Stat 850 **Clayton** **May 9, 2017**

Final Exam

Name:____

Instructions:

1. This exam is open book. You may use textbooks, notebooks, class notes, and a calculator.
2. Laptops, tablets, phones and other devices with wireless capability must be turned off and put away.
3. Do all your work in the spaces provided. If you need additional space, use the back of the preceding page, indicating *clearly* that you have done so.
4. To get full credit, you must show your work. Partial credit will be awarded.
5. Some partial computations have been provided on some questions. You may find some *but not necessarily all* of these computations useful. You may assume that these computations are correct.
6. Do not dwell too long on any one question. Answer as many questions as you can.
7. Note that some questions have multiple parts. For some questions, these parts are independent, and so you can work on part (b) or (c) separately from part (a).

For grader's use:

Question	Possible Points	Score
1	36	
2	25	
3	24	
4	15	
Total	100	

1. An experiment was conducted to study soil organic matter (SOM) concentrations in response to various cropping approaches. Specifically of interest was: (A) the use of organic fertilizer, chemical fertilizer, or a blend of organic and chemical fertilizer; (B) whether the cover crop was genetically modified (also called GM corn), conventional corn (non-GM corn), alfalfa, or wheat.

   The experiment was conducted as follows. A large field was divided into 2 areas. Within each area, 12 plots were located, and at random, one of the combinations of fertilizer treatment and cover crop was assigned to each plot. This was done in such a way that each of the 12 combinations of fertilizer and cover crop appeared exactly once within each of the 2 areas.

   The fertilizer was applied at the beginning of the growing season (in the spring) and each plot was planted according to the cover crop assigned to it. At the end of the growing season (in the fall) the investigators went to the center of each plot and removed some soil. They did this by digging a small hole to a depth of 30 cm, and removed about 100 gm of soil. Then, right next to that hole, they dug another hole to a depth of 70 cm, and removed 100 gm of soil from that hole. Finally, next to those two holes another was dug to 100 cm and about 100 gm of soil were removed from that hole, too. Each of these samples of soil were analyzed in a standard way to measure soil organic matter.

   In a separate handout is the data, some summary statistics including Type 1 Sums of Squares, and sample means for each group.

   (a) Construct the ANOVA table for this experiment, indicating source and df only.

(b) Conduct a formal test for the presence of an interaction between cover crop and depth.

(c) Let μ_{CC} represent the mean soil organic matter for the combination of chemical fertilizer and conventional corn, and let μ_{OG} represent the mean soil organic matter for the combination of organic fertilizer and GM corn. Construct a 90% confidence interval for $\mu_{CC} - \mu_{OG}$.

2. Consider the model for a randomized complete block design: $Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$ where $i = 1, \ldots, K$, $j = 1, \ldots B$, and where α_i represents the effect of treatment i, β_j represents the effect of block j, and the ϵ_{ij} are all independent, with distribution $N(0, \sigma_\epsilon^2)$. This is the unconstrained model for this design — no other assumptions about the α_i or β_j are imposed.

   (a) Consider the vector $\theta^T = (\mu, \alpha_1, \ldots, \alpha_K, \beta_1, \ldots, \beta_B)$. Let λ be a vector of length $K + B + 1$, assume $K = 3$ and $B = 2$. Explain what it means to say that $\lambda^T\theta$ is "estimable."

   (b) Is $\mu + \alpha_1 + \beta_1$ estimable? Explain.

   (c) Is $\mu + \alpha_1$ estimable? Explain.

(d) Is $\alpha_1 - \alpha_2$ estimable? Explain.

(e) Suppose the observation $Y_{1,1}$ is missing. Is $\alpha_1 - \alpha_2$ estimable? Explain.

3. In class we studied an example of growth data: The data consist of 4 measurements made on each of 27 subjects; 11 are female, and 16 are male. The distance between two landmarks in the skull (the pituitary gland and the pterygomaxillary fissure) was measured at ages 8, 10, 12, and 14.

   In a separate handout is some SAS code and corresponding output for four models, called A, B, C, and D. For the pairs of models specified below, perform a formal test, if possible, to compare the models, and justify the test that you perform. If it is not possible to perform a formal test, explain why, and then make a useful comparison between the models and explain why the comparison is legitimate. If it is not possible to make a useful comparison, explain why.

   (a) Model A and Model C

   (b) Model A and Model D

(c) Model A and Model B

4. Consider the rare, but completely valid model, $Y_i = \beta x_i + \varepsilon_i$ with ε_i iid $N(0, \sigma_\varepsilon^2)$. This is a regression of a variable Y on another variable X without an intercept in the model.

   The goal in this problem is to find the REML estimate of σ_ε^2 when $\boldsymbol{Y} = (7, 4, 2)'$ and $\boldsymbol{X} = (0, 1, 2)'$.

   In general we know that one way to find the REML estimate of σ_ε^2 is to first find $\boldsymbol{K}$ of full rank such that $\boldsymbol{KX} = \mathbf{0}$.

   (a) Find a suitable $\boldsymbol{K}$ for this problem and show that it satisfies $\boldsymbol{KX} = \mathbf{0}$. Hint: You may use $\boldsymbol{K}$ of the form:

   $$\begin{pmatrix} 1 & 0 & 0 \\ 0 & k_2 & k_3 \end{pmatrix}$$

   (b) Find the REML estimate of σ_ε^2. Hint: You may use, without proof, the fact that if $W_1, W_2, \ldots, W_n$ are iid $N(0, \tau^2)$, then the MLE of τ^2 is $\sum W_i^2/n$.

(c) In class, we had the general result that if $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and we seek $\boldsymbol{K}$ such that $\boldsymbol{K}\boldsymbol{X} = \mathbf{0}$, then we can use $\boldsymbol{K} = \boldsymbol{C}[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{X}']$. Is your answer in part (a) of this form? Explain.

Enter your name here:____________________________________

SAS: Question 1

area	fert	crop	depth	som
1	org	gmcorn	30	66
1	org	gmcorn	70	57
1	org	gmcorn	100	57
1	org	convcorn	30	86
1	org	convcorn	70	80
1	org	convcorn	100	82
1	org	alf	30	87
1	org	alf	70	76
1	org	alf	100	77
1	org	wheat	30	96
1	org	wheat	70	85
1	org	wheat	100	87
1	chem	gmcorn	30	49
1	chem	gmcorn	70	42
1	chem	gmcorn	100	44
1	chem	convcorn	30	45
1	chem	convcorn	70	36
1	chem	convcorn	100	35
1	chem	alf	30	52
1	chem	alf	70	44
1	chem	alf	100	41
1	chem	wheat	30	50
1	chem	wheat	70	44
1	chem	wheat	100	45
1	blend	gmcorn	30	51
1	blend	gmcorn	70	45
1	blend	gmcorn	100	44
1	blend	convcorn	30	45
1	blend	convcorn	70	39
1	blend	convcorn	100	36
1	blend	alf	30	58
1	blend	alf	70	52
1	blend	alf	100	47
1	blend	wheat	30	48
1	blend	wheat	70	43
1	blend	wheat	100	42

area	fert	crop	depth	som
2	org	gmcorn	30	75
2	org	gmcorn	70	62
2	org	gmcorn	100	63
2	org	convcorn	30	82
2	org	convcorn	70	72
2	org	convcorn	100	74
2	org	alf	30	95
2	org	alf	70	94
2	org	alf	100	91
2	org	wheat	30	97
2	org	wheat	70	86
2	org	wheat	100	83
2	chem	gmcorn	30	54
2	chem	gmcorn	70	43
2	chem	gmcorn	100	42
2	chem	convcorn	30	58
2	chem	convcorn	70	49
2	chem	convcorn	100	50
2	chem	alf	30	57
2	chem	alf	70	48
2	chem	alf	100	48
2	chem	wheat	30	56
2	chem	wheat	70	50
2	chem	wheat	100	45
2	blend	gmcorn	30	59
2	blend	gmcorn	70	50
2	blend	gmcorn	100	52
2	blend	convcorn	30	56
2	blend	convcorn	70	50
2	blend	convcorn	100	49
2	blend	alf	30	52
2	blend	alf	70	48
2	blend	alf	100	48
2	blend	wheat	30	44
2	blend	wheat	70	38
2	blend	wheat	100	36

The GLM Procedure

Dependent Variable: som

Source	DF	Sum of Squares	Mean Square
Model	71	21968.31944	309.41295
Error	0	0.00000	.
Corrected Total	71	21968.31944	

Source	DF	Type I SS	Mean Square
area	1	284.01389	284.01389
fert	2	16922.19444	8461.09722
area*fert	2	40.86111	20.43056
crop	3	795.04167	265.01389
area*crop	3	126.81944	42.27315
fert*crop	6	1984.25000	330.70833
area*fert*crop	6	627.13889	104.52315
depth	2	1034.02778	517.01389
area*depth	2	1.36111	0.68056
fert*depth	4	19.97222	4.99306
area*fert*depth	4	13.63889	3.40972
crop*depth	6	15.41667	2.56944
area*crop*depth	6	50.30556	8.38426
fert*crop*depth	12	28.91667	2.40972
area*fert*crop*depth	12	24.36111	2.03009

Least Squares Means

fert	crop	depth	som LSMEAN
blend	alf	30	55.0000000
blend	alf	70	50.0000000
blend	alf	100	47.5000000
blend	convcorn	30	50.5000000
blend	convcorn	70	44.5000000
blend	convcorn	100	42.5000000
blend	gmcorn	30	55.0000000
blend	gmcorn	70	47.5000000
blend	gmcorn	100	48.0000000
blend	wheat	30	46.0000000
blend	wheat	70	40.5000000
blend	wheat	100	39.0000000
chem	alf	30	54.5000000
chem	alf	70	46.0000000
chem	alf	100	44.5000000
chem	convcorn	30	51.5000000
chem	convcorn	70	42.5000000
chem	convcorn	100	42.5000000
chem	gmcorn	30	51.5000000
chem	gmcorn	70	42.5000000
chem	gmcorn	100	43.0000000
chem	wheat	30	53.0000000
chem	wheat	70	47.0000000
chem	wheat	100	45.0000000
org	alf	30	91.0000000
org	alf	70	85.0000000
org	alf	100	84.0000000
org	convcorn	30	84.0000000
org	convcorn	70	76.0000000
org	convcorn	100	78.0000000
org	gmcorn	30	70.5000000
org	gmcorn	70	59.5000000
org	gmcorn	100	60.0000000
org	wheat	30	96.5000000
org	wheat	70	85.5000000
org	wheat	100	85.0000000

SAS: Question 3

```
data repeated;
    input Person Gender $ y1 y2 y3 y4 @@;
    y=y1; Age=8;  output;
    y=y2; Age=10; output;
    y=y3; Age=12; output;
    y=y4; Age=14; output;
    drop y1-y4;
    datalines;
       1   F   21.0    20.0    21.5    23.0  2   F   21.0    21.5    24.0    25.5
       3   F   20.5    24.0    24.5    26.0  4   F   23.5    24.5    25.0    26.5
[...remaining data not shown...]
       ;
```

Model A:

```
proc mixed ;
    class Person Gender ;
    model y = Gender Age ;
    repeated / type=cs subject= Person(Gender) ;

           Covariance Parameter Estimates

        Cov Parm     Subject           Estimate
        CS           Person(Gender)      3.2668
        Residual                         2.0495

         -2 Res Log Likelihood          437.5
         AIC (Smaller is Better)        441.5

            Type 3 Tests of Fixed Effects

                    Num     Den
     Effect          DF      DF    F Value    Pr > F
     Gender           1      25       9.29    0.0054
     Age              1      80     114.84    <.0001
```

Model B:

```
proc mixed ;
    class Person Gender ;
    model y = Gender Age ;
    random intercept Age / type=un subject=Person(Gender) ;

           Covariance Parameter Estimates

        Cov Parm     Subject           Estimate
        UN(1,1)      Person(Gender)      7.8233
        UN(2,1)      Person(Gender)     -0.4850
        UN(2,2)      Person(Gender)     0.05127
        Residual                         1.7162

         -2 Res Log Likelihood          435.2
         AIC (Smaller is Better)        443.2

            Type 3 Tests of Fixed Effects

                    Num     Den
     Effect          DF      DF    F Value    Pr > F
     Gender           1      25       8.02    0.0090
     Age              1      26      85.85    <.0001
```

Model C:

```
proc mixed ;
    class Person Gender ;
    model y = Gender Age ;
    repeated / type=un  subject= Person(Gender) ;

            Covariance Parameter Estimates

         Cov Parm    Subject           Estimate
         UN(1,1)     Person(Gender)      5.3746
         UN(2,1)     Person(Gender)      2.7869
         UN(2,2)     Person(Gender)      4.2151
         UN(3,1)     Person(Gender)      3.8070
         UN(3,2)     Person(Gender)      2.9096
         UN(3,3)     Person(Gender)      6.3355
         UN(4,1)     Person(Gender)      2.6284
         UN(4,2)     Person(Gender)      3.1684
         UN(4,3)     Person(Gender)      4.3014
         UN(4,4)     Person(Gender)      5.3764

         -2 Res Log Likelihood          428.7
         AIC (Smaller is Better)        448.7

            Type 3 Tests of Fixed Effects

                    Num     Den
     Effect          DF      DF    F Value    Pr > F
     Gender           1      25       7.72    0.0102
     Age              1      25      92.28    <.0001
```

Model D:

```
proc mixed ;
    class Person Gender Age;
    model y = Gender Age ;
    repeated / type=cs subject= Person(Gender) ;

           Covariance Parameter Estimates

        Cov Parm     Subject           Estimate
        CS           Person(Gender)      3.2595
        Residual                         2.0785

                  Fit Statistics

         -2 Res Log Likelihood          433.7
         AIC (Smaller is Better)        437.7

            Type 3 Tests of Fixed Effects

                    Num     Den
     Effect          DF      DF    F Value    Pr > F
     Gender           1      25       9.29    0.0054
     Age              3      78      38.04    <.0001
```