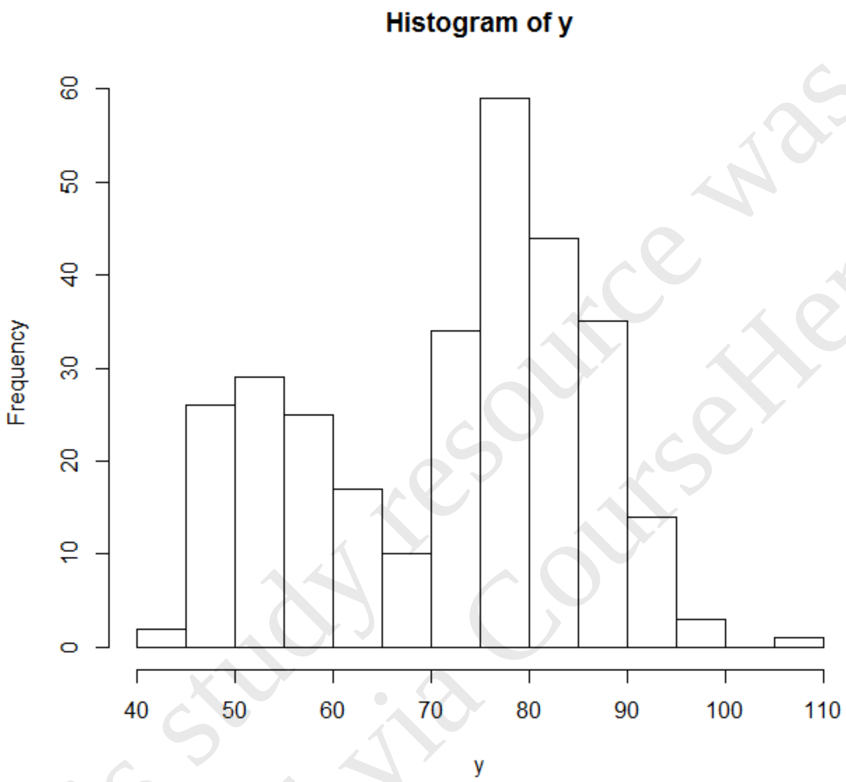
# MAST30027: Modern Applied Statistics

## Assignment 6

Due: 1:00 pm Fri 23 October (week 12)

*This assignment is worth 3 1/3% of your total mark.*

Here is a histogram of 299 observations of the time between eruptions for the Old Faithful geyser in Yellowstone National Park. The data can be found in the file `geyserdata.txt` on the subject website.






We will model this data using a mixture of two normals, which has density

$$f(x) = \pi\sigma_1^{-1}\phi((x-\mu_1)/\sigma_1) + (1-\pi)\sigma_2^{-1}\phi((x-\mu_2)/\sigma_2)$$

where $\phi$ is the standard normal density, $\pi \in [0,1]$, $\mu_i \in \mathbb{R}$ and $\sigma_i \in (0,\infty)$.

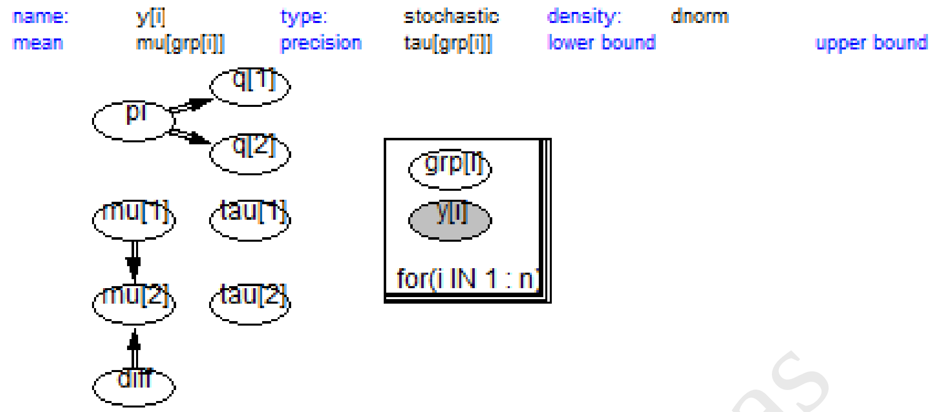
(a) Show that if $A \sim \text{bin}(1,\pi)$, $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$, all independent of each other, then $Y = AX_1 + (1-A)X_2$ has density $f$.

(b) To build a Bayesian model for $Y(i)$, the $i$-th observation, we introduce variables $G(i)$ such that if $G(i) = j$ then $Y(i)$ has mean $\mu_j$ and precision $\tau_j = 1/\sigma_j^2$, for $j = 1, 2$. Moreover $\mathbb{P}(G(i) = 1) = \pi$ and $\mathbb{P}(G(i) = 2) = 1 - \pi$. Thus, to specify the model we need priors for $\pi$, $\mu_j$ and $\tau_j$.

To avoid ambiguity, we suppose that $\mu_1 < \mu_2$. To achieve this we use a $N(60, 1000)$ prior for $\mu_1$ and a $\Gamma(1,1)$ prior for $\delta := \mu_2 - \mu_1$. (Note, $\mu_1$ acts as a location parameter and $\delta$ as a scale parameter.)

For the $\tau_j$ we use $\Gamma(0.01, 0.01)$ priors, and for $\pi$ a $\beta(3,3)$ prior.

A WinBUGS doodle for this model is given below. Based on this, or otherwise, write a BUGS model for this problem. (Note that `q[]` is intended to be a probability vector that can be used to define a distribution using `dcat`.)




(c) Fit the model using WinBUGS. Use the BGR diagnostic for $\pi$ to choose a burn-in period, and make sure that your sample size is large compared to the range of correlation in the trace.

Report the size of burn-in and sample size used, then give an estimate of $\mathbb{P}(Y > 100)$. What is the MC error for your estimate?