Calculating main effects in a factorial ANOVA on paper

Mathematically, SS_A and SS_B are also calculated by quite similar (unfortunately a little more complex) formulas as those used in the one way ANOVA:

$$SS_A = \sum_{r=1}^{R} N \cdot C \cdot (\bar{Y}_{r.} - \bar{Y}_{..})^2 \qquad SS_B = \sum_{c=1}^{C} N \cdot R \cdot (\bar{Y}_{.c} - \bar{Y}_{..})^2$$

We start with the formulas for the first (A; *drug*) and the second factor(B; *therapy*). Even though the formulas look complicated, we will see that they are quite easy to calculate. If you put it in some concentration, you might even manage to do it on paper. Otherwise, the file Clinicaltrial – Step-by-step.ods contains a spreadsheet ("Factorial ANOVA") with the calculations. These calculations are on the right side under the heading "Main effects".

Most of the values we need are in the list with the means of our six conditions in the top of p. 363 of the lsj-book. I decided to go with those. In principle, we could also have taken those in the table in the middle of p. 363 but these are rounded to two decimals, so we get to far off when using those. From the list of the six conditions, we calculate the means for the levels of factor A (drug) and factor B (therapy). I will use the nomenclature from the formulas above.

$\bar{Y}_{1.} = (0.300 + 0.600) / 2 = 0.450$

$\bar{Y}_{2.} = (0.400 + 1.033) / 2 = 0.717$

$\bar{Y}_{3.} = (1.466 + 1.500) / 2 = 1.483$

$\bar{Y}_{.1} = (0.300 + 0.400 + 1.466) / 3 = 0.722$

$\bar{Y}_{.2} = (0.600 + 1.033 + 1.500) / 3 = 1.044$

$\bar{Y}_{..} = (0.300 + 0.400 + 1.466 + 0.600 + 1.033 + 1.500) / 6 = 0.883$

Instead of the last one, you could as well have calculated the mean of the whole data column (*mood.gain*) in Excel. To make it a little easier for us, we summarize these values in a table.

	$\bar{\mathbf{Y}}_{.1}$	$\bar{\mathbf{Y}}_{.2}$	$\bar{\mathbf{Y}}_{..}$
$\bar{\mathbf{Y}}_{1.}$	0.300	0.600	0.450
$\bar{\mathbf{Y}}_{2.}$	0.400	1.033	0.717
$\bar{\mathbf{Y}}_{3.}$	1.467	1.500	1.483
$\bar{\mathbf{Y}}_{..}$	0.722	1.044	0.883

We know the we in addition to the calculations that tell us how much variance our factors explain (SS_A and SS_B), the amount of variance that is left over. In comparison to the first by-hand-calculation we did for the One way-ANOVA, we know a couple of "shortcuts". We know that the residuals are the total variance minus the variance explained by our factors (*drug* and *therapy*). We can therefore calculate according to this formula: $SS_R = SS_T – (SS_A + SS_B)$

SS_T we can simply get by using the variance of the data in *mood.gain* multiplied with the number of participants (ensure that you use VAR.P in LibreOffice Calc / Excel which doesn't apply the N – 1 correction for the variance in the sample). We could as well "cheat" a little and take it from the first calculation (the total variance is due to the data in *mood.gain* and therefore doesn't change):

$SS_T = \sigma^2 \cdot N$

$SS_T = 4.845$

As I already said, SS_A and SS_B are much easier to calculate than the formulas make believe. We have the advantage that we have a balanced design (i.e., all cells have N = 3) and therefore N = 3 mulitplied by two columns (no therapy, CBT) is 6. Otherwise, we would have had to fiddle around with determining how many participants are in each row or column. Now, we have for each of the levels of *drug* (placebo, anxifree, joyzepam to calculate the difference from the mean and square it. The results from that calculation are mulitplied by 6 ($N \cdot C$) and then summed up.

$SS_A = 3 \cdot 2 \cdot (0.450 – 0.883)^2 + 3 \cdot 2 \cdot (0.717 – 0.883)^2 + 3 \cdot 2 \cdot (1.483 – 0.883)^2$

$SS_A = 6 \cdot -0.433^2 + 6 \cdot -0.166^2 + 6 \cdot 0.600^2$
$SS_A = 1.125 + 0.165 + 2.160$
$SS_A = 3.450$
$df_A = 2$
$MS_A = (3.450 / 2) = 1.725$

We end up with 3.450 (which is reasonably close to the 3.453 that jamovi calculated). We have three levels for drug and therefore the degree of freedom are 3 – 1 = 2. We use that to calculate the mean square, which is 1.725.

The calculation of SS_B and MS_B is even simpler (given that we only have two levels for therapy). N = 3 is multiplied by the number of rows (i.e., the three levels for *drug*). We get 9 as result here. We calculate the difference of the means for our two levels from the total mean and square them. When multiplying that result with the 9 ($N \cdot R$) and summing it up, we get 0.466. The degrees of freedom for two levels are 2 – 1 = 1. Therefor, SS_B and MS_B are the same.
$SS_B = 3 \cdot 3 \cdot (0.722 - 0.883)^2 + 3 \cdot 3 \cdot (1.044 - 0.883)^2$
$SS_B = 9 \cdot -0.161^2 + 9 \cdot 0.161^2$
$SS_B = 0.233 + 0.233$
$SS_B = 0.466$
$df_B = 1$
$MS_B = (0.466 / 1) = 0.466$

I promised the calculation of SS_A and SS_B wouldn't be to tricky and here we are.
$SS_R = SS_T - (SS_A + SS_B)$
$SS_R = 4.845 - (3.450 + 0.466)$
$SS_R = 0.929$
$df_R = N - df_A - df_B - 1$
$df_R = 14$
$MS_R = (0.929 / 14) = 0.066$

To calculate the last part that we need, the sum of squares of the residuals (SS_R) we subtract the sum of SS_A and SS_B from the total sum of squares (SS_T). We get 0.929 as a result. The degrees of freedom are the total number of participants (18) minus the degrees of freedom for A (2), the degrees of freedom for B (1) and on other degree of freedom (for the total mean). We end up with df_R 14, which we can use to calculate the mean square which is 0.066. Now we are set to calculate the F-values for factor A (*drug*) and B (*therapy*). Those come pretty close to what we got from jamovi (some decimals deviation due to rounding errors).

$F_A = MS_A / MS_R = 1.725 / 0.066$
$F_A = 26.136$

$F_B = MS_B / MS_R = 0.466 / 0.066$
$F_B = 7.060$

In the Excel-sheet there are no rounding errors and the results are identical to what we get from jamovi.
So, if you got some spare time on your hands and don't fancy the entertainment the internet provides, calculating ANOVAs on paper might be a save you from dying of boredom… In memoriam Sir Ronald A. Fisher, you should not forget to do it in proper style with some Earl Grey (and don't forget to add milk). One last word: It took me an hour or two to figure out the calculations. I still believe it was worth it as I understood a couple of things through the mistakes I made on the way (mostly due to thinking to complicated in the beginning).

Calculating interaction effects in an factorial ANOVA on paper

To calculate the sum of squares for the interaction, we unfortunately need a whole zoo of variables. Fortunately, we calculated all of these before and produced a nice table which we can just copy.

	$\overline{\mathbf{Y}}_{.1}$	$\overline{\mathbf{Y}}_{.2}$	$\overline{\mathbf{Y}}_{..}$
$\overline{\mathbf{Y}}_{1.}$	0.300	0.600	0.450
$\overline{\mathbf{Y}}_{2.}$	0.400	1.033	0.717
$\overline{\mathbf{Y}}_{3.}$	1.467	1.500	1.483
$\overline{\mathbf{Y}}_{..}$	0.722	1.044	0.883

Again, with a bit of concentration, you will be able to conduct these calculations on paper. If you prefer not to do it yourself, you can use the file Clinicaltrial – Step-by-step.ods where you find them on the same spreadsheet ("Factorial ANOVA") as the calculations above. The current calculations are on the right side under the heading "Main effects + interaction".

Our null hypotheses for the two main effects were that the means for the different levels of the factor were equal. An easier way of saying that would be to assume that they are not different from the mean of the whole sample. For simplifying denoting factor A and factor B we will use greek letters. Thus, $\alpha_r = \mu_{r.} - \mu_{..}$ (or $\overline{Y}_{r.} - \overline{Y}_{..}$ when considering our sample) and $\beta_c = \mu_{.c} - \mu_{..}$ (or $\overline{Y}_{.c} - \overline{Y}_{..}$). If we "unwind" this, it means that the effect of α results from the difference between the mean in the right-most column of each row and the total mean in the bottom right corner; likewise, results the effect for β from the difference between the mean for each column in the bottom row and the total mean in the bottom right corner.
Our null hypothesis for the interaction would be that those two effects already accounted for the total mean plus the mean at each level of the factors: $\mu_{rc} = \mu_{..} + \alpha_r + \beta_c$. The alternative hypothesis is that there is some variance remaining to be explained after we took care of the total mean plus the mean at each level of the factors: $\mu_{rc} = \mu_{..} + \alpha_r + \beta_c + \boldsymbol{(\alpha\beta)_{rc}}$. In order to determine the value for that term $(\alpha\beta)_{rc}$, we take the the mean for that cell (i.e., a certain level combination of factor A and B, e.g., placebo – CBT) and subtract the total mean as well as the contributions of α_r and β_c:
$(\alpha\beta)_{rc} = \mu_{rc} - \mu_{..} - \alpha_r - \beta_c$. α_r and β_c can be replaced by $(\mu_{r.} - \mu_{..})$ and $(\mu_{.c} - \mu_{..})$ respectively:
$(\alpha\beta)_{rc} = \mu_{rc} - \mu_{..} - (\mu_{r.} - \mu_{..}) - (\mu_{.c} - \mu_{..})$. We can then remove the brackets (remember that the – inside the brackets changes to + if there is a – before the bracket; we can therefore "sum up" one $- \mu_{..}$ and two $+ \mu_{..}$ to one $+ \mu_{..}$) and end up with:
$(\alpha\beta)_{rc} = \mu_{rc} - \mu_{r.} - \mu_{.c} + \mu_{..}$ or $(\alpha\beta)_{rc} = \overline{Y}_{rc} - \overline{Y}_{r.} - \overline{Y}_{.c} + \overline{Y}_{..}$ (as long as we operate in our sample). This again, we can "convert" into a formula that we actually can use to carry out our calculations:

$$SS_{A:B} = \sum_{r=1}^{R} \sum_{c=1}^{C} N \cdot (\bar{Y}_{rc} - \bar{Y}_{r.} - \bar{Y}_{.c} + \bar{Y}_{..})^2$$

Using the numbers in the table, we can do this for the first cell: $3 \cdot (0.300 - 0.450 - 0.722 + 0.883)^2$ which gives us 0.000363. We do the same for the remaining five cells and then sum up the values. We obtain 0.271 which is the sum of squares for our interaction (SS_{AB}). The degrees of freedom are $(N - 1) \cdot (C - 1) = 2 \cdot 1 = 2$. By dividing the sum of squares by the degrees of freedom, we get the mean square for the interaction: $MS_{AB} = 0.271 / 2 = 0.136$. The sum of squares for the interaction also explains variance and further reduces the sum of squares of the residuals: $SS_R = SS_{tot} - SS_A - SS_B - SS_{AB} = 4.845 - 3.453 - 0.467 - 0.271 = 0.653$. The degrees of freedom have also to be reduced: $df_R = N - df_A - df_B - df_{AB} - 1 = 18 - 2 - 1 - 2 - 1 = 12$. The mean square for the residuals is therefore: $MS_R = SS_R / df_R = 0.653 / 12 = 0.054$. Now, we can calculate our F-statistics for AB: $F_{AB} = MS_{AB} / MS_R = 0.136 / 0.054 = 2{,}499$. The interaction is not sizeable enough to become significant ($p = 0.1238$). Again, we ended up with a value within rounding-error-distance from the calculations in jamovi.