Problem 1

(a) Generate a vector \( \mathbf{x} \) containing 1000 independent replicates of a \( N(10, 4^2) \) random variable. Draw a histogram of \( \mathbf{x} \) using the \texttt{prob=TRUE} option. I suggest that you set \texttt{breaks=20} in the \texttt{hist} command — see \texttt{help(hist)} for an explanation of what this does.

Superimpose the probability density function of the \( N(10, 4^2) \) distribution on your histogram.

(b) Generate \( \mathbf{y}_1 \) containing 400 realisations of a \( N(3, 3^2) \) random variable, \( \mathbf{y}_2 \) containing 400 realisations of a \( N(5, 4^2) \) RV, and \( \mathbf{y}_3 \) containing 400 realisations of a \( N(7, 5^2) \) RV, all variables being independent. Calculate \( \mathbf{w} = \mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3 \) and draw a histogram of \( \mathbf{w} \), using the \texttt{breaks} option to get a suitable number of bars in the histogram.

(c) Guess the distribution of \( W = Y_1 + Y_2 + Y_3 \) when \( Y_1 \sim N(3, 9) \), \( Y_2 \sim N(5, 16) \) and \( Y_3 \sim N(7, 25) \) are independent RVs.

Re-draw the histogram of \( \mathbf{w} \) using the \texttt{prob=TRUE} option and superimpose the probability density function for your chosen distribution. Remember, in the \texttt{curve} command you have to give a function of \( x \) — not \( w \) — to be plotted (see “Plotting commands” on page 1).

(d) Carry out further simulations to give a more definitive check that your chosen PDF really is correct.

In your submission: In a Word document, include the content of the R console window showing R commands and the resulting output. You should also include the histograms of \( \mathbf{x} \) and \( \mathbf{w} \) from the graphics window and an explanation for your choice of distribution in part (c).

See “Graphics”, “Writing a plot directly to a file” and “Combining material in your coursework solutions” in the Brief Introduction to R.

Problem 2

Hermione and Ron have won prizes with random values. Hermione’s prize will be \( H = 100 \exp(X) \) Sickles and Ron’s prize will be \( R = 100 \exp(Y) \) Sickles, where \( X \) and \( Y \) are independent \( N(0, 0.3^2) \) random variables.

(a) Write R commands to simulate one pair of values of \((H, R)\) and compute \( W = H/R \). Create a loop to run the above commands 200 times. Store the 200 values of \( H \) in a vector \texttt{Hsample}, the 200 values of \( R \) in a vector \texttt{Rsample} and store the ratios, \( W \), in a vector \texttt{Wsample}.

Draw histograms of the data in \texttt{Hsample}, \texttt{Rsample} and \texttt{Wsample}.

(b) Professor Slughorn claims the data generated in \texttt{Wsample} should follow a distribution with PDF \( f_W(w) = 0 \) for \( w \leq 0 \) and

\[
f_W(w) = \frac{1}{w\sqrt{0.036\pi}} \exp\left\{-\frac{(\log(w))^2}{0.36}\right\} \quad \text{for} \quad w > 0.
\]  

(1)
Investigate Professor Slughorn’s claim by superimposing the above density on a histogram of \( W_{\text{sample}} \). In order to obtain a useful plot, you may find it necessary to restrict attention to values of \( W \) below an upper limit, such as 5. You can do this with the command

\[
\text{hist}(W_{\text{sample}}[W_{\text{sample}}<5], \text{prob}=\text{TRUE}).
\]

Do your results support the Professor’s theory?

(c)

Professor Slughorn explains that the CDF corresponding to the PDF (1) is

\[
F_W(w) = \Phi \left( \frac{\log(w)}{0.18} \right)
\]

where \( \Phi \) is the standard normal CDF.

Evaluate the function \( F_W(w) \) at \( w \in \{0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8, 2\} \).

Find the proportions of values in \( W_{\text{sample}} \) less than each of these values of \( w \) and plot the proportions against \( F_W(0.2), F_W(0.4), \ldots, F_W(1.8), F_W(2) \). What does this plot show?

(d) Repeat the comparison conducted in part (c) for larger data sets and state whether you believe Professor Slughorn’s claim is correct.

In your submission: Give the content of the R console window; include the relevant plots; give clear written answers in parts (b), (c) and (d).