Chapter 10: Analysis and Interpretation of Factorial Data

Even though factorial designs are a relatively small conceptual increase in the complexity of an experimental protocol, they can pose surprisingly difficult challenges when trying to make sense of the data. Even experienced researchers can make mistakes in characterizing the effects observed and drawing inferences from factorial results. In this chapter, we will describe a systematic process of working through factorial data examining the **main effects** and then any **interactions** among the factors. Data visualizations are very helpful in providing an overview and with some practice, common outcome patterns can be recognized as visual patterns in data graphs.

In addition, we will introduce the statistical concept of **effect size** to help describe interaction effects. As we will see later (Chapter TBD), modern psychological science is working to incorporate improved models of statistical inference and shift away from a reliance on a simple rejection of the null hypothesis by the familiar standard p<.05. Here we will consider effect sizes as the simple difference in the mean performance across conditions (technically an unstandardized effect size) in order to identify some common types of interaction effects: **super-additive**, **3:1,** and **crossover**. The statistical tool of a **post-hoc t-test** is used to assess specific contrasts between conditions within a factorial design. Examples will be discussed of how to draw conclusions from main effects and different forms of interactions.

## Learning Objectives

1. Distinguish between main effects and interactions and recognize each.
2. Understand factorial data tables looking at individual conditions (cells) and marginal means
3. Interpret and understand bar graphs and line graphs showing the results of studies with factorial designs.
4. Understand types of interactions: super-additive, 3:1, crossover
5. Know the role of post hoc t-tests to further characterize data

# Main Effects

In factorial designs, the main hypotheses are tested as main effects and interactions. A main effect is the effect of one independent variable on the dependent variable—averaging across the levels of the other independent variable. Common patterns of data are illustrated here with both means tables and figures to illustrate the results.

In a means table, the average performance of participants in each of the experimental conditions is shown separately, typically with means and the standard deviations shown below in parentheses. Note that in APA format, tables are accompanied by a table note indicating what the parenthetical numbers are.

In the tables and graphs below, we have an abstract design based on two factors, Factor 1 and Factor 2. Factor 1 has two conditions, A and B. Factor 2 has two conditions, X and Y. Each participant in the study is in one of the four possibilities: AY, BY, AX, BX. For illustration, we will assume a dependent variable that is scored on a 1 to 10 scale and that 20 participants were run in each of the 4 conditions (a fully between-participants design).

## Simulated Data 1

One main effect (Factor 2) is reliable, the other main effect is not affecting the DV and no interaction between the factors occurred.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 3.93 (1.43) | 4.21 (1.22) | 4.07 (1.33) | |
| Y | 6.98 (1.28) | 6.88 (1.37) | 6.93 (1.33) | |
|  | Mean | 5.45 (2.04) | 5.55 (1.86) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In the line graph, this pattern creates two roughly parallel, nearly horizontal lines. The distance between the lines is the effect of Factor 2. The fact that they are flat (horizontal) is reflecting the absence of an effect of Factor 1. The magnitude of the Factor 2 difference is also seen in the marginal means for Factor 2, the rightmost column of the data table (M = 4.07 for X, M= 6.93 for Y). The bar plot also clearly shows the Y condition (orange) producing higher values than the X condition (blue).

With the outcome being an effect on just one factor, we can also see the magnitude of this effect in the marginal means. The score is ~3 points higher for the Y condition than the X condition. The **uncorrected effect size** for Factor 2 in these data is an increase in the DV of 2.86. Later (Chapter TBD) we will discuss how to use effect sizes for more sophisticated assessment of experimental effects and how to incorporate these into predicting the **power** and **sensitivity** of our designs to our hypothesized findings.

## Simulated Data 2

One main effect (Factor 1) is reliable, the other main effect is not affecting the DV and no interaction between the factors occurred.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 3.27 (1.51) | 7.09 (1.35) | 5.18 (2.39) | |
| Y | 3.65 (1.08) | 7.44 (1.44) | 5.55 (2.28) | |
|  | Mean | 3.46 (1.33) | 7.26 (1.41) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In the line graph, this pattern creates two nearly overlapping lines that slope up the same way. The increase across the graph (from A to B) reflects the effect of Factor 1. The lack of vertical separation is due to the non-effect from Factor 2. The magnitude of the Factor 1 difference is also seen in the marginal means for Factor 1, the bottom row of the data table (M = 3.46 for A, M= 7.26 for B). The bar plot also clearly shows the B condition (right 2 bars) producing higher values than the A condition (left 2 bars).

Again, we can characterize the effect of Factor 1 by its uncorrected effect size seen in the marginal means. Condition B is scoring an average of 3.8 points higher than condition A.

## Simulated Data 3

Both main effects are reliable but there is no interaction between them. This pattern is often mistaken for suggesting an interaction between the factors but there is none. The highest performing condition (BY) is showing the effects of both Factor 1 and Factor 2 additively.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 2.36 (1.54) | 4.70 (1.54) | 3.53 (1.93) | |
| Y | 4.00 (1.08) | 6.38 (1.44) | 5.19 (1.84) | |
|  | Mean | 3.18 (1.63) | 5.54 (1.74) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In the line graph, this pattern creates two parallel, separated lines with the same slope. As we will see below, differing slopes on a line graph is a useful visual signal of the occurrence of an interaction between factors. The increase across the graph (from A to B) reflects the effect of Factor 1. The vertical separation is due to the effect of Factor 2 (from X to Y). The marginal means for both factors show the magnitude of the two effects independent of each other (note that these would not be independent if there was an interaction). The bar plot also shows both effects but does not imply the parallel slopes quite as easily as the line graph.

Once again looking at the marginal means for uncorrected effect sizes we see the Factor 1 effect is 2.36 points higher from A to B. The effect for Factor 2 is 1.66 points higher from X to Y. The highest scoring condition, BY, is larger than the lowest scoring condition, AX, by roughly the sum of these two effects. This is synonymous with saying there is no interaction, i.e., that the two main effects simply add together when both factors are present.

# Interactions

In the next 3 simulated data presentations we will see some different common forms of interaction between the factors in a 2x2 factorial design.

## Simulated Data 4

In these data, we have two main effects and a **super-additive interaction**. The effect of both Factor 1 and Factor 2 are to increase scores on the DV. In addition, performance in the combined condition (BY) is higher than would be predicted if the two factor effects summed together. We can describe this effect as saying the effect of Factor 1 was particularly strong in the Y condition of Factor 2. It would also be correct to describe this effect as the effect of Factor 2 was particularly strong in the B condition of Factor 1. This is another example of the symmetry of factorial designs. The design does not inherently prioritize one factor over another and there are usually at least two equivalent ways to describe the results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 1.43 (1.32) | 4.43 (2.00) | 2.93 (2.26) | |
| Y | 3.56 (1.39) | 9.31 (1.38) | 6.43 (3.19) | |
|  | Mean | 2.50 (1.72) | 6.87 (2.98) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In the line graph, this pattern creates two separated lines that do not have the same slope. The different slopes are what visually signal the presence of an interaction. The main effects are visible in the same way as previous graphs (the left to right increase reflects Factor 1; the separation in lines reflects Factor 2). The marginal means for both factors show an estimate of the effects somewhat independently of each other but note how the BY condition stands out from the marginal means to show the super-additivity. The bar plot also shows both effects but the difference in slopes is again not quite as visible as in the line graph.

Compare the different here in the BY condition to the AX condition from Example 3 above. Here the BY condition is scoring even higher than the two main effects would predict independently. This interaction reflects something additional pushing up the DV score when both factors are present simultaneously that is different than either factor in isolation. The “something additional” is usually the goal in a 2x2 design and aimed to learn something new about the component variables that requires measurements across manipulations of both.

## Simulated Data 5

A common data pattern in 2x2 designs is a **3:1 interaction**. In this case, one of the conditions is producing a different score from the others, which are all roughly similar (e.g., 1 high score, 3 low scores). We could describe this result using language similar to the super-additive interaction by saying the effect of Factor 2 had a large effect in the A condition of Factor 1 but little or no effect in the B condition. Technically this is just the opposite of the super-additive case (sub-additive) where the effect of Factor 2 is smaller than predicted by the main effects in the B condition. However, it is quite common to describe this result as saying Factor 2 only matters in the A condition and does not apply to the B condition. When Factor 1 is a participants variable (e.g., an personality variable or measure like “math identification”) we are observing a case where an experimental manipulation only appears to affect one subgroup of the population. This type of design and outcome are fairly common in psychological science.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 6.70 (1.76) | 4.34 (1.58) | 5.52 (2.04) | |
| Y | 3.92 (1.42) | 4.51 (1.29) | 4.21 (1.39) | |
|  | Mean | 5.31 (2.11) | 4.43 (1.45) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In this type of interaction we see that the bar plot graph illustrates the form of the data more effectively than the line graph. The AX condition stands out from the other 3 conditions which produce similar levels of performance. We can still see that the lines are not parallel in the line graph, which reflects the fact that there is an interaction between factors. This is also a special case in that the marginal means on the table are not very helpful in understanding the data. Both factors are showing a hint of a main effect on the marginal means, but descripting the data in terms of the main effects does not necessarily contribute effectively to communicating the 3:1 interaction in the data. We could characterize this kind of effect as saying the influence of Factor 2 goes away in condition B (of Factor 1). Synonymously, we could say that the influence of Factor 2 only occurs in condition A (of Factor 1). The preferred language depends on the domain being studied and the experimental hypothesis.

## Simulated Data 6

Here we see the case where the effects of the factors are essentially inverted across conditions of the other factor, which produces a **cross-over interaction**. This are somewhat complex to describe fully. We can say that the effect of Factor 2 in the A condition of Factor 1 is to increase performance while the effect of Factor 2 decreases performance in the B condition. Written descriptions tend to be wordy and these interactions are examples of where data visualizations are very valuable.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Means Table | | Factor 1 | |  |
| A | B | Mean |
| Factor 2 | X | 3.29 (1.07) | 5.87 (1.54) | 4.58 (1.85) | |
| Y | 6.59 (1.63) | 3.20 (1.31) | 4.90 (2.25) | |
|  | Mean | 4.94 (2.15) | 4.53 (1.96) |  | |

|  |  |
| --- | --- |
|  |  |
|  |  |

In this type of interaction, we see that the line graph makes a clear visual signal of the crossover (although note that this does not always happen if there is also a large main effect of Factor 2). As in all interactions, the slopes of the two lines are markedly different. The structure of the data is also quite visible on the bar plot graph. In these particular simulated data, there are no main effects of either factor as can be seen in the similar numbers across all the marginal means. This kind of data is often of theoretical importance because it suggests that neither factor has simple effects on the dependent variable and that prior work with simpler 2-group designs might have produced inconsistent or null results. As a reminder, studies that produce a null result do not establish that the IV had no effect on the DV, but only tell us that the study did not work – the IV did not affect the DV under the procedure and sampling conditions of that study.

For these types of findings, we would not focus on the effect sizes characterized by the main effects. Instead, we might characterize the differences in specific conditions contrasted by use of a post-hoc t-test.

## Post-hoc t-tests

The analytical tool used to identify reliable main effects and interactions is the ANOVA (analysis of variance). As we shall see, the output of the ANOVA will give us a statistical measure (the F ratio) and a p value (the probability of the data under the null hypothesis, as always) for each of the two main effects and interaction. However, the ANOVA itself does not indicate anything about the direction of the effects or the form of the interaction. For the main effects, we can use the descriptive statistics like the marginal means to support the inferential statement about differences in conditions. For the interaction term, it is often necessary to actually graph the data and visualize the form of the interaction in order to describe it accurately in the text. Nothing in the output of the ANOVA procedure indicates the form of the interaction.

In addition, for some interaction types, particularly 3:1 and cross-over interactions, it may be theoretically important to further characterize the data by targeted analysis of a subset of the data. For example, in the 3:1 interaction above, we may want to know whether the X condition is producing reliably higher numbers than the Y condition when only considering participants from the A condition of Factor 1. The tool for this is a post-hoc t-test on just the participants from condition A comparing scores in the X and Y conditions.

For a crossover design, we can use post-hoc t-tests to evaluate both conditions separately to see if the effect of Factor 2 is reliable in just the A condition (Factor 1) and also is it reliable in the B condition. It should be noted that the fact that the interaction is statistically reliable does not automatically require that these post-hoc t-tests are individually reliable. A reliable interaction means the effects of one factor (e.g., Factor 2) are reliably different across conditions of the other factor (e.g., Factor 1). That does not mean the effects are reliably within each condition considered separately. A reliable post-hoc t-test allows for a slightly stronger description of the results, which is often theoretically relevant.

We should also be aware when using these follow-up t-tests that there is some risk of weakening our conclusions through running too many **parallel comparisons**. Care needs to be taken in larger factorial designs where there can be a lot of individual conditions that we might be interested in specific comparisons between. For complex designs, we might need to consider using a **Bonferroni correction** for multiple comparisons or restricting our analysis to **pre-specified hypotheses** of specific importance.