This question concerns the analysis of some data like those collected in a study (by forestry/wood products researchers Mantuana and Li) of factors affecting "void fraction" of foamed high density polyethylene/wood-flour composites. (The original data were not provided in the published paper, but I have simulated data from a model the researchers fit.)

Suppose that an experiment is done in 3 stages (at 3 different times), varying two formulation variables (the chemical foaming agent (CFA) concentration and the wood flour moisture content) and two processing conditions (the extruder's die temperature and screw speed rate) run-to-run. Within a stage, order of the manufacture and testing is randomized (and so the order of data in the `R` data frame listed below is not the time order of data collection).

30 values of the response variable

$$y = \text{Void} = \text{percent of sample void}$$

and corresponding coded formulation and processing variables

$$x_1 = \text{CFA} = \frac{\text{raw CFA (\%)} - .5}{.25}, \quad x_2 = \text{Moisture} = \frac{\text{raw moisture (\%)} - 6}{3},$$

$$x_3 = \text{Temp} = \frac{\text{raw die temperature (°C)} - 190}{10}, \text{ and } \quad x_4 = \text{Speed} = \frac{\text{raw screw speed (rpm)} - 80}{20}$$

are as below (the listing continues on page 2).

```
> HDPE.Wood
```

	Void	Stage	CFA	Moisture	Temp	Speed	CFASq	MoistureSq	TempSq	SpeedSq
1	16.51	1	-1	-1	-1	1	1	1	1	1
2	16.72	1	1	-1	-1	-1	1	1	1	1
3	15.61	1	-1	1	-1	-1	1	1	1	1
4	16.68	1	1	1	-1	1	1	1	1	1
5	14.08	1	-1	-1	1	-1	1	1	1	1
6	12.17	1	1	-1	1	1	1	1	1	1
7	13.43	1	-1	1	1	1	1	1	1	1
8	16.15	1	1	1	1	-1	1	1	1	1
9	13.06	1	0	0	0	0	0	0	0	0
10	14.09	1	0	0	0	0	0	0	0	0
11	13.37	2	-1	-1	-1	-1	1	1	1	1
12	13.63	2	1	-1	-1	1	1	1	1	1
13	14.28	2	-1	1	-1	1	1	1	1	1
14	15.37	2	1	1	-1	-1	1	1	1	1
15	13.36	2	-1	-1	1	1	1	1	1	1
16	10.64	2	1	-1	1	-1	1	1	1	1
17	12.81	2	-1	1	1	-1	1	1	1	1
18	13.91	2	1	1	1	1	1	1	1	1
19	13.67	2	0	0	0	0	0	0	0	0
20	13.98	2	0	0	0	0	0	0	0	0

21	13.52	3	-2	0	0	0	4	0	0	0
22	15.68	3	2	0	0	0	4	0	0	0
23	10.61	3	0	-2	0	0	0	4	0	0
24	13.96	3	0	2	0	0	0	4	0	0
25	13.17	3	0	0	-2	0	0	0	4	0
26	13.84	3	0	0	2	0	0	0	4	0
27	17.23	3	0	0	0	-2	0	0	0	4
28	13.71	3	0	0	0	2	0	0	0	4
29	15.91	3	0	0	0	0	0	0	0	0
30	13.09	3	0	0	0	0	0	0	0	0

1. Notice first that data points #9,#10,#19,#20,#29 and #30 are all from the same conditions (x_1, x_2, x_3, x_4). Suppose that observations are modeled as realizations of independent normal variables with means depending upon (x_1, x_2, x_3, x_4) and experimental stage, and with standard deviation σ. Make 95% confidence limits for σ (based on all 6 of these data points) under two sets of assumptions:

- **a)** Mean responses for a given set of conditions (x_1, x_2, x_3, x_4) do not change experimental stage to experimental stage.
- **b)** Mean responses for a given set of conditions (x_1, x_2, x_3, x_4) are potentially different stage-to-stage.

There is an R printout at the end of this question. Use it as appropriate in answering the following questions.

If one completely ignores the fact that the data are collected in stages, a standard (Gauss-Markov normal) regression model linear in the formulation and processing variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \varepsilon \qquad (*)$$

or a full quadratic regression model

$$\begin{aligned} y = {} & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1^2 + \beta_6 x_2^2 + \beta_7 x_3^2 + \beta_8 x_4^2 \\ & + \beta_9 x_1 x_2 + \beta_{10} x_1 x_3 + \beta_{11} x_1 x_4 + \beta_{12} x_2 x_3 + \beta_{13} x_2 x_4 + \beta_{14} x_3 x_4 + \varepsilon \end{aligned} \qquad (**)$$

or models "in between" these two are potentially useful as descriptions of percent void.

2. Under model (*) find 95% prediction limits for the difference in two future responses with (x_1, x_2, x_3, x_4) respectively $(1,1,1,1)$ and $(-1,1,1,-1)$.

3. Consider model (*) and the responses from the first two stages of the study. What is the distribution of

$$\bar{y}_{\text{stage 2}} - \bar{y}_{\text{stage 1}}$$

under this model? Does the observed value of this difference in the stage 1 and stage 2 sample mean responses provide clear indication that model (*) is inadequate? (Compute an appropriate p-value .)

4. Making reference to the values of appropriate F statistics (calculate them and give their degrees of freedom) say whether model (*), model (**), or a particular model "in between" these two seems most appropriate for describing percent void. (Use the `lm1`,`lm2`, and `lm3` runs.)

A model that incorporates some "curvature in response" as a function of the variables (x_1, x_2, x_3, x_4) *and* fixed "stage effects" is the instance of the linear model represented by

$$y_i = \tau_{j(i)} + \beta_0 + \beta_1 x_{2i} + \beta_2 x_{3i} + \beta_3 x_{4i} + \beta_4 x_{2i}^2 + \beta_5 x_{4i}^2 + \beta_6 x_{1i} x_{2i} + \varepsilon_i \qquad (***)$$

where $\tau_1, \tau_2,$ and τ_3 are unknown fixed constants and

$$j(i) = \begin{cases} 1 & \text{for } i = 1, \ldots, 10 \\ 2 & \text{for } i = 11, \ldots, 20 \\ 3 & \text{for } i = 21, \ldots, 30 \end{cases}$$

5. Model (***) is not a full rank linear model. Explain carefully how you know this is so. The linear combinations of parameters $\tau_2 - \tau_1$ and $\tau_3 - \tau_1$ are estimable. Show that this is true by producing linear combinations of observations with means first $\tau_2 - \tau_1$ and then $\tau_3 - \tau_1$.

Now consider a version of model (***) where $\tau_1, \tau_2,$ and τ_3 are iid $\text{N}\left(0, \sigma_\tau^2\right)$ and independent of the iid $\text{N}\left(0, \sigma^2\right)$ variables $\varepsilon_1, \ldots, \varepsilon_{30}$.

6. Give matrix and vector elements of the usual mixed linear model represented by equation (***). (That is, identify $\mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{u},$ and $\boldsymbol{\varepsilon}$. Consider the data listed in the order of the data frame **`HDPE.Wood`** listed on pages 1 and 2. You may name the columns of the data frame and use those names instead of writing out columns of numbers.) Under this model, what is the covariance matrix for $\mathbf{Y} = \left(y_1, \ldots, y_{30}\right)'$?

7. What are 95% confidence limits for the standard deviation σ? What are "95% limits" for σ_τ?

8. How do approximate BLUPs for $\tau_2 - \tau_1$ and $\tau_3 - \tau_1$ under this model compare to OLS estimators of these differences under model (***) where the τ's are unknown constants?

9. What are approximate 95% prediction limits for a future value of y for the set of conditions $(x_1, x_2, x_3, x_4) = (0,0,0,0)$ in this last model? (This will involve a 4th "stage" of data gathering.) Give numerical values for these limits.

R Printout

```
> lm1<-lm(Void~CFA+Moisture+Temp+Speed)

> summary(lm1)

Call:
lm(formula = Void ~ CFA + Moisture + Temp + Speed)

Residuals:
    Min      1Q  Median      3Q     Max
-2.8855 -0.8013 -0.1492  1.0197  2.9578

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  14.1413     0.2750  51.424   <2e-16 ***
CFA           0.2558     0.3075   0.832   0.4132
Moisture      0.6025     0.3075   1.960   0.0613 .
Temp         -0.5950     0.3075  -1.935   0.0643 .
Speed        -0.3258     0.3075  -1.060   0.2994
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.506 on 25 degrees of freedom
Multiple R-squared: 0.2733,     Adjusted R-squared: 0.157
F-statistic:  2.35 on 4 and 25 DF,  p-value: 0.08169

> vcov(lm1)
             (Intercept)           CFA      Moisture          Temp         Speed
(Intercept)  7.562208e-02  5.538600e-19  5.538600e-19 5.538600e-19 -5.538600e-19
CFA          5.538600e-19  9.452761e-02  7.426628e-20 7.426628e-20 -1.094120e-18
Moisture     5.538600e-19  7.426628e-20  9.452761e-02 1.283447e-18 -5.230020e-19
Temp         5.538600e-19  7.426628e-20  1.283447e-18 9.452761e-02  2.508317e-18
Speed       -5.538600e-19 -1.094120e-18 -5.230020e-19 2.508317e-18  9.452761e-02
```

```
> lm2<-
lm(Void~CFA+Moisture+Temp+Speed+CFASq+MoistureSq+TempSq+SpeedSq+CFA:Moisture+CFA:Temp
   +CFA:Speed+Moisture:Temp+Moisture:Speed+Temp:Speed)

> summary(lm2)

Call:
lm(formula = Void ~ CFA + Moisture + Temp + Speed + CFASq + MoistureSq +
    TempSq + SpeedSq + CFA:Moisture + CFA:Temp + CFA:Speed +
    Moisture:Temp + Moisture:Speed + Temp:Speed)

Residuals:
     Min       1Q   Median       3Q      Max
-1.96792 -0.85458  0.01188  0.74125  2.05458

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    13.96667    0.62542  22.332 6.35e-13 ***
CFA             0.25583    0.31271   0.818   0.4261
Moisture        0.60250    0.31271   1.927   0.0732 .
Temp           -0.59500    0.31271  -1.903   0.0764 .
Speed          -0.32583    0.31271  -1.042   0.3139
CFASq           0.21333    0.29251   0.729   0.4770
MoistureSq     -0.36542    0.29251  -1.249   0.2307
TempSq         -0.06042    0.29251  -0.207   0.8391
SpeedSq         0.43083    0.29251   1.473   0.1615
CFA:Moisture    0.63375    0.38299   1.655   0.1187
CFA:Temp       -0.21500    0.38299  -0.561   0.5828
CFA:Speed      -0.26250    0.38299  -0.685   0.5035
Moisture:Temp   0.27125    0.38299   0.708   0.4897
Moisture:Speed -0.15625    0.38299  -0.408   0.6891
Temp:Speed     -0.05250    0.38299  -0.137   0.8928
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.532 on 15 degrees of freedom
Multiple R-squared: 0.5489,     Adjusted R-squared: 0.1279
F-statistic: 1.304 on 14 and 15 DF,  p-value: 0.3077

> lm3<-lm(Void~CFA+Moisture+Temp+Speed+MoistureSq+SpeedSq+CFA:Moisture)

> summary(lm3)

Call:
lm(formula = Void ~ CFA + Moisture + Temp + Speed + MoistureSq +
    SpeedSq + CFA:Moisture)

Residuals:
     Min       1Q   Median       3Q      Max
-2.25719 -0.80115  0.06078  0.77823  2.63281

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.1196     0.3899  36.212   <2e-16 ***
CFA            0.2558     0.2757   0.928   0.3635
Moisture       0.6025     0.2757   2.185   0.0398 *
Temp          -0.5950     0.2757  -2.158   0.0421 *
Speed         -0.3258     0.2757  -1.182   0.2499
MoistureSq    -0.3845     0.2533  -1.518   0.1432
SpeedSq        0.4117     0.2533   1.626   0.1183
CFA:Moisture   0.6337     0.3377   1.877   0.0739 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.351 on 22 degrees of freedom
Multiple R-squared: 0.4857,     Adjusted R-squared: 0.3221
F-statistic: 2.968 on 7 and 22 DF,  p-value: 0.02378
```

```
> lm4<-lm(Void~Stage+Moisture+Temp+Speed+MoistureSq+SpeedSq+CFA:Moisture)

> summary(lm4)

Call:
lm(formula = Void ~ Stage + Moisture + Temp + Speed + MoistureSq +
    SpeedSq + CFA:Moisture)

Residuals:
     Min       1Q   Median       3Q      Max
-2.07025 -0.65722 -0.02394  0.69027  2.17998

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   14.8282     0.4822  30.752   <2e-16 ***
Stage2        -1.3480     0.5568  -2.421   0.0246 *
Stage3        -0.7780     0.5568  -1.397   0.1769
Moisture       0.6025     0.2541   2.371   0.0274 *
Temp          -0.5950     0.2541  -2.341   0.0292 *
Speed         -0.3258     0.2541  -1.282   0.2138
MoistureSq    -0.3845     0.2334  -1.647   0.1144
SpeedSq        0.4117     0.2334   1.764   0.0923 .
Moisture:CFA   0.6337     0.3112   2.036   0.0545 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.245 on 21 degrees of freedom
Multiple R-squared: 0.5829,     Adjusted R-squared: 0.424
F-statistic: 3.669 on 8 and 21 DF,  p-value: 0.007967


> vcov(lm4)
               (Intercept)        Stage2        Stage3      Moisture
(Intercept)   2.324989e-01 -1.549992e-01 -1.549992e-01  1.194207e-17
Stage2       -1.549992e-01  3.099985e-01  1.549992e-01 -6.843232e-18
Stage3       -1.549992e-01  1.549992e-01  3.099985e-01 -1.086608e-17
Moisture      1.194207e-17 -6.843232e-18 -1.086608e-17  6.458302e-02
Temp          4.724587e-18 -4.911919e-18 -7.003458e-18 -2.687082e-19
Speed        -8.925445e-18  3.916070e-18  7.268064e-18 -1.222050e-18
MoistureSq   -4.843727e-02 -2.287745e-18 -4.635872e-18 -4.705631e-18
SpeedSq      -4.843727e-02 -2.287745e-18 -4.635872e-18 -2.370068e-18
Moisture:CFA -9.256550e-18  5.874105e-18  1.090210e-17 -1.833076e-18
                     Temp         Speed    MoistureSq       SpeedSq
(Intercept)   4.724587e-18 -8.925445e-18 -4.843727e-02 -4.843727e-02
Stage2       -4.911919e-18  3.916070e-18 -2.287745e-18 -2.287745e-18
Stage3       -7.003458e-18  7.268064e-18 -4.635872e-18 -4.635872e-18
Moisture     -2.687082e-19 -1.222050e-18 -4.705631e-18 -2.370068e-18
Temp          6.458302e-02 -2.578455e-18 -1.828051e-19 -2.851786e-19
Speed        -2.578455e-18  6.458302e-02  2.528510e-18  3.495230e-18
MoistureSq   -1.828051e-19  2.528510e-18  5.449192e-02  6.054658e-03
SpeedSq      -2.851786e-19  3.495230e-18  6.054658e-03  5.449192e-02
Moisture:CFA -4.725262e-18 -2.422661e-19  3.434388e-18  4.367016e-19
              Moisture:CFA
(Intercept)  -9.256550e-18
Stage2        5.874105e-18
Stage3        1.090210e-17
Moisture     -1.833076e-18
Temp         -4.725262e-18
Speed        -2.422661e-19
MoistureSq    3.434388e-18
SpeedSq       4.367016e-19
Moisture:CFA  9.687453e-02
```

```
> mm<-lmer(Void ~ (1 | Stage)+Moisture+Temp+Speed+MoistureSq+SpeedSq+CFA:Moisture)

> summary(mm)

Linear mixed-effects model fit by REML
Formula: Void ~ (1 | Stage) + Moisture + Temp + Speed + MoistureSq + SpeedSq +
CFA:Moisture
   AIC   BIC logLik MLdeviance REMLdeviance
 115.9 127.1 -49.97      93.56        99.93

Random effects:
 Groups   Name        Variance Std.Dev.
 Stage    (Intercept) 0.30362  0.55102
 Residual             1.54976  1.24489
number of obs: 30, groups: Stage, 3

Fixed effects:
             Estimate Std. Error t value
(Intercept)   14.1196     0.4800  29.419
Moisture       0.6025     0.2541   2.371
Temp          -0.5950     0.2541  -2.341
Speed         -0.3258     0.2541  -1.282
MoistureSq    -0.3845     0.2334  -1.647
SpeedSq        0.4117     0.2334   1.764
Moisture:CFA   0.6337     0.3112   2.036

Correlation of Fixed Effects:
            (Intr) Moistr Temp   Speed  MstrSq SpedSq
Moisture     0.000
Temp         0.000  0.000
Speed        0.000  0.000  0.000
MoistureSq  -0.432  0.000  0.000  0.000
SpeedSq     -0.432  0.000  0.000  0.000  0.111
Moistur:CFA  0.000  0.000  0.000  0.000  0.000  0.000

> fitted(mm)
 [1] 14.91637 14.30054 15.50554 16.12137 14.37804 12.45887 13.66387 15.58304
 [9] 14.58877 14.58877 14.67557 12.75641 13.96141 15.88057 12.83391 12.21807
[17] 13.42307 14.03891 13.69630 13.69630 14.07368 14.07368 11.33056 13.74056
[25] 15.26368 12.88368 16.37222 15.06889 14.07368 14.07368

> ranef(mm)
An object of class "ranef.lmer"
[[1]]
  (Intercept)
1  0.46918435
2 -0.42328108
3 -0.04590327

> sim<-mcmcsamp(mm , 50000)

> HPDinterval(sim)
                       lower        upper
(Intercept)      13.73891374 15.493759518
Moisture          0.02169694  1.193919318
Temp             -1.19303720  0.002935620
Speed            -0.91646714  0.262884226
MoistureSq       -0.99402809  0.098707121
TempSq           -0.67307857  0.410095409
Moisture:CFA     -0.12211403  1.332907297
log(sigma^2)      0.14419350  1.326701122
log(Stag.(In)) -148.83380165  1.844520361
attr(,"Probability")
[1] 0.95
```