Stat 511 Exam 2

April 7, 2009
Prof. Vardeman

I have neither given nor received unauthorized assistance on this exam.

____
Name

____
Name Printed

1. A data set in *Statistical Tools for Nonlinear Regression* by Huet, Bouvier, Poursat, and Jolivet concerns the dependence of the reaction rate in catalytic conversion of n-pentane to iso-pentane upon some partial pressures. For

$y = rate =$ gm of iso-pentance produced per hour per gm of catalyst

$x_1 = H =$ hydrogen partial pressure

$x_2 = nP =$ n-pentane partial pressure

$x_3 = IP =$ iso-pentane partial pressure

the authors consider a physical model that says that

$$rate \approx \frac{\theta_0\theta_2\left(nP - IP/1.632\right)}{1+\theta_1 H + \theta_2 nP + \theta_3 IP} \quad \text{or} \quad y \approx \frac{\theta_0\theta_2\left(x_2 - x_3/1.632\right)}{1+\theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3}$$

There is an `R` printout at the end of this exam from a session based on the authors' data set and the probabilistic model

$$y_i = \frac{\theta_0\theta_2\left(x_{2i} - x_{3i}/1.632\right)}{1+\theta_1 x_{1i} + \theta_2 x_{2i} + \theta_3 x_{3i}} + \varepsilon_i \qquad (*)$$

for iid $\mathrm{N}\left(0,\sigma^2\right)$ errors ε_i, $i = 1,2,\ldots,24$. Use it as appropriate in what follows.

10 pts

a) **What interpretation** can be made of the parameter θ_0 in this model? (Answer in the context of the physical problem. Consider large n-pentane partial pressures.) **What does this suggest** about setting a starting value for θ_0 in a nonlinear search for $\hat{\boldsymbol{\theta}}_{\text{OLS}}$?

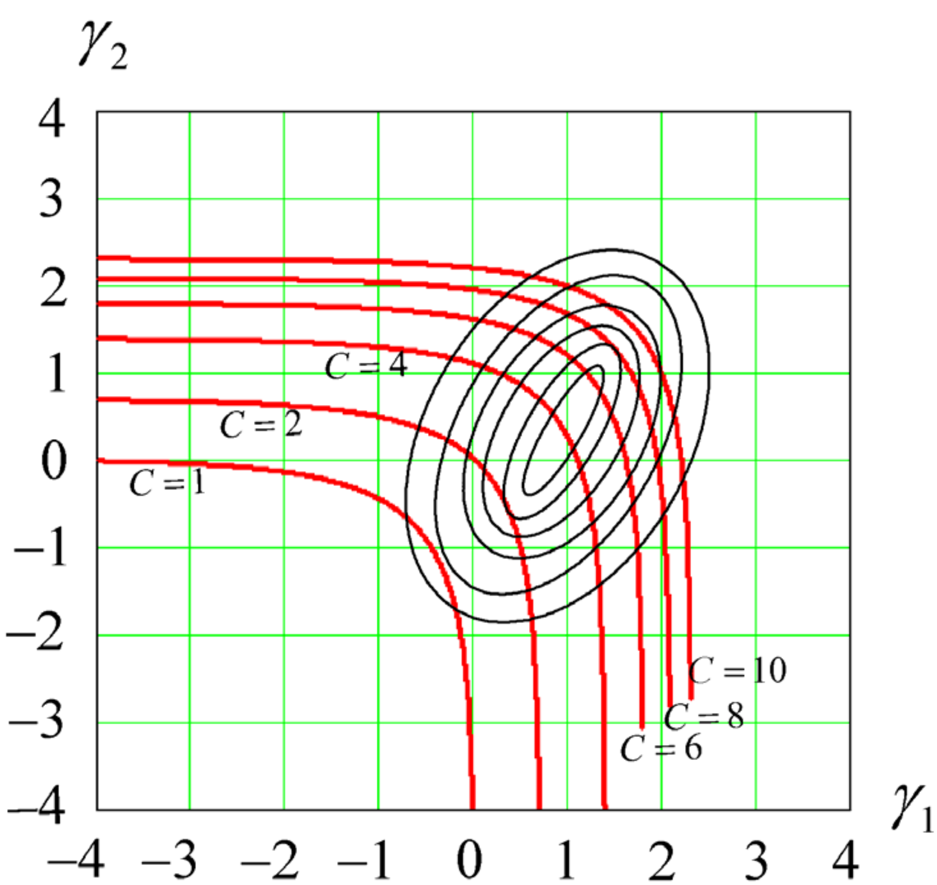
10 pts

b) **Find approximate 95% confidence limits** for the product $\theta_0\theta_2$ that is the multiplier of $\left(x_2 - x_3/1.632\right)$ in the numerator of rational function that specifies the mean reaction rate. (Produce numerical values for the limits.)

10 pts

c) Consider a set of partial pressures $x_1 = 402.6, x_2 = 102.2,$ and $x_3 = 128.9$. (Notice that this is a set that is represented in the original data.) It is possible to show (DO NOT DO SO HERE) that an appropriate standard error for the estimated mean reaction rate at this combination of partial pressures is $\text{s.e.}(\hat{y}) = .011$. Use this fact and **make approximate two-sided 95% prediction limits** for the next reaction rate at this set of partial pressures. (Produce numerical values for the limits.)

2. Below is a cartoon of a contour plot of a (restricted) log-likelihood for the estimation of $\gamma_1 = \ln(\sigma_1^2)$ and $\gamma_2 = \ln(\sigma_2^2)$. Also plotted are contours of $\exp(\gamma_1) + \exp(\gamma_2) = C$ for various values of C. The first 4 level curves for the log-likelihood are at $1.35, 1.92, 2.30,$ and 3.00 units below the maximum log-likelihood. Use the plot to answer the questions on the next page.



7 pts

a) **What**, approximately, are the REML point estimates of σ_1^2 and σ_2^2?

7 pts

b) **What** are approximate 95% confidence limits for σ_1^2 based on this plot?

6 pts

c) **What** are approximate 95% confidence limits for the quantity $\sigma_1^2 + \sigma_2^2$?

3. A famous data set of Prater concerns the performance of a crude oil refining process. Process yield, y, was treated as a function of

$SG = x_1 =$ specific gravity of the crude oil sample

$VP = x_2 =$ vapor pressure of the sample

$V10 = x_3 =$ ASTM 10% point volatility of the crude oil sample

$EP = x_4 =$ end point volatility of the desired product

In the study, 10 physical crude oil samples were employed and parts of each of those were used in from 2 to 4 runs of the process (for a total of $n = 32$ runs). For $No(i)$ the number (from 1 to 10) of the crude oil sample used on run i, we'll consider the model

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \alpha_{No(i)} + \varepsilon_i \text{ for } i = 1, 2, \ldots, 32$$

for $\alpha_1, \ldots, \alpha_{10}$ iid $\mathrm{N}\left(0, \sigma_\alpha^2\right)$ independent of $\varepsilon_1, \ldots, \varepsilon_{32}$ that are iid $\mathrm{N}\left(0, \sigma^2\right)$. (The β's are unknown parameters, as are the two variance components, σ_α^2 and σ^2.) There is a summary of an `R` session at the end of this exam that provides an analysis of the Prater data based on this model. Use it as appropriate to answer the following questions.

8 pts

a) **What is a point estimate** of the correlation between two values of y obtained from a single crude oil sample? (Give a numerical value.)

8 pts

b) **Which** of the two variance components (σ_α^2 or σ^2) is most precisely determined by the Prater data? **Explain**.

8 pts

c) Suppose that tomorrow, two new samples of crude oil with respective sets of physical properties $SG = 40, VP = 5,$ and $V10 = 220$ and then $SG = 40, VP = 6,$ and $V10 = 210$ are used to make two process runs, both with $EP = 250$. **What is** a plausible predictor of the difference in observed values of y, and **what is** a corresponding prediction standard error? (Give numerical values.)

8 pts

d) Suppose that tomorrow a single new sample of crude oil is used to make two process runs, the first with $EP = x_4 = 200$ and the second with $EP = x_4 = 300$. (Note that a single sample of oil must have a fixed set of values of $SG, VP,$ and $V10$.) **What is** a plausible predictor of the difference in observed values of y, and **what is** a corresponding prediction standard error? (Give numerical values.)

8 pts

e) **Which** crude oil sample seems to have produced the consistently lowest process yields after taking account of its physical properties described by $SG, VP,$ and $V10$ and the experimentally set end points, EP? **Explain**.

10 pts

4. Consider a very small normal mixed linear model with

$$\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}\mu + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \end{pmatrix}$$

where as usual, the ε_{ij} are iid $\mathrm{N}\left(0,\sigma^2\right)$ independent of the α_i that are iid $\mathrm{N}\left(0,\sigma_\alpha^2\right)$. Two ANOVA sums of squares for this model can respectively be built from the random vectors

$$\begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}\mathbf{Y} \text{ and } \begin{pmatrix} 1 & 1 & -1 & -1 \end{pmatrix}\mathbf{Y}$$

Argue very carefully that these two vectors (one that is 2×1 and one that is 1×1) are independent.

For Problem #1

```
> Reaction
       H    nP    IP   Rate
1  205.8  90.9  37.1  3.541
2  404.8  92.9  36.3  2.397
3  209.7 174.9  49.4  6.694
4  401.6 187.2  44.9  4.722
5  224.9  92.7 116.3  0.593
6  402.6 102.2 128.9  0.268
7  212.7 186.9 134.4  2.797
8  406.2 192.6 134.9  2.451
9  133.3 140.8  87.6  3.196
10 470.9 144.2  86.9  2.021
11 300.0  68.3  81.7  0.896
12 301.6 214.6 101.7  5.084
13 297.3 142.2  10.5  5.686
14 314.0 146.7 157.1  1.193
15 305.7 142.0  86.0  2.648
16 300.1 143.7  90.2  3.303
17 305.4 141.1  87.4  3.054
18 305.2 141.5  87.0  3.302
19 300.1  83.0  66.4  1.271
20 106.6 209.6  33.0 11.648
21 417.2  83.9  32.9  2.002
22 251.0 294.4  41.5  9.604
23 250.3 148.0  14.7  7.754
24 145.1 291.0  50.2 11.590

> nlrfit<-nls(formula=Rate~theta0*theta2*(nP-
IP/1.632)/(1+theta1*H+theta2*nP+theta3*IP),start=c(theta0=10,theta1=.1,theta2
=.1,theta3=.1),trace=T)
174.8512 :  10.0  0.1  0.1  0.1
169.6269 :  12.43506544  0.06028672  0.04793838  0.09389552
120.7087 :  17.59353606  0.05377972  0.03440488  0.10882831
41.15171 :  26.53074089  0.05938979  0.03270767  0.13688238
3.244297 :  35.97395861  0.07047099  0.03731626  0.16695303
3.234487 :  35.90200224  0.07118139  0.03792012  0.16787327
3.234482 :  35.92166923  0.07081521  0.03771402  0.16707606
3.234482 :  35.92005806  0.07084638  0.03773170  0.16714116

> summary(nlrfit)

Formula: Rate ~ theta0 * theta2 * (nP - IP/1.632)/(1 + theta1 * H +
                theta2 * nP + theta3 * IP)

Parameters:
       Estimate Std. Error t value Pr(>|t|)
theta0 35.92006    8.21227   4.374 0.000294 ***
theta1  0.07085    0.17870   0.396 0.695964
theta2  0.03773    0.10007   0.377 0.710102
theta3  0.16714    0.41600   0.402 0.692108
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4021 on 20 degrees of freedom

Number of iterations to convergence: 7
Achieved convergence tolerance: 6.935e-06
```

```
> vcov(nlrfit)
           theta0      theta1      theta2      theta3
theta0 67.4413434 -1.18124994 -0.69042262 -2.69784401
theta1 -1.1812499  0.03193273  0.01784324  0.07415911
theta2 -0.6904226  0.01784324  0.01001379  0.04143425
theta3 -2.6978440  0.07415911  0.04143425  0.17305849

> fitted(nlrfit)
 [1]  3.6646259  2.4397694  6.3824485  4.9166099  0.7287535  0.5729211
 [7]  3.1083840  2.5004140  3.8844871  2.2690380  0.6422888  4.3485483
[13]  6.3052770  1.2420225  2.8553067  2.8029517  2.7873473  2.8122421
[19]  1.5715929 11.6793504  2.2025878  9.8987988  7.0359069 11.5075844
attr(,"label")
[1] "Fitted values"
```

For Problem #2

```
> Petrol
   No   SG  VP V10  EP    Y
1   A 50.8 8.6 190 205 12.2
2   A 50.8 8.6 190 275 22.3
3   A 50.8 8.6 190 345 34.7
4   A 50.8 8.6 190 407 45.7
5   B 40.8 3.5 210 218  8.0
6   B 40.8 3.5 210 273 13.1
7   B 40.8 3.5 210 347 26.6
8   C 40.0 6.1 217 212  7.4
9   C 40.0 6.1 217 272 18.2
10  C 40.0 6.1 217 340 30.4
11  D 38.4 6.1 220 235  6.9
12  D 38.4 6.1 220 300 15.2
13  D 38.4 6.1 220 365 26.0
14  D 38.4 6.1 220 410 33.6
15  E 40.3 4.8 231 307 14.4
16  E 40.3 4.8 231 367 26.8
17  E 40.3 4.8 231 395 34.9
18  F 32.2 5.2 236 267 10.0
19  F 32.2 5.2 236 360 24.8
20  F 32.2 5.2 236 402 31.7
21  G 41.3 1.8 267 235  2.8
22  G 41.3 1.8 267 275  6.4
23  G 41.3 1.8 267 358 16.1
24  G 41.3 1.8 267 416 27.8
25  H 38.1 1.2 274 285  5.0
26  H 38.1 1.2 274 365 17.6
27  H 38.1 1.2 274 444 32.1
28  I 32.2 2.4 284 351 14.0
29  I 32.2 2.4 284 424 23.2
30  J 31.8 0.2 316 365  8.5
31  J 31.8 0.2 316 379 14.7
32  J 31.8 0.2 316 428 18.0

> Prater.lme<-lme(Y ~ SG+VP+V10+EP, random = ~ 1 | No, data=Petrol)

> summary(Prater.lme)
Linear mixed-effects model fit by REML
 Data: Petrol
       AIC      BIC    logLik
  166.3820 175.4528 -76.19098

Random effects:
 Formula: ~1 | No
        (Intercept) Residual
StdDev:    1.444741 1.872208

Fixed effects: Y ~ SG + VP + V10 + EP
                Value Std.Error DF   t-value p-value
(Intercept) -6.134791 14.554171 21 -0.421514  0.6777
SG           0.219398  0.146938  6  1.493136  0.1860
VP           0.545863  0.520528  6  1.048673  0.3347
V10         -0.154243  0.039962  6 -3.859697  0.0084
EP           0.157177  0.005588 21 28.127561  0.0000
 Correlation:
    (Intr) SG     VP     V10
SG  -0.694
VP  -0.715  0.067
V10 -0.934  0.433  0.836
EP   0.011  0.023 -0.116 -0.197
```

```
Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max 
-1.7806793 -0.6064446 -0.1068210  0.4571694  1.7811272 

Number of Observations: 32
Number of Groups: 10 

> vcov(Prater.lme)
              (Intercept)            SG           VP           V10
(Intercept)  2.118239e+02 -1.484138e+00 -5.414847657 -5.432469e-01
SG          -1.484138e+00  2.159071e-02  0.005089335  2.542959e-03
VP          -5.414848e+00  5.089335e-03  0.270948888  1.738364e-02
V10         -5.432469e-01  2.542959e-03  0.017383640  1.596993e-03
EP           9.014698e-04  1.883113e-05 -0.000337413 -4.393934e-05
                       EP
(Intercept)  9.014698e-04
SG           1.883113e-05
VP          -3.374130e-04
V10         -4.393934e-05
EP           3.122573e-05

> intervals(Prater.lme)
Approximate 95% confidence intervals

 Fixed effects:
                  lower       est.       upper
(Intercept) -36.4018468 -6.1347910 24.13226478
SG           -0.1401457  0.2193982  0.57894198
VP           -0.7278218  0.5458631  1.81954805
V10          -0.2520272 -0.1542427 -0.05645827
EP            0.1455559  0.1571768  0.16879767
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: No 
                    lower     est.    upper
sd((Intercept)) 0.6047843 1.444741 3.451274

 Within-group standard error:
   lower     est.    upper 
1.386645 1.872208 2.527801 

> fixed.effects(Prater.lme)
(Intercept)          SG          VP         V10          EP 
 -6.1347910   0.2193982   0.5458631  -0.1542427   0.1571768 

> random.effects(Prater.lme)
  (Intercept)
A -0.05943851
B -0.21857100
C  1.92029510
D -1.92760603
E -0.21650143
F  0.56931101
G  0.06701580
H  0.19194449
I -0.40278000
```

```
> predict(Prater.lme,level=0:1)
   No predict.fixed predict.No
1   A    12.6201830 12.5607445
2   A    23.6225582 23.5631197
3   A    34.6249335 34.5654950
4   A    44.3698944 44.3104559
5   B     6.6007433  6.3821723
6   B    15.2454667 15.0268957
7   B    26.8765492 26.6579782
8   C     5.8217091  7.7420042
9   C    15.2523165 17.1726116
10  C    25.9403382 27.8606333
11  D     8.6230101  6.6954040
12  D    18.8395014 16.9118953
13  D    29.0559927 27.1283866
14  D    36.1289482 34.2013422
15  E    17.9503034 17.7338020
16  E    27.3809108 27.1644093
17  E    31.7818609 31.5653594
18  F     9.3332384  9.9025494
19  F    23.9506798 24.5199908
20  F    30.5521050 31.1214160
21  G    -0.3373546 -0.2703388
22  G     5.9497170  6.0167328
23  G    18.9953905 19.0624063
24  G    28.1116443 28.1786601
25  H     5.4121939  5.6041383
26  H    17.9863370 18.1782815
27  H    30.4033034 30.5952479
28  I    13.6040214 13.2012414
29  I    25.0779270 24.6751470
30  J     9.5800712  9.6564018
31  J    11.7805463 11.8568768
32  J    19.4822089 19.5585395

> fitted(Prater.lme)
         A          A          A          A          B          B          B 
12.5607445 23.5631197 34.5654950 44.3104559  6.3821723 15.0268957 26.6579782 
         C          C          C          D          D          D          D 
 7.7420042 17.1726116 27.8606333  6.6954040 16.9118953 27.1283866 34.2013422 
         E          E          E          F          F          F          G 
17.7338020 27.1644093 31.5653594  9.9025494 24.5199908 31.1214160 -0.2703388 
         G          G          G          H          H          H          I 
 6.0167328 19.0624063 28.1786601  5.6041383 18.1782815 30.5952479 13.2012414 
         I          J          J          J 
24.6751470  9.6564018 11.8568768 19.5585395 
attr(,"label")
[1] "Fitted values"


> Prater.lmer<-lmer(Y ~ SG+VP+V10+EP+(1|No), Petrol)

> summary(Prater.lmer)
Linear mixed model fit by REML 
Formula: Y ~ SG + VP + V10 + EP + (1 | No) 
   Data: Petrol 
   AIC   BIC logLik deviance REMLdev
 166.4 176.6 -76.19    136.1   152.4
Random effects:
 Groups   Name        Variance Std.Dev.
 No       (Intercept) 2.0873   1.4447  
 Residual             3.5052   1.8722  
Number of obs: 32, groups: No, 10
```

```
Fixed effects:
             Estimate Std. Error t value
(Intercept) -6.134828  14.553350  -0.422
SG           0.219398   0.146929   1.493
VP           0.545864   0.520499   1.049
V10         -0.154242   0.039960  -3.860
EP           0.157177   0.005588  28.128

Correlation of Fixed Effects:
    (Intr) SG     VP     V10
SG  -0.694
VP  -0.715  0.067
V10 -0.934  0.433  0.836
EP   0.011  0.023 -0.116 -0.197

> vcov(Prater.lmer)
5 x 5 Matrix of class "dpoMatrix"
              [,1]          [,2]          [,3]          [,4]          [,5]
[1,] 211.800009268 -1.483965e+00 -5.4142438849 -5.431864e-01  9.014619e-04
[2,]  -1.483965337  2.158823e-02  0.0050886162  2.542658e-03  1.883112e-05
[3,]  -5.414243885  5.088616e-03  0.2709190399  1.738177e-02 -3.374120e-04
[4,]  -0.543186368  2.542658e-03  0.0173817691  1.596823e-03 -4.393911e-05
[5,]   0.000901462  1.883112e-05 -0.0003374120 -4.393911e-05  3.122559e-05


> fixef(Prater.lmer)
(Intercept)          SG          VP         V10          EP
 -6.1348280   0.2193985   0.5458639  -0.1542424   0.1571766

> ranef(Prater.lmer)
$No
  (Intercept)
A -0.05943596
B -0.21857463
C  1.92034362
D -1.92767172
E -0.21650156
F  0.56933023
G  0.06701389
H  0.19194964
I -0.40278262
J  0.07632910

> fitted(Prater.lmer)
 [1] 12.5607572 23.5631210 34.5654848 44.3104356  6.3821756 15.0268900
 [7] 26.6579603  7.7420645 17.1726621 27.8606726  6.6953467 16.9118274
[13] 27.1283081 34.2012563 17.7338014 27.1643990 31.5653445  9.9025740
[19] 24.5200002 31.1214185 -0.2703204  6.0167446 19.0624045 28.1786488
[25]  5.6041562 18.1782863 30.5952397 13.2012427 24.6751364  9.6564099
[31] 11.8568826 19.5585373
```