FES 524: Natural Resources Data Analysis

Reading 4.1: Factorial designs

Contents

1 Multiple factors of interest **1**

1.1 Crossed factors 1

1.2 Simple effects 1

2 Factorial designs **2**

2.1 Interaction effects 2

2.2 Main (overall) effects 3

1 Multiple factors of interest

Up to this point, we have been studying relatively simple study designs with a single factor of interest, but many studies involve interest in the effect of more than one factor on the mean of the response variable. There are several reasons we might want to study multiple factors at once. First, it is often practically efficient in terms of costs to investigate multiple factors at once rather than doing separate studies to investigate separate factors. In addition, the analysis can often be done as a single analysis instead of separate analyses, which is also efficient. Finally, investigators may expect the factors to act together, in which case they can explore the interaction between the two factors. You will learn more about interactions later in this reading. While you may see research involving many factors of interest, in this class we focus on studies where there are only two factors of interest.

1.1 Crossed factors

Factors are crossed if all combinations of factor levels are present. This means every level of one factor is present with every level of another factor. To check for crossing we check that all combinations are present in our data. If they are, we will describe the two factors as crossed or fully crossed.

Table 1 shows an example of crossed factors, A and B. The columns show the three levels of factor A (for example three levels of herbicide concentrations) while the rows show the two levels of a second factor, B (for example, ungulate exclosure or not). The cells of the table show the combined factor levels.

Table 2 another example. This example involves two different fertilizers (A and B) and two different fertilizer concentrations (0.5 kg/ha and 1.0 kg/ha). All combinations of the two factors are present.

1.2 Simple effects

When there are multiple factors of interest and we are interested in the mean response for each combination of the levels of each, one analysis option is to continue to use the separate means model approach we used in

Table 1: Example of a factorial design crossing the three levels of factor A with the two levels of factor B.

	$A1$	$A2$	$A3$
$B1$	$A1$-$B1$	$A2$-$B1$	$A3$-$B1$
$B2$	$A1$-$B2$	$A2$-$B2$	$A3$-$B2$

Table 2: Example of a factorial design crossing the two fertilizers with the two concentrations of interest.

	Fertilizer A	Fertilizer B
0.5 kg/ha	0.5 kg/ha A	0.5 kg/ha B
1.0 kg/ha	1.0 kg/ha A	1.0 kg/ha B

weeks 2 and 3. This involves making a new, *combined* factor variable. The levels of this new factor are made by combining the levels of the original factors. Using Table 1 as an example, the new factor would have six levels, one for each cell of the table.

Doing an analysis like this can be referred to as a *simple effects* analysis. There is nothing wrong with such an analysis, and in some cases this may actually be the preferred analysis. The usefulness of this approach can depend on the research question. Setting up the comparisons of interest can be more complicated in a simple effects analysis unless there is interest in all pairwise comparisons.

The simple effects analysis approach is especially useful when factors are only partially crossed rather than fully crossed. The most common example of partial crossing that I see is when there is a control group for one of the factors that involves doing nothing. The example in Table 2 could be a situation where this could come up. For example, perhaps there is a third level of the fertilizer factor that is a "do nothing" treatment where neither fertilizer is applied. Obviously we can't have 0.5 kg/ha of no fertilizer, so the last level of no fertilizer would not be crossed with the concentration factor.

2 Factorial designs

It is important to know about the simple effects analysis as an option. However, the new concept we are learning about this week is factorial designs and analyzing multiple factors simultaneously rather than combining multiple factors into a single factor.

A fully factorial design is one with fully crossed factors. Such designs are often described by the number of levels of the factors. For example, the fertilizer $\times$ concentration study example without the control would be called a 2 by 2 factorial, since the fertilizer factor has 2 levels and the concentration factor has 2 levels. You may see this written as 2×2. The example in Table 1 from this week's motivating example would be called a 2x3 factorial.

2.1 Interaction effects

Over the next few weeks we are going to be focusing a lot on *interactions*, which is one of the important concepts that arises when talking about factorial designs. An interaction is when the effect of one variable depends on the value of another variable. Interactions are often specifically of interest when working with data collected from a factorial design.

Let's use the fertilizer and concentration example to explore what an interaction might look like in this case.

Figure 1 shows what is called an interaction plot for a hypothetical analysis using data collected for the fertilizer-concentration example. An interaction plot shows the estimated means for every factor combination. Such plots are most often used as exploratory plots and so should not be used to make statistical inference. In this case I have included error bars from a fitted model to show there is little variation around the estimated means. In lab we will add the raw data to our interaction plots to show variation and to make sure we recognize the plot as an exploratory plot.

The example in Figure 1 shows evidence of an interaction. The difference in mean response between the 1.0 kg/ha concentration and the 0.5 kg/ha concentration is small and negative for the A fertilizer but large and positive for the B fertilizer. The effect of one factor variable clearly depends on the value of the other factor variable.

Here are a couple of ways you could describe the interaction:

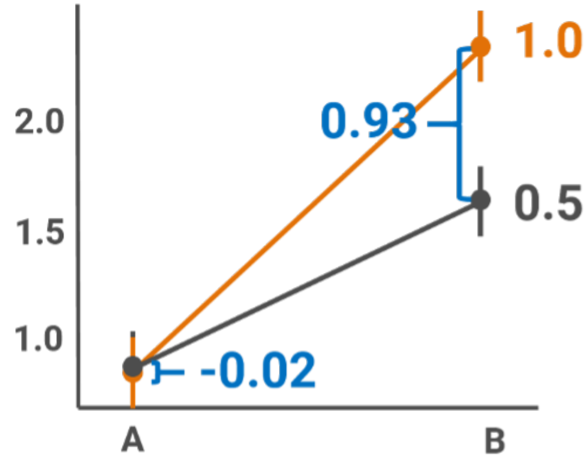

Figure 1: Example interaction plot. The orange colors represent the 1.0 kg/ha concentration treatment, while grey shows the 0.5 kg/ha concentration. The points represent the estimated mean response for each combination of the two factors of interest.

- The effect of concentration depends on fertilizer type.
- The difference in mean response among concentrations differs between fertilizers.

The above examples are focused on the concentration variable. However, we could make the same statements using the fertilizer factor as the focus. The way you describe an interaction will depend on how you worded your research question and what you want to focus on in the results.

Alternative ways to describe the interaction:

- The effect of fertilizer depends on the concentration.
- The difference in mean response among fertilizers differs between concentrations.

Below is language we can use to describe results from an analysis that focuses on the interaction. We would report estimated differences in means among concentrations (fertilizers) for each fertilizer (concentration). It wouldn't make sense to talk about some overall effect of concentration when we believed that the effect depended on which fertilizer was applied.

Effect of concentration for Fertilizer A:

Increasing the amount of Fertilizer A from 0.5 to 1.0 kg/ha is estimated to decrease mean biomass by 0.02 kg/ha (95% CI: 0.08 decrease to 0.04 increase).

Effect of concentration for Fertilizer B:

Increasing the amount of Fertilizer B from 0.5 to 1.0 kg/ha is estimated to increase the mean biomass 0.93 kg/ha (95% CI: 0.86 to 0.99 increase).

2.2 Main (overall) effects

When two factors are applied at once simply for efficiency of a study design, the investigator may be solely interested in overall effects. In such a case, the factors are not expected to interact.

We will again use the fertilizer and concentration example, this time to explore what it means to be interested in the overall effect of the factors.

Figure 2 shows another interaction plot, however, Figure 2 shows the results of a hypothetical analysis where there was no interaction. You can see in this idealized example that the difference in mean response for the two concentrations is identical for the A fertilizer and the B fertilizer.

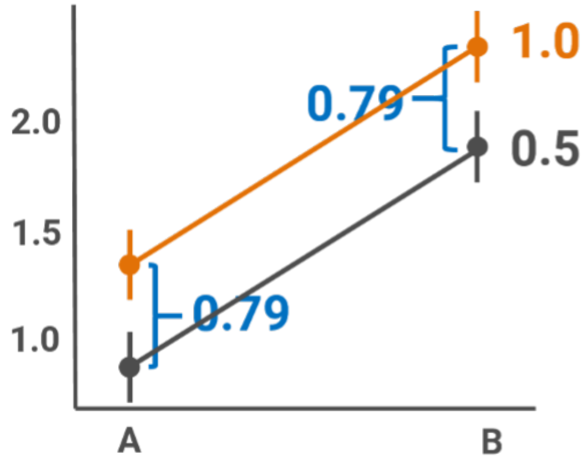

Figure 2: Example interaction plot in which there is no interaction between the fertilizer and concentration factors.