Stat 441/541 Midterm   Due Monday, March 6, 2017   Name: ________________________

Partial SAS code to read in the data for problems 1, 2, and 4 will be emailed to you. You are to work alone on this exam. Any communication with any person other than the instructor about any exam question will result in a zero grade for the entire midterm exam.

1. A manufacturer is interested in the disintegration times of seven types of tablets:

(A)	magnesium stearate	12 mesh granule size
(B)	liquid petrolatum	12 mesh granule size
(C)	magnesium stearate	16 mesh granule size
(D)	liquid petrolatum	16 mesh granule size
(E)	magnesium stearate	20 mesh granule size
(F)	liquid petrolatum	20 mesh granule size
(G)	talc powder	16 mesh granule size

Ten tablets of each type were completely randomized in manufacturing and in the testing. The following table contains the 60 disintegration times (in seconds) for this CRD:

A	56.3	61.1	60.9	53.8	59.3	56.7	60.8	55.9	60.9	55.1
B	57.3	61.8	60.8	63.5	60.6	58.7	56.5	54.1	64.2	60.8
C	62.1	63.9	67.5	65.7	65.9	61.9	62.2	65.2	70.2	65.7
D	63.6	62.0	64.4	63.1	69.5	68.6	61.8	72.1	60.7	67.1
E	69.5	69.8	70.6	68.6	66.3	64.5	66.8	66.1	71.7	66.5
F	70.8	74.7	72.3	73.6	73.0	67.1	75.8	72.7	70.1	68.0
G	64.7	62.0	65.4	64.0	70.4	69.6	62.7	70.2	62.5	68.5

(a) (2pt) The SAS code includes code for both a Box-Cox transformation and for the empirical method. What transformations (if any) are suggested by these two methods?

(b) (1.5pt) Perform Levene's Test to assess the homogeneity of variance assumption. Be sure to state the hypotheses and include the associated p-value for the test.

(c) (7pt) Perform a complete and appropriate analysis of this data. Be sure you include the model and state all hypotheses tested. Also include and interpret the results from a two-sided Dunnett's multiple comparison procedure with G talc powder being the control. You can use $\alpha = .05$ for testing.

(d) (1.5pt) If you decided to use a transformed response, then justify your decision. If you did not use a transformed response, then justify your decision.

(e) (1.5pt) Perform a test for the contrast comparing the mean of the three disintegration time means for the magnesium stearate (MS) treatments to the mean of the three disintegration time means for the three liquid petrolatum (LP) treatments. Be sure to state the hypotheses in terms of the means (the μ_i's) and include the associated p-value for the test.

(f) (1.5pt) The researcher wanted to check for a mesh size effect. To do this, she wants to perform a test for the contrast comparing the mean of the two disintegration time means for the 12 mesh granule size treatments to the mean of the two disintegration time means for the 20 mesh granule size treatments. Be sure to state the hypotheses in terms of the means (the μ_i's) and include the associated p-value for the test. 5

(g) (1.5pt) The researcher wanted to check for linear and quadratic trends for the means across the four granule sizes <u>for the MS tablets</u>. To do this, she wants to use two orthogonal contrasts. Be sure to state the hypotheses in terms of the means (the μ_i's) and include the associated p-values for the two contrast tests.

(h) (1.5pt) The researcher wanted to check for linear and quadratic trends for the means across the four granule sizes for the LP tablets. To do this, she wants to use two orthogonal contrasts. Be sure to state the hypotheses in terms of the means (the μ_i's) and include the associated p-values for the two contrast tests.

(i) (1.5pt) Compare the results of the orthogonal contrasts in (g) and (h) to the side-by-side boxplots. Briefly describe why the contrast results are or are not consistent with the pattern of the boxplots.

2. The data for this problem was collected from 15 students (subjects) in this class. Each subject was asked to guess the amount of liquid that can be held in each of 7 glasses:

Clear goblet with stem	Tall blue clear	Clear beer with stem	Short blue clear	Clear beer no stem	Tea mug not clear	Large mug
(1\CGS)	(2\TBl)	(3\CBS)	(4\Tea)	(5\SBl)	(6\CBN)	(7\Mug)
10	9	9	9	9	11	—

The SAS code converts milliliters guesses to the nearest tenth of an ounce. The Large Mug (Treatment 7\Mug) was included only to establish that all glass volumes were not equal, and will be excluded from your analysis. The true volumes in ounces of the other six "treatment" glasses are given in the last row above.

(a) (6.5pt) Perform a complete analysis comparing the means of the student guesses. Be sure you include the model and state all hypotheses tested. Also include Bonferroni's multiple comparison test. You do not need to transform the response. Use $\alpha = .10$ for testing.

(b) (1.5pt) Perform a test for the contrast comparing the mean of the two means for the stemmed glasses to the mean of the two means for the blue glasses. Be sure to include the p-value.

(c) (1.5pt) Perform a test for the contrast comparing the tea mug mean to the mean of the five means for the other glasses. Be sure to include the p-value.

(d) (.5pt) Are the two contrasts in (b) and (c) orthogonal? Why or why not?

(e) (2pt) Suppose you are asked to compare how close the guesses were to the true glass volumes across the six glass treatments. Briefly describe how (if possible) you could address this request from the data provided. No analysis is required. If you do not believe this can be addressed, briefly give a reason why.

3. (2.5pt) Look at the attached SAS output. It contains two ANOVA tables. On the top is the ANOVA for a RCBD and on the bottom is the ANOVA for the RCBD data but with blocks removed from the model. At an $\alpha = .05$ level, you would draw different conclusions regarding the equality of treatment effects. Suppose you hear the following:

> "By systematically controlling outside effects, the residual mean square error will be reduced. This increases the precision of the experiment making detection of significant effects more likely."

This example may appear to be inconsistent this statement. Explain why you reject the null hypothesis without blocks and fail to reject the null hypothesis with blocks (assuming $\alpha = .05$).

4. (6pt) An agronomist wanted to compare the effect of five different sources of nitrogen on dry matter yield of barley used as a forage crop. The five sources of nitrogen were:

1. $(NH_4)_2SO_4$	3. $CO(NH_2)_2$	5. $NaNO_3$
2. NH_4NO_3	4. $Ca(NO_3)_2$	6. Control (no nitrogen source)

For his experimental design he chose a completely randomized design (CRD) with four replicates per nitrogen treatment. At maturity he clipped each plot and measured dry matter of forage produced. He obtained the yields (kilograms per plot). The data are summarized below.

Treatment	Yield			
1	32.1	35.6	41.9	35.4
2	30.1	31.5	37.1	30.8
3	25.4	27.4	33.8	31.1
4	24.1	33.0	35.6	31.4
5	26.1	31.0	33.8	31.9
6	23.2	24.8	26.7	26.7

Use the sample means and an estimate of σ based on the oneway ANOVA to estimate the sample size needed for a CRD to achieve

(a) A power of at least .95 for the ANOVA F-test assuming $\alpha = .05$.

(b) A power of at least .90 for the ANOVA F-test assuming $\alpha = .01$.

(c) What would be the estimated power for the ANOVA F-test assuming $n = 2$ replicates are to be taken for each treatment and assuming $\alpha = .05$?

5. (5pt) **Stat 541 Students**: Read Section 4.3 (pages 165-168) on Graeco-Latin squares. Consider the following experiment. A food processor wants to determine the effect of package design on the sale of one of her company's products. She had five designs to be tested: A, B, C, D, E. There were several sources of variation, however, whose possible effects were to be accounted for in the study. These included (1) day of the week, (2) differences among stores, and (3) effect of shelf height. She decided to use a Graeco-Latin square with five weekdays (M, Tu, W, Th, F,) for the row blocks, five stores for the column blocks, and five shelf heights for the Greek letter blocks. The experimental design and sales in dollars are shown in the following table.

(a) Perform an ANOVA. State the null and alternative hypotheses of interest, and your conclusions. You can use $\alpha = .05$ for testing.

(b) Include a Tukey multiple comparison procedure, and your conclusions from it.

DAY	STORE I	II	III	IV	V
MON.	E α 238	C δ 228	B γ 158	D ϵ 188	A β 74
TUES.	D δ 149	B β 220	A α 92	C γ 169	E ϵ 282
WED.	B ϵ 222	E γ 295	D β 104	A δ 54	C α 213
THUR.	C β 187	A ϵ 66	E δ 242	B α 122	D γ 90
FRI.	A γ 65	D α 118	C ϵ 279	E β 278	B δ 176

6. (5pt) **For Stat 541 students**: Suppose you have a RCBD with a treatments and b blocks and with constraint $\sum_{i=1}^{a} \tau_i = 0$.

   (a) Show that you get the same $\widehat{\tau}_i$ estimates whether we use $\beta_1 = 0$ or use $\beta_b = 0$ for the second constraint.

   (b) What would be the $\widehat{\mu}$ estimate for each of the two cases in (a)?

PROBLEM # 3 ---- TABLE I

RCBD ANOVA RESULTS

The GLM Procedure

Dependent Variable: response

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	9	0.32600000	0.03622222	1.22	0.3358
Error	20	0.59266667	0.02963333		
Corrected Total	29	0.91866667			

R-Square	Coeff Var	Root MSE	response Mean
0.354862	1.597865	0.172143	10.77333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	4	0.29533333	0.07383333	2.49	0.0759
block	5	0.03066667	0.00613333	0.21	0.9557

==

PROBLEM # 3 ---- TABLE II

ANOVA WITHOUT BLOCKS

The GLM Procedure

Dependent Variable: response

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	0.29533333	0.07383333	2.96	0.0394
Error	25	0.62333333	0.02493333		
Corrected Total	29	0.91866667			

R-Square	Coeff Var	Root MSE	response Mean
0.321480	1.465683	0.157903	10.77333

Source	DF	Type III SS	Mean Square	F Value	Pr > F
trt	4	0.29533333	0.07383333	2.96	0.0394