# Homework Model selection AIC and BIC

1. Perform a small simulation study to investigate the frequency by which models are chosen by AIC, BIC and the Hannan–Quinn criterion.

   **Note**: The Hannan–Quinn criterion has not been discussed in class. It uses a $\log\log(n)$ term as penalty, see for instance `https://en.wikipedia.org/wiki/Hannan-Quinn_information_criterion`

   Generate (independently for $i = 1, \ldots, n$)

   $$x_{1i} \sim \text{Uniform}(0, 1),\ x_{2i} \sim N(5, 1),\ Y_i \sim N(2 + 3x_{1i}, (1.5)^2).$$

   (If you wish to use `R`, then uniform and normal data can be generated via `runif, rnorm`, type `?runif` and `?rnorm` for help.)

   Consider 4 normal regression models to fit:

   $$M_1 : Y = \beta_0 + \sigma Z$$
   $$M_2 : Y = \beta_0 + \beta_1 x_1 + \sigma Z$$
   $$M_3 : Y = \beta_0 + \beta_2 x_2 + \sigma Z$$
   $$M_4 : Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \sigma Z$$

   For sample sizes $n = 50, 100, 200, 500$ and 1500, and 1000 simulation runs, construct a table which for each sample size, shows the number of times (out of 1000 simulation runs) that each model has been chosen. Do this for each of AIC, BIC and Hannan–Quinn. Discuss.

   **Hint for `R` users**: Loops are easily constructed in `R` via

   ```
   for(i in 1:nsim){ }
   ```

   where the commands of what to do appear in between the curly parentheses. Other ways are possible, for example via the `apply` function.)

2. Now repeat the smulation study for the nested models used in the illustration of the Prediction Error and Mallows's Cp:

   $$Y_i = \beta_{p-1} \sum_{k=1}^{p-1} (x_i - \widetilde{x}_k) + \varepsilon_i,$$

   where the knots $\widetilde{x}_k$ are given by

   ```
   knots = [-0.1000 0.1555 0.3143 0.5469 0.6903 0.8730
   1.1]
   ```

   Obviously, these knots are parameters that correspond to a parameter vector $\boldsymbol{\beta}$ in the equivalent model

   $$Y_i = \sum_{k=0}^{p-1} \beta_k x_i^k + \varepsilon_i.$$

Consider the nested models

$$M_p : Y_i = \sum_{k=0}^{p-1} \beta_k x_i^k + \varepsilon_i.$$

and compute AIC, BIC

## Available software

Matlab-software is provided. The following zip-file:
http://homepages.ulb.ac.be/∼majansen/teaching/STAT-F-408/mfilesforAICBIC.zip contains a file `AICBICconsistefficientSTATF408.m`, which may serve as a starting file for this homework. It also contains a file `illustrateCpnestedmodels.m`, which generates the Cp plots for the nested models.

Good luck!