Stat 642 Spring 2022 - Solutions for Homework 6

1. (60 points) Soil Porosity Experiment:

   a. $Y_{ijk} = \mu + A_i + e_{ij} + d_{ijk},\ i = 1, \cdots, t = 15,\ j = 1, n_i = 2,\ k = 1, \cdots, m_{ij} = 1$ or 2, where μ is overall mean, A_i is the random effect due to the selected field, e_{ij} is the random effect due to the selected section from fields, d_{ijk} is the random effect due to variation in Soil subsamples within the same Field-Section.

   $A_i \sim iid\ N(0, \sigma_A^2),\ e_{ij} \sim iid\ N(0, \sigma_e^2),\ d_{ijk} \sim iid\ N(0, \sigma_d^2)$, and $A_i,\ e_{ij}$ and d_{ijk} are indep.

   b. $t = 15,\ n_i = r = 2,\ m_{ij} = 1$ or 2. Use the following SAS program

```
ods html; ods graphics on;
option ls=80 ps=55 nocenter nodate;
data porosity;
INFILE 'u:meth2\kuehl\ex5-8.dat';
INPUT FIELD $ SECTION $ Y @@;
LABEL Y= 'POROSITY';
RUN;
PROC PRINT;
RUN;
 title 'ANALYSIS USING PROC MIXED-TYPE1';
PROC MIXED METHOD=TYPE1;
CLASS FIELD SECTION;
MODEL  Y = /DDFM=SAT RESIDUAL;
RANDOM FIELD SECTION(FIELD);
RUN;
title 'ANALYSIS USING PROC MIXED-REML';
PROC MIXED METHOD=REML;
CLASS FIELD SECTION;
MODEL  Y = /E3;
RANDOM FIELD SECTION(FIELD);
RUN;
ods graphics off; ods html close;
```

   The following is obtained from METHOD=TYPE1 output:

Source	df	SS	MS	EMS	Error Term	Error DF	F	$Pr > F$
FIELD	14	14.432887	1.030920	$\sigma_d^2 + 1.1905\sigma_e^2 + 2.381\sigma_A^2$	0.9921MS(SEC(Field)+ .0079MS(Res)	15.452	1.33	0.2919
SEC(FIELD)	15	11.534705	0.768980	$\sigma_d^2 + 1.2\sigma_e^2$	MS(Res)	6	0.52	0.8546
SUBS	6	8.798863	1.466477	σ_d^2				
Total	35	34.76645475						

   From the above we have $C_1 = 1.1905, \quad C_2 = 2.381, \quad C_3 = 1.2$

   c. Variance for Fields=σ_A^2, Variance for Sections=σ_e^2, Variance for Subs=σ_d^2, :

   $E(MS_{Subs}) = \sigma_d^2 \quad \Rightarrow \quad \hat{\sigma}_d^2 = MS_{Subs} = 1.466477$

   $E(MS_{SEC(FIELD)}) = \sigma_d^2 + C_3\sigma_e^2 \quad \Rightarrow \quad \hat{\sigma}_e^2 = \frac{MS_{SEC(FIELD)} - \hat{\sigma}_d^2}{C_3} = \frac{.76898 - 1.466477}{1.2} = -0.5812$ (questionable)

   $E(MS_{FIELD}) = \sigma_d^2 + C_1\sigma_e^2 + C_2\sigma_A^2 \ \Rightarrow\ \hat{\sigma}_A^2 = \frac{MS_{FIELD} - \hat{\sigma}_d^2 - C1\hat{\sigma}_e^2}{C_2} = \frac{1.03092 - 1.466477 - 1.1905*(-.5812)}{2.381} = 0.10767$

   With $\hat{\sigma}_e^2 < 0$, the estimation of $\hat{\sigma}_A^2$ is not meaningful.

   With $\hat{\sigma}_e^2 < 0$ by AOV moment matching, REML (Restricted Maximum Likelihood) estimators should be used in place of the AOV MOM's. From SAS PROC MIXED-METHOD=REML,

$$\hat{\sigma}_d^2 = 0.936, \qquad \hat{\sigma}_e^2 = 0, \qquad \hat{\sigma}_A^2 = 0.05944$$

There is a considerable difference in the two sets of estimators which is confirmation that the moment matching estimators from AOV should not be used when the data is unbalanced or when one or more of the moment matching estimators from AOV are negative.

$\hat{\sigma}_y^2 = \hat{\sigma}_d^2 + \sigma_e^2 + \hat{\sigma}_A^2 = 0.9360 + 0 + 0.05944 = .99544 \;\Rightarrow$

Proportion of variability due to Fields is 100(.05944/.99544) = 5.97% and

due to Samples is 100(.936/.99544) = 94.03%

d. $E(MS_{FIELD}) = \sigma_d^2 + c_1\sigma_e^2 + c_2\sigma_A^2$ $(c_1 = 1.1905,\ c_2 = 2.381)$, $E(MS_{SEC(FIELD)}) = \sigma_d^2 + c_3\sigma_e^2$ $(c_3 = 1.2)$. Since $c_1 \neq c_3$, construct the mean square with a linear function of

MS_{SUB} and $MS_{SEC(FIELD)}$ as $M = a_1 MS_{SEC(FIELD)} + a_2 MS_{SUB}$. Then,

$E(M) = a_1(\sigma_d^2 + c_3\sigma_e^2) + a_2\sigma_d^2 = \sigma_d^2 + c_1\sigma_e^2 \;\Rightarrow\; a_1 = \frac{c_1}{c_3} = \frac{1.1905}{1.2} = 0.99208,\ a_2 = 1 - a_1 = 0.00792.$

Thus, $M = (0.99208)MS_{SEC(FIELD)} + (0.00792)MS_{SUB} = (0.99208)(.768980) + (0.00792)(1.466477) = 0.7745$ with approximate degrees of freedom using the Satterthwaite Approximation:

$$\upsilon = \frac{M^2}{\frac{(a_1 MS_{SEC(FIELD)})^2}{15} + \frac{(a_2 MS_{SUB})^2}{6}} = \frac{(.7745)^2}{\frac{((.99208)(.76898033))^2}{15} + \frac{((0.00792)(1.46647717))^2}{6}} = 15.45$$

For testing $H_0 : \sigma_A^2 = 0$ vs $H_1 : \sigma_A^2 > 0$, test statistic is $F_0 = \frac{MS_{FIELD}}{M} = \frac{1.03092}{0.7745} = 1.33$.

Since $F_0 < F_{0.05,14,\upsilon} = F_{0.05,14,15.45} = qf(.95, 14, 15.45) = 2.4005$ and $p - value = 1 - pf(1.33, 14, 15.45) = 0.2924 > 0.05$, fail to reject H_0. Thus, we conclude that there is not significant evidence of a difference in the populations of fields.

e. For testing $H_0 : \sigma_e^2 = 0$ vs $H_1 : \sigma_e^2 > 0$, test statistic is $F_0 = \frac{MS_{SEC(FIELD)}}{MS_{SUB}} = 0.52$. Since $F_0 < F_{0.05,15,6} = 3.98$ and $p - value = 0.8546 > 0.05$, fail to reject H_0. Thus, we conclude that there is not significant evidence of a difference in the populations of sections within fields.

The following R program can be used to obtain the above results:

```
install.packages("lsmeans")
install.packages("ggplot2")
install.packages("lme4")
library(lsmeans)
library(ggplot2)
library(lme4)
field = c(rep(1,4), rep(2,4), rep(3,4), rep(4,4), rep(5,4), rep(6,4), rep(7,4), rep(8,4), rep(9,4),
          rep(10,4), rep(11,4), rep(12,4), rep(13,4), rep(14,4), rep(15,4))
field = as.factor(field)
section = c(rep(1,2), rep(2,2), rep(3,2), rep(4,2), rep(5,2), rep(6,2), rep(7,2), rep(8,2), rep(9,2),
          rep(10,2), rep(11,2), rep(12,2), rep(13,2), rep(14,2), rep(15,2), rep(16,2), rep(17,2),
          rep(18,2), rep(19,2), rep(20,2), rep(21,2), rep(22,2), rep(23,2), rep(24,2),
          rep(25,2), rep(26,2), rep(27,2), rep(28,2), rep(29,2), rep(30,2)
          )
section = as.factor(section)
poro =

c(3.846,3.712,5.629,2.021,5.087,NA,4.621,NA,4.411,NA,3.357,NA,3.991,NA,5.766,NA,5.677,NA,3.333,NA,
  4.355,6.292,4.940,4.810,2.983,NA,4.396,NA,5.603,NA,3.683,NA,5.942,NA,5.014,NA,5.143,NA,4.061,NA,
  3.835,2.964,4.584,4.398,4.193,NA,4.125,NA,3.074,NA,3.483,NA,3.867,NA,4.212,NA,6.247,NA,4.730,NA)
data <- data.frame(poro,field,section)

#treat Field as a fixed effect to obtain Sum of Squares

fixfield = lm(poro ~ field + field:section)
summary(fixfield)
anova(fixfield)

#treat Field as a random effect to obtain estimates of variances

ranfield = lmer(poro ~ 1+(1| field) + (1|section:field),data)
summary(ranfield)
```

2. (20 points) Cottonseed Problem:

   a. Because the levels of the treatment factor Cotton Gins were randomly selected, it would not be appropriate to test for differences in the means of the eight Cotton Gins

   b. Determine the number of reps, r, for a fixed number of random treatment levels, t.

      With $\alpha = .01$, $\gamma_o = .90$, $t = 8$, and $\tau_o = \frac{\sigma_A^2}{\sigma_e^2} = 2$, we obtain the value of the number of reps, r, as follows:

      i. $\nu_1 = t - 1 = 7,\ \nu_2 = t(r-1) = 8(r-1), \quad \lambda = \sqrt{1 + r\tau_o} = \sqrt{1 + 2r}$

      ii. Power at λ is given by $\gamma(\lambda) = 1 - pf(qf(1 - .01, 7, 8(r-1))/(1 + r\tau_o), 7, 8(r-1))$ using the R-function for the cdf of an F-distribution.

      iii. Iteratively determine the smallest value of r such that $\gamma(\lambda) \geq .90$

      iv. Using either Table X in the textbook, the SAS program, **repsize,randomeffects,fixedt.sas**, or the following R code we obtain:

```
r=seq(4,8,1)
t=8
nu1=t-1
nu2=t*(r-1)
a=.01
tau=2
L = sqrt(1+2*r)
Fcr = qf(1-a,nu1,nu2)
Power = 1-pf(Fcr/(1+tau*r),nu1,nu2)
out = cbind(r,nu1,nu2,Fcr,L,Power)
 out
     r nu1 nu2    Fcr        L     Power
[3,] 4   7  24 3.495928 3.000000 0.8999179
[4,] 5   7  32 3.258338 3.316625 0.9503449
[5,] 6   7  40 3.123757 3.605551 0.9724452
[6,] 7   7  48 3.037188 3.872983 0.9833860
[7,] 8   7  56 2.976845 4.123106 0.9893302
```

      From the above results, we have that $r \geq 5$, although, $r = 4$ yields almost a power of 0.90.

3. (20 points) Determine the number of random treatment levels, t, for a fixed number of reps, $r = 5$.

   With $\alpha = .01$, $\gamma_o = .90$, $r = 5$, and $\tau_o = \frac{\sigma_A^2}{\sigma_e^2} = (2.1)^2/(2)^2 = 1.1025$, we obtain the value of the number of circuits, t, as follows:

   i. $\nu_1 = t - 1,\ \nu_2 = t(r-1) = t(5-1), \quad \lambda = \sqrt{1 + r\tau_o} = \sqrt{1 + 5(1.1025)} = 2.55196$

   ii. Power at τ_o is given by $\gamma(\tau_o) = 1 - pf(qf(1 - .01, t-1, t(5-1))/(1 + r\tau_o), t-1, t(5-1))$ using the R-function for the cdf of an F-distribution.

   iii. Iteratively determine the smallest value of r such that $\gamma(\tau_o = 1.1025) \geq .90$

   iv. Using either Table X in the textbook, the SAS program, **trtsize,randomeffects,fixedr.sas**, or the following R code we obtain:

```
r=5
t=seq(2,12)
nu1=t-1
nu2=t*(r-1)
a=.01
tau=1.1025
L = sqrt(1+tau*r)
Fcr = qf(1-a,nu1,nu2)
Power = 1-pf(Fcr/(1+tau*r),nu1,nu2)
out = cbind(t,nu1,nu2,Fcr,L,Power)
out

       t     nu1    nu2      Fcr          L          Power
 [7,]  8      7     32     3.258338     2.55196     0.8271899
 [8,]  9      8     36     3.051726     2.55196     0.8700052
 [9,] 10      9     40     2.887560     2.55196     0.9028619
[10,] 11     10     44     2.753570     2.55196     0.9278456
```

From the above results, we have that $t \geq 10$ that is, a random sample of at least 10 different Types of Circuits is needed to achieve the specifications in the design.

From Table X on page 620 in the Book with $\nu_1 = 8,\ \nu_2 = 36,\ \alpha = .01,\ \lambda = 2.55196$ the power is determined to be approximately .87 which is less than .90. The book does not have tables for $\nu_1 = 9, 10, 11$ so the power at $\nu_1 = 9$ cannot be determined from these graphs.