In a table, show the average length of stay of patients at every hospital in Westchester county in 2016. Create a column for Medicare, Medicaid, Commercially Insured and Other. List the hospitals vertically in rows and make the insurance types the columns and include totals both vertically and horizontally. Show as both absolute counts and percentages.

*Note1: Use the table below to better classify insurance type*

|  |  |
| --- | --- |
| **SPARCS Description** | **HPM 8015 Description** |
| Blue Cross/Blue Shield | Commercial |
| Department of Corrections | Other |
| Federal/State/Local/VA | Other |
| Managed Care, Unspecified | Commercial |
| Medicaid | Medicaid |
| Medicare | Medicare |
| Miscellaneous/Other | Other |
| Private Health Insurance | Commercial |

*Note 2: Length of Stay is a “String” variable and must be converted into a numeric variable*

*The code for this is*

**destring lengthofstay, replace force**

1. What are the license numbers for the top 3 physicians (attending physicians) delivering newborns in Westchester hospitals in 2016? (Hint: There is a shortcut to figuring out who is a newborn…read the data codebook!)
2. What were the average total costs of the deliveries with those three physicians by facility?

You will need to use the code below because for some reason the “cost” data is recorded as a string (meaning STATA reads it as a word not a number)

\*First line of code eliminates the decimal point

**replace totalcosts=subinstr(totalcosts,".","",.)**

\*Second line of code eliminates the “comma”

**replace totalcosts=subinstr(totalcosts,",","",.)**

\*Third line of code changes the “string” to a numeric variable so we can add/subtract/multiply and the second line

**destring totalcosts, replace force**

Fourth line of code divides by 100 since we eliminated the decimal point.

**replace totalcosts=totalcosts/100**

1. What are the average costs for newborns by hospital? Which hospital is the outlier? Why do you think that is? What portion were over $100,000 and what portion were over $1,000,000?
2. Tell me something I didn’t know about the data and why it is important.