AMS 597 Spring 2021
Homework 3 Due 03/11/2021 10:00 AM

Instruction: Submit your homework via Blackboard. If you have difficulty submitting via Blackboard, email the TA directly. No late homework will be accepted, based on the time the email is sent out. If you are submitting via email (not recommended), put this as the subject of your email

AMS597 2021: Homework/Exam #3. ID: XXXXXXXXX. Name: XXX XXX

1. Use the following dataset to answer the questions.
   `http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/d_logret_6stocks.txt`

   (a) Regress the return of Pfizer on the returns of Exxon and Citigroup (with intercept). Report the estimated coefficients

   (b) Use `matlines` to plot the fitted values and the corresponding confidence bands.

   (c) Generate an ANOVA table to conclude if regression effects are significant.

   (d) Regress the return of Pfizer on the returns of Exxon and Citigroup (without intercept). Report the estimated coefficients.

   (e) Compute the correlation of Pfizer and Exxon, and test if their correlation is zero.

2. Consider the dataset in Problem 1, we now ignore the time series features of all returns, and consider them independent. We also treat all returns of 'Citigroup' and 'AmerExp' as Group 1, returns of 'Exxon' and 'GenMotor' as Group 2, and returns of 'Intel' as Group 3.

   (a) Write your own function `my.oneway.anova` which performs one-way ANOVA (i.e., your function will computes MSB, MSW, F and the p-value)

   (b) Perform one-way ANOVA for Groups 1 and 2 using your function, and compare the results obtained using R built-in function for one-way ANOVA.

   (c) Perform one-way ANOVA for Groups 1-3 using your function, and compare the results obtained using R built-in function for one-way ANOVA.

3. Using the `ChickWeight` dataset in `R`,

   (a) Perform a two way ANOVA comparing the weights to Time and Diet.

   (b) For subset of Time 2, perform a one way ANOVA comparing the weights to Diet. If necessary, perform the post hoc analysis (all pairwise comparisons via R function `TukeyHSD`) to identify which diet groups have different weight.
   Post-Hoc Analysis Reference:
   `https://stats.idre.ucla.edu/r/faq/how-can-i-do-post-hoc-pairwise-comparisons-in-r`

4. Consider the dataset in Problem 1, perform the following test.

   (a) Test if the proportion of positive returns of Pfizer is 0.55.

(b) Test if the proportion of positive returns of Intel is larger than 0.55.

(c) Test if the proportions of positive returns of Pfizer and Intel are same.

(d) We treat all returns of Citigroup and AmerExp as Group 1, returns of Exxon and GenMotor as Group 2, and returns of Intel as Group 3. We also consider the following 4 ranges of their returns r: $r < -0.1$, $-0.1 \leq r < 0$, $0 \leq r < 0.1$, $r \geq 0.1$. Use chi-square test to conclude if the group and return range effects are independent.

5. We will use the `mcycle` data in the `MASS` package for this problem. The data set consists of two variables, namely acceleration and measurement times from a simulated motorcycle accident. Our objective is to investigate the relationship between this two variables. Using acceleration as dependent variable and times as covariate, fit your "best" polynomial regression model to describe the relationship between this two variables. Check for the assumptions and justify how you choose the final model.

```
library(MASS)
attach(mcycle)
```

6. Read the following data into `R`
   http://www.ams.sunysb.edu/~pfkuan/Teaching/AMS597/Data/HW3Qn6Data.txt
   The `HW3Qn6Data.txt` dataset consists of $n = 200$ (sample size), a response variable `y` and 6 covariates `x1,...,x6`.
   Using the best subset selection method, i.e., consider all possible combination of covariates (y~1, y~x1, y~x2, ..., y~x1+x2, ..., y~x1+x2+x3+x4+x5+x6), choose the best model based on the BIC criterion. Write down your final model. Note that BIC stands for Bayesian Information Criterion, which can be used to compare models.

$$BIC = -2 * \text{log-likelihood} + p * \log(n)$$

where p is the number of parameters. The smaller the BIC, the better is the model fit.