

# Homework to Week 8

## Statistics: Principle, Methods and R (II)

GAO FENGAN

17 April 2017

The homework is due on Monday, 24 April 2017. Please hand in the solutions to the teaching assistant He Siyuan at the beginning of the lecture.

1. Read Section 14.4 on Multinomial of *All of Statistics*. Find the Fisher information matrix for the MLE of a Multinomial.
2. Please attach your **R** code. Generate 100 random vectors from a  $N(\mu, \Sigma)$  distribution where

$$\mu = \begin{pmatrix} 3 \\ 8 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

Plot the simulations as a scatterplot. Estimate the mean and covariance matrix  $\Sigma$ . Find the correlation  $\rho$  between  $X_1$  and  $X_2$ . Compare this with the sample correlations from your simulation. Find a 95% confidence interval for  $\rho$ . Use two methods—the bootstrap and Fisher's method. Compare.

3. Using the same dataset as the first problem in the last homework, apply the Zheng-Loh model selection method. Please attach your **R** code, in particular please write your own implementation of Zheng-Loh selection algorithm.
4. This question should be answered using the **Weekly** dataset, which is part of the **ISLR** package. Read the documentation of **Weekly** to find out what the dataset is about. Please attach your **R** code.
  - (a) Produce some numerical and graphical summaries of the **Weekly** data. Do there appear to be any patterns?
  - (b) Use the full dataset to perform a logistic regression with **Direction** as the response and the five lag variables plus **Volume** as predictors. Use the **summary** function to print the results. Do any of the predictors appear to be statistically significant? If so, which ones?