Statistics 496, Applied Statistics for Industry II
Exam 2, Spring 1997

Name:_____
Site: _____

INSTRUCTIONS: You will have 1 hour and 30 minutes to complete the exam. There are 5 questions worth a total of 100 points. Not all questions have the same point value so gauge your time appropriately. Read the questions carefully and completely. Answer each question and show work in the space provided on the exam. Turn in the **entire** exam when you are done or when time is up. For essay questions, think before you write.

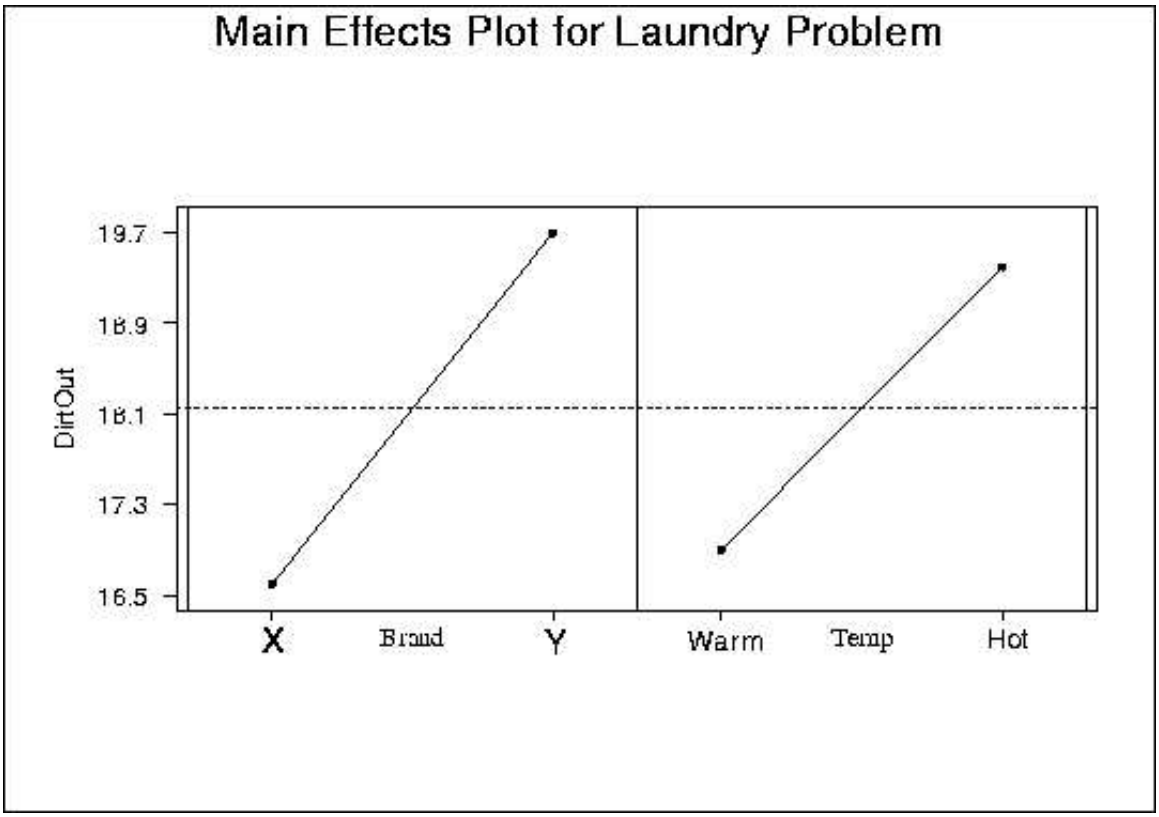
1. **[30 pts]** In an experiment on the effectiveness of laundry detergents, 2 brands of detergent and 2 water temperatures are used in a 2^2 factorial experiment. Five standard laundry loads are washed for each of the combinations. The data are given below.

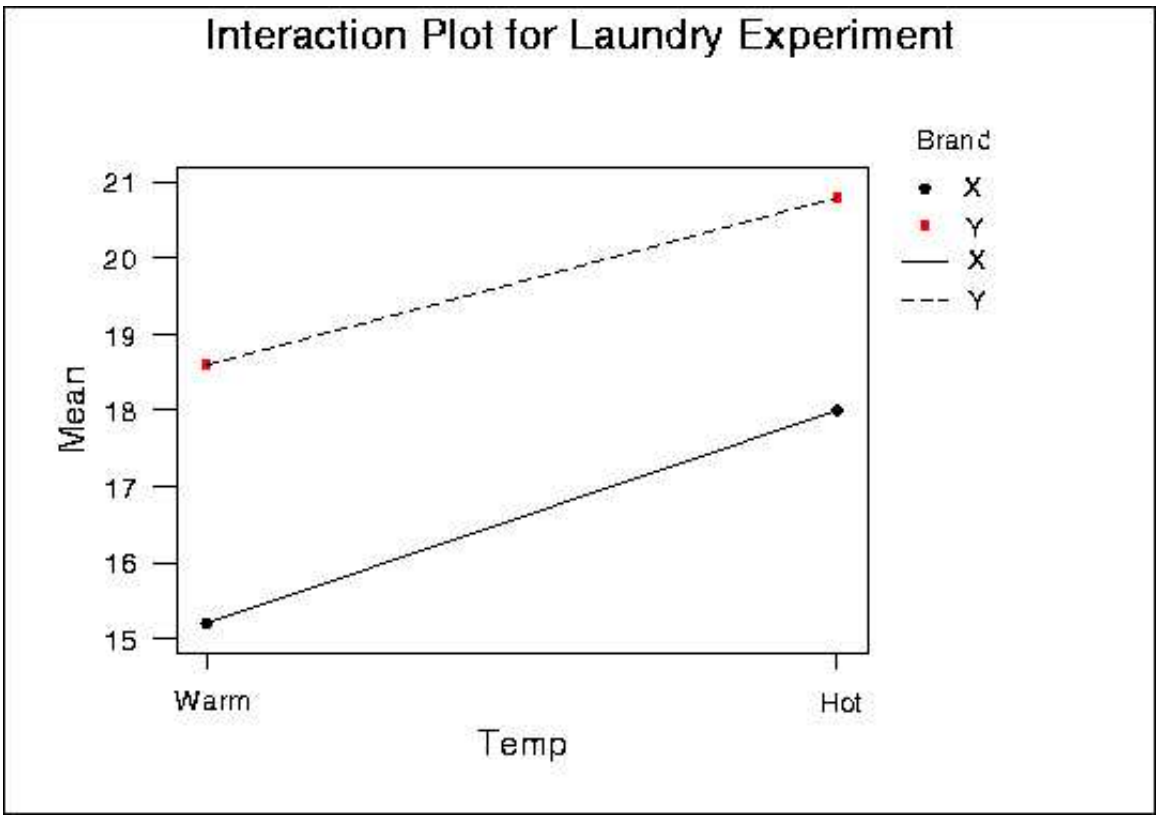
Treatment	Detergent Brand	Water Temp	Amount of Dirt (mg) Removed					Mean	Variance
1	X	Warm	13	16	16	14	17	15.2	2.7
2	Y	Warm	17	19	20	17	20	18.6	2.3
3	X	Hot	18	16	17	19	20	18.0	2.5
4	Y	Hot	21	20	18	22	23	20.8	3.7

(a) **[4]** Looking at the plots on the next page, comment on the apparent main effects and whether or not there is an interaction between detergent brand and water temperature. Be sure to state what you see in the plot that supports your comments.

(b) **[6]** Calculate the full effects for detergent brand, water temperature and the interaction between the two.

(c) **[4]** What is the value of the $MS_{repError}$ for this experiment?


Main Effects Plot for Laundry Problem
DirtOut
19.7
18.9
18.1
17.3
16.5
X
Brand
Y
Warm
Temp
Hot


Interaction Plot for Laundry Experiment
Brand
X
Y
X
Y
Mean
21
20
19
18
17
16
15
Warm
Hot
Temp

(d) **[6]** Which of the effects in (b) are statistically significant? How do you know?

(e) **[2]** What recommendation would you make to get the most dirt out?

(f) **[4]** We could also look at this as a "one-factor" experiment, where the "treatments" for the "one-factor" are combinations of detergent and water temperature. Calculate the LSD (use t=2) for such an experiment.

(g) **[4]** Based on the LSD, which "treatments" are significantly different? not significantly different? Based on this analysis which combination of detergent brand and water temperature would you recommend?

2. **[10 pts]** You will need the sample size tables in order to answer the following questions. A copy is attached at the end of the exam. We wish to run an experiment with three factors each at two levels.

   (a) **[2]** How many treatment combinations are there?

   (b) **[3]** Using the sample size tables, with ALPHA=0.05 and BETA=0.10, how many experimental units (total) do we need if we wish to detect a difference in treatment means as small as 1.4σ?

   (c) **[5]** Resources limit us to 5 experimental units per treatment combination. What size difference in **factor** means (*e.g.* $\bar{Y}_{A-High} - \bar{Y}_{A-Low}$) can we detect with ALPHA=0.05 and BETA=0.05?

3. **[10 pts]** We have discussed two randomization schemes in the course so far. These are the completely randomized design (CRD) and the randomized complete block design (RCBD). Discuss advantages and disadvantages of each design and indicate when you would use each.

4. **[35 pts]** The following problem is adapted from a problem given in *Statistical Quality Design and Control* by DeVor, Chang and Sutherland, p. 609-610. A fisherman conducted an experiment using 5 factors each at 2 levels. The factors and levels are given below. The response is the number of fish (bluegills) caught in a four hour period. All experimental trails were performed at a lake stocked with bluegills in northern Wisconsin. Refer to the Minitab analysis and the normal plot of estimated effects.

Variable	Low Level (–1)	High Level (+1)
A: Fishing location	Pier	Boat
B: Bait type	Worms	Nightcrawlers
C: Time of day	Day	Night
D: Weather	No rain	Rain
E: Hook size	Small	Large

(a) **[3]** What are the experimental units in this experiment?

(b) **[4]** Do you see any possible difficulties with performing a completely randomized experiment? Explain briefly. Hint: it may help to explain what a completely randomize experiment is.

(c) **[5]** Looking at the normal plot, which of the estimated effects appear to be significant? List them here and label them on the normal plot.

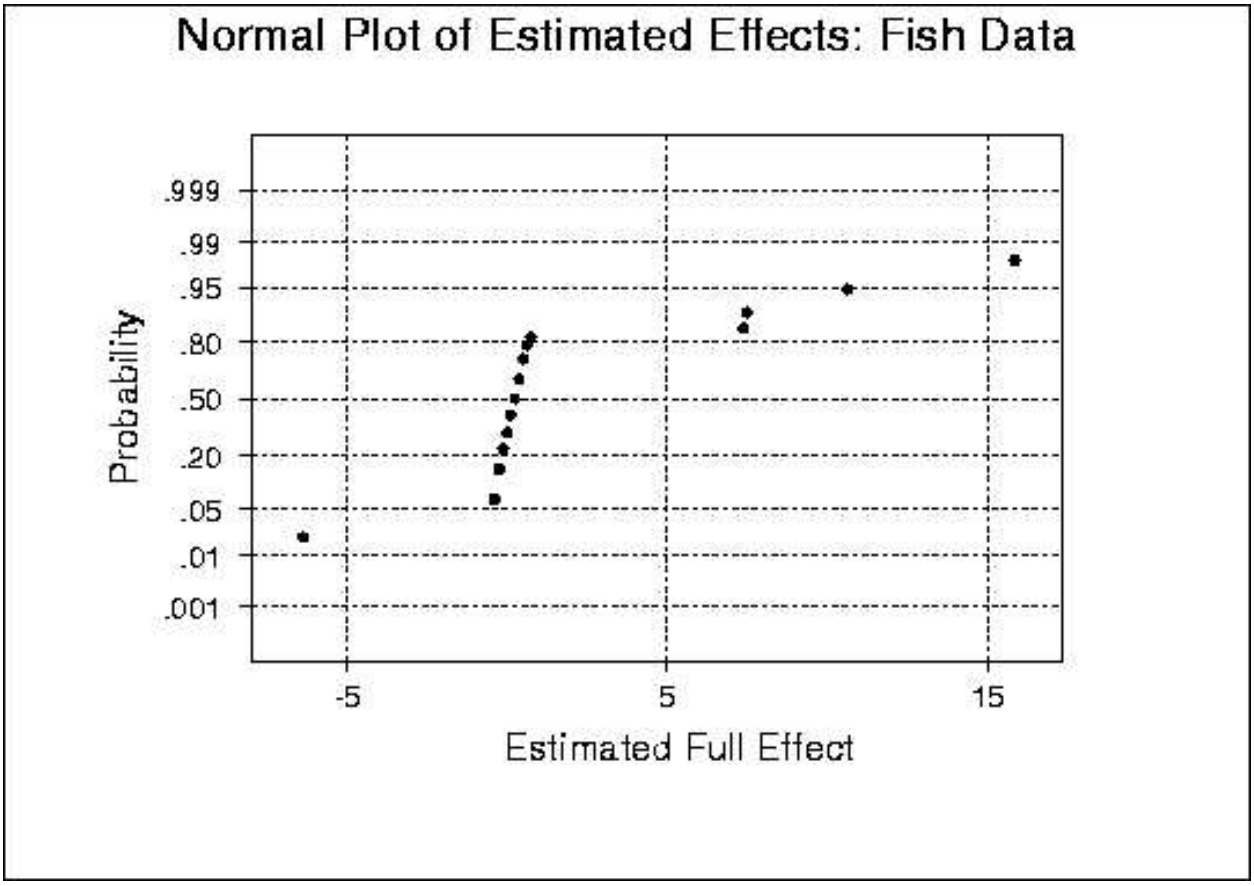
Fractional Factorial Fit

Estimated Effects and Coefficients for Fish

Term	Effect	Coef
Constant		39.875
A	10.625	5.313
B	15.875	7.938
C	-0.000	-0.000
D	7.375	3.688
E	-6.375	-3.187
A*B	0.250	0.125
A*C	0.125	0.063
A*D	0.500	0.250
A*E	7.500	3.750
B*C	0.375	0.187
B*D	0.500	0.250
B*E	7.500	3.750
C*D	-0.375	-0.187
C*E	0.125	0.063
D*E	-0.250	-0.125
A*B*C	0.500	0.250
A*B*D	-0.375	-0.187
A*B*E	0.375	0.187
A*C*D	-0.000	-0.000
A*C*E	0.250	0.125
A*D*E	-0.125	-0.063
B*C*D	-0.000	-0.000
B*C*E	-0.250	-0.125
B*D*E	0.125	0.063
C*D*E	0.250	0.125
A*B*C*D	-0.125	-0.063
A*B*C*E	0.375	0.187
A*B*D*E	0.750	0.375
A*C*D*E	0.625	0.313
B*C*D*E	0.375	0.187
A*B*C*D*E	-0.250	-0.125

Analysis of Variance for Fish

Source	DF	Seq SS	Adj SS	Adj MS	F	P
Main Effects	5	3679.50	3679.50	735.900	**	
2-Way Interactions	10	907.50	907.50	90.750	**	
3-Way Interactions	10	6.00	6.00	0.600	**	
4-Way Interactions	5	10.00	10.00	2.000	**	
5-Way Interactions	1	0.50	0.50	0.500	**	
Residual Error	0	0.00	0.00	0.000		
Total	31	4603.50				


Normal Plot of Estimated Effects: Fish Data
Probability
.999
.99
.95
.80
.50
.20
.05
.01
.001
-5
5
15
Estimated Full Effect

(d) **[5]** If we had decided before hand to use 4-way and 5-way interaction terms to estimate the "Error" variability, what are the values of $SS_{\text{“}Error''}$, $df_{\text{“}Error''}$, and $MS_{\text{“}Error''}$?

(e) **[5]** Give the prediction equation that includes only those terms that are statistically significant.

(f) **[4]** What would your recommendation be if we wished to maximize the number of fish (bluegills) caught?

(g) **[5]** What is the predicted value at that "best" combination and the prediction interval?

(h) **[4]** Could you also estimate "Error" by dropping a single factor from the analysis and getting pseudo-replication? If not, why not? If so, which factor would you drop and how many degrees of freedom would you have for pseudo-replication error?

5. **[15 pts]** An experiment is performed on a nitride etch process on a single-wafer plasma etcher. This process is used in the manufacture of computer chips. Two variables, the gap between anode and cathode and the power applied to the cathode are investigated. Two observations at each of the four treatment combinations along with four center point observations are taken in a completely random order. The response variable is the etch rate for silicon nitride. Refer to the Minitab output on the next page.

   (a) **[5]** What effects are statistically significant? You must justify your answer by making a comparison to a critical effect size or test statistic.

   (b) **[5]** Is there a significant curvature effect? Again, justify your answer with the value of the appropriate test statistic.

   (c) **[5]** Another analysis uses the linear and quadratic coded values of power to come up with a regression model. According to the Minitab output the t-statistic indicates that the quadratic term is not significant. Does this agree with your finding in (b). If so, explain why. If not, explain why not.

```
MTB > %NormPlot 'EFFE1';
SUBC>   Title 'Normal Plot of Estimated Effects: Fish Data'.
Executing from file: C:\MTBWIN\MACROS\NormPlot.MAC
Macro is running ... please wait
MTB > FFactorial  'EtchRate' =  Gap|Power.
```

Fractional Factorial Fit

```
Estimated Effects and Coefficients for EtchRate

Term           Effect      Coef  Std Coef t-value      P
Constant                1186.17     18.03   65.84  0.000
Gap             36.50     18.25     32.07    0.83  0.433
Power          299.00    149.50     32.07    6.78  0.000
Gap*Power      147.50     73.75     32.07    3.34  0.010

Analysis of Variance for EtchRate

Source                 DF      Seq SS     Adj SS     Adj MS      F      P
Main Effects            2      181467     181467    90733.2  33.29  0.000
2-Way Interactions      1       43513      43513    43513.5  11.17  0.010
Residual Error          8       31161      31161     3895.1
  Curvature             1       28017      28017    28016.7  63.38  0.000
  Pure Error            7        3144       3144      449.1
Total                  11      356140

MTB > Regress 'EtchRate' 2 'C1' 'C2';
SUBC>   Constant.
```

Regression Analysis

```
The regression equation is
EtchRate = 1186 + 150 C1 - 34.2 C2

Predictor        Coef       Stdev    t-ratio        p
Constant      1186.17       31.37      55.51    0.000
C1             149.50       26.17       5.71    0.000
C2             -34.17       15.11      -2.26    0.050

s = 74.03        R-sq = 80.7%     R-sq(adj) = 76.5%

Analysis of Variance

SOURCE        DF          SS          MS         F        P
Regression     2      306819      103409     18.87    0.001
Error          9       49321        5480
Total         11      356140

SOURCE        DF      SEQ SS
C1             1      178803
C2             1       28017


MTB >
```