STAT 5200 Handout #15b: EMS Calculations (enrichment only)

Example: (Same as Handout #15, $5 \times 3 \times 5$ design with fixed factors Tryptone and Temperature, and random factor Strain)

In mixed models, the expected values of the mean squares (EMS, or expected mean squares) are helpful when constructing F-ratio tests for model terms. If a certain ratio of EMS values equals one when the null hypothesis is true, then the corresponding ratio of MS values will be the F-ratio test statistic for the null hypothesis. The basic reason for this is that this ratio generally represents

(variability between factor levels) / (variability within factor levels),

and if those variabilities are roughly the same (so the ratio is one), then there is no real difference between factor levels (i.e., the null hypothesis is true).

The Cornfield-Tukey algorithm shows how to obtain the EMS for every term in a mixed model. For details on why this algorithm works, see Cornfield and Tukey (1956) "Average Values of Mean Squares in Factorials" 27(4):907-949. *Annals of Mathematical Statistics*.

Steps (but just see example on following page):

0. Write out ANOVA model, including subscripts. List each model term as a "Source" in a row of the table.
1. List each subscript (i, j, k, …) as a column header in new columns of the table, and indicate whether it is fixed or random, and how many levels it has.
2. In the elements (rows) within the subscript columns, for each subscript, mark a 1 if the Source for the row contains the subscript; otherwise mark the number of factor levels for the subscript.
3. Add a "Product" column to the table representing the product of the subscript column values.
4. Add a "Variance Component" column to the table, representing for each row (Source) either Q_{Source} (for fixed effects) or σ^2_{Source} (for random effects).
5. Define each Source's "contribution" as the Product times the Variance Component columns, and calculate the EMS for each Source as the sum of the "contributions" for the Source and for every term below it (and connected by line paths) in the Hasse Diagram.

(EMS Calculation, after first constructing Hasse Diagram)

Q(A*B)
Q(B, A*B)
Q(A, A*B)
} SAS notation

(0)	(1) *For each subscript, indicate fixed or random, and number of levels*					
	F-3	F-5	R-5		**Variance**	
Source	**i**	**J**	**k**	**Product**	**Component**	**EMS**
A_i	1	5	5	25	Q_A	$25Q_A + 5Q_{AB} + 5\sigma^2_{AC} + \sigma^2$
B_j	3	1	5	15	Q_B	$15Q_B + 5Q_{AB} + 3\sigma^2_{BC} + \sigma^2$
AB_{ij}	1	1	5	5	Q_{AB}	$5Q_{AB} + \sigma^2$
C_k	3	5	1	15	σ^2_C	$15\sigma^2_C + 5\sigma^2_{AC} + 3\sigma^2_{BC} + \sigma^2$
AC_{ik}	1	5	1	5	σ^2_{AC}	$5\sigma^2_{AC} + \sigma^2$
BC_{jk}	3	1	1	3	σ^2_{BC}	$3\sigma^2_{BC} + \sigma^2$
E_{ijk}	1	1	1	1	σ^2	σ^2
(2)	*1 if row contains subscript; number of levels otherwise*			(3) *across all subscript columns*	(4) *Q for fixed; σ^2 for random*	(5) *sum of "contribution" for Source and every term below it in the Hasse Diagram*
				Product(Variance Component) is the "contribution" of the term*		