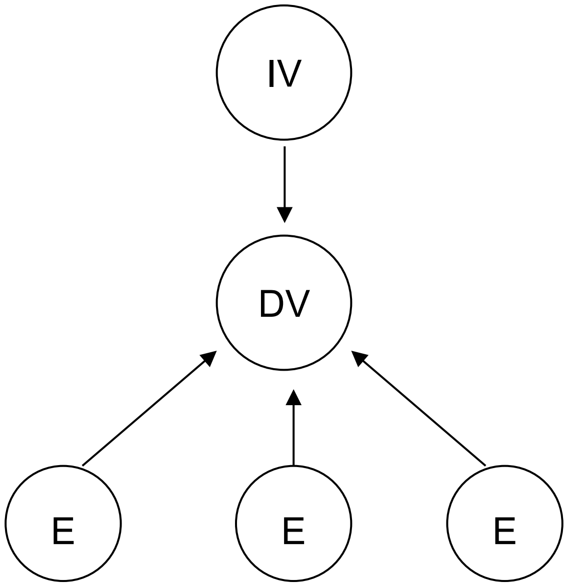
A Very Brief Introduction to Research Design©

Bivariate Experimental Research

Let me start by sketching a simple picture of a basic bivariate (focus on two variables) research paradigm.



"IV" stands for "**independent variable**" (also called the "**treatment**"), "DV" for "**dependent variable**," and "EV" for "**extraneous variable**." In **experimental research** we manipulate the IV and observe any resulting change in the DV. Because we are manipulating it experimentally, the IV will probably assume only a very few values, maybe as few as two. The DV may be categorical or may be continuous. The EVs are variables other than the IV which may affect the DV. To be able to detect the effect of the IV upon the DV, we must be able to control the EVs.

Consider the following experiment. I go to each of 100 classrooms on campus. At each, I flip a coin to determine whether I will assign the classroom to Group 1 (level 1 of the IV) or to Group 2. The classrooms are my "experimental units" or "subjects." In psychology, when our subjects are humans, we prefer to refer to them as "participants," or "respondents," but in statistics, the use of the word "subjects" is quite common, and I shall use it as a generic term for "experimental units." For subjects assigned to Group 1, I turn the room's light switch off. For Group 2 I turn it on. My DV is the brightness of the room, as measured by a photographic light meter. EVs would include factors such as time of day, season of the year, weather outside, condition of the light bulbs in the room, etc.

Think of the effect of the IV on the DV as a signal you wish to detect. EVs can make it difficult to detect the effect of the IV by contributing "**noise**" to the DV – that is, by producing variation in the DV that is not due to the IV. Consider the following experiment. A junior high school science student is conducting research on the effect of the size of a coin (dime versus silver dollar) on the height of the wave produced when the coin is tossed into a pool of water. She goes to a public pool, installs a wave measuring device, and starts tossing coins. In the pool at the time are a dozen rowdy youngsters, jumping in and out and splashing, etc. These youngsters' activities are EVs, and the noise they produce would make it pretty hard to detect the effect of the size of the coin.

Sometimes an EV is "**confounded**" with the IV. That is, it is entangled with the IV in such a way that you cannot separate the effect of the IV from that of the DV. Consider the pool example again. Suppose that the youngsters notice what the student is doing and conspire to confound her research. Every time she throws the silver dollar in, they stay still. But when she throws the dime in, they all cannonball in at the same time. The student reports back remarkable results: Dimes produce waves much higher than silver dollars.

Here is another example of a confound. When I was a graduate student at ECU, one of my professors was conducting research on a new method of instruction. He assigned one of his learning classes to be taught with method A. This class met at 0800. His other class was taught with method B. This class met at 1000. On examinations, the class taught with method B was superior. Does that mean that method B is better than method A? Perhaps not. Perhaps the difference between the two classes was due to the time the class was taught rather than the method of instruction. Maybe most students just learn better at 10 than at 8 – they certainly attend better at 8 than at 10. Maybe the two groups of students were not equivalent prior to being taught differently. Most students tend to avoid classes at 8. Upperclassmen get to register before underclassmen. Some people who hate classes at 8 are bright enough to learn how to avoid them, others not. Campbell and Stanley (1963) wrote about the importance of "achieving pre-experimental equation of groups through randomization." Note that the students in the research described here were not randomly assigned to the treatments, and thus any post-treatment differences might have been contaminated by pre-treatment differences.

Nonexperimental Research

Much research in the behavioral sciences is not experimental (no variable is manipulated), but rather "**observational**". Some use the term "correlational" to describe such a design, but that nomenclature leads to confusion, so I suggest you avoid it. Consider the following research. I recruit participants in downtown Greenville one evening. Each participant is asked whether or not e has been drinking alcohol that evening. I test each participant on a reaction time task. I find that those who report that they have been drinking have longer (slower) reaction times than those who were not drinking. I may refer to the drinking status variable as my IV, but note that it was not manipulated. In observational research like this, the variable that we think of as being a cause rather than an effect, especially if it is a grouping variable (has few values, as is generally case with the IV in experimental research), is often referred to as the IV. Also, a variable that is measured earlier in time is more likely to be called an IV than one measured later in time, since causes precede effects.

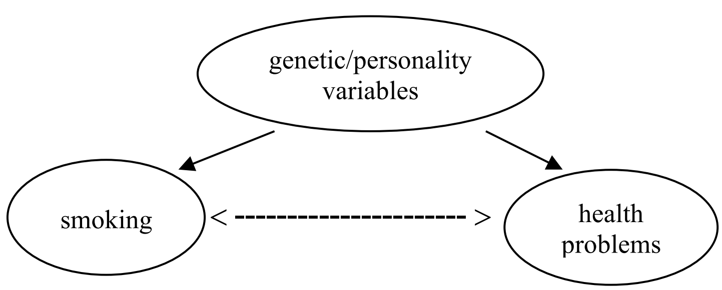
It is important, however, that you recognize that this design is observational, not experimental. With observational research like this, the results may suggest a causal relationship, but there are always alternative explanations. For example, there may be a "**third variable**" involved here. Maybe some people are, for whatever reason, mentally dull, while other people are bright. Maybe mental dullness tends to cause people to consume alcohol, and, independently of such consumption, to have slow reaction times. If that were the case, the observed relationship between drinking status and reaction time would be explained by the relationship between the third variable and the other variables, without any direct casual relationship between drinking alcohol and reaction time.

For my drinking research, I could do the statistical analysis with a method often thought of as being associated with experimental research, like a t test or an ANOVA, or with a method thought of as being associated with observational research, a correlation analysis. With the former analysis, I would compute t or F, test the null hypothesis that the two populations (drinkers and nondrinkers) have identical mean reaction times, and obtain a p, which, if low enough, would cause me to conclude that those two populations have different reaction times. With the latter analysis I would compute Pearson *r* (which is called a point biserial *r* when computed between a dichotomous variable and a continuous variable). To test the null hypothesis that there is, in the population, zero correlation between drinking status and reaction time, I would convert that *r* to a *t* and then to a *p*. If the *p* were sufficiently low, I would conclude that there is an association between drinking and reaction time. The value of *t* and of *p* would be exactly the same for these two analyses, because *t* tests and ANOVA are, quite simply, just special cases of correlation or multiple correlation analysis. Whether you can make a causal attribution or not depends not on the type of analysis done, but on how the data were collected (experimentally with adequate EV control or not). Some psychologists mistakenly think that one can never make firm causal inferences on the basis of a correlation analysis but that one always can on the basis of a *t* test or an ANOVA. These researchers have confused the "correlational" (better called "observational") research design with the correlation analysis. This is why I discourage the use of the term "correlational" when referring to a research design.

Do note that the drinking research could have been done experimentally. We could randomly assign participants to drink or not, administer the treatments, and then test their reaction time. Again, I could do the analysis via a *t* test or a Pearson *r*, and again the resulting *p* value would be identical regardless of statistical method. In this case, if I get a significant correlation between drinking and reaction time, I can conclude that drinking causes altered reaction time. In a nutshell, the demonstration of a correlation between variables X and Y is necessary, but not sufficient, to establish a causal relationship between X and Y. To establish the causal relationship, you have to rule out alternative explanations for the observed correlation.

More Examples of Third Variable Problems

Let me give you another example of a **third variable problem**. Observational research has demonstrated an association between smoking tobacco and developing a variety of health problems. One might argue that this association is due to a third variable rather than any causal relationship between smoking and ill health. Suppose that there is a constellation of third variables, think of them as genetic or personality variables, that cause some people to smoke, and, whether or not they smoke, also cause them to develop health problems. These two effects of the third variable could cause the observed associated between smoking and ill health in the absence of any direct causal relationship between smoking and ill health. The correlation between two variables that results from their being causally related to a third variable rather than causally related to each other is known as a **spurious correlation**.



How can one rule out such an explanation? It is not feasible to conduct the required experimental research on humans (randomly assigning newborns to be raised as smokers or nonsmokers), but such research has been done on rats. Rats exposed to tobacco smoke develop the same sort of health problems that are associated with smoking in humans. So the tobacco institute has promised not to market tobacco products to rats. By the way, there has been reported an interesting problem with the rats used in such research. When confined to a chamber into which tobacco smoke is pumped, some of them take their fecal boluses and stuff them into the vent from which the smoke is coming.

Side note. Researchers investigating the effects of cigarette smoke on rodents have encountered a problem – many of the subject respond "by placing feces in the smoke delivery tubing, repeatedly and in quantity. See Silverman's report on this phenomenon.

Another example of a third variable problem concerns the air traffic controllers strike that took place when Reagan was president. The controllers contended that the high stress of working in an air traffic control tower caused a variety of health problems known to be associated with stress. It is true that those working in that profession had higher incidences of such problems than did those in most other professions. The strikers wanted improved health benefits and working conditions to help with these stress related problems -- but the government alleged that it was not the job that caused the health problems, it was a constellation of third variables (personality/genetic) that on the one hand caused persons of a certain disposition (Type A folks, perfectionists) to be attracted to the air traffic controllers profession, and that same constellation of third variables caused persons with such a disposition to have these health problems, whether or not they worked in an air traffic control tower. One FAA official went so far as to say that working in an air traffic control tower is no more stressful than driving the beltway around DC. Personally, I find such driving very stressful. The government won, the union was busted. I suppose they could solve the problem by hiring as air traffic controllers only folks with a different disposition (Type B, lay back, what me worry, so what if those two little blips on the screen on headed towards one another).

- More Examples of Third Variable Problems

Karl L. Wuensch, September, 2021