HW 4- ANOVA

1. (6pt) (data in file mileage.csv) This problem is designed to review regression with categorical variables. International Oil Inc. Is attempting to a develop a reasonably priced minimum unleaded gasoline that will deliver higher gasoline mileage than can be achieved by its current premium unleaded gasolines. As part of its development process, International Oil Inc. wishes to study the effect of one qualitative variable, x_1, premium gasoline unleaded type (A, B, C) and one quantitative variable x_2 amount of gasoline additive VST (0, 1, 2, 3 units) on the gasoline mileage y obtained by an automobile called Encore. For testing purposes a sample of 22 Encores is randomly selected and driven under normal driving conditions. The combination of x_1 and x_2 used in the experiment along with the corresponding values of y are in file mileage.csv. Define $\mu_{[A,x]}, \mu_{[A,x]}$ and $\mu_{[B,x]}$ to be the mean unleaded gasoline mileage by Encore when using AST amount x and premium unleaded gasoline types A, B and C, respectively. Consider the model

$$Y_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 x_2 + \epsilon_i$$

where $D_{1i} = 1$ gas type is B and 0 otherwise and $D_{2i} = 1$ is gas type is C and 0 otherwise.

(a) (2pt) Estimate the β_is and interpret your result (see note for how to fit this model)

```
lm(y ~ factor(x1) + x2)
```

(b) (2pt) Construct a 95% confidence interval for β_1 and interpret your result

(c) (2pt) Test $H_0 : \beta_1 = \beta_2 = 0$ against H_a : Not H_0 using $\alpha = 0.05$.

2. (5pt) In this problem we study the grain yield of rice at six seeding rates (kg/ha): The seeding rates are 25, 50, 75, 100, 125 and 150 kilograms per acre. Assume that four fields were chosen and each field was divided into 6 plots and each plot was planted at a seeding rate assigned to it at random. Besides the seeding rate, all other agricultural practices are the same. The data is

	Seeding rate (kg/ha)					
Filed	25	50	75	100	125	150
1	5.1	5.3	5.3	5.2	4.8	5.3
2	5.4	6.0	5.7	4.8	4.8	4.5
3	5.3	4.7	5.5	5.0	4.4	4.9
4	4.7	4.3	4.7	4.4	4.7	4.1

Fit an appropriate model to this data and test H_0 : the average yields are the same for the 6 seeding rates against the alternative H_a : There are not the same. Use $\alpha = 0.05$.

3. (6pt) The cutting speeds of four types of tools are being compared in an experiment. Five cutting materials of varying degree of hardness are to be used as experimental blocks. The data giving the measurement of cutting time in seconds appear in the table below

	Block				
Treatment	1	2	3	4	5
1	12	2	1	8	7
2	20	14	17	12	17
3	13	7	13	8	14
4	11	5	10	3	6

(a) (2pt) Fit an appropriate model to this data and test H_0 :The mean cutting speeds are the same for the four tools H_a : There difference. Use $\alpha = 0.05$.

(b) (4pt) Use the Bonferroni method to determine where the differences are

4. (3pt) An experiment to investigate the effects of various dietary starch levels on milk production was conducted on four cows. The four diets, D1, D2, D3, and D4, (in order of increasing starch equivalent), were fed for three weeks to each cow and the total yield of milk in the third week of each period was recorded (i.e. third week to minimize carry-over effects due to the use of treatments administered in a previous period). That is, the trial lasted 12 weeks since each cow received each treatment, and each treatment required three weeks. The investigator felt strongly that time period effects might be important (i.e earlier periods in the experiment might influence milk yields differently compared to later periods). Hence, the investigator wanted to block on both cow and period. However, each cow cannot possibly receive more than one treatment during the same time period; that is, all possible cow-period blocking combinations could not logically be considered. It is decided to use a 4x4 latin square design and the data is

	Cow			
Treatment	1	2	3	4
Period 1	D4(192)	D1(195)	D3(292)	D2(249)
Period 2	D1(190)	D4(203)	D2(218)	D3(210)
Period 3	D3(214)	D2(139)	D1(245)	D4(163)
Period 4	D2(221)	D3(152)	D4(204)	D1(134)

(each cell provides the treatment applied and response between the parentheses). Fit an appropriate model to this data and test H_0 : there is no difference between the four diets against H_a there is a difference.