

Comments on Problem 7-20

```
life = read.table("Problem-6-1.txt", header = TRUE);
lifeCoded = life;
for (j in 1:3) lifeCoded[,j] = coded(lifeCoded[,j]);
life$Block = lifeCoded$A * lifeCoded$B * lifeCoded$C
lifeCoded$Block = life$Block
summary(aov(Life ~ Rep + Block:Rep + A * B * C, lifeCoded));
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Rep	2	0.58	0.29	0.0084	0.9916576	
A	1	0.67	0.67	0.0192	0.8921998	
B	1	770.67	770.67	22.1509	0.0005086	***
C	1	280.17	280.17	8.0527	0.0149597	*
Rep:Block	3	92.75	30.92	0.8886	0.4748049	
A:B	1	16.67	16.67	0.4790	0.5020299	
A:C	1	468.17	468.17	13.4563	0.0032171	**
B:C	1	48.17	48.17	1.3844	0.2621609	
Residuals	12	417.50	34.79			

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

If we fit a reduced model with only the significant effects:

```
> summary(aov(Life ~ Rep + Block:Rep + B + C + A:C, life))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Rep	2	0.58	0.29	0.0085	0.9915750	
B	1	770.67	770.67	22.3690	0.0003226	***
C	1	280.17	280.17	8.1320	0.0128078	*
Rep:Block	3	92.75	30.92	0.8974	0.4669564	
C:A	2	468.83	234.42	6.8041	0.0086225	**
Residuals	14	482.33	34.45			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```
> summary(aov(Life ~ Rep + Block:Rep + B + C + A:C, lifeCoded))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Rep	2	0.58	0.29	0.0091	0.9909883	
B	1	770.67	770.67	23.9337	0.0001953	***
C	1	280.17	280.17	8.7008	0.0099391	**
Rep:Block	3	92.75	30.92	0.9601	0.4370014	
C:A	1	468.17	468.17	14.5393	0.0016973	**
Residuals	15	483.00	32.20			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Note the different degrees of freedom for C:A, and hence for the residuals:

- in the *non-coded* version, variables are symbolic, and each is represented by two dummy variables, A- and A+, etc.
- C:A is represented by the 4 products of the dummy variables for A and C;
- if A and C were both in the model, only 1 of these 4 would be included, but since A is *not* in the model, two are;
- so including C:A effectively includes A, even though it's not explicitly included;

- in the *coded* version, A and C are numeric, so $C:A$ is simply their product, and including it doesn't imply including A .
- Which is correct? The *coded* version fits precisely the specified model, so it more closely follows what was requested. But...
- Note: this model isn't hierarchical; we would usually include A in the model simply because the AC interaction is significant.
- For a hierarchical model, the difference between the coded and non-coded analyses goes away.