BAN 250 Multivariate Homework 2.

1. Complete the following

   Let **X** be $\mathbf{N}_3(\boldsymbol{\mu},\boldsymbol{\Sigma})$ with $\boldsymbol{\mu}' = [-3,1,4]$ and

$$\sum = \begin{bmatrix} 1 & -2 & 0 \\ -2 & 5 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

   Which of the following random variables are independent? Explain.

   (a) X1 and X 2
   (b) X2 and X3
   (c) (X1,X2) and X3
   (d) $\frac{(X1+X2)}{2}$ and X3
   (e) X2 and $X2 - \frac{5}{2}X1 - X3$

2. Using an example show why the following is true.

   a) $\begin{vmatrix} A & 0 \\ 0' & B \end{vmatrix} = |A||B|$
   b) $\begin{vmatrix} A & C \\ 0' & B \end{vmatrix} = |A||B| \ for \ |A| \neq 0$

   To do this make a two by two matrix for A and B and see if you get the correct result. If you prefer to prove it mathematically (e.g. proof) that is acceptable of course.

3. Using the Nuclear dataset and only columns 1:5 provide the means of each of the columns and identify whether each column is normal. Conduct qqplots for each of these variables to confirm your conclusion. . Then conduct 3 separate multivariate normal tests. Provide the correct hypothesis and conclusions (i.e. H_0, and H_a.).

4. Using the dataset (propval.txt)

   a) Test the normality of each of the variables individually, then run the three tests of multivariate normality. What is your conclusion.
   b) Run a regression with all the variables and y as the response. Does the model have good utility?
   c) Does the model meet all regression assumptions? If not what do you suppose is the cause?
   d) What additional tests would you conduct on your model. Then run them an interpret the results.

5. (Applicative Example) A school district trying to determine its budget needs to predict the number of English Language Learners coming into the district. Therefore, your job as a data scientist will be to conduct an analysis and create a model. Use “Elementary.txt” dataset.

   Create a model that will predict the number of English language learners based on the percentage of free meals, year Round school, mobility, average class size in k-3 , average class size in 4-6, pct of full credential, pct of emer credentials, and number of students enrolled.

   Summarize all your findings and be sure to indicate all assumptions and show all validations you performed.