Math 3080 § 1.	**Cap Data: Two-Way Mixed Effects**	Name: Example
Treibergs	**Data Only Gives Sums Per Cell**	April 11, 2010

Data File Used in this Analysis:

```
# Math 3081-1         Cap Data        March 10, 2010
# Treibergs
#
# Devore "Probability and Statistics for Engineering and the Sciences, 5th ed"
# An experiment to determine whether compressive strength of concrete cylinders
# is influenced by the capping material used, "The Effect of Type of Capping
# Material on Compressive Strength of Concrete Cylinders," (Proceedings ASTM,
# 1958) Each number is the sum of K=3 strength observations. For each batch
# (from a random population of batches) there are I=3 caps (fixed treatments).
# We are given that   \sum_{ijk} y_{ijk}^2 = 16815853
#
#
"Cap" "B1" "B2" "B3" "B4" "B5"
"C1" 1847 1942 1935 1891 1795
"C2" 1779 1850 1795 1785 1626
"C3" 1806 1892 1889 1891 1756
```

This is a mixed effects model: the treatment, the capping material for the concrete cylinders is a fixed effect and the concrete batch is a random effect. We will be interested whether the differences in capping material affect the compressive strength as well as whether there is significant variance due to the batch mixture.

We have $I = 3$ levels of the fixed factor $A =$ cap and $J = 5$ levels of the random factor $B =$ batch. For each cap and batch, there were $K = 3$ cylinders made and tested. Thus there are $K > 1$ replications per cell.

Note that the data given does not specify the individual values, only the cell sums. Thus, I have computed the ANOVA tables "by hand." I'll derive the relevant formulas and then give the **R** computation.

R Session:

```
R version 2.10.1 (2009-12-14)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
```

```
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[R.app GUI 1.31 (5538) powerpc-apple-darwin8.11.1]

[Workspace restored from /Users/andrejstreibergs/.RData]


> yy <- read.table("M3081DataCap.txt",header=TRUE)
Warning message:
In read.table("M3081DataCap.txt", header = TRUE) :
  incomplete final line found by readTableHeader on 'M3081DataCap.txt'
> yy <- read.table("M3081DataCap.txt",header=TRUE)
> yy
  Cap   B1   B2   B3   B4   B5
1  C1 1847 1942 1935 1891 1795
2  C2 1779 1850 1795 1785 1626
3  C3 1806 1892 1889 1891 1756
> attach(yy)

>#============CELL MEANS=================================================
> yy/3
  Cap       B1       B2       B3       B4       B5
1  NA 615.6667 647.3333 645.0000 630.3333 598.3333
2  NA 593.0000 616.6667 598.3333 595.0000 542.0000
3  NA 602.0000 630.6667 629.6667 630.3333 585.3333

> x <- c(B1,B2,B3,B4,B5)


> A<- factor(rep(yy[,1],times=5));A
 [1] C1 C2 C3 C1 C2 C3 C1 C2 C3 C1 C2 C3 C1 C2 C3
Levels: C1 C2 C3


> B <- factor(rep(names(yy)[2:6],each=3));B
 [1] B1 B1 B1 B2 B2 B2 B3 B3 B3 B4 B4 B4 B5 B5 B5
Levels: B1 B2 B3 B4 B5
```

The only place in the entire analysis where the individual replicates matter is in the computation of SST. However, they do provide

$$\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} {y_{ijk}}^2 = 16815853.$$

It follows that

$$SST = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} (y_{ijk} - \overline{y_{\cdots}})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}^2 - IJK\left(\frac{1}{IJK}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} y_{ijk}\right)^2 = 35954.31.$$

The sums $x_{ij} = \sum_{k=1}^{K} y_{ijk}$ are given in the table. Hence the grand sample mean is

$$\hat{\mu} = \overline{y_{\cdots}} = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \left(\sum_{k=1}^{K} y_{ijk} \right) = \frac{1}{K} \left(\frac{1}{IJ} \sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij} \right) = 610.6444.$$

The vector x includes all fifteen sums.

```
> I<-3;J<-5;K<-3;c(I,J,K)
[1] 3 5 3
> n<-I*J*K;n
[1] 45
> muhat <- mean(x)/3
> muhat
[1] 610.6444
> SST <- 16815853-muhat^2*n;SST
[1] 35954.31
```

To get the estimators for means, we take the factor means and divide by $K = 3$.

$$\hat{\mu}_{i\cdot\cdot} = \overline{y_{i\cdot\cdot}} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk} = \frac{1}{K} \left(\frac{1}{J} \sum_{j=1}^{J} \sum_{k=1}^{K} y_{ijk} \right) = \frac{1}{K} \left(\frac{1}{J} \sum_{j=1}^{J} x_{ij} \right).$$

Similarly

$$\hat{\mu}_{\cdot j\cdot} = \overline{y_{\cdot j\cdot}} = \frac{1}{IK} \sum_{i=1}^{I} \sum_{k=1}^{K} y_{ijk} = \frac{1}{K} \left(\frac{1}{I} \sum_{i=1}^{I} \sum_{k=1}^{K} y_{ijk} \right) = \frac{1}{K} \left(\frac{1}{I} \sum_{i=1}^{I} x_{ij} \right).$$

```
> muAhat <-tapply(x,A,mean)/3
> muAhat
      C1       C2       C3
627.3333 589.0000 615.6000
> muBhat <-tapply(x,B,mean)/3;muBhat
      B1       B2       B3       B4       B5
603.5556 631.5556 624.3333 618.5556 575.2222
```

To get SSA, we compute

$$SSA = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\overline{y_{i\cdot\cdot}} - \overline{y_{\cdots}} \right)^2 = JK \sum_{i=1}^{I} \left(\overline{y_{i\cdot\cdot}} - \overline{y_{\cdots}} \right)^2 = 11573.38.$$

Similarly

$$SSB = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\overline{y_{\cdot j\cdot}} - \overline{y_{\cdots}} \right)^2 = IK \sum_{i=1}^{I} \left(\overline{y_{\cdot j\cdot}} - \overline{y_{\cdots}} \right)^2 = 17930.09.$$

```
> muAhat-muhat
        C1         C2         C3
 16.688889 -21.644444   4.955556
> (muAhat-muhat)^2
       C1        C2        C3
278.51901 468.48198  24.55753

>  SSA <- J*K*sum((muAhat-muhat)^2);SSA
[1] 11573.38

>  SSB <- I*K*sum((muBhat-muhat)^2);SSB
[1] 17930.09
```

The *outer product* $\text{outer}(\overline{y_{i\cdot\cdot}}, \overline{y_{\cdot j\cdot}})$ takes a pair of vectors, say $\overline{y_{i\cdot\cdot}}$ and $\overline{y_{\cdot j\cdot}}$, views them as an $I \times 1$ matrix and a $1 \times J$ matrix and outputs the matrix product, an $I \times J$ matrix whose (i, j) coefficient is $\overline{y_{i\cdot\cdot}}\,\overline{y_{\cdot j\cdot}}$. Let $U_I = (1, 1, 1, \ldots, 1)$ be the vector of ones in I dimensions. Thus the $I \times J$ matrix

$$\overline{y_{\cdots}}\text{outer}(U_I, U_J) - \text{outer}(\overline{y_{i\cdot\cdot}}, U_J) - \text{outer}(U_I, \overline{y_{\cdot j\cdot}}) + \overline{y_{ij\cdot}}$$

has (i, j) coefficient is $\overline{y_{\cdots}} - \overline{y_{i\cdot\cdot}} - \overline{y_{\cdot j\cdot}} + \overline{y_{ij\cdot}}$.

To get $SSAB$, we compute

$$SSAB = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\left(\overline{y_{\cdots}} - \overline{y_{i\cdot\cdot}} - \overline{y_{\cdot j\cdot}} + \overline{y_{ij\cdot}}\right)^2 = K\sum_{i=1}^{I}\sum_{j=1}^{J}\left(\overline{y_{\cdots}} - \overline{y_{i\cdot\cdot}} - \overline{y_{\cdot j\cdot}} + \overline{y_{ij\cdot}}\right)^2 = 1734.178.$$

Then we find $SSE = SST - SSA - SSB - SSAB = 4716.667$.

```
> muAhat
      C1       C2       C3
627.3333 589.0000 615.6000

> -outer(muAhat,c(1,1,1,1,1))
        [,1]      [,2]      [,3]      [,4]      [,5]
C1 -627.3333 -627.3333 -627.3333 -627.3333 -627.3333
C2 -589.0000 -589.0000 -589.0000 -589.0000 -589.0000
C3 -615.6000 -615.6000 -615.6000 -615.6000 -615.6000

> -outer(c(1,1,1),muBhat)
            B1        B2        B3        B4        B5
[1,] -603.5556 -631.5556 -624.3333 -618.5556 -575.2222
[2,] -603.5556 -631.5556 -624.3333 -618.5556 -575.2222
[3,] -603.5556 -631.5556 -624.3333 -618.5556 -575.2222

> matrix(x/3,ncol=5)
         [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 615.6667 647.3333 645.0000 630.3333 598.3333
[2,] 593.0000 616.6667 598.3333 595.0000 542.0000
[3,] 602.0000 630.6667 629.6667 630.3333 585.3333
```

```
> w<- -outer(muAhat,c(1,1,1,1,1))-outer(c(1,1,1),muBhat)+matrix(x/3,ncol=5);w
        [,1]      [,2]      [,3]      [,4]      [,5]
C1 -615.2222 -611.5556 -606.6667 -615.5556 -604.2222
C2 -599.5556 -603.8889 -615.0000 -612.5556 -622.2222
C3 -617.1556 -616.4889 -610.2667 -603.8222 -605.4889

> w+muhat
        [,1]       [,2]       [,3]      [,4]       [,5]
C1 -4.577778 -0.9111111  3.9777778 -4.911111   6.422222
C2 11.088889  6.7555556 -4.3555556 -1.911111 -11.577778
C3 -6.511111 -5.8444444  0.3777778  6.822222   5.155556

> (w+muhat)^2
        [,1]       [,2]       [,3]      [,4]      [,5]
C1  20.95605  0.8301235 15.8227160 24.119012  41.24494
C2 122.96346 45.6375309 18.9708642  3.652346 134.04494
C3  42.39457 34.1575309  0.1427160 46.542716  26.57975


> SSAB<-sum((w+muhat)^2)*K
> SSAB
[1] 1734.178
> SSE=SST-SSA-SSB-SSAB;SSE
[1] 4716.667
```

We construct the ANOVA table. $MSA = SSA/df(A) = MSA/(I-1) = 5786.6889$ and so on. Because we have a mixed model, the F ratios have to be adjusted accordingly. Because A is a fixed effect, and B is a random effect, the underlying model is

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

where μ and α_i are constants with $\sum_{i=1}^{I} \alpha_i = 0$ and B_j, G_{ij} and ϵ_{ijk} are normally distributed random variables with expected value 0 and with variances σ_B^2, σ_G^2 and σ^2, respectively. The B_j, G_{ij} and ϵ_{ijk} are mutually independent except, since the I levels of factor A is of primary importance we also assume $\sum_i G_{ij} = 0$ which implies that for fixed j, the G_{ij} are not independent of one another but are negatively correlated. The expected sums of squares are

$$\begin{aligned}
\mathbf{E}(MSE) &= \sigma^2 \\
\mathbf{E}(MSA) &= \sigma^2 + \frac{IK}{I-1}\sigma_G^2 + \frac{JK}{I-1}\sum_{i=1}^{I}\alpha_i^2 \\
\mathbf{E}(MSB) &= \sigma^2 + IK\sigma_B^2 \\
\mathbf{E}(MSAB) &= \sigma^2 + \frac{IK}{I-1}\sigma_G^2.
\end{aligned}$$

One first tests the interaction hypothesis

$$\begin{aligned}
\mathcal{H}_{G0}: &\quad \sigma_G^2 = 0; \qquad \text{versus} \\
\mathcal{H}_{G1}: &\quad \sigma_G^2 > 0.
\end{aligned}$$

Under the null hypothesis, $F_G = MSAB/MSE \sim f_{(I-1)(J-1),IJ(K-1)}$. Thus the null hypothesis is rejected if $F_G > f_{(I-1)(J-1),IJ(K-1)}(\alpha)$. One does not test for the main effects if $\mathcal{H}_{G0}$ is rejected.

If $\mathcal{H}_{G0}$ cannot be rejected then we test the random factor

$$\begin{aligned}\mathcal{H}_{B0}: &\quad \sigma_B^2 = 0; \qquad \text{versus}\\ \mathcal{H}_{B1}: &\quad \sigma_B^2 > 0.\end{aligned}$$

Under the null hypothesis, $F_B = MSB/MSE \sim f_{J-1,IJ(K-1)}$. Thus the null hypothesis is rejected if $F_B > f_{J-1,IJ(K-1)}(\alpha)$. The test for the fixed factor in the mixed model is different

$$\begin{aligned}\mathcal{H}_{A0}: &\quad \alpha_1 = \cdots = \alpha_I = 0; \qquad \text{versus}\\ \mathcal{H}_{A1}: &\quad \alpha_i \neq 0 \qquad \text{for at least one } i.\end{aligned}$$

Under the null hypothesis, $F_A = MSA/MSAB \sim f_{J-1,(I-1)(J-1)}$. Thus the null hypothesis is rejected if $F_A > f_{J-1,(I-1)(J-1)}(\alpha)$. Using the denominator $MSAB$ detects whether it is the $\sum_{i=1}^I \alpha_i^2$ that is large.

```
> MSA<-SSA/2; MSB <- SSB/4; MSAB <- SSAB/8; MSE <- SSE/30

> ct<-c(2,4,8,30,n-1,SSA,SSB,SSAB,SSE,SST,MSA,MSB,MSAB,MSE,-1)

> ct2<-c(ct,MSA/MSAB,MSB/MSE,MSAB/MSE,-1,-1)

> ct3<-c(pf(MSA/MSAB,2,8,lower.tail=FALSE) ,
  pf(MSB/MSE,4,30,lower.tail=FALSE),pf(MSAB/MSE,8,30,lower.tail=FALSE),-1,-1 )

> alpha=.01

> cr<-function(z,n1,n2)qf(z,n1,n2,lower.tail=FALSE)
> ct4<-c(cr(alpha,2,8) , cr(alpha,4,30),cr(alpha,8,30),-1,-1 )

> matrix(c(ct2,ct3 ,ct4 ),ncol=6)
     [,1]      [,2]      [,3]      [,4]          [,5]      [,6]
[1,]    2 11573.378 5786.6889 26.694790  2.883907e-04  8.649111
[2,]    4 17930.089 4482.5222 28.510742  7.748481e-10  4.017877
[3,]    8  1734.178  216.7722  1.378763  2.456338e-01  3.172624
[4,]   30  4716.667  157.2222 -1.000000 -1.000000e+00 -1.000000
[5,]   44 35954.311   -1.0000 -1.000000 -1.000000e+00 -1.000000
>
```

Source	D.F.	Sum Sq.	Mean Sq.	F	$P(f > F)$	Critical F
A	2	11573.378	5786.6889	26.694790	2.883907e-04	8.649111
B	4	17930.089	4482.5222	28.510742	7.748481e-10	4.017877
Interaction	8	1734.178	216.7722	1.378763	2.456338e-01	3.172624
Error	30	4716.667	157.2222			
Total	44	35954.311				

Table 1: ANOVA table for this data.

The $F_G = 1.378763 < f_{8,30}(0.01) = 3.172624$ (P-value is 0.2456338) so we cannot reject $\mathcal{H}_G$. $F_A = 26.694790 > f_{2,8}(0.01) = 8.649111$ (P-value is 0.00029) so we reject $\mathcal{H}_A$ at the $\alpha = 0.01$ level: the data shows that the fixed effect A is significant. $F_B = 28.510742 > f_{4,30}(0.01) = 4.017877$ (P-value is $7.748481e - 10$) so we reject $\mathcal{H}_B$ at the $\alpha = 0.01$ level: the data shows that the there is significant variation in the random effect B.