ANOVA

Jonathan M. Lees
University of North Carolina, Chapel Hill
Department of Geological Sciences
CB #3315, Mitchell Hall
Chapel Hill, NC 27599-3315
email: jonathan.lees@unc.edu
ph: (919) 962-0695

November 5, 2008

1 ANOVA

To see how analysis of variance breaks down a matrix into its constiuent parts consider the following matrix,

```
> PEN = c(89, 88, 97, 94, 84, 77, 92, 79, 81, 87, 87, 85, 87, 92,
+     89, 84, 79, 81, 80, 88)
> pen = matrix(PEN, byrow = TRUE, ncol = 4)
> print(pen)
```

```
     [,1] [,2] [,3] [,4]
[1,]   89   88   97   94
[2,]   84   77   92   79
[3,]   81   87   87   85
[4,]   87   92   89   84
[5,]   79   81   80   88
```

```
> N = length(pen)
> n = dim(pen)[1]
> m = dim(pen)[2]
```

This is a 5 by 4 matrix which we can now decompose to see how the information is distributed through the breakdown of the matrix.

The grand average can be reconstructed by taking the gr

```
> ga = sum(pen)/N
> GA = matrix(rep(ga, length = N), ncol = 4)
> print(GA)
```

```
     [,1] [,2] [,3] [,4]
[1,]   86   86   86   86
[2,]   86   86   86   86
[3,]   86   86   86   86
[4,]   86   86   86   86
[5,]   86   86   86   86
```

likewise the block deviations and thre treatment deviations can be extracted,

```
> blox = matrix(rep(apply(pen - GA, 1, mean), m), byrow = FALSE,
+     ncol = 4)
> print(blox)
```

```
     [,1] [,2] [,3] [,4]
[1,]    6    6    6    6
[2,]   -3   -3   -3   -3
[3,]   -1   -1   -1   -1
[4,]    2    2    2    2
[5,]   -4   -4   -4   -4
```

```
> treatz = matrix(rep(apply(pen - GA, 2, mean), n), byrow = TRUE,
+     ncol = 4)
> print(treatz)
```

```
     [,1] [,2] [,3] [,4]
[1,]   -2   -1    3    0
[2,]   -2   -1    3    0
[3,]   -2   -1    3    0
[4,]   -2   -1    3    0
[5,]   -2   -1    3    0
```

Finally the residual is determined by subtracting the grand average, the block and treatment deviations. By definition the residuals are what is left over when the row and column means are extracted.

```
> res = pen - (GA + blox + treatz)
> print(res)
```

```
     [,1] [,2] [,3] [,4]
[1,]   -1   -3    2    2
[2,]    3   -5    6   -4
[3,]   -2    3   -1    0
[4,]    1    5   -2   -4
[5,]   -1    0   -5    6
```

```
> SUMS = c(sum(pen^2), sum(GA^2), sum(blox^2), sum(treatz^2), sum(res^2))
> print(SUMS)
```

```
[1] 148480 147920    264     70    226
```

```
> dof = c(20, 1, 4, 3, 12)
> print(dof)
```

```
[1] 20  1  4  3 12
```

This can be conveniently summarized in the following equation:

$$\begin{bmatrix} 89 & 88 & 97 & 94 \\ 84 & 77 & 92 & 79 \\ 81 & 87 & 87 & 85 \\ 87 & 92 & 89 & 84 \\ 79 & 81 & 80 & 88 \end{bmatrix} = \begin{bmatrix} 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \\ 86 & 86 & 86 & 86 \end{bmatrix} + \begin{bmatrix} 6 & 6 & 6 & 6 \\ -3 & -3 & -3 & -3 \\ -1 & -1 & -1 & -1 \\ 2 & 2 & 2 & 2 \\ -4 & -4 & -4 & -4 \end{bmatrix} + \begin{bmatrix} -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \\ -2 & -1 & 3 & 0 \end{bmatrix} + \begin{bmatrix} -1 & -3 & 2 & 2 \\ 3 & -5 & 6 & -4 \\ -2 & 3 & -1 & 0 \\ 1 & 5 & -2 & -4 \\ -1 & 0 & -5 & 6 \end{bmatrix} \tag{1.1}$$

where the vectors, sums of squares and degrees of freedom are summarized by,

$$\mathbf{Y} = \mathbf{A} + \mathbf{B} + \mathbf{T} + \mathbf{R} \tag{1.2}$$

The Sums of squares are summarized by,

$$\mathbf{S} = \mathbf{S_A} + \mathbf{S_B} + \mathbf{S_T} + \mathbf{S_R} \tag{1.3}$$

and the degrees of freedom are,

$$20 = 1 + 4 + 3 + 12 \tag{1.4}$$

Or, written a different way as a table,

matrix	Y=	A	+ B	+ T	+ R
SS	148480=	147920	+ 264	+ 70	+ 226
SS	S=	S_A	+ S_B	+ S_T	+ S_R
DOF	20 =	1	+ 4	+ 3	+ 12