This question concerns the analysis of some data taken from *Methods and Applications of Linear Models* by Hocking on viscosity testing of a liquid lubricant. The testing procedure was as follows. One of 4 small vessels was mounted on a test rig and filled with the lubricant. Two viscosity measurements were made, after which the vessel was removed from the rig, emptied and cleaned before another (possibly the same, possibly different) vessel was mounted and more data collected. All samples of the lubricant were drawn from a single reservoir. Each vessel was mounted on the rig and used in viscosity testing 4 different times. Hocking's data and some summary statistics are in the table below.

Vessel (i)	Trial (j)	Set-Up	Measured Viscosity	Sample Mean	Sample Variance
1	1	1	4.92, 4.98	4.950	.0018
1	2	2	5.40, 5.36	5.380	.0008
1	3	3	5.33, 5.17	5.250	.0128
1	4	4	5.18, 5.25	5.215	.0025
2	1	5	5.01, 5.07	5.040	.0018
2	2	6	5.31, 5.55	5.430	.0288
2	3	7	5.45, 5.68	5.565	.0265
2	4	8	5.00, 4.92	4.960	.0032
3	1	9	4.87, 4.96	4.915	.0041
3	2	10	4.95, 4.81	4.880	.0098
3	3	11	4.96, 4.38	4.670	.1682
3	4	12	4.61, 4.27	4.440	.0587
4	1	13	4.99, 4.39	4.690	.1800
4	2	14	4.52, 4.59	4.555	.0025
4	3	15	4.63, 4.94	4.785	.0481
4	4	16	4.60, 4.71	4.655	.0061

The 16 "Set-Ups" above were *not* run in the order "1" through "16," but rather in a (completely) randomly chosen order (so that, for example, not all Vessel 1 tests were made before all Vessel 2 tests).

We will use the notation

y_{ijk} = the kth measured viscosity for vessel i when mounted for the jth trial on the test rig

and in this question consider several different models and analyses for the y_{ijk}.

To begin, consider only the test results for vessel #1 and the model

$$y_{1jk} = \mu_1 + \tau_{1j} + \varepsilon_{1jk} \qquad (*)$$

where μ_1 is an unknown constant, $\tau_{11}, \tau_{12}, \tau_{13}, \tau_{14}$ are iid $\mathrm{N}\left(0, \sigma_\tau^2\right)$ random variables independent of $\varepsilon_{111}, \varepsilon_{112}, \varepsilon_{121}, \varepsilon_{122}, \varepsilon_{131}, \varepsilon_{132}, \varepsilon_{141}, \varepsilon_{142}$ that are iid $\mathrm{N}\left(0, \sigma^2\right)$.

a) Write the model (*) in the standard matrix form $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon}$. (Write out all of the matrices and vectors indicated here.)

b) What is the distribution of $\bar{y}_{1j.} = \frac{1}{2}\left(y_{1j1} + y_{1j2}\right)$ according to model (*)? Notice that $\bar{y}_{11.}, \bar{y}_{12.}, \bar{y}_{13.},$ and $\bar{y}_{14.}$ are iid. Let $\bar{y}_{1..} = \frac{1}{4}\left(\bar{y}_{11.} + \bar{y}_{12.} + \bar{y}_{13.} + \bar{y}_{14.}\right)$ and $s_1^2 = \frac{1}{4-1}\sum_{j=1}^{4}\left(\bar{y}_{1j.} - \bar{y}_{1..}\right)^2$. Give (numerical) 95% confidence limits for μ_1 based on $\bar{y}_{1..}$ and s_1^2.

c) According to the model (*), what is $\mathrm{E}\, s_1^2$? According to the same model, what is the expected value of a pooled sample variance from the 4 "samples" of 2 viscosity determinations each for vessel #1? Use the realized values of s_1^2 and this pooled sample variance to make a point estimate of σ_τ^2.

Now consider the test results for all 4 vessels and the model

$$y_{ijk} = \mu_i + \tau_{ij} + \varepsilon_{ijk} \qquad (**)$$

under two different sets of assumptions on $\mu_1, \mu_2, \mu_3, \mu_4$, namely

$$\mu_1, \mu_2, \mu_3, \mu_4 \text{ are unknown constants} \qquad (**F)$$

and

$$\mu_1, \mu_2, \mu_3, \mu_4 \text{ are iid } \mathrm{N}\left(\mu, \sigma_\mu^2\right) \text{ for unknown constants } \mu \text{ and } \sigma_\mu^2 \qquad (**R)$$

(In these models, we assume that the 16 variables τ_{ij} are iid $\mathrm{N}\left(0, \sigma_\tau^2\right)$ independent of the 32 iid $\mathrm{N}\left(0, \sigma^2\right)$ variables ε_{ijk}.)

d) In the lubricant viscosity testing context, what are circumstances in which the first of these assumptions (**F) might be appropriate? What are circumstances in which the second of these assumptions (**R) might be appropriate?

Consider analysis under the assumptions (**F). The most straightforward means of analysis under these assumptions is using R's `lme` function, and the **first printout** attached to this question takes this approach. Another possibility is to note that according to (**)

$$\bar{y}_{ij.} = \mu_i + \left(\tau_{ij} + \bar{\varepsilon}_{ij.}\right) = \mu_i + \varepsilon_{ij}^*$$

where $\bar{y}_{ij}$ and $\bar{\varepsilon}_{ij}$ are the obvious averages, and the $\varepsilon_{ij}^* = \left(\tau_{ij} + \bar{\varepsilon}_{ij.}\right)$ are iid $\mathrm{N}\left(0, \gamma^2\right)$ for $\gamma^2 = \sigma_\tau^2 + \frac{1}{2}\sigma^2$. So as long as the fixed effects are the ones of primary interest, one could use the $\bar{y}_{ij.}$ as responses in an ordinary linear model analysis using the R function `lm`. The **second printout** attached to this question takes this approach.

e) Give two sets of 95% confidence limits for $\mu_1 - \mu_2$ based on the two R printouts referred to above.

f) Give approximate (large sample) 95% prediction limits for a single viscosity measurement if vessel #3 is mounted on the test rig a 5th time and testing is done.

g) Under the assumptions (**F), what appears to be the biggest source of variation in measured viscosity for a given vessel,

i) variation among results for a given mounting of the vessel on the test rig, or

ii) variation between results for different mountings of the vessel on the test rig?

How sure are you about this comparison? Explain.

Now consider analysis under the assumptions (**R). The **third R printout** attached to this question enables such an analysis.

h) What are approximate 95% confidence limits for σ_μ?

i) Give approximate 95% prediction limits for a single measured viscosity made using a vessel not represented in the data on page 1 (say y_{511}) assuming that the model (**R) extends to this new observation.

j) Give a point prediction of $\mu_1 - \mu_2$ that is appropriate under assumptions (**R).

Printout 1

```
> RawData
   vessel trial setup viscosity
1       1     1     1      4.92
2       1     1     1      4.98
3       1     2     2      5.40
4       1     2     2      5.36
5       1     3     3      5.33
6       1     3     3      5.17
7       1     4     4      5.18
8       1     4     4      5.25
9       2     1     5      5.01
10      2     1     5      5.07
11      2     2     6      5.31
12      2     2     6      5.55
13      2     3     7      5.45
14      2     3     7      5.68
15      2     4     8      5.00
16      2     4     8      4.92
17      3     1     9      4.87
18      3     1     9      4.96
19      3     2    10      4.95
20      3     2    10      4.81
21      3     3    11      4.96
22      3     3    11      4.38
23      3     4    12      4.61
24      3     4    12      4.27
25      4     1    13      4.99
26      4     1    13      4.39
27      4     2    14      4.52
28      4     2    14      4.59
29      4     3    15      4.63
30      4     3    15      4.94
31      4     4    16      4.60
32      4     4    16      4.71

> Vessel<-as.factor(vessel)

> fit1<-lme(viscosity~Vessel-1,random=~1|trial)

> summary(fit1)
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  18.48831 26.48153 -3.244153

Random effects:
 Formula: ~1 | trial
        (Intercept)  Residual
StdDev:  0.09514364 0.2231988
```

```
Fixed effects: viscosity ~ Vessel - 1
          Value  Std.Error DF  t-value p-value
Vessel1 5.19875 0.09214278 25 56.42059       0
Vessel2 5.24875 0.09214278 25 56.96323       0
Vessel3 4.72625 0.09214278 25 51.29268       0
Vessel4 4.67125 0.09214278 25 50.69578       0
 Correlation:
        Vessl1 Vessl2 Vessl3
Vessel2 0.267
Vessel3 0.267  0.267
Vessel4 0.267  0.267  0.267

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-1.83333191 -0.75322890  0.08606579  0.66142949  1.65010949

Number of Observations: 32
Number of Groups: 4

> vcov(fit1)
            Vessel1     Vessel2     Vessel3     Vessel4
Vessel1 0.008490292 0.002263078 0.002263078 0.002263078
Vessel2 0.002263078 0.008490292 0.002263078 0.002263078
Vessel3 0.002263078 0.002263078 0.008490292 0.002263078
Vessel4 0.002263078 0.002263078 0.002263078 0.008490292

> intervals(fit1)
Approximate 95% confidence intervals

 Fixed effects:
           lower    est.    upper
Vessel1 5.008978 5.19875 5.388522
Vessel2 5.058978 5.24875 5.438522
Vessel3 4.536478 4.72625 4.916022
Vessel4 4.481478 4.67125 4.861022
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: trial
                    lower       est.     upper
sd((Intercept)) 0.0243249 0.09514364 0.3721419

 Within-group standard error:
    lower      est.     upper
0.1691642 0.2231988 0.2944932
```

```
> predict(fit1)
       1        1        2        2        3        3        4        4
5.161722 5.161722 5.257995 5.257995 5.261698 5.261698 5.113586 5.113586
       1        1        2        2        3        3        4        4
5.211722 5.211722 5.307995 5.307995 5.311698 5.311698 5.163586 5.163586
       1        1        2        2        3        3        4        4
4.689222 4.689222 4.785495 4.785495 4.789198 4.789198 4.641086 4.641086
       1        1        2        2        3        3        4        4
4.634222 4.634222 4.730495 4.730495 4.734198 4.734198 4.586086 4.586086
attr(,"label")
[1] "Fitted values"
```

Printout 2

```
> MeansData
   vess  mean
1     1 4.950
2     1 5.380
3     1 5.250
4     1 5.215
5     2 5.040
6     2 5.430
7     2 5.565
8     2 4.960
9     3 4.915
10    3 4.880
11    3 4.670
12    3 4.440
13    4 4.690
14    4 4.555
15    4 4.785
16    4 4.655

> Vess<-as.factor(vess)

> fit2<-lm(mean~Vess-1)

> summary(fit2)

Call:
lm(formula = mean ~ Vess - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-0.2888 -0.1394  0.0175  0.1606  0.3163

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
Vess1    5.199      0.105   49.53 3.01e-15 ***
Vess2    5.249      0.105   50.00 2.68e-15 ***
Vess3    4.726      0.105   45.03 9.39e-15 ***
Vess4    4.671      0.105   44.50 1.08e-14 ***
---
Signif. codes:  0 `***' 0.001 `**' 0.01 `*' 0.05 `.' 0.1 ` ' 1
```

```
Residual standard error: 0.2099 on 12 degrees of freedom
Multiple R-Squared: 0.9987,     Adjusted R-squared: 0.9982
F-statistic:  2240 on 4 and 12 DF,  p-value: < 2.2e-16

> vcov(fit2)
           Vess1      Vess2      Vess3      Vess4
Vess1 0.01101823 0.00000000 0.00000000 0.00000000
Vess2 0.00000000 0.01101823 0.00000000 0.00000000
Vess3 0.00000000 0.00000000 0.01101823 0.00000000
Vess4 0.00000000 0.00000000 0.00000000 0.01101823

> predict(fit2)
      1       2       3       4       5       6       7       8       9      10
5.19875 5.19875 5.19875 5.19875 5.24875 5.24875 5.24875 5.24875 4.72625 4.72625
     11      12      13      14      15      16
4.72625 4.72625 4.67125 4.67125 4.67125 4.67125
```

Printout 3

```
> fit3<-lme(viscosity~1,random=~1|vessel/trial)

> summary(fit3)
Linear mixed-effects model fit by REML
 Data: NULL
       AIC      BIC    logLik
  15.60460 21.34055 -3.802301

Random effects:
 Formula: ~1 | vessel
        (Intercept)
StdDev:   0.2859678

 Formula: ~1 | trial %in% vessel
        (Intercept)  Residual
StdDev:   0.1635383 0.1861619

Fixed effects: viscosity ~ 1
              Value Std.Error DF  t-value p-value
(Intercept) 4.96125  0.152312 16 32.57294       0

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-1.7355129 -0.4284252 -0.1021811  0.5576975  1.5783772

Number of Observations: 32
Number of Groups:
           vessel trial %in% vessel
                4                16

> vcov(fit3)
            (Intercept)
(Intercept)  0.02319895
```

```
> intervals(fit3)
Approximate 95% confidence intervals

 Fixed effects:
              lower    est.    upper
(Intercept) 4.638363 4.96125 5.284137
attr(,"label")
[1] "Fixed effects:"

 Random Effects:
  Level: vessel
                    lower      est.     upper
sd((Intercept)) 0.1151573 0.2859678 0.7101377
  Level: trial
                     lower      est.     upper
sd((Intercept)) 0.08149846 0.1635383 0.3281630

 Within-group standard error:
    lower      est.     upper
0.1316491 0.1861619 0.2632472

> predict(fit3)
     1/1      1/1      1/2      1/2      1/3      1/3      1/4      1/4
5.036714 5.036714 5.297651 5.297651 5.218763 5.218763 5.197524 5.197524
     2/1      2/1      2/2      2/2      2/3      2/3      2/4      2/4
5.108653 5.108653 5.345316 5.345316 5.427239 5.427239 5.060106 5.060106
     3/1      3/1      3/2      3/2      3/3      3/3      3/4      3/4
4.851760 4.851760 4.830521 4.830521 4.703086 4.703086 4.563515 4.563515
     4/1      4/1      4/2      4/2      4/3      4/3      4/4      4/4
4.696166 4.696166 4.614244 4.614244 4.753815 4.753815 4.674927 4.674927
attr(,"label")
[1] "Fitted values"

> predict(fit3,level=0:2)
   vessel trial predict.fixed predict.vessel predict.trial
1       1   1/1       4.96125       5.170550      5.036714
2       1   1/1       4.96125       5.170550      5.036714
3       1   1/2       4.96125       5.170550      5.297651
4       1   1/2       4.96125       5.170550      5.297651
5       1   1/3       4.96125       5.170550      5.218763
6       1   1/3       4.96125       5.170550      5.218763
7       1   1/4       4.96125       5.170550      5.197524
8       1   1/4       4.96125       5.170550      5.197524
9       2   2/1       4.96125       5.214613      5.108653
10      2   2/1       4.96125       5.214613      5.108653
11      2   2/2       4.96125       5.214613      5.345316
12      2   2/2       4.96125       5.214613      5.345316
13      2   2/3       4.96125       5.214613      5.427239
14      2   2/3       4.96125       5.214613      5.427239
15      2   2/4       4.96125       5.214613      5.060106
16      2   2/4       4.96125       5.214613      5.060106
17      3   3/1       4.96125       4.754153      4.851760
18      3   3/1       4.96125       4.754153      4.851760
19      3   3/2       4.96125       4.754153      4.830521
```

20	3	3/2	4.96125	4.754153	4.830521
21	3	3/3	4.96125	4.754153	4.703086
22	3	3/3	4.96125	4.754153	4.703086
23	3	3/4	4.96125	4.754153	4.563515
24	3	3/4	4.96125	4.754153	4.563515
25	4	4/1	4.96125	4.705684	4.696166
26	4	4/1	4.96125	4.705684	4.696166
27	4	4/2	4.96125	4.705684	4.614244
28	4	4/2	4.96125	4.705684	4.614244
29	4	4/3	4.96125	4.705684	4.753815
30	4	4/3	4.96125	4.705684	4.753815
31	4	4/4	4.96125	4.705684	4.674927
32	4	4/4	4.96125	4.705684	4.674927