Homework 6:

1. Here is a set of data with the model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \epsilon_i, i = 1, \cdots, 7, \epsilon_i \sim N(0, \sigma^2),$$

Y_i	5	10	10	25	10	30	20
X_{i1}	-1	1	1	0	-1	-1	1
X_{i2}	5	0	-3	-4	-3	0	5

Please find the values of AIC_p, SBC_p, and C_p for all possible models.

2.

	Observations
Population 1	$(0.96, 1.51, 1.59), (2.10, -1.30, 0.02), (0.87, 1.84, 2.71)$ $(2.94, 1.06, 0.03), (0.73, 0.57, -0.52)$
Population 2	$(1.02, 1.51, 0.80), (2.07, 1.59, 0.72), (-0.38, 1.33, 0.75)$ $(-0.76, -0.16, 0.78), (1.00, 0.82, -0.68)$
Population 3	$(1.59, 3.04, -0.22), (4.21, 3.66, 6.73), (4.68, 2.57, 1.62)$ $(1.47, 0.85, -1.29), (0.59, 2.66, -0.72)$

Please do the following:

(a) For all data in the three populations, please produce pairwise scatter plots for the variables and find the correlation matrix for these variables.

(b) Please use classification tree method to allocate the observations $(1.46, 1.06, 1.35)$, $(0.91, 1.02, 0.17)$ and $(2.52, 3.22, 1.71)$. Also, use the classification tree to allocate the observations in population 2.

(c) Also, please use X_1 as response and X_2, X_3 as covariates. as covariates. Then, for an observation with $(X_2, X_3) = (1.2, 1.2)$, use the regression tree method to predict the value of X_1. Also, please use X_2 as response and X_1 as covariates. Then, for an observation with $X_1 = 1$, use the regression tree method to predict the value of X_2.

(d) Please use K-means method to find the clusters corresponding to each observation with number of clusters equal to 3.

(e) For the data in all populations, please use Fisher's discrimination method to find $\hat{a}_1$ and $\hat{a}_2$. Then, classify the data $(0, 0, 0)$, $(2, 2, 2)$, and $(5, 5, 5)$.

(f) Find the error rate for the 15 observations in the table based on $\hat{a}_1$ and $\hat{a}_2$.

3. First, please write a function with two arguments for an input vector (data) and the choices for obtaining statistical quantities, the choice "MLE" for calculating MLEs of mean and variance of normal population and the other "Unbiased" for calculating unbiased estimates of mean and variance of the population. Secondly,

please create a menu with the text “Point Estimation” and a menu item with the text “Location-Dispersion Estimators”. Finally, implement the corresponding dialog which allows the user to input the data, select the location and dispersion estimates, and save the result.