3rd set of SAS assignments

1. Create density function of a normally distributed variable

i) Create a data set with values of the density function of a normal distribution

\[ f(x|\mu, \sigma^2) = \frac{1}{\sigma \sqrt{2\pi}} e^{-1/2(x-\mu)^2/\sigma^2} \]

with parameters \(\mu = 0.8\) and \(\sigma^2 = 0.6\). Therefore, use a do loop in a data step.

\[ \text{DO } x=\text{start} \text{ TO } \text{end} \text{ BY } \text{step}; \]

\[ \text{CALCULATE VALUES OF A NORMAL DENSITY FUNCTION} \]

\[ \text{output;} \]

\[ \text{END;} \]

Let \(x\) run from \(-4\) to 4 in steps of 0.01. Label the created variable.

ii) Read out the maximum value of the density function into a macro variable named \(max\) (Use PROC Means or PROC SQL). Then, add 0.1 to the rounded value of \(max\). Use the following statement:

\[ \text{DATA } \_\text{NULL}_-; \]

\[ \text{call symlput(’max’,round(max_normal,0.1)+0.1);} \]

\[ \text{RUN;} \]

The word \_NULL_- appears in the DATA statement instead of a SAS data set name. \_NULL_- is a special keyword that tells SAS not to bother making a new data set. The SYMPUT command creates macro variables either from variables in a data step or from previous macro variables. Here, the variable \(max\_normal\) is rounded and read into the macro variable \(max\).

Print the value contained in this variable into the log window (use the PUT command).
iii) Plot the created density function. Use the \texttt{goptions} provided for last week’s assignment sheet. Label the axis and save the plot as an Encapsulated Postscript (eps). The range of the vertical axis should be from 0 to $\texttt{max}$.

iv) Create values from a standard normal distribution ($\mu = 0$ and $\sigma^2 = 1$) and plot them together with the values from Task (i) into a graph. Use the \texttt{symbol} options to create differently coloured lines for the plots.

v) Create a SAS \texttt{MACRO} for the steps i) to iv) with the arguments

\texttt{%MACRO(path, dataset, startx, endx, step, mu, sigma)}.

\texttt{path} denotes the path where the .eps graph is written out to, \texttt{dataset} is any name for your data set, \texttt{startx} (\texttt{endx}) are any starting (ending) values for which to compute the density function and \texttt{mu} and \texttt{sigma} are the parameters of your normal distribution.