

STAT 102B: Homework 1

George Michailidis

Due electronically through **Gradescope**
on April 17 at 11:59 pm

Problem 1: Consider the correlation coefficient r between two random variables X and Y . Recall that it is a measure of *linear association* between X and Y and takes values in the interval $[-1, 1]$; i.e., $r(X, Y) \in [-1, 1]$.

(a) Write code to simulate data from a *bivariate normal distribution* with mean vector $\mu = [0 \ 0]$ and correlation matrix

$$R = \begin{bmatrix} 1 & r \\ r & 1 \end{bmatrix}$$

Hint: Use the `mvrnorm` function in the R library `MASS`

(b) Generate simulated data from a bivariate multivariate normal distribution with mean vector $\mu = [00]$ and correlation matrix R for the following cases:

1. Sample size $n \in \{20, 50, 100, 200\}$ and correlation coefficient $r = 0$
2. Sample size $n \in \{20, 50, 100, 200\}$ and correlation coefficient $r = 0.5$
3. Sample size $n \in \{20, 50, 100, 200\}$ and correlation coefficient $r = 0.85$

(c) Obtain the *bootstrap sampling distribution* of the sample correlation coefficient \hat{r} for the three cases in part (b), for the following number of bootstrap replicates $B \in \{200, 1000, 5000, 10000\}$.

Comment on the results; in particular how the bootstrap sampling distribution behaves as a function of the sample size n , the number of bootstrap replicates B and the value of the correlation coefficient r .

Problem 2:

The data set “cats” can be obtain through the following R code:

```
> library(MASS)
> data(cats)
> summary(cats)
```

Sex	Bwt	Hwt
F:47	Min. :2.000	Min. : 6.30
M:97	1st Qu.:2.300	1st Qu.: 8.95
	Median :2.700	Median :10.10
	Mean :2.724	Mean :10.63
	3rd Qu.:3.025	3rd Qu.:12.12
	Max. :3.900	Max. :20.50

It contains the body weight (Bwt) in kilograms and the (Hwt) in grams of 47 female and 97 male cats.

Part (a):

- Obtain the bootstrap sampling distribution of the difference of sample means for [body weight](#) between female and male cats
- Obtain the bootstrap sampling distribution of the difference of sample means for [heart weight](#) between female and male cats

Explain how many bootstrap replicates you decided to use and comment on the results.

Part (a):

- Obtain the bootstrap sampling distribution of the t -statistic when testing for mean differences for [body weight](#) between female and male cats

- Obtain the bootstrap sampling distribution of the t -statistics when testing for mean differences for [heart weight](#) between female and male cats

Explain how many bootstrap replicates you decided to use and comment on the results.

Part (c):

Using your code from Problem 1(c):

- Obtain the bootstrap sampling distribution of the sample correlation coefficient between body weight and heart weight for [female cats](#)
- Obtain the bootstrap sampling distribution of the sample correlation coefficient between body weight and heart weight for [male cats](#)

Explain how many bootstrap replicates you decided to use and comment on the results.