Topic 9: Factorial treatment structures (continued)

9.7.4.3 Another example of partitioning the Interaction SS

A factorial experiment was carried out to determine the effects of vernalization genes ***Vrn1*** and ***Vrn2*** on flowering time (days to flowering) in wheat. 102 plants from a segregating population (parents A and B) were characterized with molecular markers and the number of alleles of parent A indicated for each of the two genes (BB = 0, AB = 1, AA = 2). In the R code below, the auxiliary variable "**Type**" represents each combination of ***Vrn1*** and ***Vrn2*** classes.

The data (column format: Type Vrn1 Vrn2 Days):

```
1  0 0  89      1  0 0  97      1  0 0 101      1  0 0 100
1  0 0  98      2  0 1 133      2  0 1 144      2  0 1 148
2  0 1 148      2  0 1 138      2  0 1 130      2  0 1 133
2  0 1 128      2  0 1 130      2  0 1 137      2  0 1 141
2  0 1 134      2  0 1 133      2  0 1 138      2  0 1 131
2  0 1 148      3  0 2 163      3  0 2 153      3  0 2 161
3  0 2 153      3  0 2 156      3  0 2 148      4  1 0 109
4  1 0  83      4  1 0  87      4  1 0 103      4  1 0 110
4  1 0  81      4  1 0  99      4  1 0  98      4  1 0  83
4  1 0  78      4  1 0  92      4  1 0  92      4  1 0  91
4  1 0  85      4  1 0  83      4  1 0  66      5  1 1 122
5  1 1 121      5  1 1 121      5  1 1 122      5  1 1 125
5  1 1 118      5  1 1 123      5  1 1 124      5  1 1 125
5  1 1 108      5  1 1 112      5  1 1 126      5  1 1 118
5  1 1  98      5  1 1 116      5  1 1 106      5  1 1 117
5  1 1 110      5  1 1 113      5  1 1 129      5  1 1 116
6  1 2 140      6  1 2 125      6  1 2 178      6  1 2 136
6  1 2 132      6  1 2 133      6  1 2 135      6  1 2 134
6  1 2 125      6  1 2 125      6  1 2 128      6  1 2 121
6  1 2 128      6  1 2 135      7  2 0  91      7  2 0 103
7  2 0  81      7  2 0  99      7  2 0  88      7  2 0  99
7  2 0  73      8  2 1 137      8  2 1 118      8  2 1 120
8  2 1 153      8  2 1  86      8  2 1 114      8  2 1 126
8  2 1 120      8  2 1 120      8  2 1 118      8  2 1 119
8  2 1 106      8  2 1 112      8  2 1 111      8  2 1 117
9  2 2 124      9  2 2 124
```

The unbalanced nature of this dataset causes some problems, which we will get to in future lectures. For now, let's run the analysis as we've learned until now and take a look at the Vrn1*Vrn2 interaction:

```
#The ANOVA (the problem with unbalanced data)
#Notice how the order of factors in these two models affects the results)
vrn1_mod<-lm(Days ~ Vrn1 + Vrn2 + Vrn1*Vrn2, vrn_dat)
anova(vrn1_mod)

vrn2_mod<-lm(Days ~ Vrn2 + Vrn1 + Vrn1*Vrn2, vrn_dat)
anova(vrn2_mod)
```

Results of analysis as a 3x3 factorial experiment

```
           Df  Sum Sq Mean Sq  F value    Pr(>F)
Vrn1:Vrn2   4   807.9   202.0   1.8267    0.1303 <-- NS
Residuals  93 10282.4   110.6
```

Is there any significance lurking inside here? We can partition this interaction in various ways. One way is to keep both factors (Vrn1 and Vrn2) in the model and create contrasts for each:

```
#The same contrasts are used for factors Vrn1 and Vrn2
# Contrast 'Linear'          -1  0  1
# Contrast 'Quadratic'        1 -2  1
contrastmatrix<-cbind(c(-1,0,1),c(1,-2,1))
contrastmatrix

contrasts(vrn_dat$Vrn1)<-contrastmatrix
contrasts(vrn_dat$Vrn2)<-contrastmatrix
vrn_dat$Vrn1
vrn_dat$Vrn2

vrn_Rcon_mod<-lm(Days ~ Vrn1 + Vrn2 + Vrn1*Vrn2, vrn_dat)
summary(vrn_Rcon_mod)
```

The output:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.26415    1.34242  88.097  < 2e-16 ***
Vrn11        -9.45893    1.87007  -5.058 2.12e-06 ***
Vrn12         2.20747    0.79774   2.767 0.006822 **
Vrn21        22.67063    1.87398  12.098  < 2e-16 ***
Vrn22        -3.06971    0.79467  -3.863 0.000207 ***
Vrn11:Vrn21  -6.30952    2.64122  -2.389 0.018920 *  <- !!!
Vrn12:Vrn21   0.35317    1.08924   0.324 0.746484
Vrn11:Vrn22  -0.06488    1.08250  -0.060 0.952335
Vrn12:Vrn22  -0.59239    0.49082  -1.207 0.230512
```

The other way is to collapse the factorial structure into the single classification variable "Type" and use contrasts to dissect the sums of squares of the Type factor:

```
#The same analysis using the Type variable
type_mod<-lm(Days ~ Type, vrn_dat)
anova(type_mod)

#Type                        1    2    3    4    5    6    7    8    9
#Vrn1-Vrn2                  00   01   02   10   11   12   20   21   22
# Contrast 'Vrn1 Linear'    -1   -1   -1    0    0    0    1    1    1
# Contrast 'Vrn1 Quad'       1    1    1   -2   -2   -2    1    1    1
# Contrast 'Vrn2 Linear'    -1    0    1   -1    0    1   -1    0    1
# Contrast 'Vrn2 Quad'       1   -2    1    1   -2    1    1   -2    1
# Contrast '1Lin*2Lin'       1    0   -1    0    0    0   -1    0    1
# Contrast '1Lin*2Quad'     -1    2   -1    0    0    0    1   -2    1
# Contrast '1Quad*2Lin'     -1    0    1    2    0   -2   -1    0    1
# Contrast '1Quad*2Quad'     1   -2    1   -2    4   -2    1   -2    1

contrastmatrix2<-cbind(c(-1,-1,-1, 0, 0, 0, 1, 1, 1),
                       c(1, 1, 1,-2,-2,-2, 1, 1, 1),
                       c(-1, 0, 1,-1, 0, 1,-1, 0, 1),
                       c(1,-2, 1, 1,-2, 1, 1,-2, 1),
                       c(1, 0,-1, 0, 0, 0,-1, 0, 1),
                       c(-1, 2,-1, 0, 0, 0, 1,-2, 1),
                       c(-1, 0, 1, 2, 0,-2,-1, 0, 1),
                       c(1,-2, 1,-2, 4,-2, 1,-2, 1))
contrasts(vrn_dat$Type)<-contrastmatrix2

vrn_con2_mod<-lm(Days ~ Type, vrn_dat)
summary(vrn_con2_mod)
```

The output from this approach is exactly the same:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 118.26415    1.34242  88.097  < 2e-16 ***
Type1        -9.45893    1.87007  -5.058 2.12e-06 ***
Type2         2.20747    0.79774   2.767 0.006822 **
Type3        22.67063    1.87398  12.098  < 2e-16 ***
Type4        -3.06971    0.79467  -3.863 0.000207 ***
Type5        -6.30952    2.64122  -2.389 0.018920 *  <- !!!
Type6        -0.06488    1.08250  -0.060 0.952335
Type7         0.35317    1.08924   0.324 0.746484
Type8        -0.59239    0.49082  -1.207 0.230512
```

Note that even though the interaction in the 3x3 factorial is declared not significant, **the linear by linear component of that interaction *is* significant**. What does that mean? A look at the contrast coefficients for this interaction reveals the null hypothesis of this interaction component (Note: Here, "Type i" indicates the mean of treatment level i):

Type 1 - Type 3 - Type 7 + Type 9 = 0

Plugging in:

(No *Vrn1* or *Vrn2*) – (Full *Vrn1* but no *Vrn2*) – (Full *Vrn2* but no *Vrn1*) + (Full *Vrn1* and *Vrn2*) = 0

This can be rewritten in two ways:

(No *Vrn1* or *Vrn2*) – (Full *Vrn1* but no *Vrn2*) = (Full *Vrn2* but no *Vrn1*) - (Full *Vrn1* and *Vrn2*)

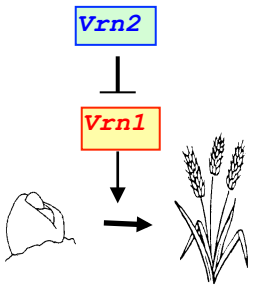
i.e. The effect of *Vrn1* in the absence of *Vrn2* = The effect of *Vrn1* in the presence of *Vrn2*

OR

(No *Vrn1* or *Vrn2*) – (Full *Vrn2* but no *Vrn1*) = (Full *Vrn1* but no *Vrn2*) - (Full *Vrn1* and *Vrn2*)

i.e. The effect of *Vrn2* in the absence of *Vrn1* = The effect of *Vrn2* in the presence of *Vrn1*

So, a linear x linear interaction is quite understandable. It means that the effect of one factor (full dose compared to no dose) is different depending on the level (full dose vs. no dose) of the other factor. Vrn2, in fact, is a repressor of Vrn1 (see schematic to the right); and the detection of this interaction opened the way for a great deal of interesting research.



9.7.4.4. Example of a factorial experiment with subsamples

Assume that in the quack-grass shoots experiment (9.7.4.1), two random areas of 1 square foot each were evaluated in each plot (each R-D combination). In the dataset below, the values for these two subsamples (one example of such a pair are highlighted in blue) were created to give an average identical to the values in the previous exercise.

The data (column format: D R Block Plot Number):

```
 3 0 1 1 14.7  3 4 1 9   8.8  3 8 1 17  6.9    3 0 1 1 16.7  3 4 1 9  10.8  3 8 1 17  8.9
 3 0 2 2 13.6  3 4 2 10 13.6  3 8 2 18  9.3    3 0 2 2 15.6  3 4 2 10 15.6  3 8 2 18 11.3
 3 0 3 3 15.5  3 4 3 11 10.9  3 8 3 19  8.7    3 0 3 3 17.5  3 4 3 11 12.9  3 8 3 19 10.7
 3 0 4 4 13.7  3 4 4 12 11.4  3 8 4 20  8.6    3 0 4 4 15.7  3 4 4 12 13.4  3 8 4 20 10.6
10 0 1 5 17.0 10 4 1 13 12.6 10 8 1 21  7.8   10 0 1 5 19.0 10 4 1 13 14.6 10 8 1 21  9.8
10 0 2 6 16.4 10 4 2 14  9.6 10 8 2 22  7.2   10 0 2 6 18.4 10 4 2 14 11.6 10 8 2 22  9.2
10 0 3 7 14.1 10 4 3 15 10.8 10 8 3 23 10.3   10 0 3 7 16.1 10 4 3 15 12.8 10 8 3 23 12.3
10 0 4 8 13.4 10 4 4 16 12.3 10 8 4 24 10.2   10 0 4 8 15.4 10 4 4 16 14.3 10 8 4 24 12.2
```

Because we have subsamples, the experimental unit (Plot) needs to enter our linear model. Each plot is located within a specific D-R-Block combination; so the proper way to code for this is as follows:

```
#The ANOVA
quack1_mod<-lm(Number ~ D + R + D*R + Block + (D:R:Block)/Plot, quack_dat)
anova(quack1_mod)
```

Notice when you do this that R drops "/Plot" from the output, reporting the sums of squares for the D:R:Block interaction:

```
           Df  Sum Sq Mean Sq F value    Pr(>F)
D           1   3.000   3.000  1.5000   0.23256
R           2 307.327 153.663 76.8317 3.693e-11 ***
Block       3   1.163   0.388  0.1939   0.89952
D:R         2   0.980   0.490  0.2450   0.78464
D:R:Block  15  78.767   5.251  2.6256   0.01698 *
Residuals  24  48.000   2.000
```

The reason R does this is because it is this interaction that is used to estimate the differences among experimental units within specific D-R-Block combinations (we have no true replication for this). Thus it is also correct to code things this way:

```
quack2_mod<-lm(Number ~ D + R + D*R + Block + D:R:Block, quack_dat)
anova(quack2_mod)
```

It gives you the same results.

Either way you code it, the results are wrong for the majority of the ANOVA table because R automatically uses the subsampling error, unspecified in the model, as the denominator for all F tests. You must specify the correct error (**D:R:Block**) for testing the three main effects (D, R, and Block) as well as the interaction (D:R). Carrying out these custom F tests produces the results in the table below. Notice that these agree perfectly with the results of this analysis from before (page 9).

```
           Df  Sum Sq Mean Sq F value    Pr(>F)
D           1   3.000   3.000  0.5713   0.46144
R           2 307.327 153.663 29.2636 6.642e-06 ***
Block       3   1.163   0.388  0.0739   0.97309
D:R         2   0.980   0.490  0.0933   0.91143
D:R:Block  15  78.767   5.251  2.6256   0.01698 *
Residuals  24  48.000   2.000
```

If you are confused by this analysis, you are not alone. Simply average the subsamples and analyze the experiment as a non-nested RCBD; you will get the correct results. The objective of analyzing the experiment with the individual subsample values is simply to understand the sources of variation in the experimental error. This information can be used later to optimize the distribution of resources between the number of samples and subsamples, as you've seen before. To find the variance components of this experiment:

```
#library(lme4)
quackCV_mod<-lmer(Number ~ D + R + D*R + Block + (1|Block:D:R), quack_dat)
summary(quackCV_mod)
```

The output:

```
Random effects:
 Groups     Name        Variance
 Block:D:R (Intercept) 1.626    <-- variation among reps
 Residual              2.000    <-- variation b/w subsamples
```

This output shows the relative contribution of each component in the model to the total variance. In this case, the major component is the significant R factor. Within the error term, the variance between subsamples is similar to the variance among replications.

If one plot (EU) cost \$50 and one subsample cost \$5…

$$n_{sub} = \sqrt{\frac{C_{e.u.} * s^2_{sub}}{C_{sub} * s^2_{e.u.}}} = \sqrt{\frac{50 * 2.00}{5 * 1.63}} = 3.5$$

…the optimum allocation of resources would be to take more than 3 subsamples per plot (i.e. 4).

9.7.5 Two-way factorial within a CRD, one replication per cell

Analogous to the RCBD case with one replication per block-treatment combination, when there is only one observation per factor-factor combination, there is no source of variation to estimate the true experimental error. Due to a lack of degrees of freedom, the interaction effect cannot be partitioned from the error. As in the RCBD case, Tukey's 1 df test for nonadditivity can be used to probe the significance of that interaction, though only in an approximate way. In the end, because the interaction cannot be removed from the error, all tests for significance assume additivity of the factor (i.e. no interaction). In the following code, only the first block from the maleic hydrazide example is used as an example of a CRD:

The data:

D	R	Number
3	0	15.7
10	0	18
3	4	9.8
10	4	13.6
3	8	7.9
10	8	8.8

```
#The ANOVA
quack_mod<-lm(Number ~ D + R, quack_dat)
anova(quack_mod)
```

The output:

```
          Df Sum Sq Mean Sq F value  Pr(>F)
D          1  8.167   8.167  7.7655 0.10826
R          2 73.330  36.665 34.8637 0.02788 *
Residuals  2  2.103   1.052
```

Note that the error (2 df) is estimated by the interaction. If the interaction is non significant, SSE and SS(A*B) estimate the same quantity and the conclusions are valid.

So the situation is very similar to an RCBD with one rep per cell. The difference is one of intention. *The primary objective of a factorial experiment is to investigate potential interactions among factors of interest. For that reason, it makes very little sense to conduct a factorial experiment with only one replication per factor-factor combination.* In an RCBD, blocks are merely an error control strategy and the assumption is that their effect on the response variable is independent of the treatment; it is this assumption that justifies a design with a single rep per cell.

9.7.6. 2x2 CRD with a significant interaction (ST&D, p. 358)

The interpretation of factorial experiments is often complicated when the interactions are large. This is especially true if the effects change direction, as they do in this example. In this CRD with five replications per treatment combination, Factor A is time of bleeding of a lamb (AM vs. PM) and Factor B is treatment with the synthetic estrogen diethylstilbestrol (no DES vs. DES). The response variable is level of plasma phospholipid in the blood. A data summary:

Factor	A			
B	**AM**	**PM**	**Means**	**Effects**
No DES	13.28	36.53	24.905	0.660
DES	19.36	27.81	23.585	-0.660
Means	16.32	32.17	24.245	
Effects	-7.925	7.925		

The data (column format: ID Time DES Phos):

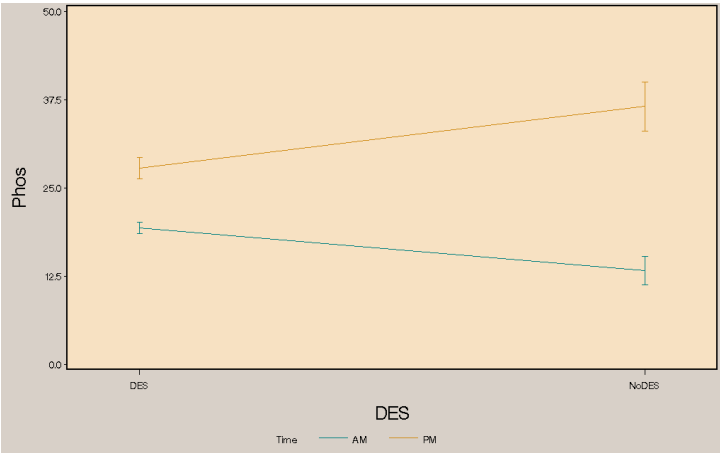
```
1 AM NoDES  8.53  2 AM DES 17.53  3 PM NoDES 39.14  4 PM DES 32.00
1 AM NoDES 20.53  2 AM DES 21.07  3 PM NoDES 26.20  4 PM DES 23.80
1 AM NoDES 12.53  2 AM DES 20.80  3 PM NoDES 31.33  4 PM DES 28.87
1 AM NoDES 14.00  2 AM DES 17.33  3 PM NoDES 45.80  4 PM DES 25.06
1 AM NoDES 10.80  2 AM DES 20.07  3 PM NoDES 40.20  4 PM DES 29.33
```

```
#The ANOVA
phos_mod<-lm(Phos ~ Time*DES, phos_dat)
anova(phos_mod)
```

The output:

```
          Df  Sum Sq Mean Sq F value    Pr(>F)
Time       1 1256.75 1256.75 52.9263 1.859e-06 ***
DES        1    8.71    8.71  0.3669  0.553198
Time:DES   1  273.95  273.95 11.5370  0.003686 **
Residuals 16  379.92   23.75
```

The interaction is significant, which means that the simple effects are heterogeneous. It make no sense to talk about the main (or general) effect of DES on phospholipid levels unless you specify a time of day. Conversely, it makes no sense to talk about the mean effect of time of day on phospholipid levels unless you specify a DES treatment level. The non-parallel nature of the lines in the interaction plot on the next page illustrates this interaction.



If an interaction is present in a fixed-effects model, the next step is to analyze the simple effects.

```
#Analyze the simple effects of Time by subsetting the data...
phos_AM_dat<-subset(phos_dat, phos_dat$Time == "AM")
phos_PM_dat<-subset(phos_dat, phos_dat$Time == "PM")

#...and then performing multiple ANOVAs
anova(lm(Phos ~ DES, phos_AM_dat))
anova(lm(Phos ~ DES, phos_PM_dat))

#Analyze the simple effects of DES by subsetting the data...
phos_DES_dat<-subset(phos_dat, phos_dat$DES == "DES")
phos_noDES_dat<-subset(phos_dat, phos_dat$DES == "NoDES")

#...and then performing multiple ANOVAs
anova(lm(Phos ~ Time, phos_DES_dat))
anova(lm(Phos ~ Time, phos_noDES_dat))
```

What the above code accomplishes are *four* separate ANOVAs. The first half breaks the data into two separate experiments, one testing the effect of DES in the morning and one testing the effect of DES in the evening. The second half breaks the data into two separate experiments, one testing the effect of time in the absence of DES and one testing the effect of time in the presence of DES. A summary of these results:

```
Response: Phos
                  Df  Sum Sq Mean Sq F value    Pr(>F)
DES (in AM)        1  92.477  92.477  7.7607   0.02371 *
DES (in PM)        1  190.18 190.183  5.3461   0.04952 *
Time (DES)         1 178.591 178.591  24.802  0.001079 **
Time (no DES)      1 1352.10 1352.10  33.559 0.0004084 ***
```

The significant effect of DES is now seen. It was hidden in the overall ANOVA ($p = 0.5532$) because the simple effects of DES are in opposite directions, depending on the time of day. These opposite effects cancelled one another out in the analysis of main effects. Note: For *each* of these four ANOVAs, you must test all assumptions.

9.8. Three-way ANOVA (Model I, fixed-effects model)

There is no reason to restrict the a factorial experiment to a consideration of only two factors. Three or more factors may be analyzed simultaneously, each at different levels. However, as the number of factors increases, even without replication within subgroups, the required number of experimental units becomes very large. It is frequently prohibitive in terms of resources to carry out such experiments. For example, a 4x4x4 factorial requires 64 experimental units to represent each combination of factors. Moreover, if only 64 EU's are used, there will not be sufficient replication to estimate the true experimental error. In this case, the three-way interaction (A:B:C) would have to be used as an estimate of experimental error; and this replies on the assumption no significant three-way interaction effect is present.

There are also logistical difficulties with such large experiments. It may not be possible to run all the tests on the same day or to hold all materials in a single controlled environmental chamber. If this is case, treatment effects may become confounded with one another or irrelevant sources of variation, either of which, ultimately, may reduce the power of the tests.

The third problem that accompanies a factorial ANOVA with several main effects is the large number of possible interactions. A two-way ANOVA has only one interaction, A:B. A three-way factorial has three **first-order (or two-way) interactions** (A:B, A:C, and B:C) and one **second-order (or three-way) interaction** (A:B:C).

The linear model for a three-way factorial is:

$$\mu_{ijkl} = \mu + \tau_{Ai} + \tau_{Bj} + \tau_{Ck} + (\tau_A\tau_B)_{ij} + (\tau_A\tau_C)_{ik} + (\tau_B\tau_C)_{jk} + (\tau_A\tau_B\tau_C)_{ijk} + \varepsilon_{ijklm}$$

A four-way factorial has 6 first-order interactions, four second-order interactions, and one **third-order (or four-way) interaction** (A*B*C*D). The number of possible interactions increases rapidly as the number of factors increases. The testing of their significance, and more importantly, their *interpretation* becomes complex.

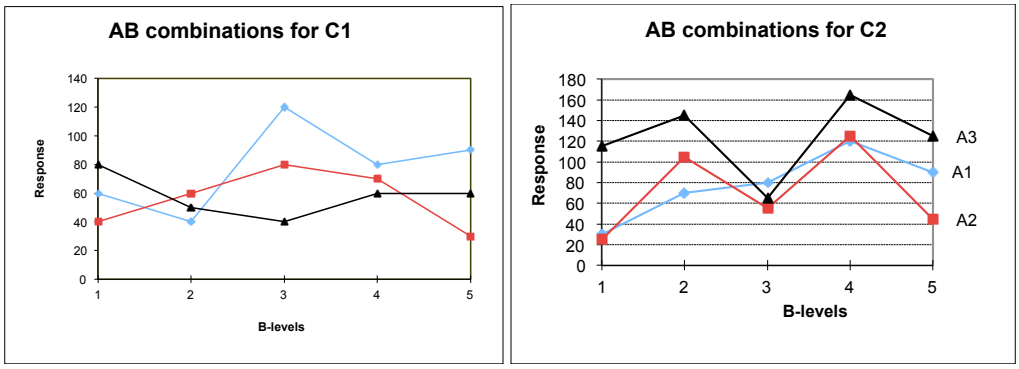
9.8.1. Example of a three-way factorial ANOVA

C.J. Monlezun (1979) Two-dimensional plots for interpreting interactions in the three-factor analysis of variance model. *The American Statistician* **33**: 63-69

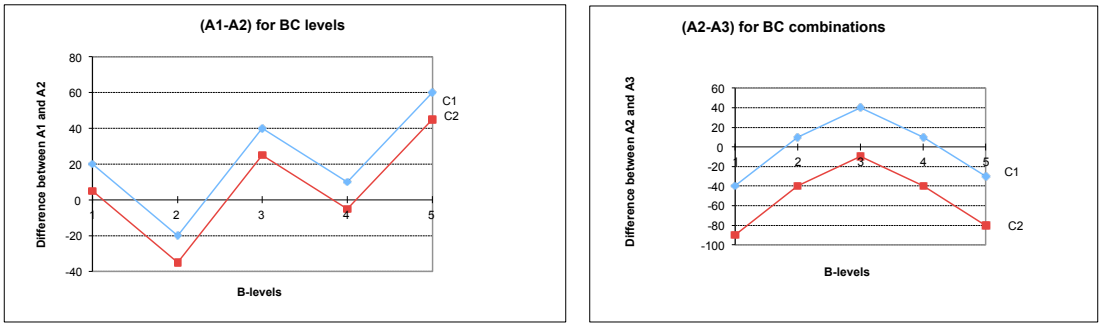
The following group means from a hypothetical 3x5x2 experiment are used to illustrate an example with **no three-way interactions**. A graphical technique to show the non-significant three-way interactions is discussed.

	A1C1	A2C1	A3C1	A1C2	A2C2	A3C2
B1	61	38	81	31	27	113
B2	39	61	49	68	103	143
B3	121	82	41	78	57	63
B4	79	68	59	122	127	167
B5	91	31	61	92	43	128

The lines of mean values for fixed C1 (below, left) and C2 (right) levels are not parallel. This indicates the presence of a two-way interaction between factors A and B at both levels of factor C. The ***first order interaction*** (A:B) now has two values: (A:B at C1) and (A:B at C2). The interaction term (A:B) is the average of these values.



If, however, **the differences between different levels of A** are determined for all combinations of B and C, the plot of these differences reveals no interaction between B and C. The lack of B*C interaction when the response variable is the *differences* between A levels indicates that no significant A*B*C interaction is present (i.e. $(\alpha\beta\gamma)_{ijk} = 0$). A graphical check of whether $(\alpha\beta\gamma)_{ijk} = 0$ is satisfied in the general situation would require (a-1) different graphs, like those below:



Phrasing these results in words, we can say that the factors A and B *interact in the same way* at all levels of factor C.

The interpretation of a three-factor interaction is that the effect of factor A depends on the precise combination of factors B and C. Take, for example, A to be nitrogen level and B to be plow depth. In a simple two-factor experiment, a significant A:B interaction indicates that the crop has a different response to N depending on plow depth. Now introduce a third factor C, which could be soil type. A nonzero three-way interaction (A:B:C) means that the effect of plow depth on the crop's response to nitrogen varies, depending on the soil type.