**2.2 – Introducing generalized linear models**

Exercise for afternoon session:

*Goal*: Students should be able to demonstrate an understanding of the statistical properties of maximum likelihood and Bayesian estimators.

*Task:* Please develop code using R that conducts a simulation experiment. This experiments involves multiple replicates for different sample sizes, using the following steps:

1. Simulate a data set involving 100 samples from a negative binomial distribution with log-mean of 3 and “size” of 1. You can modify code from the R script “2-2 afternoon -- intro to linear models.R”.
2. Fit a negative binomial distribution to those 100 samples, and record the posterior mean estimate of the mean and size parameter. You can use any of the estimation models provided in the R script “2-2 afternoon -- intro to linear models.R”. Also, please record the 80% confidence interval for the posterior mean parameter.
3. Repeat steps 1-2 100 times, recording parameter estimates for each replicate.
4. Repeat steps 1-3 for different sample sizes, i.e., 10, 50, 100, and 1000 samples.

For each sample size, please evaluate model performance using the following techniques:

* Plot the range of estimates of the mean parameter, and compare this with the true mean.
* Calculate the proportion of replicates for which the true mean was within the 80% credible interval.

*Interpretation*: Students should think about the following questions:

* What predictions do you have *a priori* about the outcome of this simulation experiment? Where these predictions supported?