This question concerns several analyses of a small set of data on the operation of a Butane Hydrogenolysis Reactor. The response variable

$$y = \text{percent conversion of butane}$$

is to be understood as a function of the chemical reactor process variables

flow = a total feed flow (cc/sec at STP)

ratio = feed ratio (Hydrogen/Butane)

temp = the reactor wall temperature (°F)

The data are shown in the following table.

Run, *i*	*y*	*flow*	*ratio*	*temp*	*day*	Setup, *s(i)*
1	82	115	6	495	1	1
2	91	50	4	470	1	2
3	75	180	8	520	1	3
4	98	50	4	520	2	4
5	39	180	8	470	2	5
6	77	115	6	495	2	1
7	95	50	8	520	3	6
8	61	180	4	470	3	7
9	81	115	6	495	3	1
10	76	50	8	470	4	8
11	92	180	4	520	4	9
12	82	115	6	495	4	1

Twelve runs were made over a period of four days, three runs being made each day. Nine process setups (corresponding to combinations of levels of the *flow*, *ratio*, and *temp* factors) were used. Note that one setup was repeated on each of the four days.

I. Begin by considering analysis based on a cell means model

$$y_i = \mu_{s(i)} + \varepsilon_i \text{ for } i = 1,2,\ldots,12 \qquad (*)$$

where $\mu_1, \mu_2, \ldots, \mu_9$ are unknown constants (the 9 mean responses for the different setups of the process), the ε_i are iid $\mathrm{N}\left(0,\sigma^2\right)$, and we use the notation

$s(i)$ = the setup number employed in the ith run of the process

(For example, when $i = 8$ for the 8th run, $s(8) = 7$ to indicate that setup 7 was used.)

(a) Write this linear model out in matrix form. (What are $\mathbf{X}$ and $\boldsymbol{\beta}$ here? Use the ordering of the elements of $\mathbf{Y}$ employed in the table.)

(b) Give 90% confidence limits for σ in this model. (Display a formula, insert numerical values, but don't complete the calculations.)

(c) Setup #1 is a "center point" for the set of $(flow, ratio, temp)$ combinations in the data set. The other 8 setups form a $2 \times 2 \times 2$ factorial structure. In this 3 way factorial structure, the quantity

$$\frac{1}{4}(\mu_3 + \mu_5 + \mu_7 + \mu_9) - \frac{1}{4}(\mu_2 + \mu_4 + \mu_6 + \mu_8)$$

is potentially of interest. In the jargon of factorial analysis, what is this quantity? Give 95% confidence limits for it based on the cell means model (*).

(d) Give 95% prediction limits for an additional observation under process setup #2 under this model.

There is an R printout attached to this question that you should consult for the rest of this question. When you use something off this printout in making an answer, be very explicit about where you got it (give page numbers and use the names employed on the printout).

II. A second analysis of these data can be made on the basis of a regression model

$$y_i = \beta_0 + \beta_1 flow_i + \beta_2 ratio_i + \beta_3 temp_i + \varepsilon_i \qquad (**)$$

(a) The model (**) is a "reduced model" version of model (*). Give the value and degrees of freedom for an F statistic that can be used to test the hypothesis that the regression model (**) is adequate.

(b) Under the regression model (**) the linear combination of 8 setup means considered in part **I.(c)** can be expressed in terms of the parameters β of the regression model. Do so, and make 95% confidence limits for this quantity (under model (**)).

III. To this point we have ignored the fact that these data were collected on 4 different days. It is probably sensible to think of "day" effects as random. So consider an analysis of the data based on a model equation

$$y_i = \beta_0 + \delta_{k(i)} + \beta_1 flow_i + \beta_2 ratio_i + \beta_3 temp_i + \varepsilon_i \qquad (***)$$

where

$k(i)$ = the day number on which the ith run of the process was made

under the interpretation that $\delta_1, \delta_2, \delta_3, \delta_4$ are iid $\mathrm{N}(0, \sigma_\delta^2)$, independent of the ε_i. This is a mixed linear model.

(a) What are $\mathbf{Z}$ and $\mathbf{u}$ here in the usual matrix formulation of the mixed model $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$? (Use the ordering of the elements of $\mathbf{Y}$ employed on the printout and in the table on page 1.)

(b) Give an approximate 95% confidence interval for σ_δ in the mixed model (***).

(c) What is an approximate BLUE of $\beta_0 + 50\beta_1 + 4\beta_2 + 470\beta_3$? (Give a numerical value.) Then write down a standard error for your estimate. (Display a formula, insert numerical values, but don't complete the calculations.)

(d) Consider the 4 observations from setup #1. In the mixed model (***), these all have the same mean, namely $\beta_0 + 115\beta_1 + 6\beta_2 + 495\beta_3$, and differ only in that each one has a different (δ, ε) pair added to the mean. Consider the sample variance of these 4 observations. What function of the mixed model parameters does this estimate? Find an estimate of this function of the mixed model parameters based on the REML estimates. How does this REML-based estimate compare to the sample variance? Comments?

```
> day<-c(1,1,1,2,2,2,3,3,3,4,4,4)
> flow<-c(115,50,180,50,180,115,50,180,115,50,180,115)
> temp<-c(495,470,520,520,470,495,520,470,495,470,520,495)
> ratio<-c(6,4,8,4,8,6,8,4,6,8,4,6)
> y<-c(82,91,75,98,39,77,95,61,81,76,92,82)
> Day<-as.factor(day)
> Flow<-as.factor(flow)
> Temp<-as.factor(temp)
> Ratio<-as.factor(ratio)
> options(contrasts=c("contr.sum","contr.sum"))
> setup<-c(1,2,3,4,5,1,6,7,1,8,9,1)
> D<-data.frame(y,flow,ratio,temp,day,setup)
> D<-data.frame(y,flow,ratio,temp,day,setup)
> D
    y flow ratio temp day setup
1  82  115     6  495   1     1
2  91   50     4  470   1     2
3  75  180     8  520   1     3
4  98   50     4  520   2     4
5  39  180     8  470   2     5
6  77  115     6  495   2     1
7  95   50     8  520   3     6
8  61  180     4  470   3     7
9  81  115     6  495   3     1
10 76   50     8  470   4     8
11 92  180     4  520   4     9
12 82  115     6  495   4     1

> lmout1<-lm(y~flow+ratio+temp)
> summary(lmout1)

Call:
lm(formula = y ~ flow + ratio + temp)

Residuals:
    Min      1Q  Median      3Q     Max
-11.458  -1.990   2.417   3.292   5.792

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -109.14936   46.13914  -2.366 0.045554 *
flow          -0.17885    0.03528  -5.069 0.000966 ***
ratio         -3.56250    1.14657  -3.107 0.014509 *
temp           0.46500    0.09173   5.069 0.000966 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1

Residual standard error: 6.486 on 8 degrees of freedom
Multiple R-Squared: 0.8841,     Adjusted R-squared: 0.8407
F-statistic: 20.35 on 3 and 8 DF,  p-value: 0.0004223
```

```
> anova(lmout1)
Analysis of Variance Table

Response: y
          Df  Sum Sq Mean Sq F value    Pr(>F)
flow       1 1081.12 1081.12 25.6996 0.0009656 ***
ratio      1  406.12  406.12  9.6541 0.0145090 *
temp       1 1081.13 1081.13 25.6996 0.0009656 ***
Residuals  8  336.54   42.07
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
> vcov(lmout1)
             (Intercept)         flow         ratio          temp
(Intercept) 2128.8197855 -0.143129777 -7.887695e+00 -4.164703e+00
flow          -0.1431298  0.001244607  0.000000e+00  0.000000e+00
ratio         -7.8876953  0.000000000  1.314616e+00  1.585205e-17
temp          -4.1647031  0.000000000  1.585205e-17  8.413542e-03
> predict(lmout1)
        1         2         3         4         5         6         7         8
 79.08333  86.20833  71.95833 109.45833  48.70833  79.08333  95.20833  62.95833
        9        10        11        12
 79.08333  71.95833  86.20833  79.08333


> mlmout<-lme(y~1+flow+ratio+temp,random=~1|Day)
> summary(mlmout)
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  85.84857 86.32522 -36.92429

Random effects:
 Formula: ~1 | Day
        (Intercept) Residual
StdDev:    5.107096 3.567211

Fixed effects: y ~ 1 + flow + ratio + temp
                 Value Std.Error DF   t-value p-value
(Intercept) -109.14936 25.504194  5 -4.279663  0.0079
flow          -0.17885  0.019403  5 -9.217415  0.0003
ratio         -3.56250  0.630600  5 -5.649384  0.0024
temp           0.46500  0.050448  5  9.217415  0.0003
 Correlation:
      (Intr) flow   ratio
flow  -0.087
ratio -0.148  0.000
temp  -0.979  0.000  0.000

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.34345445 -0.28756470 -0.02482258  0.49878103  1.28464913

Number of Observations: 12
Number of Groups: 4
```

```
> intervals(mlmout)
Approximate 95% confidence intervals

 Fixed effects:
                   lower         est.       upper
(Intercept) -174.7099738 -109.1493590 -43.5887442
flow          -0.2287233   -0.1788462  -0.1289690
ratio         -5.1835082   -3.5625000  -1.9414918
temp           0.3353193    0.4650000   0.5946807
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: Day
                   lower     est.    upper
sd((Intercept)) 2.003669 5.107096 13.01734

 Within-group standard error:
   lower     est.    upper
1.919401 3.567211 6.629670
> vcov(mlmout)
            (Intercept)          flow         ratio          temp
(Intercept) 650.4639276 -4.329510e-02 -2.385936e+00 -1.259774e+00
flow         -0.0432951  3.764791e-04  5.283019e-18  3.688508e-18
ratio        -2.3859364  5.283019e-18  3.976561e-01 -4.155719e-17
temp         -1.2597744  3.688508e-18 -4.155719e-17  2.544999e-03
> predict(mlmout,level=0:1)
   Day predict.fixed predict.Day
1    1      79.08333    82.16544
2    1      86.20833    89.29044
3    1      71.95833    75.04044
4    2     109.45833   102.79239
5    2      48.70833    42.04239
6    2      79.08333    72.41739
7    3      95.20833    95.13666
8    3      62.95833    62.88666
9    3      79.08333    79.01166
10   4      71.95833    75.61385
11   4      86.20833    89.86385
12   4      79.08333    82.73885
```