MATH1309 – Practice Problems 4

By hand (note that after Week 4 lecture you should be able to answer the last two questions from Practice Problem 3)

1. Let **X** be $N_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [-3,1,4]$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ Which of the following random variables are independent? Explain.

   d. $\frac{X_1+X_2}{2}$ and X_3

   e. X_2 and $X_2 - \frac{5}{2}X_1 - X_3$

Using Software

PROC IML

Adding in the line Use allows you to refer to a dataset that already exists in SAS. This may be a data set that you have read in from a file using a DATA step first.

You can also specify the variables that you want to read in by listing them in the curly brackets.

```
proc iml;
use timber;
read all var {x1 x2} into A; /*read just variables x1 and x2*/
```

Test for Bivariate Normality

Confidence ellipse

We can determine the number of observations that occur inside the contour or outside using PROC IML. This can help to determine if we have multivariate normality, due to the properties of subsets of a multivariate normally distributed random vector.

1) Read in the Data file Timber.DAT as on SAS Studio and Canvas

2) Create a scatter matrix of the variables

3) Determine if there is bivariate normality by finding where the observations lie in relation to the confidence ellipse
   a) Find the Mahalanobis distances for x1 and x2
      i) Determine how many observations are outside the ellipse, ie larger than $\chi^2_2(0.5)=1.39$
      ii) Determine how many observations are smaller (or equal to) $\chi^2_2(0.5)=1.39$ and inside the ellipse
   b) Find the Mahalanobis distances for x2 and x3
      i) Determine how many observations are outside the ellipse, ie larger than $\chi^2_2(0.5)=1.39$
      ii) Determine how many observations are smaller (or equal to) $\chi^2_2(0.5)=1.39$ and inside the ellipse

4) Create a chi-square probability plot to assess the random vector, with all four variables, for multivariate normality.

Is the data set multivariate normal?