

Independent Variable - Dependent Variable

Unedited posts from archives of CSG-L (see INTROCSG.NET):

Date: Wed Apr 28, 1993 6:20 am PST  
[From Bill Powers (930428.0700)]

General, on IV-DV: IV = Independent Variable; DV = Dependent Variable

The term "IV-DV" threatens to degenerate on this net into a stereotype of an approach to human behavior. All that this phrase means is that one variable is taken to depend on another and the degree and form of the dependence is investigated experimentally. This is a perfectly respectable scientific procedure. Want to know how the concentration of salt affects the boiling point of water? Keep the atmospheric pressure constant, carefully vary the salt content, and carefully measure the boiling point. You can find relationships like this throughout the Handbook of Chemistry and Physics, and so far nobody has suggested anything methodologically wrong with these tables and formulae.

If we're going to object to a procedure for investigating behavior, let's not indulge in synecdoche, but say exactly what it is about the method to which we object. There can be no valid objection to the IV-DV approach itself.

The basic problem with the IV-DV approach as used in the bulk of the behavioral sciences is that it is badly used; that bad or inconclusive measures of IV-DV relationships are not discarded, but are published. The basic valid approach has been turned into a cookbook procedure that substitutes crank-turning for analysis, thought, and modeling. The standards for acceptance of an apparent IV-DV relationship have been lowered to the point where practically anything that affects anything else, by however indirect and unreliable a path, for however small a proportion of the population, under however ill-defined a set of circumstances, is taken as a real measure of something important, and is thenceforth spoken of as if it were just as reliable a relationship as the dependence of the boiling point of water on the amount of dissolved salt.

While I was in Boulder, I spent some time in the library looking through a few journals. By chance, I looked first through two issues of the 1993 volume (29) of the Journal of Experimental Social Psychology. With few exceptions, the articles were of the form "the effect of A on B." One article went further: the title was "Directional questions direct self-conceptions."

All of the articles rested on some kind of ANOVA, primarily F-tests, and the justification for the conclusions was cited, for example, as " $F(1,82) = 7.88$ ,  $p < 0.01$ ." No individual data were given; it was impossible to tell how many subjects behaved contrary to the hypothesis or showed no effect. There was no indication, ever, that the conclusion was not true of all homo sapiens.

I suppose that a person who understood F-tests (how about some help, Gary) might be able to deduce the number of people in such studies who didn't show the effect cited as universal. Even I could see, in some cases, that there had to be numerous exceptions. For example, paraphrasing,

Subjects covertly primed rated John less positively ( $M = 21.32$ ) than subjects not primed ( $M=22.78$ ). Ratings were significantly correlated with the independent "priming" variable:  $r(118) = 0.35$ ,  $p < 0.001$ .

[Skowronski, J.J.; Explicit vs. implicit impression formation: The differing effects of overt labeling and covert priming on memory and impressions." J. Exp. Soc. Psychol \_29\_, 17-41 (1993)]

When means differ by only 1.46 parts out of 22, it's clear that many of the 120 students must have violated the generalization, so this conclusion would be true of something close to half of the students. The coefficient of uselessness is 0.94, showing the same thing. The authors are teasing a small effect out of an almost equal number of examples and counterexamples. In another study, "When warning succeeds ..." a rating scale ran from -5 to 5, and the mean self-ratings for one case were 0.89 and in the other -0.92. A

large number of the subjects must have given ratings in the opposite order from the one finally reported.

So what we're talking about here is not a bad methodology, but bad science based on equivocal findings.

The IV-DV approach is not incompatible with a model-based approach or with obtaining highly reliable data. In the Journal of Experimental Psychology - General, I found a gem by Mary Kay Stevenson, "Decision-making with long-term consequences: temporal discounting for single and multiple outcomes in the future" (JEP- General, 122 #1, 3-22 (1993). Mary Kay Stevenson, 1364 Dept. of Psychology, Psychological sciences building, Purdue University, W. Lafayette, Indiana 47907-1364

This paper used old stand-bys like questionnaires and rating scales, but it had some rationale in the observation that during conditioning, delaying a consequence of a behavior lowers the strength of the conditioning. It also freely postulated a thinking organism making judgements -- this was actually an experiment with high-level perceptions. Moreover, there was a systematic model behind the analysis, and an attempt to fit an analytical form to the data rather than just do a standard ANOVA.

Furthermore -- oh, unheard-of procedure -- Ms. Stevenson actually replicated the experiment with 5 randomly-selected individuals, fitting the model to each individual's data and verifying that the curve for each one was concave in the right direction.

The mathematical model predicted between 97 and 99 percent of the variance in the data.

I didn't have time to read the article carefully, but it certainly seemed to show that high standards were applied and that an IV-DV approach can yield data that anyone would call scientific. All that's required is that one think like a scientist. A LOT of work went into this paper. If only papers in psychology done to this standard were published, all the different JEPs would fit into a single issue.

In JEP-Human Perception and Performance, there was a good control-theory experiment:

Viviani, P. and Stucchi, N. Behavioral movements look uniform: evidence of perceptual-motor interactions (JEP-HPP 18 #3, 603- 623 (August 1992).

Here the authors presented subjects with spots of light moving in ellipses and "scribbles" on a PC screen, and had them press the ">" or "<" key to make the motion look uniform (as many trials as needed). The key altered an exponent in a theoretical expression used to relate tangential velocity to radius of curvature in the model. The correlation of the formula with an exponent of 2/3 (used as a generative model) with the subjects' adjustments of the exponent was 0.896, slope = 0.336, intercept 0.090.

This is just the kind of experiment a PCTer would do to explore hypotheses about what a subject is perceiving. By giving the subject control over the perception in a specified dimension, the experiment allows the subject to bring the perception to a specified state -- here, uniformity of motion -- and thus reveals a possible controlled variable (at the "transition" level?). The authors didn't explain what they were doing in that way, but this is clearly a good PCT experiment. Even the correlation was respectable, if not outstanding (the formula was rather arbitrary, so it should be possible to improve the correlation considerably by looking carefully at the way the formula misrepresented the data).

There is a world of difference between the kinds of experiments reported in J Exp. Soc. Psych and the two described above (and between the two described above and most of the others in JEP). From good experiments, even if one doesn't buy the interpretation, one can go on to better experiments. From bad experiments there is no place to go: you say "Oh" and go on to something completely different.

Best, Bill P.