OLLSCOIL NA hEIREANN, CORCAIGH
THE NATIONAL UNIVERSITY OF IRELAND, CORK

COLAISTE NA hOLLSCOILE, CORCAIGH
UNIVERSITY COLLEGE, CORK

ST4060 - ST6015 - ST6040

Continuous assessment 1 - 2020-21

Eric Wolsztynski
eric.w@ucc.ie

List of (possibly) useful R functions:

```
apply()
boxplot()
c()
coef()
fitted()
lm()
matrix()
mean()
nrow()
par()
quantile()
rgamma()
round()
sample()
set.seed()
summary()
unique()
```

Question 1

Please run `set.seed(6040)` before running the analysis below.

Implement a Monte Carlo simulation with $M = 1,000$ repetitions to demonstrate that the number of data points selected in any bootstrap resample is about 63.2%. Consider a sample size $N = 1,000$ for this Monte Carlo experiment.

Note: no dataset is required to perform this analysis. Hint: R function `unique()` may be useful here.

Question 2

Please run `set.seed(6015)` before running the analysis below.

Implement a Monte Carlo simulation of $M = 1,000$ random samples of $N = 100$ realizations of the Gamma distribution $\mathcal{G}(a, b)$ with shape $a = 3$ and rate $b = 2$. Calculate and store the M corresponding sample means.

(a) Quote the Monte Carlo estimate of the expected value of the sample mean for this probability model.

(b) Is this value of the sample mean surprising? Briefly explain why (or why not).

(c) Quote the Monte Carlo estimate of the standard error of the sample mean for this probability model.

Question 3

Please run `set.seed(4060)` before running the analysis below.

Consider `R`'s dataset `trees` of 31 felled black cherry trees. Generate 100 bootstrap estimates of the slope estimate in the linear regression of tree height (`Height`, in feet) with respect to tree diameter (`Girth`, in inches), i.e. to explain tree height in function of tree girth.

(a) Provide a boxplot of the sampling distribution of bootstrap estimates of the parameter of interest.

(b) Quote the bootstrap estimate of the expected value of the least squares estimator of the regression slope parameter.

(c) Quote the bootstrap estimate for the standard error associated with the least squares estimator of the regression slope parameter.

(d) Provide an empirical bootstrap confidence interval for the true value of the regression slope parameter.