Homework 7:

1. Suppose we have the following data for 3 populations:

	Observations
Population 1	$(-2, 5), (0, 0), (-1, 1)$
Population 2	$(0, 6), (2, 4), (1, 2)$
Population 3	$(1, -2), (0, 0), (-1, -4)$

Please do the following:

(a) Find the variance-covariance matrix and the correlation matrix for these variables.

(b) Find the principal components by 80% criterion and also give the screeplot. Please not use the command "princomp".

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

(d) Find the error rate for the 9 observations based on $\hat{a}_1$ only.

(e) Please find the smallest error rate for the above data as using K-means method with number of clusters equal to 3.

2. The density of the random variable having t-distribution with n degrees of freedom is

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1 + \frac{x^2}{n}\right)^{-(n+1)/2},$$

where $\Gamma(\cdot)$ is the gamma function.

(a) Please write a program to do the following:

Input: degree of freedom and α

Output: $t_{n,\alpha}$.

(b) Use the program in (a), generate the t-distribution table.

3. Please approximate the integral

$$\int_1^2 \frac{1}{x} dx$$

by Simpson's method to do the following.

(a) Please use 10 sub-intervals to approximate the integral.

(b) Suppose $S(N)$ is the value as using Simpson's method with N

sub-intervals. Find the small N such that

$$|S(N) - 0.693147| \leq 0.000001,$$

where 0.693147 is the value of $\log(2)$ to six decimal places.