ORIGIN := 0

Randomized Block ANOVA Designs

prepared by
Wm Stein

"Randomized Block" studies are a mixed linear model design in which one variable (variously termed "Treatment" or "Covariate") is considered fixed, and the other variable (termed "Block" or "Subject") is considered random. Within each block, sets of observations may be unique (also called "Two-Way ANOVA Without Replication or "Between Groups ANOVA - see *Biostatistics* Worksheet 300) or replicated (also called "Randomized Block with Replication" or "Repeated Measures with Replication"- see *Biostatistics* Worksheet 301). The purpose of these designs is to use the Block (Subject) as a means to control variance when assessing the effect of the Treatment (Covariate). Unlike Two-Way ANOVA, parameters for individual levels within the Block factor are not estimated since there is little interest in characterizing individual differences between them. Instead, observed Blocks are considered to be a sample from a much larger population of possible Blocks. Assuming this (and centered on the grand mean), variance between Blocks is the only parameter estimated. Shown here are examples of Randomized Blocks Without Replication drawn from Ch. 1 of Pinheiro & Bates (PB) 2004, *Mixed-Effects models in S and S-PLUS*, and from Example 12.4 in Jerrold H. Zar 2010, *Biostatistical Analysis 5th edition*. Inclusion of the latter example is designed to contrast traditional methods of handling this type of study described by Zar with newer methods involving maximum likelihood & REML estimation described by PB and here.

Example:

Ergostool Data from PB Section 1.2 p. 12.

**Each "Subject" tested each of four stool "Types" exactly once (without replication).
Ease of rising from each Type of stool is recorded in the response variable "effort".**

Loading Data for plotting in R:

```
#LMM 02 RANDOMIZED BLOCK DESIGNS
library(nlme)     # {nlme} for lme()
library(help=nlme) # prototype for finding package index

#READING DATA IN STANDARD FORMAT
setwd("c:/DATA/Models")
ergoStool=read.table("ergostool.txt")
ergoStool$Subject=factor(ergoStool$Subject)
ergoStool

#CREATING GROUPED DATA OBJECT:
ES=groupedData(effort~Type |Subject,data=ergoStool)
ES
```

^ PB provide their datasets as part of the R package {nlme} and after being loaded it may be accessed directly using in this case the command:

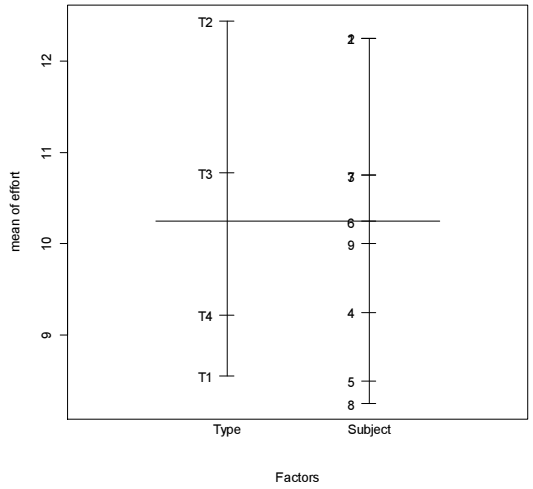
```
data(ergoStool)
```

Typically in this package data are formed within the "groupedData" data class. This class includes an embedded formula that may be utilized as the default formula by {nlme} functions such as lmList(), lme() & gls(), etc, as well as quite useful plot() routines employing Trellis graphics. PB use the embedded formula in groupedData as a means of simplifying formulas within commands, but unless great care is taken resultant partial formulas are very difficult to read and often confusing. On the other hand, the default plots utilizing groupedData are very useful.

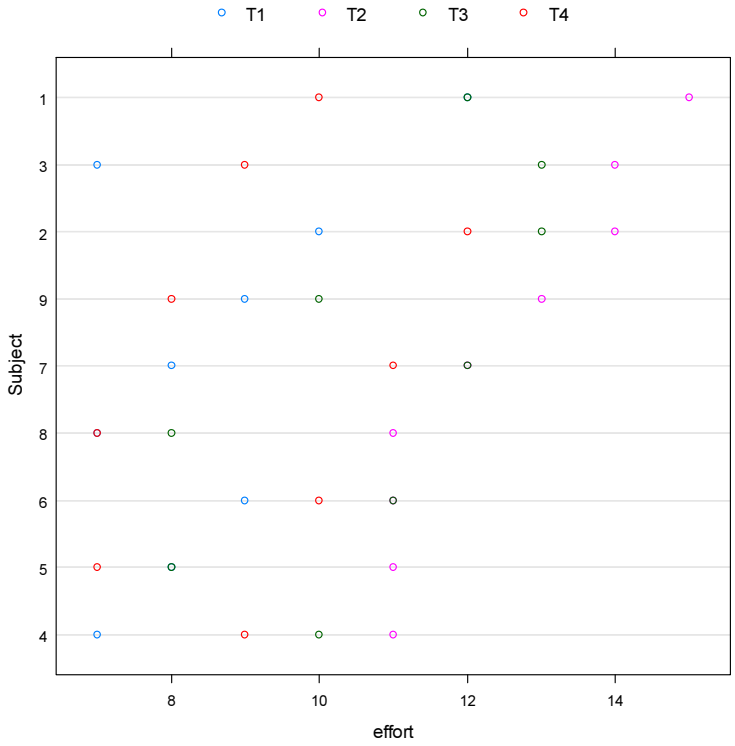
```
> ES
Grouped Data: effort ~ Type | Subject
   effort Type Subject
1      12   T1       1
2      15   T2       1
3      12   T3       1
4      10   T4       1
5      10   T1       2
6      14   T2       2
7      13   T3       2
8      12   T4       2
9       7   T1       3
10     14   T2       3
11     13   T3       3
12      9   T4       3
13      7   T1       4
14     11   T2       4
15     10   T3       4
16      9   T4       4
17      8   T1       5
18     11   T2       5
19      8   T3       5
20      7   T4       5
21      9   T1       6
22     11   T2       6
23     11   T3       6
24     10   T4       6
25      8   T1       7
26     12   T2       7
27     12   T3       7
28     11   T4       7
29      7   T1       8
30     11   T2       8
31      8   T3       8
32      7   T4       8
33      9   T1       9
34     13   T2       9
35     10   T3       9
36      8   T4       9
```

```
#DIAGNOSTIC PLOTS OF DATA
plot(ES)
plot.design(ES)
```

plot.design(ES)



plot(ES)



Linear Fixed Model ANOVA:

$Y_{ij} = \beta_i + \varepsilon_{ij}$ < Cell Means model

$Y_{ij} = \beta + \tau_i + \varepsilon_{ij}$ < Treatments Effect model

$\Sigma\tau_i = 0$, $\varepsilon_{ij} \sim N(0,\sigma^2)$

^ Observations for each Subject are seen on a separate horizontal line.

Y ~ X
effort ~ Type

where:
X = *fixed* treatment factor coded according to some contrast system

where: Y_{ij} is the response variable effort, β = overall mean (intercept), β_i = cell means in cell means model, τ_i = treatment effects for each fixed factor Type, ε_{ij} = error, with i as index of the fixed factor, j = is index of all observations within each group i up to n_i.

```
#SIMPLE LINEAR MODEL - "treatments" CONTRASTS IN R
LM1=lm(effort~Type,data=ergoStool)
summary(LM1)
anova(LM1)
```

```
#SIMPLE LINEAR MODEL - "cell means" CONTRASTS IN R
LM2=lm(effort~Type-1,data=ergoStool)
summary(LM2)
anova(LM2)
```

PB describe the use of different contrasts in coding factors with two or more levels in ANOVA. See also *Biostatistics* worksheet 390.

In R, "treatments" contrasts (the default) results in a summary() command showing "estimates" for a *baseline level* for each factor (otherwise missing from the report) called "intercept" followed by differences between the named levels of the factor and this baseline. The "cell means" contrasts gives means for each level of the factor, and all levels are named. For each line in the report, t-tests give probabilities for a null hypothesis that each "estimate" is zero. These tests are "marginal" tests in the since of comparing models with all other factors and levels still in the model. See *Biostatistics* 360, 400 & 402 for further details.

```
> summary(LM1)

Call:
lm(formula = effort ~ Type, data = ergoStool)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7778 -1.4444 -0.2222  1.2778  3.4444

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   8.5556     0.5760  14.853 6.56e-16 ***
TypeT2        3.8889     0.8146   4.774 3.83e-05 ***
TypeT3        2.2222     0.8146   2.728   0.0103 *
TypeT4        0.6667     0.8146   0.818   0.4192
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1

Residual standard error: 1.728 on 32 degrees of freedom
Multiple R-squared: 0.4594,     Adjusted R-squared:
0.4087
F-statistic: 9.064 on 3 and 32 DF,  p-value: 0.0001723
```

```
> anova(LM1)
Analysis of Variance Table

Response: effort
          Df Sum Sq Mean Sq F value    Pr(>F)
Type       3 81.194 27.0648  9.0636 0.0001723 ***
Residuals 32 95.556  2.9861
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.'
0.1 ' ' 1
```

```
> summary(LM2)

Call:
lm(formula = effort ~ Type - 1, data = ergoStool)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7778 -1.4444 -0.2222  1.2778  3.4444

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
TypeT1    8.556      0.576   14.85 6.56e-16 ***
TypeT2   12.444      0.576   21.60  < 2e-16 ***
TypeT3   10.778      0.576   18.71  < 2e-16 ***
TypeT4    9.222      0.576   16.01  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1

Residual standard error: 1.728 on 32 degrees of freedom
Multiple R-squared: 0.9759,     Adjusted R-squared:
0.9728
F-statistic: 323.5 on 4 and 32 DF,  p-value: < 2.2e-16
```

```
> anova(LM2)
Analysis of Variance Table

Response: effort
          Df Sum Sq Mean Sq F value    Pr(>F)
Type       4 3863.4  965.86  323.45 < 2.2e-16 ***
Residuals 32   95.6    2.99
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
' ' 1
```

The anova() summaries in R provide an F-test of significance of a model combining all levels of a factor. This is a serial or "Type 1" report following the order of the formula in lm(). See *Biostatistics* 400 for further description.

Linear Mixed Model - Random Block ANOVA:

$Y_{ij} = \beta_j + b_i + \varepsilon_{ij}$

where: Y_{ij} is the response variable effort, β_j = fixed cell means for different levels j of Treatment, b_i = random deviation from overall mean attributable to each Block (Subject), ε_{ij} = error, with i as index of the Block, and j is index of Treatment levels.

$b_i \sim N(0,\sigma_b^2)$, $\varepsilon_{ij} \sim N(0,\sigma^2)$

$y_i = X_i\beta + Z_i b_i + \varepsilon_i$

< Matrix formulation in terms of each Block i, PB p. 14, with X_i being the fixed contrasts, Z_i the matrix of random contrasts, and I the Identity matrix.

$\varepsilon_{ij} \sim N(0,\sigma^2 I)$

$Y \sim X + (1 \mid B)$

effort ~ Type + (1 | Subject)

< formula representation with Y the response variable, X the fixed Treatment variable coded as contrasts, and random component (1|B) showing Z as a vector of 1, with Block as the "group" variable.

```
#MIXED MODEL - "treatments" CONTRASTS IN R
LMe1=lme(effort~Type,random=~1|Subject, data=ergoStool)
summary(LMe1)
anova(LMe1)
coef(LMe1)
fitted(LMe1)

#MIXED MODEL - "cell means" CONTRASTS IN R
FMe2=lme(effort~Type-1,random=~1|Subject, data=ergoStool)
summary(FMe2)
anova(LMe)
```

> summary(LMe1)

```
Linear mixed-effects model fit by REML
 Data: ergoStool
       AIC      BIC    logLik
  133.1308 141.9252 -60.56539

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.332465 1.100295

Fixed effects: effort ~ Type
               Value Std.Error DF   t-value p-value
(Intercept) 8.555556 0.5760123 24 14.853079  0.0000
TypeT2      3.888889 0.5186838 24  7.497610  0.0000
TypeT3      2.222222 0.5186838 24  4.284348  0.0003
TypeT4      0.666667 0.5186838 24  1.285304  0.2110
 Correlation:
       (Intr) TypeT2 TypeT3
TypeT2 -0.45
TypeT3 -0.45   0.50
TypeT4 -0.45   0.50   0.50

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3
    Max
-1.80200345 -0.64316591  0.05783115  0.70099706
1.63142054

Number of Observations: 36
Number of Groups: 9
```

> anova(LMe)

```
            numDF denDF  F-value p-value
(Intercept)     1    24 455.0075  <.0001
Type            3    24  22.3556  <.0001
```

> summary(LMe2)

```
Linear mixed-effects model fit by REML
 Data: ergoStool
       AIC      BIC    logLik
  133.1308 141.9252 -60.56539

Random effects:
 Formula: ~1 | Subject
        (Intercept) Residual
StdDev:    1.332465 1.100295

Fixed effects: effort ~ Type - 1
           Value Std.Error DF  t-value p-value
TypeT1  8.555556 0.5760123 24 14.85308       0
TypeT2 12.444444 0.5760123 24 21.60448       0
TypeT3 10.777778 0.5760123 24 18.71102       0
TypeT4  9.222222 0.5760123 24 16.01046       0
 Correlation:
       TypeT1 TypeT2 TypeT3
TypeT2 0.595
TypeT3 0.595  0.595
TypeT4 0.595  0.595  0.595

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3
    Max
-1.80200345 -0.64316591  0.05783115  0.70099706
1.63142054

Number of Observations: 36
Number of Groups: 9
```

> anova(LMe2)

```
     numDF denDF  F-value p-value
Type     4    24 130.5186  <.0001
```

#COMPARISON OF MODELS
anova(LMe1,LM1)
GLS1=gls(effort~Type,data=ergoStool)
summary(GLS1)
anova(GLS1)
anova(GLS1,LMe1)

> anova(LMe1,LM1)

```
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
LMe1     1  6 133.1308 141.9252 -60.56539
LM1      2  5 144.6081 151.9367 -67.30403 1 vs 2 13.47728   2e-04
```

Comparisons of nested models, such as here allows one to determine whether fitting a random component (and therefore a mixed linear model) is desirable, or whether a straightforward linear model will suffice. These tests are handled by R's anova() command. Depending upon the context, either an F-ratio test (see *Biostatistics* 402) or log-Liklihood test (PB p. 83 & Worksheet LM 05) will be reported. High probability values indicates *failure* to reject the null hypothesis asserting that the *more parsimonious model* (also indicated by smaller AIC value) is sufficient. Here we see strong support for rejection of the null, indicating that the mixed linear model is to be preferred.

Comparisons of the results of lm() with lme() will work in anova() so long as the lme() model is specified first in the parenthesis. Alternatively the function gls() provides equivalent results to lm() in this context.

> anova(GLS1,LMe1)

```
     Model df      AIC      BIC    logLik   Test  L.Ratio p-value
GLS1     1  5 144.6081 151.9367 -67.30403
LMe1     2  6 133.1308 141.9252 -60.56539 1 vs 2 13.47728   2e-04
```

The gls() function in R's {nlme} package is an extension of lm() in {stats} (part of the base installation of R). Coefficient estimates and other values are equivalent to LM1 above. Both are coded with "treatments" contrasts.

```
> summary(GLS1)
Generalized least squares fit by REML
  Model: effort ~ Type
  Data: ergoStool
       AIC      BIC    logLik
  144.6081 151.9367 -67.30403

Coefficients:
             Value Std.Error   t-value p-value
(Intercept) 8.555556 0.5760123 14.853079  0.0000
TypeT2      3.888889 0.8146043  4.773960  0.0000
TypeT3      2.222222 0.8146043  2.727977  0.0103
TypeT4      0.666667 0.8146043  0.818393  0.4192

 Correlation:
       (Intr) TypeT2 TypeT3
TypeT2 -0.707
TypeT3 -0.707  0.500
TypeT4 -0.707  0.500  0.500

Standardized residuals:
       Min         Q1        Med         Q3        Max
-1.6074761 -0.8358876 -0.1285981  0.7394390  1.9932703

Residual standard error: 1.728037
Degrees of freedom: 36 total; 32 residual
> anova(GLS1)
Denom. DF: 32
            numDF   F-value p-value
(Intercept)     1 1266.6140  <.0001
Type            3    9.0636   2e-04
```

Zar's Example 12.4

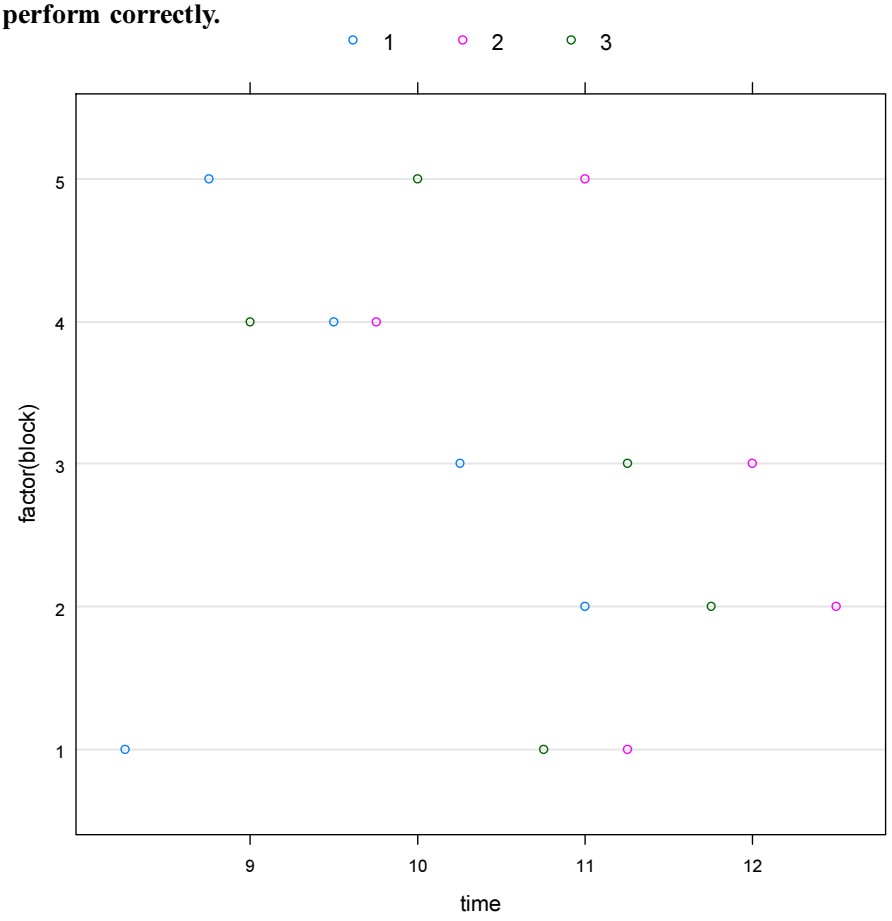
```
#ZAR'S EXAMPLE 12.4
setwd("c:/DATA/Models")
Z=read.table("ZarEX12.4R.txt")
Z
```

Zar Example 12.4:

	time	treatment	block
1	8.25	1	1
2	11	1	2
3	10.25	1	3
4	9.5	1	4
5	8.75	1	5
6	11.25	2	1
7	12.5	2	2
8	12	2	3
9	9.75	2	4
10	11	2	5
11	10.75	3	1
12	11.75	3	2
13	11.25	3	3
14	9	3	4
15	10	3	5

```
ZG=groupedData(time~factor(treatment)|factor(block),data=Z)
plot(ZG)
```

^ **Note: factors need to be correctly indicated in order for plot() to perform correctly.**




```
#TRADITIONAL LINEAR MODEL TEST OF TREATMENT EFFECT
anova(lm(time~factor(treatment)*factor(block),data=Z))
F=3.8542/0.3698
F
P=1-pf(F,2,8)
P
```

```
> anova(lm(time~factor(treatment)*factor(block),data=Z))
Analysis of Variance Table

Response: time
                                 Df  Sum Sq Mean Sq F value Pr(>F)
factor(treatment)                 2  7.7083  3.8542
factor(block)                     4 11.0667  2.7667
factor(treatment):factor(block)   8  2.9583  0.3698
Residuals                         0  0.0000
Warning message:
In anova.lm(lm(time ~ factor(treatment) * factor(block), data =
Z)) :
  ANOVA F-tests on an essentially perfect fit are unreliable

> F=3.8542/0.3698
> F
[1] 10.42239
> P=1-pf(F,2,8)
> P
[1] 0.005916856
```

Results here are identical to those in *Biostatistics* > Worksheet 300, shown in more detail.

< suggests significant treatment effect

```
#LINEAR MIXED MODEL TEST OF TREATMENT EFFECT
LMZ=lme(time~factor(treatment),random=~1|factor(block),data=Z)
summary(LMZ)
anova(LMZ)
```

Results are nearly identical here...

```
> summary(LMZ)
Linear mixed-effects model fit by REML
 Data: Z
       AIC      BIC    logLik
  44.99489 47.41942 -17.49744

Random effects:
 Formula: ~1 | factor(block)
        (Intercept) Residual
StdDev:   0.8938448 0.608105

Fixed effects: time ~ factor(treatment)
                   Value Std.Error DF   t-value p-value
(Intercept)         9.55 0.4834770  8 19.752749  0.0000
factor(treatment)2  1.75 0.3845993  8  4.550190  0.0019
factor(treatment)3  1.00 0.3845993  8  2.600108  0.0316
 Correlation:
                   (Intr) fct()2
factor(treatment)2 -0.398
factor(treatment)3 -0.398  0.500

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-1.5916707 -0.3264449  0.1538579  0.3770602  1.4136662

Number of Observations: 15
Number of Groups: 5
> anova(LMZ)
                   numDF denDF  F-value p-value
(Intercept)            1     8 593.9517  <.0001
factor(treatment)      2     8  10.4225  0.0059
```