STAT 8004 – Statistical Methods II
Spring 2015
Homework Assignment 7 – Solutions

1. This is Problem 3 of Faraway (2006), Chapter 8.

   The eggprod dataset concerns an experiment where six pullets were placed into each of 12 pens. Four blocks were formed from groups of three pens based on location. Three treatments were applied. The number of eggs produced was recorded.

```
> library(faraway)
> data(eggprod)
```

   (a) Fit a model for the number of eggs produced with the treatments as fixed effects and the blocks as random effects. Describe the estimated differences between the treatments.

   Solutions:

   It is clear that the treatment is the fixed effect, while the location blocks contribute as random effect. Thus a linear mixed effect model is

$$y_{ij} = \mu + \alpha_i + u_j + \epsilon_{ij}$$

   for the number of eggs corresponding to the pen in the jth block ($j = 1, 2, 3, 4$) with ith treatment ($i = 1, 2, 3$).

```
> egg.m=lmer(eggs~treat+(1|block),data=eggprod)
> summary(egg.m)
Linear mixed model fit by REML ['lmerMod']
Formula: eggs ~ treat + (1 | block)
   Data: eggprod

REML criterion at convergence: 85.4

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.71233 -0.47453 -0.02845  0.64196  1.42942

Random effects:
 Groups   Name        Variance Std.Dev.
 block    (Intercept) 129.9    11.40
 Residual             386.9    19.67
Number of obs: 12, groups:  block, 4

Fixed effects:
```

```
            Estimate Std. Error t value
(Intercept)   349.00      11.37  30.702
treatF         -6.25      13.91  -0.449
treatO        -42.50      13.91  -3.056

Correlation of Fixed Effects:
       (Intr) treatF
treatF -0.612
treatO -0.612  0.500

> model.matrix(egg.m)
   (Intercept) treatF treatO
1            1      0      1
2            1      0      1
3            1      0      1
4            1      0      1
5            1      0      0
6            1      0      0
7            1      0      0
8            1      0      0
9            1      1      0
10           1      1      0
11           1      1      0
12           1      1      0
```

From the design matrix associated with the default R option, the design matrix is

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

Therefore the difference between treatment "F" and "E " is -6.25, and the difference between treatment "O" and "E" is -42.50.

(b) Test for the significance of the treatment.

Solutions:

The overall treatment effect is significant with p-value 0.01481.

```
> egg.m.reduced=lmer(eggs~1+(1|block),data=eggprod)
> anova(egg.m,egg.m.reduced)
refitting model(s) with ML (instead of REML)
Data: eggprod
Models:
egg.m.reduced: eggs ~ 1 + (1 | block)
egg.m: eggs ~ treat + (1 | block)
              Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
egg.m.reduced  3 119.31 120.77 -56.657   113.31
egg.m          5 114.89 117.31 -52.444   104.89 8.4245      2    0.01481 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

2. This is Problem 4 of Faraway (2006), Chapter 8.

Data on the cutoff times of lawnmowers may be found in the dataset lawn. 3 machines were randomly selected from those produced by manufacturers A and B. Each machine was tested twice at low speed and high speed.

```
> library(faraway)
> data(lawn)
```

(a) Fit a mixed effects model with manufacturer and speed as main effects along with their interaction and machine nested in manufacturer as random effects. Write down the formula for the model. In the summary output for the model, you will find that fixed manufacturer effect has zero degrees of freedom. Explain why this is so (check your model formula).

Solutions:

I think the text book exercise meant to ask a random effect model

$$y_{ijkl} = \mu + \alpha_i + \beta_j + \delta_{ij} + u_i + v_{ik} + \epsilon_{ijkl}$$

for the lth ($l = 1, 2$) test of kth ($k = 1, 2, 3$) machine with jth ($j = 1, 2$) speed manufactured by the ith ($i = 1, 2$) lawn maker where α_i and β_j are the maker and speed fixed effect with interactions δ_{ij}, and u_i denotes the random effect of manufacturer and v_{ik} is the machine random effect nested in manufacturer.

The fundamental reason here is that there is no way to separate the fixed effect and random effect α_i and u_i in the model, leading to the so-called identifiability issue. Or, intuitively, it is very clear that one may either use manufacturer as a fixed effect or a random effect depending on the objective of the study – being specific for the two manufacturers, or they are

just sampled from many potential manufacturers where the main concern is the association between the speed and time. But, more importantly, data would not support using it for both purposes.

Do not ignore warning message in the output. Additionally, with the package `lmerTest` (Thanks Google!), p-values with the Cochran-Satterthwaite approximated d.f. for the degrees of freedoms are calculated for outputs in `lmer`.

```
> require(lmerTest)
> lawn.m=lmer(time~manufact*speed+(1|manufact)+(1|manufact:machine),data=lawn)
Warning message:
In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model is nearly unidentifiable: large eigenvalue ratio
 - Rescale variables?
> summary(lawn.m)
[1] "Asymptotic covariance matrix A is not positive!"
Linear mixed model fit by REML
t-tests use  Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: time ~ manufact * speed + (1 | manufact) + (1 | manufact:machine)
   Data: lawn

REML criterion at convergence: 168.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.0909 -0.6740 -0.1291  0.6661  1.5405

Random effects:
 Groups           Name        Variance Std.Dev.
 manufact:machine (Intercept) 145.2    12.05
 manufact         (Intercept) 181.1    13.46
 Residual                     132.3    11.50
Number of obs: 24, groups:  manufact:machine, 6; manufact, 2

Fixed effects:
                Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       270.500     15.861   0.000  17.055     1.00
manufactB         -21.833     22.431   0.000  -0.973     1.00
speedL            -60.333      6.641  16.000  -9.085 1.03e-07 ***
manufactB:speedL    2.667      9.392  16.000   0.284     0.78
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Correlation of Fixed Effects:
            (Intr) mnfctB speedL
manufactB   -0.707
speedL      -0.209  0.148
mnfctB:spdL  0.148 -0.209 -0.707
```

(b) Show why the manufacturer term may be removed from the fixed effect part of the model.

Solutions

Removing the manufacturer as a random effect, and the test will conclude that it is not significant as a fixed effect.

```
> lawn.m2=lmer(time~manufact*speed+(1|machine),data=lawn)
> summary(lawn.m2)
Linear mixed model fit by REML
t-tests use  Satterthwaite approximations to degrees of freedom ['merModLmerTest']
Formula: time ~ manufact * speed + (1 | machine)
   Data: lawn

REML criterion at convergence: 168.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.0909 -0.6740 -0.1291  0.6661  1.5405

Random effects:
 Groups   Name        Variance Std.Dev.
 machine  (Intercept) 145.2    12.05
 Residual             132.3    11.50
Number of obs: 24, groups:  machine, 6

Fixed effects:
                 Estimate Std. Error      df t value Pr(>|t|)
(Intercept)       270.500      8.394   5.574  32.225 1.50e-07 ***
manufactB         -21.833     11.871   5.574  -1.839    0.119
speedL            -60.333      6.641  16.000  -9.085 1.03e-07 ***
manufactB:speedL    2.667      9.392  16.000   0.284    0.780
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Correlation of Fixed Effects:
            (Intr) mnfctB speedL
```

```
manufactB   -0.707
speedL      -0.396  0.280
mnfctB:spdL  0.280 -0.396 -0.707
```

(c) Determine if the manufacturer term can be removed from the random part of the model.

Solutions

Likelihood ratio will conclude that the manufacturer effect is also not significant in the random component.

```
> lawn.m3=lmer(time~speed+(1|manufact)+(1|manufact:machine),data=lawn)
> rand(lawn.m3)
Analysis of Random effects Table:
                 Chi.sq Chi.DF p.value
manufact          0.788      1    0.37
manufact:machine  6.463      1    0.01 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```