

Experimental Design

Copyright © 2000, 2011, 2016, J. Toby Mordkoff

The difference between experiments and correlational studies is both very clear and the object of occasionally acrimonious debate. For some reason, experiments are often seen as “better” than correlational studies, so some people who run experiments are pretty snotty about it, and some people who use correlations don’t like to admit it. Hopefully, as you learn more about these two methods, you will see that the question of which is better is naïve at best. The best design is the one that addresses the question of interest in the most direct and defensible manner -- sometimes this means an experiment; sometimes this means a correlational study.

By definition, an experiment involves at least one IV. Conversely, if there are no IVs, then the project is either a quasi-experiment or a correlational study. It’s a quasi-experiment if it includes at least one SV that is going to be treated as if it were an IV; it’s a correlational study if all of the variables are either DVs or SVs being treated as covariates.

As defined in the chapter on data, an IV is a property, characteristic, or quality that is set by the experimenter. An SV is a stable difference between people that is difficult or even impossible to manipulate. A DV is probably best thought of as a response made by the subject or a behavior observed by the experimenter.

Experimental Factors

A **factor** is an IV that is being used within a certain kind of statistical analysis; *viz.*, ANOVA. The two terms, *factor* and *IV*, are not exactly synonymous: *IV* describes the role of some variable in the experiment (i.e., during data collection); *factor* describes the role of some variable during the analysis (i.e., after data collection). In 99% of the cases, the two are the same. It’s that last 1% that forces the two different labels.

By definition, a factor must have more than one level. (If it had only one level, then it wouldn’t be varying, so it wouldn’t be a variable, so it couldn’t be a factor.) As above, *level* is close but not quite identical to *value*. Sometimes two or more values of an IV are collapsed to create one level of a factor.

In general, there are two types of factor: within- and between-subjects. A **within-subjects factor** is one where each subject experiences all of the levels; a **between-subjects factor** is one where each subject experiences only one of the levels. The labels come from the fact that all of the comparisons across levels of a within-subjects factor occur *within* each of the subjects, while all of the comparisons across levels of a between-subjects factor must occur *between* subjects.

One complication to all this occurs when the level or levels of one factor depend on the level or levels of a different factor. In this case we say that the factors are **nested**; going farther, we often say that one factor is nested inside the other. For example, assume an experimental design with two factors, under which the first factor has three levels: a control condition and two experimental conditions that we’ll call A and B. Now assume that the specific manipulations that create Conditions A and B can be applied any number of times between one and four, and that this was done in the experiment. This creates the second factor -- that can be called “applications” -- with

four levels. However, the control condition does not involve applications, because nothing is done to the subjects in the control condition. Therefore, the second factor (applications) is nested inside of the first factor, because only the two experimental conditions involve this factor. Nested designs of this sort are complicated and rare, and will not be directly addressed. We will only be discussing **completely crossed** designs (which is a fancy way of saying “not nested”).

Another complication is when an IV or factor is partly within and partly between. An example would be an experiment with three conditions with each subject being run in two of the three. There is no technical name for this for this design. This is no generally accepted way to analyze the data from such an experiment. For your own sake, don't do this.

Design Types

If an experimental design includes only within-subjects factors, then it is referred to as being a **repeated-measures** or **within-subjects design**. If an experiment only includes between-subjects factors, then it is said to be using a **factorial** or **between-subjects design**. Finally, if there is at least one of each type of factor, then the experiment is said to employ a **mixed-factor design**.

One last set of labels applies only to between-subject factors and, therefore, to factorial and mixed-factor designs. If all of the groups defined by a factor have the same number of subjects, then the factor is **balanced**. If the group sizes are not all equal, then the factor is **unbalanced**. More generally, if all of the between-subject factors in the experimental design are balanced, then the design as a whole is said to be balanced. (Note: repeated-measures designs are balanced by definition, so we don't apply the term to them.) As we shall see, the results from unbalanced designs are often more difficult to analyze than those from balanced designs. Therefore, recognizing that a design is unbalanced is important.

Multiple Observations per Subject

If one is using a between-subjects design with only one DV (and no SVs), then one takes one and only one measure of each subject. This is increasingly rare in psychology. More often than not, there are at least two measures of each subject (and usually a lot more than that). In other words, we usually have multiple pieces of information about each of our subjects. Because this issue is best thought of as being an aspect of experimental design, and because it will become very important when we get to data management using SPSS (which is next), it will be covered here.

There are three, non-exclusive reasons why an experimenter can end up with more than one piece of information about a subject: (1) there was a within-subjects factor in the experiment (i.e., it used a repeated-measures or mixed-factor design), (2) there was more than one DV (e.g., both response time and accuracy were measured), and (3) at least one SV was measured in addition to the DV (e.g., in order to perform a covariance analysis).

In the case of within-subject factors, each of the pieces of information are of the same type. For example, if one used a two-level, within-subjects design to measure the effects of chocolate on anxiety, then there would probably be one measure of anxiety while the subject was “on” chocolate, and another measure of anxiety when the subject was “clean.” Because these are two

measures of the same thing (*viz.*, anxiety), they will have the same units of measurement and be directly comparable.

If there are two different DVs (e.g., anxiety and depression, or RT and accuracy), then there are two different pieces of information. These could easily have very different units of measurement and, therefore, be non-comparable. But this isn't often a problem, because more often than not the two different DVs are analyzed completely separately.

If there is an SV in addition to the DV (e.g., sex as an SV in a study of the effects of chocolate on anxiety), then there are, again, two very different pieces of information about the subject. As above, these could easily have very different units. However, similar to the above (but for a very different reason) this isn't often a problem, because SVs are almost never compared (directly) to DVs. Instead, they are used as covariates to "control for" differences between the subjects assigned to the different groups in a between-subjects experiment, or are used as an IV under a quasi-experimental design.

Correlational Studies

In many situations of great interest to psychologists, it is either impossible or unethical to manipulate a variable that is believed to have an important (maybe even causative) relationship with another variable. Under these conditions, one always conducts a correlational study (aka, a "natural experiment").

In a correlational study, each subject is measured on anywhere from two to several hundred variables, and the (often complex) pattern of relationships between the variables are quantified. This usually requires the data analyst to conduct a series of separate analyses where, in any given analysis, one or more variables are used to "predict" the value of one other variable. Across the whole series of analyses, this can become quite complicated, because the rôle that any given variable plays can change from being a "predictor" to being the "predicted" (and back again several times).

☞ At this point, we must face one of the most petty and annoying quibbles in all of statistics: how does one label the variables employed in correlational studies? By my definitions of *IV*, *DV*, and *SV*, nearly all of the variables in most correlational studies are DVs (with the rest being SVs), because all of the variables are measured, instead of being manipulated. Recall here that, by definition, an IV is a manipulated variable, while a DV or SV is a measured variable. The labels are set during data collection; they do not depend on the type of analysis. Going farther, it is quite possible to analyze the same data using both ANOVA and multiple regression/correlation (MRC); one sometimes even combines the two methods in a single set of analyses. For all of these reasons, I argue that we should label the variables in terms of what they are, not how they are being used.

⊕ Following from the last two paragraphs, I would suggest that we use the labels *predicted* and *predictor* to describe the rôle that a given variable is playing in a specific correlational analysis. Unfortunately (from my perspective, at least), very few people seem to agree with me. Both SPSS and the Cohens refer to predictor variables as "independent variables" and to the predicted

variable as the “dependent variable.” Out-numbered on this one (not to mention out-ranked), I hereby concede. Feel free to refer to predictors as IVs and to the predicted as the DV, if you wish. But be warned that I will probably continue to use my own terms. And please note why I do what I do: how a variable is being used in a particular analysis cannot change what the variable actually is; it was either set by the experimenter or it was measured, and nothing will ever change this.

Preview

As a quick, verbal preview of some topics to come: when the predicted variable is regressed onto only one predictor, we have bivariate correlation and/or simple linear regression (SLR). When more than one predictor is used, we have multiple regression/correlation (MRC). Many of the weaknesses that you may have heard attributed to correlational studies can be dodged, eliminated, and/or directly addressed using MRC. It is an extremely powerful technique. Returning, then, to the question of *what's better?* that was raised earlier: most people who say that experiments are always better than correlational studies probably don't know enough about statistics to be making any such general claims.

It might also surprise such people to learn that the statistical procedures that underlie ANOVA and MRC are *exactly* the same.