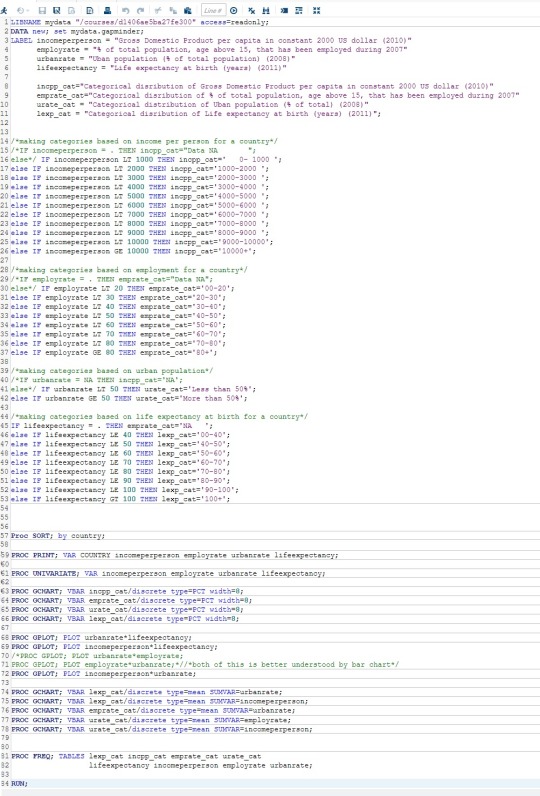
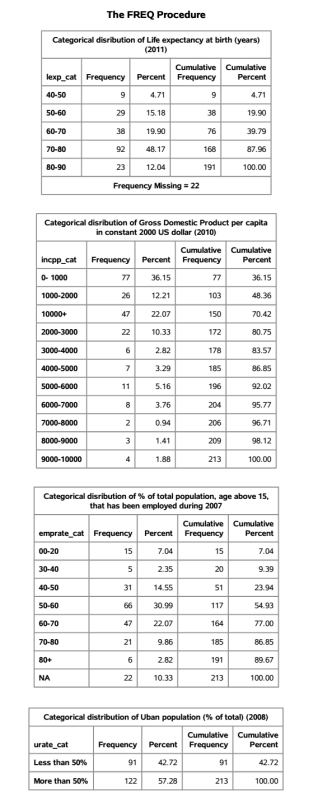
Assignment 3

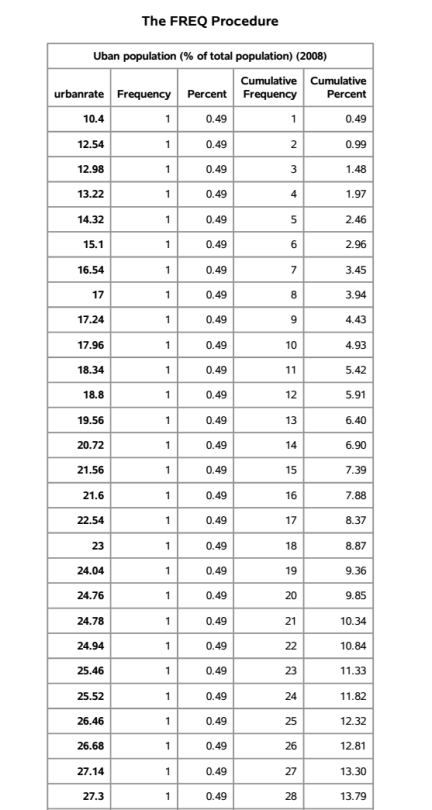
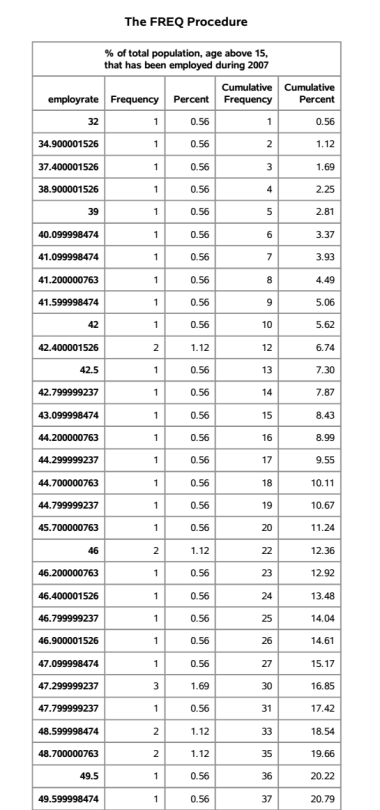
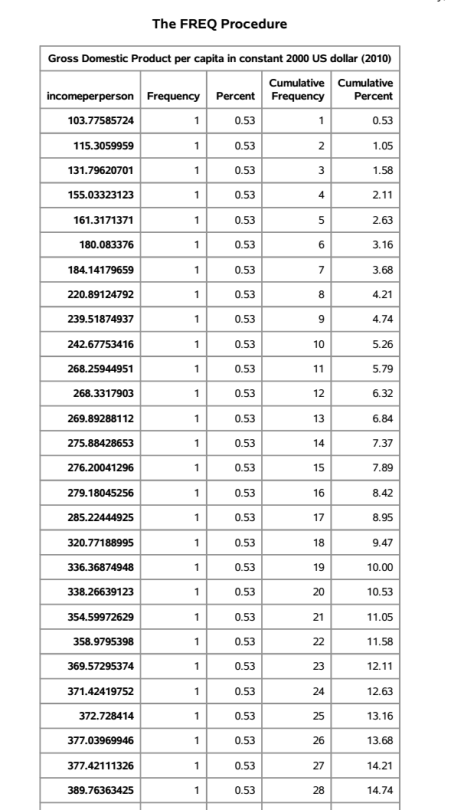
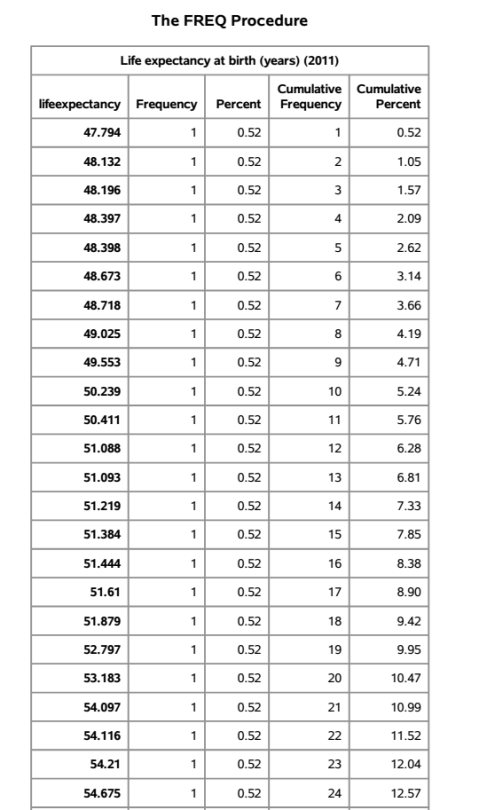
**Remark**: While working I wrote down enough code and for this assignment I haven’t shorten it and though this program will give various result, I will only be showing the asked results for this particular assignment.



Note: As I copied the image from PDF its color scheme has changed and also for some variable I won’t be able to showcase their full freq.

Procedure.





Data management includes such things as coding out missing data, coding in valid data, recoding variables, creating secondary variables and binning or grouping variables. Not everyone does all of these, but some is required

As for the data management objective I collapsed the responses for **incomeperperson, employrate, urbanrate,** and **lifeexpectancy** to create four **new secondary categorical variables** for each of them: **incpp\_cat**, **emprate\_cat**, **urate\_cat**, and **lexp\_cat**. These variable categorize the dataset for the variable into categories. Also the missing data has been coded out wherever its present. Also for more better interpretation I am adding some plotted charts at the end.

For **lexp\_cat**, the most commonly endorsed response was ‘*70-80*′ (48.17%), meaning that most countries have an life expectancy in the age group of 70 and 80 years. Also among the 213 countries we are missing the data for 22 of them for this variable.

For **incpp\_cat**, the most commonly endorsed response was ‘*0-1000*′ (36.15%), meaning that most countries GDP per capital lies in constant 2000 $ lies in 0 - 1000 $ group. Also looking at the data it can be said that the plot will have a skewed right distribution.

For **emprate\_cat**, the most common endorsed response was ‘*50-60*′ (30.99%), meaning that for many countries comes in the range of having in between 50 to 60 % of total population, age above 15, that has been employed

during 2007. By looking at the data it seem this variable is having an approximate Gaussian curve with about center symmetric around 50-60 range. Also from among the 213 countries, dataset is missing the values for 22 countries.

For **urate\_cat**, the most commonly endorsed response was ‘*More than 50%*’ (57.28%), meaning that among the 213 countries, 122 countries are having majority of its population living in urban areas and the majority population for the rest 91 countries is living in rural areas.

The above discussed variable significantly help in understanding the result and making the result concise that otherwise wold have been difficult to be discussed with the freq procedure for incomeperperson, employrate, urbanrate, and lifeexpectancy.

