STAT 8004 – Statistical Methods II
Spring 2015
Homework Assignment 8 – Solutions

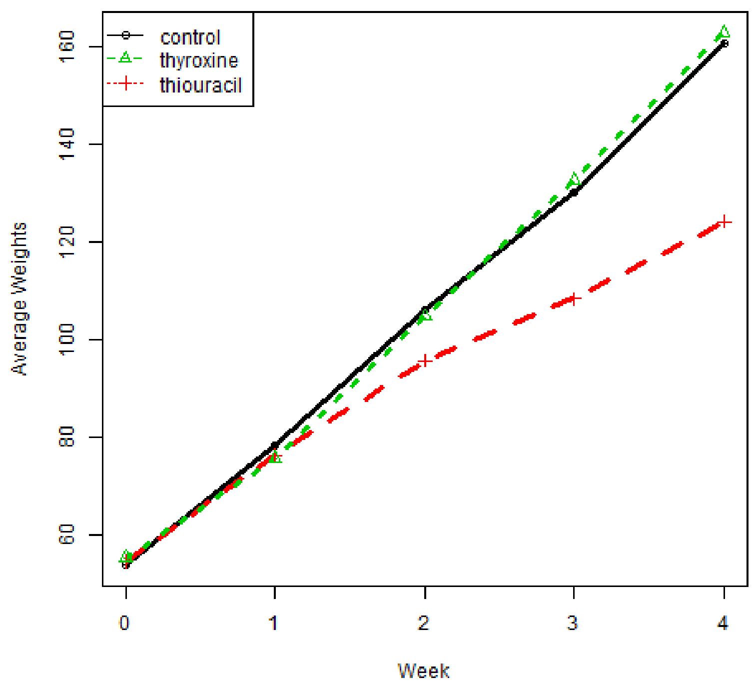
1. This is Problem 3 of Faraway (2006), Chapter 9.

   The ratdrink data consist of five weekly measurements of body weight for 27 rats. The first 10 rats are on a control treatment while seven rats have thyroxine added to their drinking water. Ten rats have thiouracil added to their water. Build a model for the rat weights that shows the effect of the treatment.

```
> library(faraway)
> data(ratdrink)
```

   (a) Model the weights of the rate, incorporating the treatment effects and random effect. Use R to fit the model.

   The data suggest trend in the weights over time, and treatment effect of thiouracil.



   Below is the estimation of the fixed effect with assessment of statistical evidence, taking time as a continuous variable.

```
> ratm=lmer(wt~treat*weeks+(1|subject),data=ratdrink)
> summary(ratm)
Linear mixed model fit by REML t-tests use Satterthwaite approximations to
```

```
  degrees of freedom [merModLmerTest]
Formula: wt ~ treat * weeks + (1 | subject)
   Data: ratdrink

REML criterion at convergence: 948.4

Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.05506 -0.65511 -0.04848  0.57702  2.80847

Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept) 71.21    8.438
 Residual             51.22    7.157
Number of obs: 135, groups:  subject, 27

Fixed effects:
                      Estimate Std. Error       df t value Pr(>|t|)
(Intercept)            52.8800     3.1928  37.0600  16.562  < 2e-16 ***
treatthiouracil         4.7800     4.5153  37.0600   1.059    0.297
treatthyroxine         -0.7943     4.9756  37.0600  -0.160    0.874
weeks                  26.4800     0.7157 105.0000  36.999  < 2e-16 ***
treatthiouracil:weeks  -9.3700     1.0121 105.0000  -9.258 2.89e-15 ***
treatthyroxine:weeks    0.6629     1.1153 105.0000   0.594    0.554
```

(b) What is the implication of the random effect on the correlations between weights of the same rat? Is that implication reasonable? It would be nice to support your argument with data evidence.

The random effect will introduce correlations in the exchangeable structure, i.e, the same correlation between each pair observations. A test will indicate significance for the existence of such an effect.

```
> rand(ratm)
Analysis of Random effects Table:
        Chi.sq Chi.DF p.value
subject   57.3      1   4e-14 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

But it may not be fully supported by data, for example, take the residuals corresponding to the 0,1,2 week respectively, their sample correlation suggests something else.

```
> rr=residuals(ratm)
```

```
> rr1=rr[which(ratdrink$weeks==0)]
> rr2=rr[which(ratdrink$weeks==1)]
> rr3=rr[which(ratdrink$weeks==2)]
> cor(cbind(rr1,rr2,rr3))
           rr1       rr2        rr3
rr1  1.0000000 0.4068410 -0.4015909
rr2  0.4068410 1.0000000  0.5460809
rr3 -0.4015909 0.5460809  1.0000000
```

2. The article "Variability of Sliver Weights at Different Carding Stages and a Suggested Sampling Plan for Jute Processing" by A. Lahiri (Journal of the Textile Institute, 1990) concerns the partitioning of variability in "sliver weight." (A sliver is a continuous strand of loose, untwisted wool, cotton, etc., produced along the way to making yarn.) For a particular mill, 3 (of many) machines were studied, using 5 (10 mm) pieces of sliver cut from each of 5 rolls produced on the machines. The weights of the (75) pieces of sliver were determined and a standard hierarchical (balanced data) ANOVA table was produced as below. (The units of weight were not given in the original article.)

Source	SS	df
Machines	1966	2
Rolls	644	12
Pieces	280	60
Total	2890	74

The model is

$$y_{ijk} = \mu + \alpha_i + u_{ij} + \varepsilon_{ijk}$$

for the kth piece of the jth roll on the ith machine, where $\alpha_i \stackrel{iid}{\sim} N(0, \sigma_a^2)$, $u_{ij} \stackrel{iid}{\sim} N(0, \sigma_u^2)$, and $\varepsilon_{ijk} \stackrel{iid}{\sim} N(0, \sigma_e^2)$.

(a) Find estimates for σ_a^2, σ_u^2, and σ_e^2.

The following the moment identity, taking A to be the machine factor, B to be the Rolls nested within machine, and error corresponding to the replicaiton of pieces:

$$\begin{aligned} E(\text{MSE}) &= \sigma_e^2 \\ E(\text{MSB(A)}) &= \sigma_e^2 + 5\sigma_u^2 \\ E(\text{MSA}) &= \sigma_e^2 + 5\sigma_u^2 + 25\sigma_a^2 \end{aligned}$$

This will give MOM estimates:

$$\hat{\sigma}_e^2 = 4.667$$
$$\hat{\sigma}_u^2 = \frac{1}{5}(53.667 - 4.667) = 9.8$$
$$\hat{\sigma}_a^2 = \frac{1}{25}(983 - 53.667) = 37.1733$$

(b) Make 95% confidence intervals for each of the 3 standard deviations σ_a, σ_u, and σ_u. Based on these, where do you judge the largest part of variation in measured weight to come from? You need to use the Cochran- Satterthwaite approximation for σ_a and σ_u.

The $(SSE/\chi^2_{60,0.975}, SSE/\chi^2_{60,0.025})$ will given an exact confidence interval for σ_e^2, taking a square root to get (1.83, 2.63) as an confidence interval for σ_e.

The Cochran-Saterthwaite approximation for the degress of freedom of $\hat{\sigma}_u^2$ is

$$\hat{\nu} = \frac{(\hat{\sigma}_u^2)^2}{\frac{(\text{MSB(A)}/5)^2}{12} + \frac{(\text{MSE}/5)^2}{60}} = 9.988675$$

Then the approximated confidence interval for σ_u^2 is

$$\left(\frac{\hat{\nu}\hat{\sigma}_u^2}{\chi^2_{\hat{\nu},0.975,}}, \frac{\hat{\nu}\hat{\sigma}_u^2}{\chi^2_{\hat{\nu},0.025,}} \right) = (4.782825, 30.20658)$$

taking square root to get confidence interval for σ_u as (2.19, 5.50).

The Cochran-Saterthwaite approximation for the degress of freedom of $\hat{\sigma}_a^2$ is

$$\hat{\nu} = \frac{(\hat{\sigma}_a^2)^2}{\frac{(\text{MSA}/25)^2}{2} + \frac{(\text{MSB(A)}/25)^2}{12}} = 1.786695$$

Then the approximated confidence interval for σ_a^2 is

$$\left(\frac{\hat{\nu}\hat{\sigma}_a^2}{\chi^2_{\hat{\nu},0.975,}}, \frac{\hat{\nu}\hat{\sigma}_a^2}{\chi^2_{\hat{\nu},0.025,}} \right) = (9.596762, 2143.497)$$

taking square root to get confidence interval for σ_a as (3.10, 46.30).

(c) Suppose for sake of illustration that the grand average of all 75 weight measurements was in fact $\bar{y}_{...} = 1/75 \sum_{ijk} y_{ijk} = 35.0$. Use this and information from the ANOVA table to make a 95% confidence interval for the model parameter μ.

The 95% Confidence Interval for μ is

$$\left(\; \bar{y}_{...} - t_{2,0.975}\sqrt{MSA/75}, \quad \bar{y}_{...} + t_{2,0.975}\sqrt{MSA/75} \; \right),$$

giving (19.42, 50.58).