Advanced Applied Statistics A computer session
Jianqing Fan

All data are recorded in the ./Data directory. Creating a directory .Data in a subdirectory will enable you to record data in this subdirectory. This is accomplished by issuing the command splus CHAPTER in the subdirectory where you want to work. Change .Data/.Audit to be unwritable if you want to turn off Auditing.

Basics

splus
> q()  #exit
> labor <- matrix(scan("labor.suppl.dat", skip=17),byrow=T,ncol=9)
#reading data from "labor.DATA"
> ?matrix
#this is the same as typing 'help(matrix)'.
> dimnames(labor) <- list(NULL, c("hours", "age", "earning", "job.pres",  
"edu", "rent", "hearning", "child", "unemploy"))
> labor[1:2,]  
#take the first two rows
hours age earning job.pres edu rent hearning child unemploy
 21  36   8.269    55   12  1010  2800   1  16.8
 40  35   6.059    29   10   268  2500   1  16.8
> edu
Problem: Object "edu" not found

> edu <- labor[,5]  
#take a column out
> rm(edu)
#rm the vector edu from the data base

> labor.df <- data.frame(labor)  
# create data frame
> attach(labor.df)  
# each vector is now recognizable
> edu
 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
36 35 33 30 43 45 39 33 48 55 30 47 54 51 52 37 36 33 36 52 33 36 48 32 30

> eduf <- rep("A",length(edu))  
# create data vector of length 607
> eduf[edu < 16] <- "B"
> eduf[edu < 13] <- "C"
> eduf[edu < 12] <- "D"
> laborf <- data.frame(earning, eduf)
> laborf[1:3,]
# Show the first 3 cases
 1  8.269   C
 2  6.059   D
 3 11.500   C

> sink("result")  
> edu
> sink(on.exit=T)
One-Way ANOVA

We now use the female labor supply data in the East Germany. 607 women with job who live together with a partner have been asked for their weekly number of working hours. Furthermore, it has been recorded, if they have children less than 16 years old, the unemployment rate in the Land of the Federal Republic of Germany where she lives, the age of the woman, her wage per hour, the "Treiman prestige index" of her job (see Treiman, 1978), her years of education [introduction of this covariate makes sense because of the strongly regulated system of education in the former state of East Germany], her rent or redemption the monthly net income of her partner. For simplicity of illustration, we use only two variables. Hourly earning rate and education level (A = 16, B = 13-15, C = 12, D < 12. One naturally try to exam the data. Here are a few commands related to ANOVA.

Figure 1: Visual inspection of the data

> postscript("labor.anova1.ps",width=5.5, height = 5,horizontal=F, pointsize=8);
#record the results in the file instead of on the screen
> par(mfrow=c(2,2),mar=c(5,3,3,1)+0.1)
#setting some paramgers
> plot.design(laborf)
#graphic plot using mean
> title("Comparison of the means")
> plot.design(laborf, fun=median)
#setting some paramgers
> title("Comparison of the medians")
```r
# graphic plot using median
> plot.design(laborf, fun=var)
# graphic plot for comparing variance
> title("Comparison of the variance")
> plot.factor(laborf)
> title("Boxplots for each treatment")
> dev.off()
# force to create the postscript now
> lm(earning ~ eduf)
# run a linear model fit
Call:
  lm(formula = earning ~ eduf)
Coefficients:
(Intercept)   eduf1   eduf2   eduf3
  10.51736  -1.664151  -1.098402  -0.7939263
Degrees of freedom: 607 total; 603 residual
Residual standard error: 3.991209
> options()$contrast
  factor     ordered
"contr.helmert" "contr.poly"
> options(contrasts=c("contr.treatment"))
> options()$contrast
[1] "contr.treatment"
> lm(earning ~ eduf)
Call:
  lm(formula = earning ~ eduf)
Coefficients:
(Intercept)   edufB   edufC   edufD
  14.07384  -3.328302  -4.959358  -5.938258
Degrees of freedom: 607 total; 603 residual
Residual standard error: 3.991209
> options(contrasts=c("contr.sum"))
> lm(earning~eduf)
Call:
  lm(formula = earning ~ eduf)
Coefficients:
(Intercept)   eduf1   eduf2   eduf3
  10.51736   3.556479  0.2281777 -1.402878
Degrees of freedom: 607 total; 603 residual
Residual standard error: 3.991209
>aov.labor <- aov(earning~eduf)
# run anova for the data
> summary(aov.labor)
    Df Sum Sq Mean Sq  F value Pr(>F)
  eduf   3 1884.84  628.28 39.44076 0
Residuals 603 9605.63  15.9297
```
> multcomp(aov.labor,focus="eduf",method="bon")
    # Bonferroni method

> postscript("labor.anova2.ps",width=5.5, height = 5,horizontal=F, pointsize=8);
#record the results in the file instead of on the screen
> par(mfrow=c(2,2),mar=c(5,3,3,1)+0.1)
> residuals <- resid(aov.labor)
> hist(residuals)
#plot the histogram of the residuals
> plot(density(residuals),type="l")
#plot the density of the residuals
> qqnorm(residuals)
#Q-Q plot of the residuals

Figure 2: Visual inspection of the data

Two-way ANOVA models
We now try to fit a two-way layout ANOVA model with interactions

```r
> attach(laborf)
> options(contrasts=c("contr.treatment","contr.treatment"))
> lm(earning~childf*eduf)
### fitting a two-term interaction model
```

Coefficients:
```
(Intercept)  childf  edufB  edufC  edufD  childfedufB  childfedufC
  14.40383 -0.44422 -3.87124 -4.07230 -5.96529  0.75907 -1.18940
```

```
childfedufD
-0.6810723
```

```r
> options(contrasts=c("contr.sum","contr.sum"))
> lm(earning~childf*eduf)
```

Call: lm(formula = earning ~ childf + eduf)

Coefficients:
```
(Intercept)  childf  eduf1  eduf2  eduf3  childfeduf1  childfeduf2
  10.56559  0.36103  3.61613  0.12443 -1.05087 -0.13893 -0.51846
```

```
childfeduf3
  0.4557743
```
Figure 3: Visual inspection of the data with two factors
Degrees of freedom: 607 total; 599 residual
Residual standard error: 3.975031
> labor.aov <- aov(earning~childf*eduf)
> summary(labor.aov)

summary(labor.aov)

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F Value</th>
<th>Pr(F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>childf</td>
<td>1</td>
<td>4.057</td>
<td>4.0575</td>
<td>0.25679</td>
</tr>
<tr>
<td>eduf</td>
<td>3</td>
<td>1936.489</td>
<td>645.4964</td>
<td>40.85195</td>
</tr>
<tr>
<td>childf:eduf</td>
<td>3</td>
<td>85.213</td>
<td>28.4042</td>
<td>1.79764</td>
</tr>
<tr>
<td>Residuals</td>
<td>599</td>
<td>9464.721</td>
<td>15.8009</td>
<td></td>
</tr>
</tbody>
</table>

We now try to fit a two-way layout ANOVA model with only the main effect

> options(contrasts=c("contr.treatment","contr.treatment"))
> lm(earning~childf+eduf)

Call:

lm(formula = earning ~ childf + eduf)

Coefficients:

(Intercept)  childf  edufB  edufC  edufD
14.57527 -0.6749966 -3.373208 -4.957911 -6.257954

Degrees of freedom: 607 total; 602 residual
Residual standard error: 3.982923

> summary(aov(earning~childf+eduf))

Df   Sum of Sq  Mean Sq   F Value  Pr(F)
childf 1     4.057   4.0575  0.25577 0.6132244
eduf    3    1936.489  645.4964 40.69021 0.0000000
childf:eduf 3    85.213   28.4042  1.79764 0.1464194
Residuals 602    9549.934  15.8637

ANOVA Decompositions: Balanced vs Unbalanced

Comparison two ANOVA decompositions. For balanced design, the decomposition is unconditional. For inbalanced design, the decomposition is conditional.

> battery

#balanced design data set

<table>
<thead>
<tr>
<th>Brand</th>
<th>Duty</th>
<th>resp</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>named</td>
<td>Alkaline</td>
</tr>
<tr>
<td>2</td>
<td>named</td>
<td>Alkaline</td>
</tr>
<tr>
<td>3</td>
<td>named</td>
<td>Alkaline</td>
</tr>
<tr>
<td>4</td>
<td>named</td>
<td>Alkaline</td>
</tr>
<tr>
<td>5</td>
<td>named</td>
<td>Heavyduty</td>
</tr>
<tr>
<td>6</td>
<td>named</td>
<td>Heavyduty</td>
</tr>
<tr>
<td>7</td>
<td>named</td>
<td>Heavyduty</td>
</tr>
<tr>
<td>8</td>
<td>named</td>
<td>Heavyduty</td>
</tr>
<tr>
<td>9</td>
<td>store</td>
<td>Alkaline</td>
</tr>
<tr>
<td>10</td>
<td>store</td>
<td>Alkaline</td>
</tr>
</tbody>
</table>
> attach(battery)
> summary(aov(resp~Brand))
#ANOVA decomposition with one factor

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>1</td>
<td>124609.0</td>
<td>124609.0</td>
<td>5.259052</td>
</tr>
<tr>
<td>Residuals</td>
<td>14</td>
<td>331718.7</td>
<td>23694.2</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(resp~Brand+Duty))
#ANOVA decomposition with two factors. Note that the contribution due
# to BRAND does not change so does it in the following two models.

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brand</td>
<td>1</td>
<td>124609.0</td>
<td>124609.0</td>
<td>20.32142</td>
</tr>
<tr>
<td>Duty</td>
<td>1</td>
<td>252004.0</td>
<td>252004.0</td>
<td>41.09719</td>
</tr>
<tr>
<td>Residuals</td>
<td>13</td>
<td>79714.7</td>
<td>6131.9</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(resp~Duty+Brand))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duty</td>
<td>1</td>
<td>252004.0</td>
<td>252004.0</td>
<td>41.09719</td>
</tr>
<tr>
<td>Brand</td>
<td>1</td>
<td>124609.0</td>
<td>124609.0</td>
<td>20.32142</td>
</tr>
<tr>
<td>Residuals</td>
<td>13</td>
<td>79714.8</td>
<td>6131.9</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(resp~Duty*Brand))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duty</td>
<td>1</td>
<td>252004.0</td>
<td>252004.0</td>
<td>106.4337</td>
</tr>
<tr>
<td>Brand</td>
<td>1</td>
<td>124609.0</td>
<td>124609.0</td>
<td>52.6285</td>
</tr>
<tr>
<td>Duty:Brand</td>
<td>1</td>
<td>51302.2</td>
<td>51302.2</td>
<td>21.6675</td>
</tr>
<tr>
<td>Residuals</td>
<td>12</td>
<td>28412.5</td>
<td>2367.7</td>
<td></td>
</tr>
</tbody>
</table>

#The situations changes dramatically for unbalanced design in
#the earning data. Note that the different sum of squares reductions
#due to eduf. They are conditional reductions.

> summary(aov(earning~eduf))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eduf</td>
<td>3</td>
<td>1884.844</td>
<td>628.2813</td>
<td>39.44076</td>
</tr>
<tr>
<td>Residuals</td>
<td>603</td>
<td>9605.637</td>
<td>15.9297</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(earning~eduf+childf))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>eduf</td>
<td>3</td>
<td>1884.844</td>
<td>628.2813</td>
<td>39.60502</td>
</tr>
<tr>
<td>childf</td>
<td>1</td>
<td>55.703</td>
<td>55.703</td>
<td>3.51135</td>
</tr>
<tr>
<td>Residuals</td>
<td>602</td>
<td>9549.934</td>
<td>15.8637</td>
<td></td>
</tr>
</tbody>
</table>

> summary(aov(earning~childf+eduf))

<table>
<thead>
<tr>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>childf</td>
<td>1</td>
<td>4.057</td>
<td>4.057</td>
<td>0.25577</td>
</tr>
</tbody>
</table>
eduf 3 1936.489 645.4964 40.69021 0.0000000
Residuals 602 9549.934 15.8637

> summary(aov(earning~childf*eduf))
Df Sum of Sq Mean Sq F Value Pr(F)
childf 1 4.057 4.0575 0.25679 0.6125209
eduf 3 1936.489 645.4964 40.85195 0.0000000
childf:eduf 3 85.213 28.4042 1.79764 0.1464194
Residuals 599 9464.721 15.8009

> summary(aov(earning~eduf*childf))
Df Sum of Sq Mean Sq F Value Pr(F)
eduf 3 1884.844 628.2813 39.76245 0.0000000
childf 1 55.703 55.7030 3.52531 0.0609242
eduf:childf 3 85.213 28.4042 1.79764 0.1464194
Residuals 599 9464.721 15.8009

Analysis of Covariance

The analysis of Covariance does not make any extra difficulty. Please note that the ANOVA decomposition is valid in conditional sense. The unconditional decomposition holds only when linear spaces are orthogonal (design matrix are blockwise orthogonal). Compare the ANOVA decomposition for different orders. Compare different meaning of contrasts in different models.

> attach(labor.df)
> lm(earning ~ eduf + age + job.pres)
Call:
lm(formula = earning ~ eduf + age + job.pres)

Coefficients:
(Intercept)   eduf1   eduf2   eduf3     age    job.pres
4.522455 -1.16764 -0.5829174 -0.4750492 0.02124989 0.1185623

Degrees of freedom: 607 total; 601 residual
Residual standard error: 3.787025

> lm(earning ~ age + job.pres)
Call:
lm(formula = earning ~ age + job.pres)

Coefficients:
(Intercept)     age    job.pres
3.224799 0.003113607 0.1590534

Degrees of freedom: 607 total; 604 residual
Residual standard error: 3.880508

> summary(aov(earning ~ eduf + age + job.pres))
Df Sum of Sq Mean Sq F Value Pr(F)
eduf 3 1884.844 628.2813 43.80844 0.0000000
age 1 32.268 32.2683 2.24999 0.1341406
job.pres 1 954.093 954.0929 66.52645 0.0000000
```r
> summary(aov(earning ~age + eduf + job.pres))

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F Value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1</td>
<td>2.484</td>
<td>2.4838</td>
<td>0.1732</td>
<td>0.6774</td>
</tr>
<tr>
<td>eduf</td>
<td>3</td>
<td>1914.628</td>
<td>638.2094</td>
<td>44.50</td>
<td>0.0000</td>
</tr>
<tr>
<td>job.pres</td>
<td>1</td>
<td>954.093</td>
<td>954.0929</td>
<td>66.53</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Residuals 601 8619.276 14.3416

> summary(aov(earning ~job.pres + eduf + age))

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>Sum of Sq</th>
<th>Mean Sq</th>
<th>F Value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>job.pres</td>
<td>1</td>
<td>2394.842</td>
<td>2394.842</td>
<td>166.99</td>
<td>0.0000</td>
</tr>
<tr>
<td>eduf</td>
<td>3</td>
<td>461.556</td>
<td>153.852</td>
<td>10.73</td>
<td>0.0000</td>
</tr>
<tr>
<td>age</td>
<td>1</td>
<td>14.806</td>
<td>14.806</td>
<td>1.03</td>
<td>0.31</td>
</tr>
</tbody>
</table>

Residuals 601 8619.276 14.342

> options(contrasts=c("contr.treatment"))
> summary(lm(earning ~job.pres + eduf + age))

Call: lm(formula = earning ~ job.pres + eduf + age)

Residuals:
Min 1Q Median 3Q Max
-9.492 -1.758 -0.1898 1.575 57.09

Coefficients:

|             | Value | Std. Error | t value | Pr(>|t|) |
|-------------|-------|------------|---------|---------|
| (Intercept) | 6.7481| 1.1958     | 5.6430  | 0.0000  |
| job.pres    | 0.1186| 0.0145     | 8.1564  | 0.0000  |
| edufB       | -2.3353| 0.5373    | -4.3465 | 0.0000  |
| edufC       | -2.9164| 0.5809    | -5.0205 | 0.0000  |
| edufD       | -3.6508| 0.6582    | -5.5470 | 0.0000  |
| age         | 0.0212| 0.0209     | 1.0161  | 0.3100  |

Residual standard error: 3.787 on 601 degrees of freedom
Multiple R-Squared: 0.2499
F-statistic: 40.04 on 5 and 601 degrees of freedom, the p-value is 0

Correlation of Coefficients:

<table>
<thead>
<tr>
<th></th>
<th>(Intercept)</th>
<th>job.pres</th>
<th>edufB</th>
<th>edufC</th>
<th>edufD</th>
</tr>
</thead>
<tbody>
<tr>
<td>job.pres</td>
<td>-0.6256</td>
<td>0.1186</td>
<td>-0.4702</td>
<td>-0.6205</td>
<td>-0.3965</td>
</tr>
<tr>
<td>edufB</td>
<td>-0.4702</td>
<td>0.2235</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>edufC</td>
<td>-0.6205</td>
<td>0.4131</td>
<td>0.7504</td>
<td></td>
<td></td>
</tr>
<tr>
<td>edufD</td>
<td>-0.3965</td>
<td>0.4684</td>
<td>0.6802</td>
<td>0.7068</td>
<td></td>
</tr>
<tr>
<td>age</td>
<td>-0.6440</td>
<td>-0.0590</td>
<td>0.0036</td>
<td>0.0738</td>
<td>-0.2573</td>
</tr>
</tbody>
</table>

> lm(earning ~job.pres + eduf)

Call:
  lm(formula = earning ~ job.pres + eduf)

Coefficients:

|             | Value | Std. Error | t value | Pr(>|t|) |
|-------------|-------|------------|---------|---------|
| (Intercept) | 7.530575| 0.1194338  | -2.337265| 0.0003490 |
| job.pres    | -0.95995| 0.0738   | -4.1079 | 0.00008  |
| edufB       | -3.478709| 0.2573   | -1.3497 | 0.1790   |
```

Residuals 601 8619.276 14.3416
Degrees of freedom: 607 total; 602 residual
Residual standard error: 3.787127

> lm(earning~eduf)
Call:
lm(formula = earning ~ eduf)

Coefficients:
(Intercept) edufB edufC edufD
  14.07384 -3.328302 -4.959358 -5.938258

Degrees of freedom: 607 total; 603 residual
Residual standard error: 3.991209

Mixed effect models:

############### Analyzing the battery data using the random effect model
> repl <- c(rep(1,4), rep(2,4), rep(3,4), rep(4,4))
###### create index for cluster
> battery <- cbind(repl, battery)
> attach(battery)
> options(contrasts=c("contr.treatment","contr.treatment"))
> Fit1 <- lme(fixed= resp ~ Brand * Duty, random = ~Brand*Duty,
cluster = ~repl, data=battery)
> summary(Fit1)
Call:
 Fixed: resp ~ Brand * Duty
 Random: ~ Brand * Duty
 Cluster: ~ repl
 Data: battery

Estimation Method: RML
Convergence at iteration: 1
Restricted Loglikelihood: -66.41792
Restricted AIC: 162.8358
Restricted BIC: 174.4247

Variance/Covariance Components Estimate(s):
Structure: unstructured
Parametrization: matrixlog
Standard Deviation(s) of Random Effect(s)
(Intercept) Brand Duty Brand:Duty
  48.6591 48.6591 48.6591 48.6591
Correlation of Random Effects
 (Intercept) Brand Duty
       Brand 0
       Duty 0 0
Brand:Duty 0 0 0

11
Cluster Residual Variance: 2367.708

Fixed Effects Estimate(s):
   Value  Approx. Std.Error  z ratio(C)  
(Intercept)  570.75  54.40253  10.491240 
   Brand     289.75  91.03285   3.182917  
   Duty     -137.75  91.03285   -1.513190 
   Brand:Duty -226.50 153.87360  -1.471987 

Conditional Correlation(s) of Fixed Effects Estimates
   (Intercept)  Brand  Duty 
   Brand -0.5976143 
   Duty -0.5976143  0.3571429 
   Brand:Duty  0.3535534 -0.5916080 -0.5916080 

Random Effects (Conditional Modes):
   (Intercept)  Brand  Duty  Brand:Duty 
   1 -9.094947e-14  0  0  0 
   2 0.000000e+00  0  0  0 
   3 0.000000e+00  0  0  0 
   4 0.000000e+00  0  0  0 

Standardized Population-Average Residuals:
   Min  Q1  Med  Q3  Max 
-1.366651 -0.6897476 -0.3724894 0.7847966 1.495095 

Number of Observations: 16 
Number of Clusters: 4 

> Fit2 <- lme(fixed= resp ~ Brand + Duty, 
   random = ~Brand+Duty, 
   cluster = ~repl, data=battery) 
> anova(Fit1,Fit2) 
Response: resp 
Fit1 
   fixed: (Intercept), Brand, Duty, Brand:Duty 
   random: (Intercept), Brand, Duty, Brand:Duty 
   block: list(1:4) 
   covariance structure: unstructured 
   serial correlation structure: identity 
   variance function: identity 
Fit2 
   fixed: (Intercept), Brand, Duty 
   random: (Intercept), Brand, Duty 
   block: list(1:3) 
   covariance structure: unstructured 
   serial correlation structure: identity 
   variance function: identity 

Model Df  AIC  BIC  Loglik Test Lik.Ratio  P value 
Fit1  1 15 162.84 174.42 -66.418 
Fit2  2 10 166.52 174.25 -73.260 1 vs. 2 13.683 0.017751 

# the reduced model fits reasonably 
> summary(Fit2) 
Call:
Fixed: resp ~ Brand + Duty
Random: ~ Brand + Duty
Cluster: ~ repl
Data: battery

Estimation Method: RML
Convergence at iteration: 4
Restricted Loglikelihood: -73.2596
Restricted AIC: 166.5192
Restricted BIC: 174.2451

Variance/Covariance Components Estimate(s):
Structure: unstructured
Parametrization: matrixlog
Standard Deviation(s) of Random Effect(s)
(Intercept) Brand Duty
67.09202 58.69031 58.69031
Correlation of Random Effects
(Intercept) Brand
Brand 0.4713093
Duty 0.4713093 0.3343843
Cluster Residual Variance: 2367.71

Fixed Effects Estimate(s):
Value Approx. Std.Error z ratio(C)
(Intercept) 593.2371 67.73186 8.758612
Brand 213.1806 107.14465 1.989652
Duty -214.3194 107.14465 -2.000281

Conditional Correlation(s) of Fixed Effects Estimates
(Intercept) Brand
Brand -0.46457409
Duty -0.46457409 -0.06703392

Random Effects (Conditional Modes):
(Intercept) Brand Duty
1 -19.87368 -8.193694 -8.193694
2 28.06737 13.278977 23.401594
3 28.06737 23.401594 13.278977
4 -36.26106 -28.486877 -28.486877

Standardized Population-Average Residuals:
Min Q1 Med Q3 Max
-1.312942 -0.6702421 -0.3724892 0.7862 1.441386

Number of Observations: 16
Number of Clusters: 4

# Now analyzing earning data using the fixed effect model
> subj <- 1:607  # index of subjects
> labor.df <- cbind(subj, labor.df)
# create the data structure. Now, fitting the mixed effect model.

> options(contrasts=c("contr.treatment"))

> fit1 <- lme(fixed = earning ~ job.pres + eduf, random = ~job.pres,
cluster = ~subj, data = labor.df)

> fit1
Call:
  Fixed: earning ~ job.pres + eduf
  Random: ~ job.pres
  Cluster: ~ subj
  Data: labor.df
Variance/Covariance Components Estimate(s):
  Structure: unstructured
  Parametrization: matrixlog
  Standard Deviation(s) of Random Effect(s)
    (Intercept) job.pres
    3.954585 0.1562136
  Correlation of Random Effects
    (Intercept)
    job.pres -0.9995454

  Cluster Residual Variance: 4.688236
Fixed Effects Estimate(s):
  (Intercept) job.pres edufB edufC edufD
  6.981957 0.1196808 -1.923901 -2.420932 -2.891313

Number of Observations: 607
Number of Clusters: 607

> summary(fit1)
Call:
  Fixed: earning ~ job.pres + eduf
  Random: ~ job.pres
  Cluster: ~ subj
  Data: labor.df
Estimation Method: RML
Convergence at iteration: 21
Restricted Loglikelihood: -1614.309
Restricted AIC: 3246.619
Restricted BIC: 3286.295
Variance/Covariance Components Estimate(s):
  Structure: unstructured
  Parametrization: matrixlog
  Standard Deviation(s) of Random Effect(s)
Correlation of Random Effects

Cluster Residual Variance: 4.688236

Fixed Effects Estimate(s):

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Approx. Std.Error</th>
<th>z ratio(C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>6.9819575</td>
<td>0.81369392</td>
<td>8.580570</td>
</tr>
<tr>
<td>job.pres</td>
<td>0.1196808</td>
<td>0.01404981</td>
<td>8.518320</td>
</tr>
<tr>
<td>edufB</td>
<td>-1.9239006</td>
<td>0.56941088</td>
<td>-3.378756</td>
</tr>
<tr>
<td>edufC</td>
<td>-2.4209323</td>
<td>0.56969742</td>
<td>-4.249506</td>
</tr>
<tr>
<td>edufD</td>
<td>-2.8913125</td>
<td>0.59338188</td>
<td>-4.872600</td>
</tr>
</tbody>
</table>

Conditional Correlation(s) of Fixed Effects Estimates

Random Effects (Conditional Modes):
numeric matrix: 607 rows, 2 columns.

Standardized Population-Average Residuals:

Min Q1 Med Q3 Max
-2.862184 -0.348693 -0.011304 0.263574 6.325188

Number of Observations: 607
Number of Clusters: 607

# Now I am fitting the model with only random intercepts and
# see if the model can be reduced to this simplified model.

> fit2 <- lme(fixed = earning ~ job.pres + eduf, random = ~1,
           cluster = ~subj, data = labor.df)

> fit2
Call:
  Fixed: earning ~ job.pres + eduf
  Random: ~ 1
  Cluster: ~ subj
  Data: labor.df

Variance/Covariance Components Estimate(s):

  Structure: unstructured
Parametrization: matrixlog
Standard Deviation(s) of Random Effect(s)
(Intercept)  3.651806

Cluster Residual Variance: 1.006645
Fixed Effects Estimate(s):
(Intercept) job.pres edufB edufC edufD
  7.530575  0.1194338  -2.337265  -2.95995  -3.478709

Number of Observations: 607
Number of Clusters: 607

> anova(fit1,fit2)
Response: earning
fit1
  fixed: (Intercept), job.pres, edufB, edufC, edufD
  random: (Intercept), job.pres
  block: list(1:2)
  covariance structure: unstructured
  serial correlation structure: identity
  variance function: identity
fit2
  fixed: (Intercept), job.pres, edufB, edufC, edufD
  random: (Intercept)
  block: list(1:1)
  covariance structure: unstructured
  serial correlation structure: identity
  variance function: identity

Model Df  AIC  BIC  Loglik  Test Lik.Ratio P value
fit1   1  9 3246.6 3286.3  -1614.3
fit2   2  7 3356.5 3387.4  -1671.3  1 vs. 2 113.91   0

####### So the model can not be simplified further #######