**R assignment 3 -- Epi 510, Autumn 2020**

For this assignment, please submit two files: (1) **a file containing text, tables and figures** (.doc, .docx, .pdf or .txt) responding to questions posed in the assignment, and (2) **an R script** that performs the requested operations (.R).

1. There are five files called “gbdChildMortality\_1970s.csv” through “gbdChildMortality\_2010s.csv”. Use a for loop and the rbind function to loop over and append these into a single data frame. **(10 points)**
2. Read in “countryCovars.csv” and merge it to the data frame that you created in question 1. When you merge, keep all observations from the child mortality datasets, and don’t keep those observations that were found only in the countryCovars dataset. You’ll need to take a look at the structure of these two, and the variable names to determine the “by” variables on which to merge. **(5 points)**

You now have a full analysis dataset. There are no special missing codes (i.e. all missing values should be coded as NA) and there’s no other cleaning to do.

1. We want to make a histogram for each mortality rate variable, but we’re ~~lazy~~ efficient coders, so we don’t want to have to type out the variable list or copy and paste the histogram code.

A note: You can choose to get either the column numbers or names from the grep command in 3a (depends on how you set the values option). Either one will work in the loop for 3b. The way you create the labels for the histogram in 3b will differ depending on the approach you choose. You might choose to experiment with both approaches to get more practice and to ensure that you really understand the material (but you need only submit one with your HW).

* 1. Use a grep command to find all of the variables that contain “MR” and assign these to an object. Now we have our list of variables to loop through. **(6 points)**
  2. Use a for loop to loop over the variables in the object that you created in 3a. Have the loop make a histogram for each MR variable. Note: the R function to produce histograms is hist (we’ll cover this more next week). Paste the histograms into your word document. **(6 points)**

1. We’re going to use sapply functions to build a table of means and SDs:
2. Use an sapply function to find the mean value for all MR and Death variables (i.e. columns 5 through 12). Assign the result to an object called “means”. **(6 points)**
3. Use an sapply function to find the SD for all MR and Death variables (i.e. columns 5 through 12). Assign the result to an object called “SDs”. **(6 points)**
4. Use a cbind function to combine the contents of “means” and “SDs” into a single table. Paste the table into your Word document. **(5 points)**

A screenshot of a computer

Description automatically generated with low confidence

1. Now we’re going to practice using tapply to summarize a variable by levels of another variable.
2. Use the tapply function to find the mean value of neoMR for each year. Paste the output into your word document. **(6 points)**

1970 1971 1972 1973 1974 1975 1976 1977 1978

40.07703 39.09589 38.12427 37.06532 36.13720 35.22667 34.31812 33.43048 32.58387

1979 1980 1981 1982 1983 1984 1985 1986 1987

31.74715 30.94866 30.13677 29.35812 28.61586 27.91887 27.25263 26.61973 26.03070

1988 1989 1990 1991 1992 1993 1994 1995 1996

25.54376 24.98457 24.50844 24.07021 23.62620 23.23658 22.70674 22.24209 21.79390

1997 1998 1999 2000 2001 2002 2003 2004 2005

21.28850 20.79433 20.27513 19.74802 19.23070 18.73230 18.24786 17.77642 17.34824

2006 2007 2008 2009 2010

16.92647 16.53011 16.15316 15.79658 15.50155

1. **Bonus question** (this isn’t required to get full credit on the HW, but is a good exercise to test your understanding of apply functions, and see their power): Nest a tapply function inside of an sapply function to make a table that gives the mean value for each MR and Death variable in each year. **(5 points extra credit)**