Math 3080 § 1. **Final Exam** Name:_____
Treibergs April 28, 2014

This is an **open book test**. You are allowed your textbook, notes and a calculator. Other books, laptops, or messaging devices are not permitted. Give complete solutions. Be clear about the order of logic and state the theorems and definitions that you use. There are [120] total points. **Do SIX of nine problems.** If you do more than six problems, only the first six will be graded. Cross out the problems you don't wish to be graded.

1.	/20
2.	/20
3.	/20
4.	/20
5.	/20
6.	/20
7.	/18
8.	/20
9.	/20
Total	/120

1. [20] In the article "Daily Weigh-Ins Can Help You Keep Off Lost Pounds, Experts Say" (Associated Press, Oct. 17, 2005) describes an experiment in which 291 people had lost at least 10% of their body weight in a medical weight loss program were assigned at random to one of three groups for follow-up. One group met monthly in person, one group "met" monthly on line in a chat room, and one group received a monthly newsletter by mail. After 18 months, participants in each group were classified according to whether or not they had regained more than five pounds. Does appear to be a difference in the weight regained proportions for the three follow-up methods? State the null and alternative hypotheses. State the test statistic and rejection region for a significance level 0.01. Give formulas for the expected cell counts. What are your conclusions?

	Amount of Weight Gained		
	Regained 5 lb or Less	Regained More Than 5 lb	Total
In Person	52	45	97
Online	44	53	97
Newsletter	27	70	97
Total	123	168	291

Your grades will be posted at my office according to

Secret Id. :

2. In a study by Casey, May and Morgan in *Journal of Experimental Biology,* 1985, the wing stroke frequencies of two species of Euglossine bees were recorded for a sample of $m = 4$ *Euglossa mandibularis* Friese, and $n = 6$ *Euglossa imperialis* Cockrell. Can you conclude that the distibution of wing strokes frequencies differ in these two species? Analyze the samples using a Wilcoxon non-parametric test at the $\alpha = .05$ level of significance.

Species X:	235	225	182	188		
Species Y:	180	169	180	185	178	190

(a) [7] What assumptions are you making on the data? State the null and alternative hypotheses. State the test statistic and the rejection region.

(b) [13] Perform the test of hypothesis. What is your conclusion?

3. [20] The paper "... Protocols for Mobile Ad Hoc Networks," *Proceedings 2002 International Conference on Wireless Networks*, tried to predict network performance measured by y data overhead (in kB) in terms of x_1 speed of computers (m/s), x_2 pause time at each link (s) and x_3 the link change rate (100/s). Consider fitting the quadratic model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_1 x_2 + \beta_5 x_1^2 + \beta_6 x_2^2 + \beta_7 x_3^2 + \epsilon$. Here is the data and **R** output of the analysis of variance.

Speed	Pause	LCR	Overhead	Speed	Pause	LCR	Overhead	Speed	Pause	LCR	Overhead
5	10	9.43	428.90	10	50	8.31	498.77	30	30	16.70	506.23
5	20	8.32	443.68	20	10	26.31	452.24	30	40	13.26	516.27
5	30	7.37	452.38	20	20	19.01	475.97	30	50	11.11	508.18
5	40	6.74	461.24	20	30	14.73	499.67	40	10	37.82	444.41
5	50	6.06	475.07	20	40	12.12	501.48	40	20	24.14	490.58
10	10	16.46	446.06	20	50	10.28	519.20	40	30	17.70	511.35
10	20	13.28	465.89	30	10	33.01	445.45	40	40	14.06	523.12
10	30	11.16	477.07	30	20	22.13	489.02	40	50	11.69	523.36
10	40	9.51	488.73								

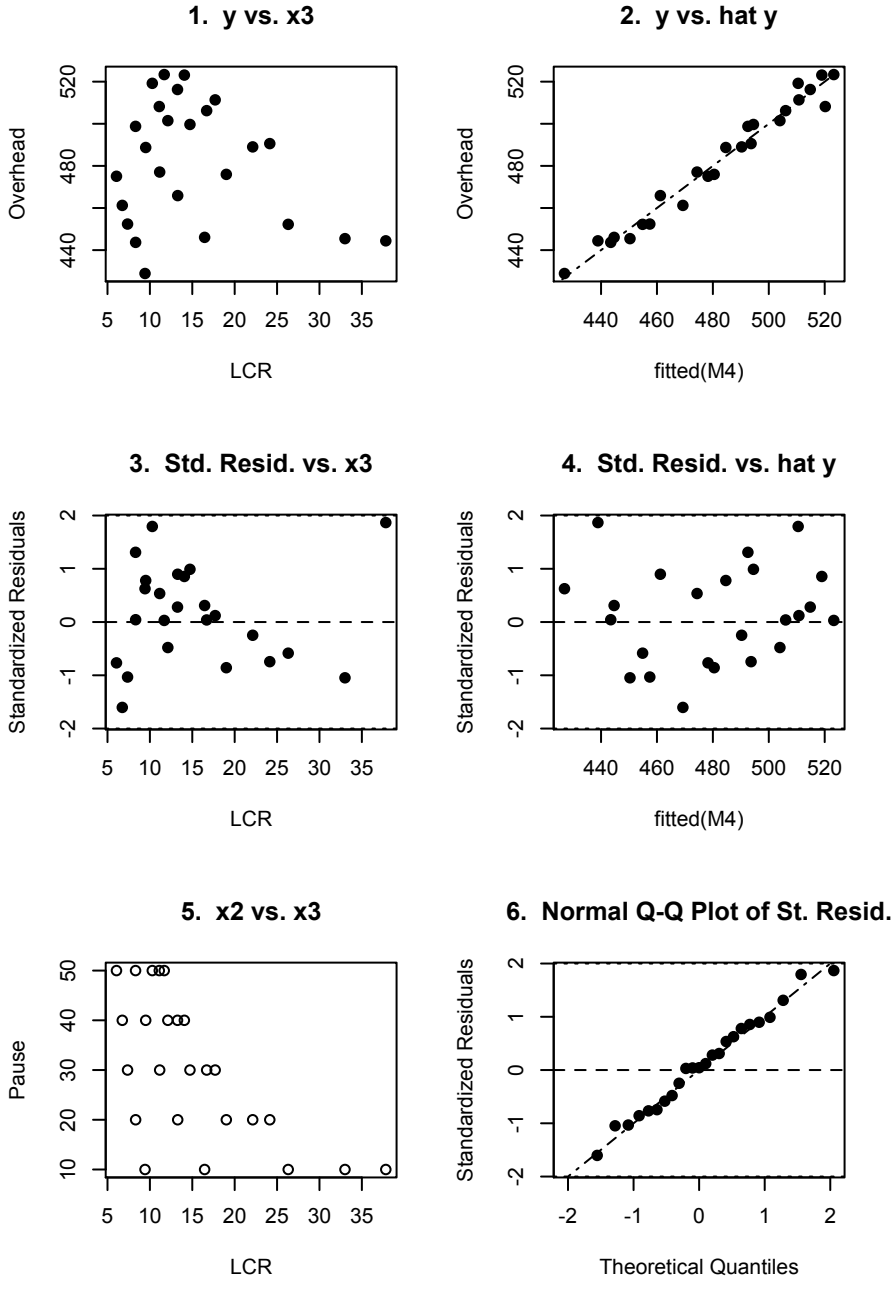
```
> M4=lm(Overhead~x1+x2+x3+x1:x2+I(x1^2)+I(x2^2)+I(x3^2)); summary(M4); anova(M4)

Call:
lm(formula = Overhead ~ x1 + x2 + x3 + x1:x2 + I(x1^2) + I(x2^2) +
    I(x3^2))
Residuals:
     Min       1Q   Median       3Q      Max
-12.0242  -3.0847   0.2109   4.0988   8.6939
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 367.96413   19.40264  18.965 7.12e-13
x1            3.04382    1.59133   1.913  0.07278
x2            2.29237    0.69838   3.282  0.00439
x3            3.47669    2.12913   1.633  0.12087
I(x1^2)      -0.03131    0.01906  -1.643  0.11885
I(x2^2)      -0.01318    0.01045  -1.261  0.22442
I(x3^2)      -0.10412    0.03192  -3.262  0.00459
x1:x2        -0.01222    0.01534  -0.797  0.43663

Residual standard error: 5.723 on 17 degrees of freedom
Multiple R-squared: 0.9723,Adjusted R-squared: 0.9609
F-statistic: 85.33 on 7 and 17 DF,  p-value: 5.409e-12

Analysis of Variance Table
Response: Overhead
          Df  Sum Sq Mean Sq  F value    Pr(>F)
x1         1  5571.2  5571.2 170.0732 2.789e-10
x2         1 10973.9 10973.9 335.0019 1.268e-12
x3         1   559.2   559.2  17.0708 0.0006973
I(x1^2)    1  1714.9  1714.9  52.3500 1.394e-06
I(x2^2)    1   316.7   316.7   9.6676 0.0063766
I(x3^2)    1   410.8   410.8  12.5400 0.0025096
x1:x2      1    20.8    20.8   0.6347 0.4366304
Residuals 17   556.9    32.8
```

(Prob. 3 Continued.) Six diagnostic plots were produced by **R**. For each of the six plots shown, briefly explain what information about the data, the analysis or the appropriateness of the model can be concluded from that plot.



4. (a) [3] If you were to remove a variable from the model in Problem 3, which one would you remove and why?

   (b) [17] Another model was fitted to the data in Problem 3. $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_6 x_2^2 + \beta_7 x_3^2 + \epsilon$. Is the inclusion of the extra variables as in Problem 3 justified? State the null and alternative hypotheses. State the test statistic and the rejection region. Perform the test and state your conclusion.

```
> M5=lm(Overhead~x1+x2+x3+I(x2^2)+I(x3^2))
> summary(M5);anova(M5)
Call:
lm(formula = Overhead ~ x1 + x2 + x3 + I(x2^2) + I(x3^2))
Residuals:
    Min      1Q  Median      3Q     Max
-9.6678 -4.2616  0.0772  3.0904 11.4229
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 345.416002  13.185266  26.197 2.24e-16
x1            0.707245   0.188269   3.757  0.00134
x2            2.853696   0.533710   5.347 3.69e-05
x3            6.484002   1.050376   6.173 6.23e-06
I(x2^2)      -0.018334   0.007879  -2.327  0.03118
I(x3^2)      -0.144816   0.019965  -7.254 6.95e-07

Residual standard error: 5.832 on 19 degrees of freedom
Multiple R-squared: 0.9679,Adjusted R-squared: 0.9594
F-statistic: 114.6 on 5 and 19 DF,  p-value: 1.647e-13

Analysis of Variance Table
Response: Overhead
          Df  Sum Sq Mean Sq F value    Pr(>F)
x1         1  5571.2  5571.2 163.826 8.652e-11
x2         1 10973.9 10973.9 322.697 2.221e-13
x3         1   559.2   559.2  16.444 0.0006756
I(x2^2)    1   584.6   584.6  17.191 0.0005487
I(x3^2)    1  1789.3  1789.3  52.615 6.954e-07
Residuals 19   646.1    34.0
```

5. [20] The Corinne Concrete Company took measurements relating x, the age of certain concrete pipes in years and y, the corresponding load necessary to obtain the first crack (in 1000 lb/ft).

x	1	3	3	4	5	5	6	7	7	9
y	9	7	7	9	5	7	4	5	7	5

Fill in the missing boxes in the analysis of variance and summary tables for a simple regression on this data.

```
> c(sum(x), sum(y), sum(x*x), sum(x*y), sum(y*y))
[1]   50      65     300       300       449
```

Response: y	Df	Sum Sq	Mean Sq	F value
x	1			
Residuals				
Total				

Coefficients:	Estimate	Std. Error	t value
(Intercept)			
x			

6. In an article by Zenaitis and Duff in *Ozone Science and Engineering,* 2002, runoff water from saw mills in British Columbia was measured. Included were pH for six water specimens. Analyze the data using a non-parametric method.

```
5.9     5.0     6.5     5.6     5.9     6.5
```

(a) [7] What assumptions are you making on the data?

(b) [13] Construct a non-parametric two sided 90% confidence interval for the mean pH.

7. A test of the strength of bread wrapper stock under 16 different conditions, represented by two levels of each of four factors was conducted. An operator effect was introduced into the model, since it was necessary to obtain half the experimental runs under operator 1 and half under operator 2. It was felt that the operators do have an effect on the product.

Operator 1					Operator 2				
a	b	c	d	Str	a	b	c	d	Str
-1	-1	-1	-1	18.8	1	-1	-1	-1	14.7
1	1	-1	-1	16.5	-1	1	-1	-1	15.1
1	-1	1	-1	17.8	-1	-1	1	-1	14.7
-1	1	1	-1	17.3	1	1	1	-1	19.0
-1	-1	-1	1	13.5	1	-1	-1	1	16.9
1	1	-1	1	17.6	-1	1	-1	1	17.5
1	-1	1	1	18.5	-1	-1	1	1	18.2
-1	1	1	1	17.6	1	1	1	1	20.1

Analysis of Variance Table
Response: Str

	Df	SumSq
a	1	4.4100
b	1	3.6100
c	1	9.9225
d	1	2.2500
Oper	1	0.1225
a:b	1	0.5625
a:c	1	2.8900
b:c	1	0.2500
a:d	1	1.1025
b:d	1	0.9025
c:d	1	1.6900
d:Oper	1	9.6100
a:b:d	1	1.6900
a:c:d	1	4.2025
b:c:d	1	5.5225

(a) [5] In order to make significance tests on the factors, assume that all interactions are negligible. State the assumptions on the model.

(b) [5] What interaction is confounded with operators?

(c) [10] Test for significance of the factors at a $\alpha = .10$ level. Only a part of the **R**© output is shown. `> anova(lm(Str ~ a * b * c * d * Oper))`

8. An experiment was run to study the effect of two factors on the amplification of a stereo recording, type of receiver (two brands) and type of amplifier (four brands). For each combination of factor levels, three tests are performed to measure the decibel output.

	Amplifiers											
	A			B			C			D		
Receiver 1	9	4	12	8	11	16	8	7	1	10	15	9
Receiver 2	7	1	4	6	10	7	0	1	7	6	7	5

(a) [5] State the assumptions on the model. To test the interaction between receivers and amplifiers, state the null and alternative hypotheses. State the test statistic and rejection region.

(b) [5] Is there an interaction between receivers and amplifiers? Is there an effect due to receivers? Is there an effect due to amplifiers?

```
>  summary( aov(Decibel ~ Receiver * Amplifier) )
                   Df Sum Sq Mean Sq F value  Pr(>F)
Receiver            1 100.04  100.04   9.306 0.00763
Amplifier           3 117.12   39.04   3.632 0.03588
Receiver:Amplifier  3   5.46    1.82   0.169 0.91557
Residuals          16 172.00   10.75
```

(c) [10] If appropriate, use the Tukey procedure to determine which amplifiers differ in average decibel output.

```
> tapply(Decibel, Amplifier, mean)
       A        B        C        D
6.166667 9.666667 4.000000 8.666667
```

9. [20] The owners of the Ivins Ice Cream shop suspect that the probability that day of the week that a random ice cream cone is purchased is the same for any weekday and the same for Saturday and Sunday, but not, perhaps, the same for a weekday and a weekend day. If $\pi(x)$ is the probability that the a cone is purchased on day $x \in \{1, 2, \dots, 7\}$, then $\pi(1) = \cdots = \pi(5) = p$ and $\pi(6) = \pi(7) = q$ where $5p + 2q = 1$ or $q = .5 - 2.5p$. Suppose that n_i is the number of cones sold on the ith day of the week. Then the maximum likelihood estimator (MLE) is

$$\hat{p} = \frac{n_1 + \cdots + n_5}{5(n_1 + \cdots + n_7)}.$$

Assume that the number of cones sold this week is given. Test the null hypothesis that the probability of a cone sale on the xth day of the week is given by $\pi(x; p)$.

Day	1	2	3	4	5	6	7	Total
Number of Cones Sold	248	237	214	226	217	440	418	2000