1. Robustness of correlation (written):

Suppose you have two data vectors \( x \) and \( y \) for which the correlation \( r_{xy} \) is strictly between -1 and 1. For example, suppose it is zero. Suppose you are allowed to change one data pair, say \((x_1, y_1)\), to be any two numbers. What is the range of possible values of the correlation coefficient as \((x_1, y_1)\) are allowed to vary? Show your result mathematically. Non-mathematically, how would you describe the robustness of the correlation coefficient to outliers?

2. Invariance of correlation (written):

Let \( x \) and \( y \) be two data vectors. Show that \( \text{Cor}(ax + b, cy + d) = \text{Cor}(x, y) \) if \( a \) and \( c \) are both greater than zero. What happens if \( a > 0 \) but \( c < 0 \)?

3. GDP and mortality (Rmd): The data file \( u5mortality_2013.csv \) includes data on childhood mortality rates in 2013 (per 1000 live births) and GDP per capita in 2003 (in US dollars). The data are lagged this way to assess the effect of current GDP on future mortality. These data were originally obtained from http://data.worldbank.org/ and may be read into R with the command

\[ \text{read.csv}("http://www.stat.washington.edu/~pdhoff/courses/423/Data/u5mortality_2013.csv") \]

(a) Make univariate histograms of the raw data, and a bivariate scatterplot with GDP on the \( x \)-axis and mortality on the \( y \) axis. Briefly describe in words what these plots show in terms of the distribution of the two variables and their association.

(b) Compute the sample correlation of the raw data. Also, letting \( x^\alpha = (x_1^\alpha, \ldots, x_n^\alpha) \) and defining \( y^\alpha \) similarly, compute \( \text{Cor}(x^\alpha, y^\alpha) \) for each \( \alpha \in \{1, .5, .25, .125\} \). Describe what is happening to the correlation.

(c) Make univariate histograms and a bivariate scatterplot of the data on the log scale, i.e. of \( \log x_i \) versus \( \log y_i \), and briefly describe in words what these plots show. Compute the sample correlation of these logged variables, and compare to the correlations you obtained in (b). How would you describe the association between GDP and child mortality?