

16.2: Factorial ANOVA 2- Balanced Designs, Interactions Allowed

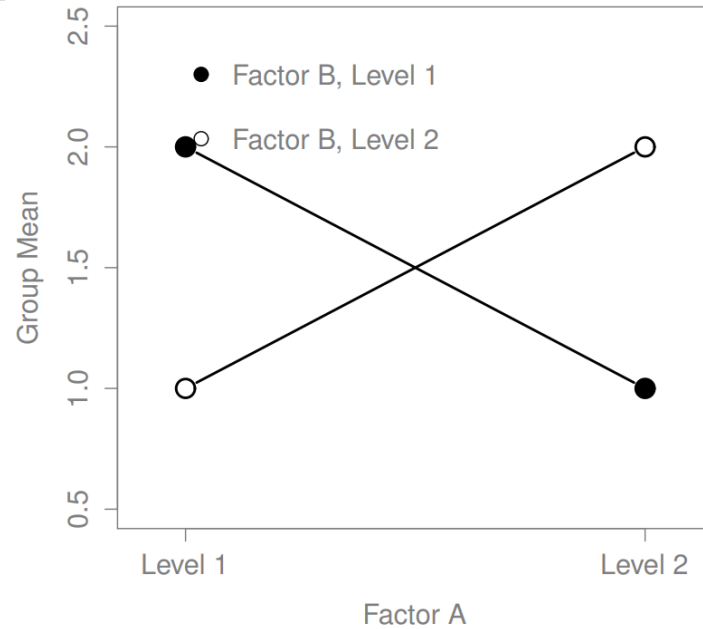


Figure 16.5: Qualitatively different interactions for a 2imes2 ANOVA

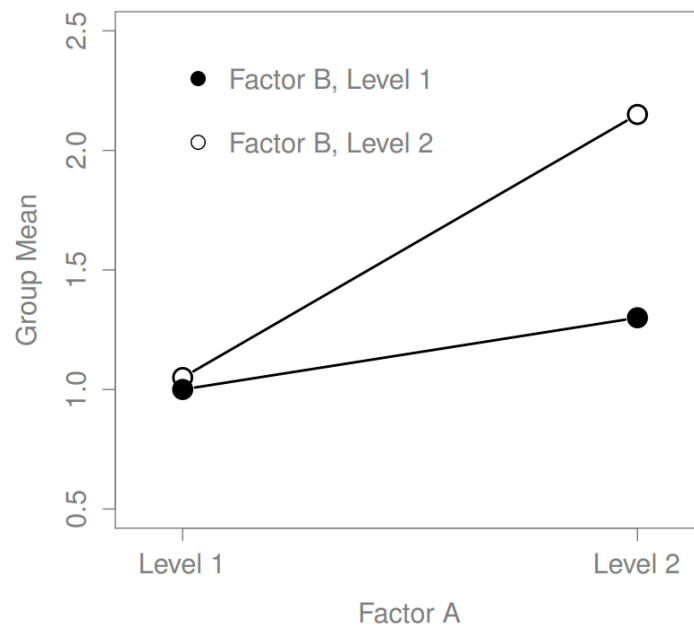


Figure 16.6: Qualitatively different interactions for a 2imes2 ANOVA

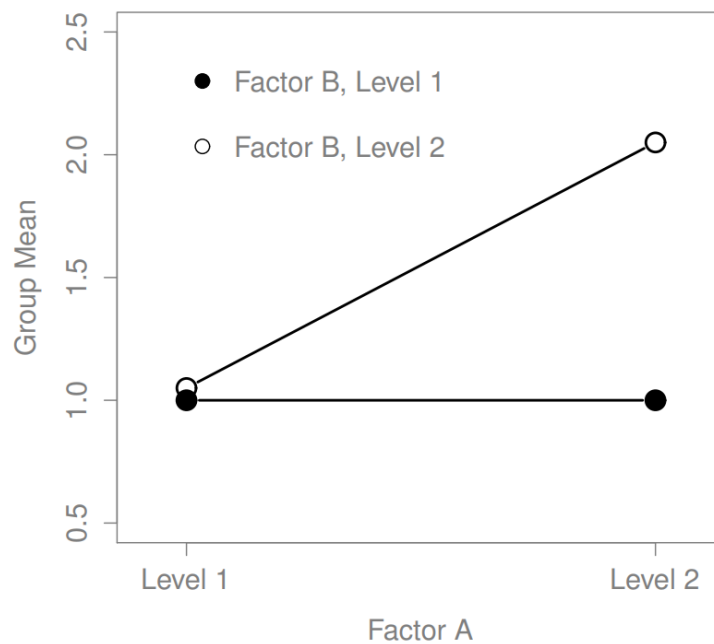


Figure 16.7: Qualitatively different interactions for a 2imes2 ANOVA

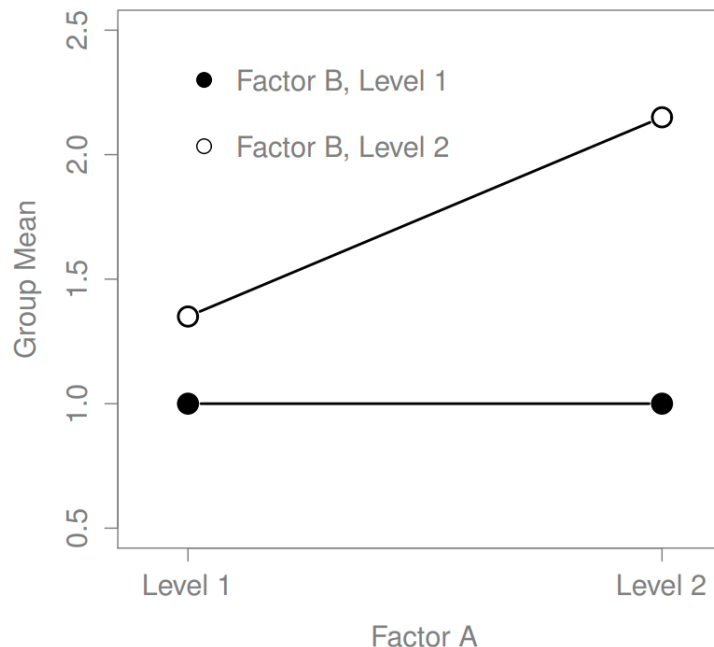


Figure 16.8: Qualitatively different interactions for a 2imes2 ANOVA

The four patterns of data shown in Figure ?? are all quite realistic: there are a great many data sets that produce exactly those patterns. However, they are not the whole story, and the ANOVA model that we have been talking about up to this point is not sufficient to fully account for a table of group means. Why not? Well, so far we have the ability to talk about the idea that drugs can influence mood, and therapy can influence mood, but no way of talking about the possibility of an *interaction* between the two. An interaction between A and B is said to occur whenever the effect of Factor A is *different*, depending on which level of Factor B we're talking about. Several examples of an interaction effect with the context of a 2 x 2 ANOVA are shown in Figure ??. To give a more concrete example, suppose that the operation of Anxifree and Joyzepam is governed quite different physiological mechanisms, and one consequence of this is that while Joyzepam has more or less the same effect on mood regardless of whether one is in therapy, Anxifree is actually much more effective when administered in conjunction with CBT. The ANOVA that we developed in the previous section does not capture this idea. To get some idea of whether an interaction is actually happening here, it helps to plot the various group means. There are quite a few different ways draw these plots in R. One easy way is to use the

`interaction.plot()` function, but this function won't draw error bars for you. A fairly simple function that will include error bars for you is the `lineplot.CI()` function in the `sciplots` package (see Section 10.5.4). The command

```
library(sciplot)
library(lsr)
lineplot.CI( x.factor = clin.trial$drug,
             response = clin.trial$mood.gain,
             group = clin.trial$therapy,
             ci.fun = ciMean,
             xlab = "drug",
             ylab = "mood gain" )
```

produces the output is shown in Figure 16.9 (don't forget that the `ciMean` function is in the `lsr` package, so you need to have `lsr` loaded!). Our main concern relates to the fact that the two lines aren't parallel. The effect of CBT (difference between solid line and dotted line) when the drug is Joyzepam (right side) appears to be near zero, even smaller than the effect of CBT when a placebo is used (left side). However, when Anxifree is administered, the effect of CBT is larger than the placebo (middle). Is this effect real, or is this just random variation due to chance? Our original ANOVA cannot answer this question, because we make no allowances for the idea that interactions even exist! In this section, we'll fix this problem.

16.2.1 What exactly *interaction effect*?

The key idea that we're going to introduce in this section is that of an interaction effect. What that means for our R formulas is that we'll write down models like so although there are only two *factors* involved in our model (i.e., `drug` and `therapy`), there are three distinct *terms* (i.e., `drug`, `therapy` and `drug:therapy`). That is, in addition to the main effects of `drug` and `therapy`, we have a new component to the model, which is our interaction term `drug:therapy`. Intuitively, the idea behind an interaction effect is fairly simple: it just means that the effect of Factor A is different, depending on which level of Factor B we're talking about. But what does that actually mean in terms of our data? Figure ?? depicts several different patterns that, although quite different to each other, would all count as an interaction effect. So it's not entirely straightforward to translate this qualitative idea into something mathematical that a statistician can work with. As a consequence, the way that the idea of an interaction effect is formalised in terms of null and alternative hypotheses is slightly difficult, and I'm guessing that a lot of readers of this book probably won't be all that interested. Even so, I'll try to give the basic idea here.

To start with, we need to be a little more explicit about our main effects. Consider the main effect of Factor A (`drug` in our running example). We originally formulated this in terms of the null hypothesis that the two marginal means $\mu_{.r}$ are all equal to each other. Obviously, if all of these are equal to each other, then they must also be equal to the grand mean $\mu_{..}$ as well, right? So what we can do is define the *effect* of Factor A at level r to be equal to the difference between the marginal mean $\mu_{.r}$ and the grand mean $\mu_{..}$.

Let's denote this effect by α_r , and note that

$$\alpha_r = \mu_{.r} - \mu_{..}$$

Now, by definition all of the α_r values must sum to zero, for the same reason that the average of the marginal means $\mu_{.r}$ must be the grand mean $\mu_{..}$. We can similarly define the effect of Factor B at level i to be the difference between the column marginal mean $\mu_{.c}$ and the grand mean $\mu_{..}$.

$$\beta_c = \mu_{.c} - \mu_{..}$$

and once again, these β_c values must sum to zero. The reason that statisticians sometimes like to talk about the main effects in terms of these α_r and β_c values is that it allows them to be precise about what it means to say that there is no interaction effect. If there is no interaction at all, then these α_r and β_c values will perfectly describe the group means μ_{rc} . Specifically, it means that

$$\mu_{rc} = \mu_{..} + \alpha_r + \beta_c$$

That is, there's nothing *special* about the group means that you couldn't predict perfectly by knowing all the marginal means. And that's our null hypothesis, right there. The alternative hypothesis is that

$$\mu_{rc} \neq \mu_{..} + \alpha_r + \beta_c$$

for at least one group r_c in our table. However, statisticians often like to write this slightly differently. They'll usually define the specific interaction associated with group r_c to be some number, awkwardly referred to as $(\alpha\beta)_{rc}$, and then they will say that the alternative hypothesis is that

$$\mu_{rc} = \mu_{..} + \alpha_r + \beta_c + (\alpha\beta)_{rc}$$

where $(\alpha\beta)_{rc}$ is non-zero for at least one group. This notation is kind of ugly to look at, but it is handy as we'll see in the next section when discussing how to calculate the sum of squares.

```
## Warning: package 'sciplot' was built under R version 3.5.2
```

```
## Warning: package 'lsr' was built under R version 3.5.2
```

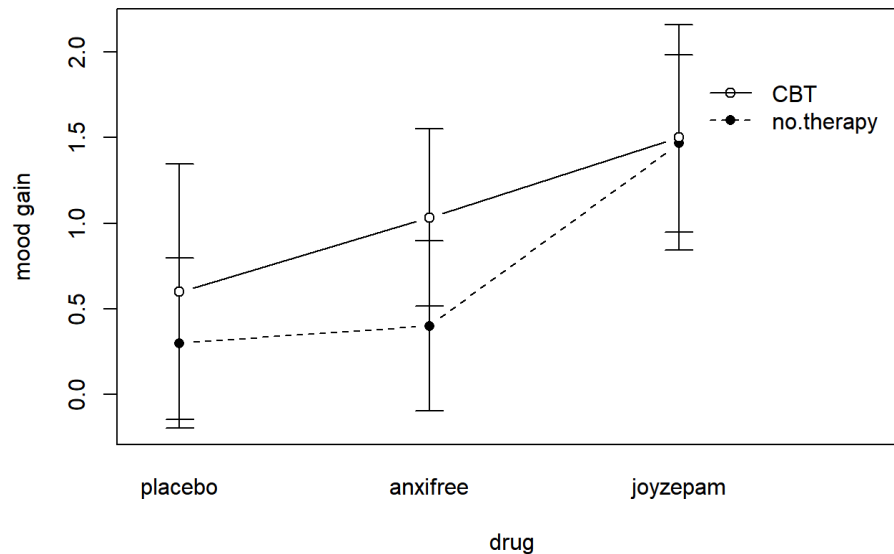


Figure 16.9: An interaction plot for the group means in the clinical trial data. The command to produce it is included in the main text. You'll notice that the legend doesn't quite fit properly. You can fix this by playing around with the `x.legend` and `y.legend` arguments: type `?lineplot.CI` for details.

16.2.2 Calculating sums of squares for the interaction

How should we calculate the sum of squares for the interaction terms, $SS_{A:B}$? Well, first off, it helps to notice how the previous section defined the interaction effect in terms of the extent to which the actual group means differ from what you'd expect by just looking at the marginal means. Of course, all of those formulas refer to population parameters rather than sample statistics, so we don't actually know what they are. However, we can estimate them by using sample means in place of population means. So for Factor A, a good way to estimate the main effect at level r as the difference between the *sample* marginal mean $\bar{Y}_{r.}$ and the sample grand mean $\bar{Y}_{..}$. That is, we would use this as our estimate of the effect:

$$\hat{\alpha}_r = \bar{Y}_{r.} - \bar{Y}_{..}$$

Similarly, our estimate of the main effect of Factor B at level c can be defined as follows:

$$\hat{\beta}_c = \bar{Y}_{.c} - \bar{Y}_{..}$$

Now, if you go back to the formulas that I used to describe the SS values for the two main effects, you'll notice that these effect terms are exactly the quantities that we were squaring and summing! So what's the analog of this for interaction terms? The answer to this can be found by first rearranging the formula for the group means μ_{rc} under the alternative hypothesis, so that we get this:

$$\begin{aligned} (\alpha\beta)_{rc} &= \mu_{rc} - \mu_{..} - \alpha_r - \beta_c \\ &= \mu_{rc} - \mu_{..} - (\mu_{r.} - \mu_{..}) - (\mu_{.c} - \mu_{..}) \\ &= \mu_{rc} - \mu_{r.} - \mu_{.c} + \mu_{..} \end{aligned}$$

So, once again, if we substitute our sample statistics in place of the population means, we get the following as our estimate of the interaction effect for group rc , which is

$$(\hat{\alpha}\hat{\beta})_{rc} = \bar{Y}_{rc} - \bar{Y}_{r.} - \bar{Y}_{.c} + \bar{Y}_{..}$$

Now all we have to do is sum all of these estimates across all R levels of Factor A and all C levels of Factor B, and we obtain the following formula for the sum of squares associated with the interaction as a whole:

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

where, we multiply by N because there are N observations in each of the groups, and we want our SS values to reflect the variation among *observations* accounted for by the interaction, not the variation among groups.

Now that we have a formula for calculating $SS_{A:B}$, it's important to recognise that the interaction term is part of the model (of course), so the total sum of squares associated with the model, SS_M is now equal to the sum of the three relevant SS values, $SS_A + SS_B + SS_{A:B}$. The residual sum of squares SS_R is still defined as the leftover variation, namely $SS_T - SS_M$, but now that we have the interaction term this becomes

$$SS_R = SS_T - (SS_A + SS_B + SS_{A:B})$$

As a consequence, the residual sum of squares SS_R will be smaller than in our original ANOVA that didn't include interactions.

16.2.3 Degrees of freedom for the interaction

Calculating the degrees of freedom for the interaction is, once again, slightly trickier than the corresponding calculation for the main effects. To start with, let's think about the ANOVA model as a whole. Once we include interaction effects in the model, we're allowing every single group has a unique mean, μ_{rc} . For an $R \times C$ factorial ANOVA, this means that there are $R \times C$ quantities of interest in the model, and only the one constraint: all of the group means need to average out to the grand mean. So the model as a whole needs to have $(R \times C) - 1$ degrees of freedom. But the main effect of Factor A has $R - 1$ degrees of freedom, and the main effect of Factor B has $C - 1$ degrees of freedom. Which means that the degrees of freedom associated with the interaction is

$$\begin{aligned} df_{A:B} &= (R \times C - 1) - (R - 1) - (C - 1) \\ &= RC - R - C + 1 \\ &= (R - 1)(C - 1) \end{aligned}$$

which is just the product of the degrees of freedom associated with the row factor and the column factor.

What about the residual degrees of freedom? Because we've added interaction terms, which absorb some degrees of freedom, there are fewer residual degrees of freedom left over. Specifically, note that if the model with interaction has a total of $(R \times C) - 1$, and there are N observations in your data set that are constrained to satisfy 1 grand mean, your residual degrees of freedom now become $N - (R \times C) - 1 + 1$, or just $N - (R \times C)$.

16.2.4 Running the ANOVA in R

Adding interaction terms to the ANOVA model in R is straightforward. Returning to our running example of the clinical trial, in addition to the main effect terms of `drug` and `therapy`, we include the interaction term `drug:therapy`. So the R command to create the ANOVA model now looks like this:

```
model.3 <- aov( mood.gain ~ drug + therapy + drug:therapy, clin.trial )
```

However, R allows a convenient shorthand. Instead of typing out all three terms, you can shorten the right hand side of the formula to `drug*therapy`. The `*` operator inside the formula is taken to indicate that you want both main effects and the interaction. So we can also run our ANOVA like this, and get the same answer:

```
model.3 <- aov( mood.gain ~ drug * therapy, clin.trial )
summary( model.3 )
```

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## drug           2   3.453   1.7267   31.714 1.62e-05 ***
## therapy        1   0.467   0.4672    8.582  0.0126  *
## drug:therapy    2   0.271   0.1356    2.490  0.1246
## Residuals     12   0.653   0.0544
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

As it turns out, while we do have a significant main effect of drug ($F_{2,12}=31.7, p<.001$) and therapy type ($F_{1,12}=8.6, p=.013$), there is no significant interaction between the two ($F_{2,12}=2.5, p=0.125$).

16.2.5 Interpreting the results

There's a couple of very important things to consider when interpreting the results of factorial ANOVA. Firstly, there's the same issue that we had with one-way ANOVA, which is that if you obtain a significant main effect of (say) `drug`, it doesn't tell you anything about which drugs are different to one another. To find that out, you need to run additional analyses. We'll talk about some analyses that you can run in Sections 16.7 and ???. The same is true for interaction effects: knowing that there's a significant interaction doesn't tell you anything about what kind of interaction exists. Again, you'll need to run additional analyses.

Secondly, there's a very peculiar interpretation issue that arises when you obtain a significant interaction effect but no corresponding main effect. This happens sometimes. For instance, in the crossover interaction shown in Figure ??, this is exactly what you'd find: in this case, neither of the main effects would be significant, but the interaction effect would be. This is a difficult situation to interpret, and people often get a bit confused about it. The general advice that statisticians like to give in this situation is that you shouldn't pay much attention to the main effects when an interaction is present. The reason they say this is that, although the tests of the main effects are perfectly valid from a mathematical point of view, when there is a significant interaction effect the main effects rarely test interesting hypotheses. Recall from Section 16.1.1 that the null hypothesis for a main effect is that the *marginal means* are equal to each other, and that a marginal mean is formed by averaging across several different groups. But if you have a significant interaction effect, then you *know* that the groups that comprise the marginal mean aren't homogeneous, so it's not really obvious why you would even care about those marginal means.

Here's what I mean. Again, let's stick with a clinical example. Suppose that we had a 2×2 design comparing two different treatments for phobias (e.g., systematic desensitisation vs flooding), and two different anxiety reducing drugs (e.g., Anxifree vs Joyzepam). Now suppose what we found was that Anxifree had no effect when desensitisation was the treatment, and Joyzepam had no effect when flooding was the treatment. But both were pretty effective for the other treatment. This is a classic crossover interaction, and what we'd find when running the ANOVA is that there is no main effect of drug, but a significant interaction. Now, what does it actually *mean* to say that there's no main effect? Well, it means that, if we average over the two different psychological treatments, then the *average* effect of Anxifree and Joyzepam is the same. But why would anyone care about that? When treating someone for phobias, it is never the case that a person can be treated using an “average” of flooding and desensitisation: that doesn't make a lot of sense. You either get one or the other. For one treatment, one drug is effective; and for the other treatment, the other drug is effective. The interaction is the important thing; the main effect is kind of irrelevant.

This sort of thing happens a lot: the main effect are tests of marginal means, and when an interaction is present we often find ourselves not being terribly interested in marginal means, because they imply averaging over things that the interaction tells us shouldn't be averaged! Of course, it's not always the case that a main effect is meaningless when an interaction is present. Often you can get a big main effect and a very small interaction, in which case you can still say things like “drug A is generally more effective than drug B” (because there was a big effect of drug), but you'd need to modify it a bit by adding that “the difference in effectiveness was different for different psychological treatments”. In any case, the main point here is that whenever you get a significant interaction you should stop and *think* about what the main effect actually means in this context. Don't automatically assume that the main effect is interesting.

This page titled [16.2: Factorial ANOVA 2- Balanced Designs, Interactions Allowed](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Danielle Navarro](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.