8004 Homework 8

Nooreen Dabbish
April 2, 2015

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.

We write y_{ijk} to represent the kth rat in the jth treatment group on the ith week, where (i=1,2,3,4), (j=1,2,3), and (k= 1-10 for control, k = 1-7 thyroxine, and k = 1-10 for thiouracil). μ represents the overall mean weight, α_i represents the fixed effect contribution of the ith week, β_j repsents the fixed effect contribution of the jth treatment, and δ_{ij} is the interaction of weeks and treatment. The random effect u_{jk} incorporates the repeated measures of the same rat.

$$y_{ijk} = \mu + \alpha_i + \beta_j + \delta_{ij} + u_{jk} + \epsilon_{ijk}$$

To fit the model in R we write:
`rat.lme <- lmer(wt ~ weeks+ treat+ weeks*treat+ (1|subject))`
The command `summary(rat.lme)` gives:

```
Linear mixed model fit by REML ['lmerMod']
Formula: wt ~ weeks + treat + weeks * treat + (1 | subject)

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 t value
(Intercept)            52.8800     3.1928   16.56
weeks                  26.4800     0.7157   37.00
treatthiouracil         4.7800     4.5153    1.06
treatthyroxine         -0.7943     4.9756   -0.16
weeks:treatthiouracil  -9.3700     1.0121   -9.26
weeks:treatthyroxine    0.6629     1.1153    0.59

Correlation of Fixed Effects:
```

```
             (Intr) weeks  trtthr trtthy wks:trtthr
weeks        -0.448
treatthircl -0.707  0.317
treatthyrxn -0.642  0.288  0.454
wks:trtthrc  0.317 -0.707 -0.448 -0.203
wks:trtthyr  0.288 -0.642 -0.203 -0.448  0.454
```

(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 captures the correlation between measuring the same rat on multiple weeks. Because that rat has a starting weight at week zero, subsequent measurements will be correlated with the previous weight.

We fit a completely fixed model and call `summary(rat.lm)` for comparison. As shown below, and as we would expect, the fixed coefficient estimates are the same without the random effect from the same rat. However, above, we were able to obtain a correlation matrix showing the correlation of the fixed effects. The correlation across weeks for the overall mean obtained was -.448, which is relatively low (highly correlated would be close to 1, or inversely correlated close to -1).

```
> rat.lm <- lm(wt ~weeks+treat+weeks*treat)
> summary(rat.lm)

Call:
lm(formula = wt ~ weeks + treat + weeks * treat)

Residuals:
    Min      1Q  Median      3Q     Max
-23.514  -6.660   0.230   6.914  28.343

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)            52.8800     2.6547  19.919  < 2e-16 ***
weeks                  26.4800     1.0838  24.433  < 2e-16 ***
treatthiouracil         4.7800     3.7544   1.273    0.205
treatthyroxine         -0.7943     4.1371  -0.192    0.848
weeks:treatthiouracil  -9.3700     1.5327  -6.113 1.08e-08 ***
weeks:treatthyroxine    0.6629     1.6890   0.392    0.695
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.84 on 129 degrees of freedom
Multiple R-squared:  0.9121,    Adjusted R-squared:  0.9087
F-statistic: 267.8 on 5 and 129 DF,  p-value: < 2.2e-16
```

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

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

for the kth piece of the jth roll on the ith machine, where

$$\alpha_i \overset{iid}{\sim} N(0, \sigma^2_\alpha),\ u_{ij} \overset{iid}{\sim} N(0, \sigma^2_u), \text{and}\, \epsilon_i \overset{iid}{\sim} N(0, \sigma^2_\epsilon).$$

(a) Find estimates for $\sigma^2{}_\alpha$, $\sigma^2{}_u$ and $\sigma^2{}_\epsilon$.

Source	SS	df	Term	MS		E(MS)
Machines	1966	2		MSA	983	$\sigma^2_\epsilon + 5\sigma^2_u + 25\sigma^2_\alpha$
Rolls	644	12		MSB]A	53.666667	$\sigma^2_\epsilon + 5\sigma^2_u$
Pieces	280	60	error	MSE	4.6666667	σ^2_ϵ
Total	2890	74				

Solving for the expectations above gives the following estimates:

- $\widehat{\sigma^2_\epsilon} = 4.6667$
- $\widehat{\sigma^2_u} = 9.8$
- $\widehat{\sigma^2_\alpha} = 37.17333$

(b) Make 95% confidence intervals for each of the 3 standard deviations

σ_α, σ_u, and σ_ϵ. Based on theses, where do you judge the largest part of the variation in measured weight to come from? You need to use the Cochran-Satterthwaite approximation for σ_α and σ_u.

For σ_ϵ we note that

$$\frac{SSE}{\sigma_\epsilon} \sim \chi^2_{60}.$$

We write:

$$c(\sigma_\epsilon) = \left(\sigma_\epsilon : \chi^2_{60,.05} < \frac{SSE}{\sigma^2_\epsilon} < \chi^2_{60,.95}\right)$$

$$c(\sigma_\epsilon) = \left(\sigma_\epsilon : \sqrt{\frac{SSE}{\chi^2_{60,.95}}} < \sigma_\epsilon < \sqrt{\frac{SSE}{\chi^2_{60,.05}}}\right)$$

$$c(\sigma_\epsilon) = (\sigma_\epsilon : 1.833423 < \sigma_\epsilon < 2.629961)$$

We use the Cochran-Satterwaithe approximation for $\sigma\alpha$ and σ_u. There is used to determine the degrees of freedom of a Chi-squared distribution which is approximately:

$$\frac{\nu(S^2)}{E(S^2)} \sim \chi^2_\nu$$

This gives us a 1-α confidence interval for E(S^2) determined by:

$$P\left(\frac{\nu \cdot S^2}{\chi^2_{\nu,upper}} < E(S^2) < \frac{\nu \cdot S^2}{\chi^2_{\nu,lower}}\right) = 1-\alpha$$

$$\hat{\nu}_u = \frac{(\sigma_u^2)^2}{\frac{((MSB|A)/5)^2}{df=12} + \frac{(-MSE/5)^2}{df=60}} = 9.988675$$

$$\hat{\nu}_\alpha = \frac{(\sigma_\alpha^2)^2}{\frac{((-MSB|A)/25)^2}{df=12} + \frac{(MSA/25)^2}{df=2}} = 1.786695$$

Solving for our confidence intervals gives

$$c(\sigma_u) = (\sigma_u : 2.186967 < \sigma_u < 5.496051)$$

$$c(\sigma_\alpha) = (\sigma_\alpha : 3.097864 < \sigma_\alpha < 46.29789)$$

Clearly, the largest contribution to the variability in the measured weight of the silver comes from the differences between machines. This is because the estimate of the standard deviation associated with $\sigma\alpha$ isthe largest, and its confidence interval also ranges over the largest values.

(c) Suppose for the 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.

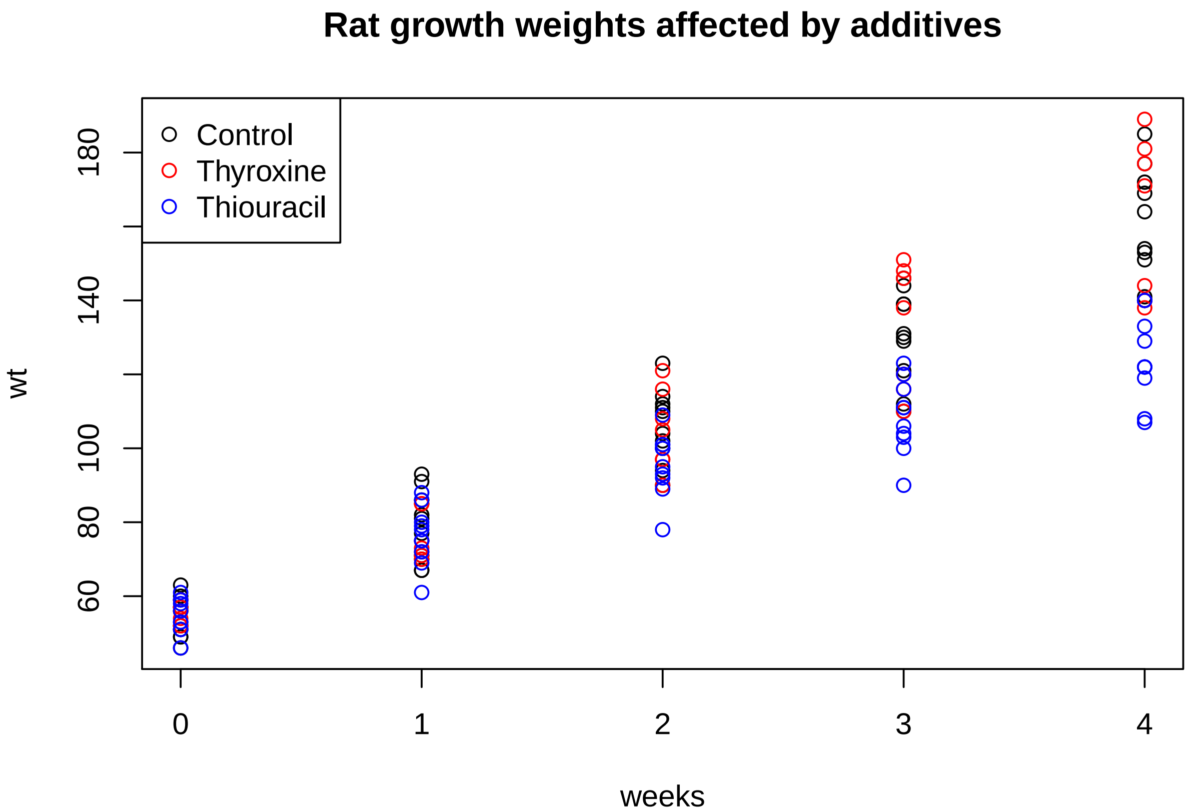
The 95% confidence interval we are interested in is $\bar{y}_{...} \pm t_{.975,df=2} \sqrt{\frac{MSA}{3\cdot 5\cdot 5}} = (19.42305, 50.57695)$.

3 Appendix: Tangled R Code

```
library(MASS); library(xtable); library(nlme)
  lvector <- function(x, dig = 2, dsply=rep("f",ncol(x)+1)) {
   x <- xtable(x, align=rep("",ncol(x)+1),display=dsply, digits=dig) # We repeat empty string 6 time
   print(x, floating=FALSE, tabular.environment="pmatrix",
     hline.after=NULL, include.rownames=FALSE, include.colnames=FALSE)
   }

library(faraway)
data(ratdrink)

help(ratdrink)
library(lattice)

ratdrink$thecolor = "black"
ratdrink$thecolor[ratdrink$treat == "thyroxine"] = "red"
ratdrink$thecolor[ratdrink$treat == "thiouracil"] = "blue"
attach(ratdrink)
pdf("ratweights.pdf",width=7,height=5)
plot(weeks, wt, col = thecolor, main="Rat growth weights affected by additives")
legend("topleft", c("Control", "Thyroxine", "Thiouracil"), col=c("black","red","blue"), pch=1)
dev.off()
```

```
#####Fit model######
library(lme4)
rat.lme <- lmer(wt ~ weeks+ treat+ weeks*treat+ (1|subject))
summary(rat.lme)

var(ratdrink[treat=="control",]$wt)
myvars <- numeric(10);
for(i in 1:10){
    thisi <- toString(i);
    myvars[i] <- var(ratdrink[subject==thisi,]$wt)
}
mean(myvars)

typeof(weeks)


######################
sige <- 280/60
sigu <- ((644/12) - sige)/5
siga <- ((1966/2) - 5*sigu - sige)/25

sqrt(280/qchisq(.025,60))

sqrt(280/qchisq(.975,60))

MSB <- 644/12
MSA <- 1966/2
MSE <- 280/60

vhatu <- (sigu)^2/((((MSB/5)^2)/12)+((MSE/5)^2)/60)
vhatu

vhata <- (siga)^2/((((MSB/25)^2)/12)+(((MSA/25)^2)/2))
vhata

sqrt((sigu*vhatu)/qchisq(.975,vhatu))
sqrt((sigu*vhatu)/qchisq(.025,vhatu))

sqrt((siga*vhata)/qchisq(.975,vhata))
sqrt((siga*vhata)/qchisq(.025,vhata))

35 - qt(.975,2)*sqrt(MSA/(3*5*5))
35 + qt(.975,2)*sqrt(MSA/(3*5*5))
```

4 Appendix: Initial evaluation of ratdrink dataset



Plotting the ratdrink data suggested that rats that drank Thyroxine tended to have increased body weight after 5 weeks in comparison to rats drinking Thiouracil and Control. The rats that drank Thiouracil tended to have lowerbody weight than the Control and Thyroxine groups.