Question 1:

Using the SATGPA data set in Stat2Data package. Test by using $\alpha = .01$.

1) Create the following three variables and then print out all the six variables.
    a. Create a new variable "SAT", which is the sum of MathSAT and VerbalSAT.
    b. Create second new variable "SATLevel", and assign the value of "SATLevel" as 1 when SAT<=1100, 2 when 1100<SAT<=1200, 3 when 1200<SAT<=1300, and 4 when SAT>1300.
    c. Create third new variable "GPALevel" and assign the value of "GPALevel" as 1 when GPA<=2.8, 2 when 2.8<GPA<=3.3, 3 when 3.3<GPA<=3.5, and 4 when GPA>3.5
    d. Print out all the data in the ascending order of their GPALevel and the descending order of their SAT when GPALevel is the same.
2) Use the Chi-Square test to conclude if the SATLevel and GPALevel are independent.
3) Compute the mean and variance of "GPA" for each level of "GPALevel", and compute the correlation matrices for the four variables: MathSAT, VerbalSAT, GPA and SAT.
4) Do the data provide sufficient evidence to indicate that the mean of MathSAT is significantly greater than the mean of VerbalSAT.
5) Test if the proportion of MathSAT less than VerbalSAT is 0.5.

Question 2:

Analyze and interpret the effect of explanatory variables on the milk intake (dl.milk) in the kfm data set (ISwR) using a multiple regression model. Test by using $\alpha = .05$.

1) Run regression for dl.milk on all other variables. Do you find any significance that milk intake can be explained by other variables?
2) Find regression models in which fewer explanation variables should be used. i.e., select a subset of variables so that a better fit can be achieved.