Basic Research Designs & Terminology

This chapter introduces some basic terminology and concepts of experimental research design. Many research designs, both experimental and non-experimental, have been developed over the last 100 years, but this and the next chapter will focus just on basic experimental, quasi-experimental, and differential designs.

Recall that an experiment has these properties:

1. The independent variable must be completely controlled
2. The sample of research participants (subjects) must be randomly chosen
3. The research participants (subjects) must be assigned randomly to conditions

When all of these conditions are not met, the study is "less than" experimental: quasi-experimental, differential, or correlational. Property (1) requires the presence of at least two conditions in an experiment: an experimental group and a control group; or two experimental groups. This two-group design is the foundation for all other experimental designs.

Between versus Repeated Designs

One of the most fundamental distinctions in research design concerns whether the experiment compares different groups of subjects (people, rats) to each other, or whether it compares measurements of one group that were obtained at more than one time. A **between-subjects design** (also called a **between-groups design**) compares two or more groups to each other: males to females, experimental group to control group, etc. A **repeated-measures design** compares measurements taken repeatedly within a group: a pre-post measure to assess the effectiveness of psychotherapy; repeated measures of performance over time, etc. This design is also called a **within-group design**. Many experiments combine between-subjects and repeated-measures components to form a **mixed design**. One of the first issues a researcher must consider is whether to use a between or repeated design. Each has its advantages and problems.

The table below illustrates a basic between-subjects design. Each S_i is a subject. Subjects 1-10 are in the experimental group, and subjects 11-20 are in the control group. The ∂ (delta) symbol is used here to indicate a data point or observation (a datum). Each subject supplies one observation. The typical method for analyzing the between-subjects design when there are two groups and interval-level measures is the independent t-test (covered in a different chapter).

Two-group one-way between-subjects design.			
Experimental Group		Control Group	
S_1	∂	S_{11}	∂
S_2	∂	S_{12}	∂
S_3	∂	S_{13}	∂
S_4	∂	S_{14}	∂
S_5	∂	S_{15}	∂
S_6	∂	S_{16}	∂
S_7	∂	S_{17}	∂
S_8	∂	S_{18}	∂
S_9	∂	S_{19}	∂
S_{10}	∂	S_{20}	∂

The following table illustrates a repeated design, such as a pre-post design. Each subject contributes two data points, one in the pretest and one in the posttest. Altogether, there are 20 observations (as in the between design). The simplest method for analyzing the repeated measures design when there are two measures such as pre- and post-, and the data are interval-level, is the dependent t-test.

One-way pre-post repeated design.		
	Pre-test	Posttest
S_1	∂	∂
S_2	∂	∂
S_3	∂	∂
S_4	∂	∂
S_5	∂	∂
S_6	∂	∂
S_7	∂	∂
S_8	∂	∂
S_9	∂	∂
S_{10}	∂	∂

To flesh out the two designs, consider an experiment to test the effect of alcohol on driving performance. In the between design, 10 undergrads drink a large amount of alcohol and 10 drink the equivalent in water. After waiting 30 minutes for the liquid to do its special thing, the subject is let loose in the south Palm Bay wasteland behind the wheel of an old Chevy Impala. His (her) driving ability is measured from a distance and represented as a single value, the ∂ in the design. In the repeated design, 10 subjects participate. Each one drives the Impala twice, once drunk and once sober. Note that the between design uses up more undergraduates, but problems in the repeated design such as practice effects are not present.

One-way versus Factorial Designs

A one-way design is an experiment in which there is one independent variable. The examples above are one-way designs. Below is a one-way between design with three levels of the IV:

Three-group one-way between-subjects design.		
Control Group	Experimental Group 1	Experimental Group 1
S_{1} ∂	S_{11} ∂	S_{21} ∂
S_{2} ∂	S_{12} ∂	S_{22} ∂
S_{3} ∂	S_{13} ∂	S_{23} ∂
S_{4} ∂	S_{14} ∂	S_{24} ∂
S_{5} ∂	S_{15} ∂	S_{25} ∂
S_{6} ∂	S_{16} ∂	S_{26} ∂
S_{7} ∂	S_{17} ∂	S_{27} ∂
S_{8} ∂	S_{18} ∂	S_{28} ∂
S_{9} ∂	S_{19} ∂	S_{29} ∂
S_{10} ∂	S_{20} ∂	S_{30} ∂

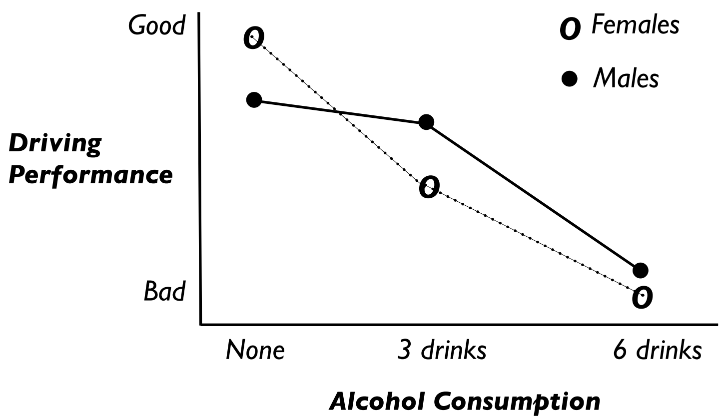
Perhaps Experimental Group 1 imbibes three drinks, and Experimental Group 2 imbibes six drinks.

Most research in modern Psychology is not this simple because we expect IVs to have complex effects that depend on other IVs, including moderator variables. The most common design looks at more than one IV at the same time in a **factorial design**. The following design is a 2x3 factorial:

Factorial between-subjects design.			
	Control Group	Experimental Group 1	Experimental Group 1
	S_{1} ∂	S_{11} ∂	S_{21} ∂
	S_{2} ∂	S_{12} ∂	S_{22} ∂
Males	S_{3} ∂	S_{13} ∂	S_{23} ∂
	S_{4} ∂	S_{14} ∂	S_{24} ∂
	S_{5} ∂	S_{15} ∂	S_{25} ∂
	S_{6} ∂	S_{16} ∂	S_{26} ∂
	S_{7} ∂	S_{17} ∂	S_{27} ∂
Females	S_{8} ∂	S_{18} ∂	S_{28} ∂
	S_{9} ∂	S_{19} ∂	S_{29} ∂
	S_{10} ∂	S_{20} ∂	S_{30} ∂

Subjects 1-5 are males in the control group, subjects 6-10 are females in the control group, etc. This design allows us to look at the overall effect of alcohol, but also to look at the differential effect of alcohol for males and females. It is a more powerful design than the one-way design because, with the same number of subjects, we obtain more information: the effect of alcohol, the relationship of Impala driving performance to gender and–this is the important part–the interaction between gender and alcohol.

An **interaction effect** occurs when the effect of one variable *depends on* the level of another. The figure below illustrates in interaction effect:

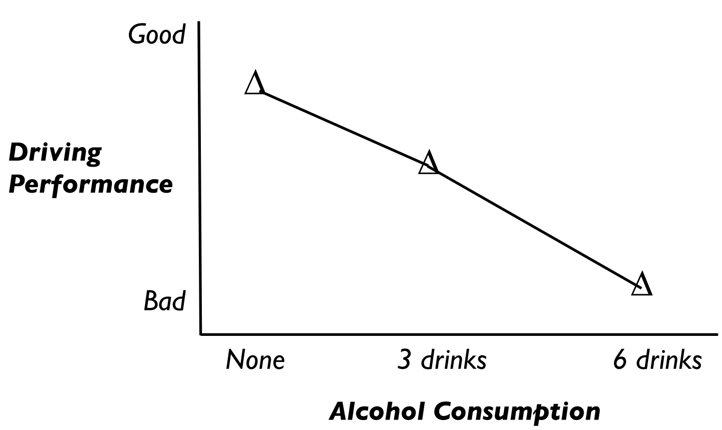


In this graph, females are better drivers than males in normal (sober) conditions, but drinking degrades their performance more than it does males' performance. In response to the question, "how does drinking affect driving?" we would have to say, "it depends" (on gender). Whenever "it depends" must qualify a statement, an interaction effect is present. Because behavior is multiply-determined and the world is complex, we often say "the truth lies in the interactions."

Effect Terminology

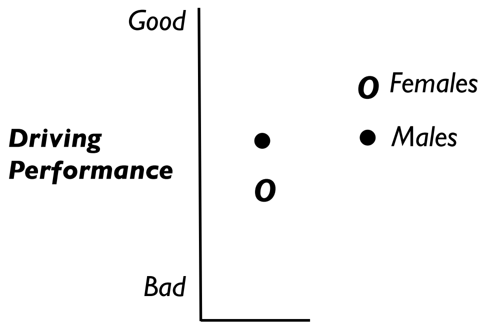
Social scientists have evolved (devolved?) a terminology for discussing the outcomes of factorial design experiments. This lingo helps us communicate with each other and has some side benefits similar to secret handshakes and codes. Any IV that has an effect on a DV is called an **effect**. We say "the effect of alcohol" or "the alcohol effect." In the graph above, because the alcohol effect depends on gender, we would say the experimental revealed an **interaction effect**.

Look at the driving graph and ask, What if we didn't know the gender of the subjects? In other words, what if we combine the male and female data together and forget about gender, just for the moment? When we ignore a variable in this way, we are **collapsing** over the variable. So if we collapse gender, what is the effect of drinking? The actual performance values will be the average of males and females at each level of the alcohol IV, looking like this:

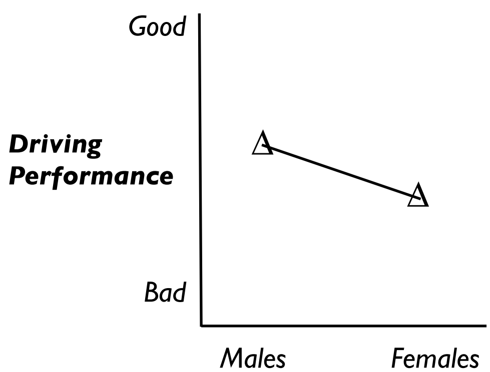


There is clearly an effect of alcohol when we collapse gender. An effect of an IV, when all other IVs are collapsed, is called a **main effect.** In our effect language, we would say "there was a main effect of alcohol."

Now let's collapse alcohol:



The graph is not designed to illustrate the collapsed alcohol situation very well, and it would normally be drawn like this:



Males are better drivers in the two alcohol conditions than females, so collapsing alcohol produces a slight advantage to males. Whether or not this difference is statistically significant depends on the details: the means and variances of the conditions. If the difference is statistically significant, we would say that there is a main effect of gender. Putting it all together, we can say:

> Main effects of alcohol and of gender were found, in addition to a significant Alcohol x Gender interaction.
>
> or
>
> Main effects of alcohol and of gender were found, but these main effects must be interpreted within the significant Alcohol x Gender interaction.

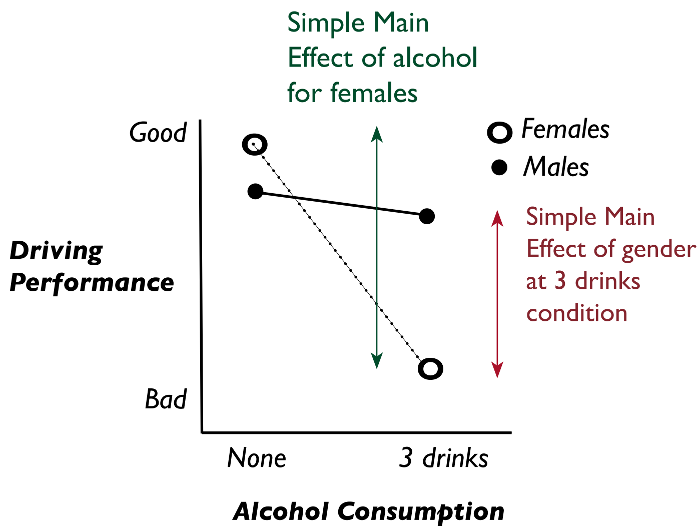
The x is spoken as "by," "alcohol by gender."

Here it is again, as you might see it in a journal (with fake numbers added):

> Main effects of alcohol, $F(2,25)=6.34$, $p<.05$, and of gender, $F(1,25)=4.55$, $p<.05$, were found, but these main effects must be interpreted within the significant Alcohol x Gender interaction, $F(2,25)=5.22$, $p<.05$.

The "F(2,25)=6.34" is the statistical analysis, Analysis of Variance (ANOVA). The "p<.05" is the familiar p-value in which the results of the statistical analysis are compared to the standard alpha level of .05 for rejecting the null hypothesis.

We aren't finished yet. Describing the results as we have done so far is still not sufficiently precise. Below is the result of the experiment, slightly changed and without the 6 drinks condition:



This graph shows a main effect of alcohol, a main effect of gender, and a Alcohol x Gender interaction. However, the real story here is that the women who had three drinks couldn't drive. The remaining three means are not especially low. One way to describe this outcome is to focus on the gender difference within the 3-drinks condition; the other way is to focus on the alcohol effect for females. The effect of one IV at just one level of another IV is called a **simple main effect**. Here, the difference between males and females when they had ingested three drinks is "the simple main effect of gender in the 3-drink condition." From the other perspective, the graph also shows a simple main effect of alcohol for females.

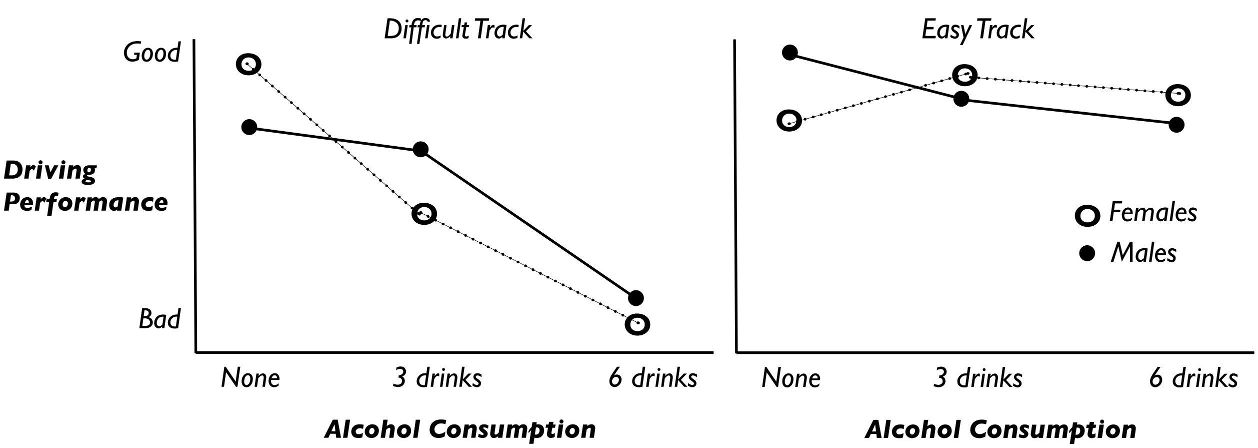
The search for specific mean differences that explain or account for more general effects such as an interaction effect is called **post-hoc analysis.** Technically, post-hoc analysis refers to digging around in the data just to understand what's there, purely after the fact and without prior expectations, hypotheses, or reasons. When the specific differences are predicted in a hypothesis, this digging around is given a more respectable name, **planned comparisons**. The difference between these two approaches to data dredging concerns Type I Error (rejecting the null hypothesis when in fact it is true). Every time you perform a post-hoc statistical test, the probability of committing a Type 1 Error is alpha (generally .05). As you perform more and more statistics (because it's so easy these days), the .05s add up and pretty soon your chance of committing a Type 1 Error is substantial. This issue is explained more fully in a statistics class.

Higher order designs

Many experiments in modern psychology involve more than two IVs. Following the current example, perhaps we should look at the gender and alcohol effects under favorable and unfavorable driving conditions. Maybe driving an Impala through the Palm Bay wastelands is not a very useful or realistic measure of alcohol effects, and instead we should have the subjects drive it on a standard test track. Maybe alcohol is a problem under difficult driving conditions but less of a problem under easy conditions. We might design a better experiment by having males and females drive under the three alcohol conditions (none, 3, 6 drinks) on a dry, well-lighted test track or a wet, poorly lighted track. This expanded design is a Gender (male/female) x Alcohol (0/3/6 drinks) x Difficulty (easy/difficult) between-groups factorial design, also called a 2 x 3 x 2 between-groups factorial design. The design looks like this, with 60 subjects:

2 x 3 x 2 between-groups factorial design.						
	Easy Test Track			Difficult Test Track		
	Control Group	3 drinks	6 drinks	Control Group	3 drinks	6 drinks
Males	S_{1} ∂ S_{2} ∂ S_{3} ∂ S_{4} ∂ S_{5} ∂	S_{11} ∂ S_{12} ∂ S_{13} ∂ S_{14} ∂ S_{15} ∂	S_{21} ∂ S_{22} ∂ S_{23} ∂ S_{24} ∂ S_{25} ∂	S_{31} ∂ S_{32} ∂ S_{33} ∂ S_{34} ∂ S_{35} ∂	S_{41} ∂ S_{42} ∂ S_{43} ∂ S_{44} ∂ S_{45} ∂	S_{51} ∂ S_{52} ∂ S_{53} ∂ S_{54} ∂ S_{55} ∂
Females	S_{6} ∂ S_{7} ∂ S_{8} ∂ S_{9} ∂ S_{10} ∂	S_{16} ∂ S_{17} ∂ S_{18} ∂ S_{19} ∂ S_{20} ∂	S_{26} ∂ S_{27} ∂ S_{28} ∂ S_{29} ∂ S_{30} ∂	S_{36} ∂ S_{37} ∂ S_{38} ∂ S_{39} ∂ S_{40} ∂	S_{46} ∂ S_{47} ∂ S_{48} ∂ S_{49} ∂ S_{50} ∂	S_{56} ∂ S_{57} ∂ S_{58} ∂ S_{59} ∂ S_{60} ∂

One possible outcome of the experiment might be:



You can see that alcohol has a large effect in the difficult track condition but not in the easy track condition. Furthermore, the Alcohol x Gender interaction in the difficult track condition is not the same pattern as the one in the easy track condition. The Alcohol x Gender interaction *depends on* the difficulty of the track. When a two-way interaction interacts with another IV, it produces a **three-way interaction**. In this case, we have discovered a Alcohol x Gender x Difficulty interaction. The Alcohol x Gender interaction in the difficult track condition is called the simple interaction. The effect of alcohol on females in the difficult track condition is called the simple-simple main effect. ("The simple-simple main effect of alcohol for females in the difficult track condition.")

If we collapse gender and alcohol, we are left with track difficulty. Which track, ignoring (collapsing) gender and alcohol, produced poorer driving? Clearly, the difficult track was...well...more difficult. This effect of track when alcohol and gender are collapsed is the main effect of track. Is there a main effect of alcohol?

Three-way designs are sometimes difficult to understand, and designs larger than this begin to tax the ability of the best researcher.