ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Current Directions in Psychological Science
20(2) 115-118

DOI: 10.1177/0963721411402670
http://cdps.sagepub.com
SAGE

Evaluating Individual Differences in Psychological Processes

Daniel J. Bauer
The University of North Carolina at Chapel Hill

Abstract

Better understanding individual differences in social, cognitive, and behavioral processes is a core goal of much psychological theory and research. Although great progress has been made toward this goal, I argue here that the classical design and analysis approach that dominates individual difference research, namely, the collection of single-time-point data and application of standard linear regression models, potentially limits further advances. In particular, the opportunity to evaluate individual differences in psychological processes is restricted by the estimation of a single effect to represent the relationship between variables. I discuss alternative analysis and design options that offer the opportunity to more fully examine individual differences in psychological processes.

Keywords

individual differences, intra-individual variability, within-person effects, multilevel models, finite mixture models, moderation, interaction

A longstanding aim of many areas of psychological research is to understand individual differences. Indeed, the field of differential psychology—concerned with the study of individual variation in personality, cognition, and behavior—was first established by Sir Francis Galton, half-cousin of Charles Darwin, more than a century ago. Many of the current subfields of psychology owe a theoretical debt to Galton, who drew on Darwin's theories of evolution and heredity to change the then-dominant view that individual differences were simply errors about the mean—random idiosyncrasies that needed to be averaged over to arrive at the truth. In contrast, the theory of evolution accorded individual differences theoretical importance as grist to the mill of evolution. Galton also developed statistical tools for conducting research on individual differences, most notably the linear regression model. Interestingly, although a century of theoretical developments have transpired since Galton's seminal contributions, psychological research on individual differences still relies quite heavily on the linear regression model (and its latent variable counterpart, the structural equation model). In part, this continued reliance on linear regression reflects the general utility of the model. But statistical models can also become entrenched in an area of research through habit and tradition, and the limitations of the models can then impede theory evaluation and scientific progress. In particular, the standard linear regression model provides limited information when one's goal is to understand individual differences in psychological processes.[1]

A key assumption of the linear regression model is that the variables are related to one another in the same way for all individuals. As typically applied, the linear regression model presumes that the same relationship between X and Y—say, negative affect and alcohol use—holds for everyone. Yet psychological theory often posits that the relationship between X and Y will be stronger for some people than for others—that is, that there will be individual differences in the effect of one variable on another. In some cases, we may think we know the source of this variation. For instance, we may think that there is a stronger relationship between negative affect and alcohol use for adolescent girls than for adolescent boys. In other cases, the source of variation may only be partly explained by theory. We may then wish to know: How much does the relationship vary? And how much of this variation do we have yet to explain? In the remainder of this article, I will describe several extensions of the linear regression model that admit individual differences not only in the level of characteristics but also in how these characteristics relate to one another. Analytically, these methods allow for individual differences in the effects of one variable on another. Put into practice, they thus offer the

Corresponding Author:
Daniel J. Bauer, Department of Psychology, CB#3270, The University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-3270
E-mail: dbauer@email.unc.edu

opportunity to obtain a more comprehensive understanding of individual differences in psychological processes.

The Linear Regression Model

Let us first more carefully examine the limitations of the linear regression model as it is typically applied. Data are collected at one point in time from a sample of N individuals on two or more characteristics. To continue the earlier example, perhaps we have measured alcohol use (Y) and negative affect (X). We might then formulate the linear regression model

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

where $i = 1, 2, \ldots, N$; β_0 is the intercept and β_1 is the slope of the regression line; and e_i is the error of prediction. Suppose we find, upon conducting the regression analysis, that negative affect predicts alcohol use—that is, our estimate for β_1 is statistically significant. The estimated value of this coefficient quantifies the relationship between our two variables. Note, however, that β_1 is identical for all individuals—it is a constant. That is, although alcohol use and negative affect vary over persons, we assume that there are no individual differences in the *relationship* between these two variables. As noted earlier, however, we often believe that the relationships between our variables differ over individuals.

Moderation Models

The most common approach to evaluating whether different relationships hold for different individuals is moderation analysis. Known or hypothesized sources of effect heterogeneity are incorporated into the model, usually via interaction terms. The resulting model permits the effect of X on Y to differ across levels of the moderator variable Z. In other words, individual differences exist in the effect of X on Y, and these differences are determined by the characteristic Z. For instance, suppose that we indeed believe that the relationship between negative affect (X) and alcohol use (Y) is stronger for girls ($Z = 1$) than for boys ($Z = 0$). We could then extend our regression model by incorporating the XZ product interaction term, as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \beta_3 X_i Z_i + e_i$$

Grouping the terms of the model illustrates how Z moderates the effect of X:

$$Y_i = (\beta_0 + \beta_2 Z_i) + (\beta_1 + \beta_3 Z_i) X_i + e_i$$

The second set of parentheses includes the terms that describe the effect of X on Y, which is dependent on the value of Z. Individuals who differ with respect to Z will thus also differ in their relationship between X and Y. For our hypothetical research study, the effect of negative affect on alcohol use for girls would be $\beta_1 + \beta_3$, whereas for boys the effect would simply be β_1. If the estimate for β_3 is statistically significant, then our hypothesis, that the effect of negative affect on alcohol use differs across these subgroups, is supported.

Moderation analysis has proven enormously fruitful in many areas of psychological research over the past several decades, a trend that has been greatly aided by the publication of several didactic books that illustrate best practice for testing and presenting interaction effects in linear regression (Aiken & West, 1991; Cohen, Cohen, West, & Aiken, 2003; Jaccard & Turrisi, 2003).[2] This approach to studying individual variability in relationships is, however, limited in one critical respect: The effect of X on Y varies only with Z (or, in more complex models, with multiple Zs). If the effect of X on Y varies across persons for reasons other than Z, the model will not reveal these differences. Thus, if we are interested in the full range of individual differences in the effect of X on Y, then an alternative modeling approach may be called for.

Finite Mixture Models

The finite mixture model allows effects to vary across persons because of unknown sources of heterogeneity. Considering a relatively simple case, the finite mixture regression model posits that there are K classes of individuals within a population and that the relationship between X and Y differs over classes (DeSarbo & Cron, 1988). Our regression equation is now

$$Y_i = \beta_{0c} + \beta_{1c} X_i + e_i$$

where $c = 1, 2, \ldots, K$. The regression coefficients are subscripted by c to indicate that they can take on different values for each class. Thus the effect of X and Y differs over classes. What distinguishes this model from a typical moderation analysis is that the classes are unobserved or latent rather than predefined by measured variables. In other words, this model posits that there are individual differences in the effect of X on Y, but it does not presume that we know the source of these differences.

Many applications of finite mixture regression have been conducted in the area of market segmentation research. For instance, food marketers might wish to differentiate a "health conscious" class of individuals that evaluate products based on nutritional information from a "hedonistic" class that is more heavily influenced by perceived taste. But mixture models are also potentially useful in other areas of psychological research. For instance, we might use such a model to investigate individual differences in the relationship between negative affect and alcohol use. We might posit that subgroups of individuals drink alcohol for different reasons. Perhaps in one class, the relationship is positive (e.g., self-medicators), in another class the relationship is negative (e.g., social drinkers), and in a third class there might be no relationship at all between the two variables (e.g., teetotalers). Class membership might be influenced by whether an individual is male or female but would not be entirely predictable or known (as a typical moderation analysis would assume). Longitudinal mixture applications, which differentiate classes of individuals that have different growth trajectories, are also becoming quite common (Bauer, 2007; Muthén, 2001; Nagin, 1999).

To use finite mixture models, we need not necessarily assume there are literally K distinct subgroups in the population.

Instead, we can view the K latent classes as points along a continuum. In other words, we might believe that individuals differ quantitatively from one another in the relationship between X and Y but that much of this variation can be succinctly summarized via K classes (Nagin, 2005). The three alcohol-use classes in our hypothetical example (i.e., self-medicators, social drinkers, and teetotalers) would then be viewed merely as an expedience, a simple way to represent the potentially vast variability that exists in the relationship between negative affect and alcohol use. Nevertheless, there is likely to be much information lost by reducing individual variation to K classes (in applications, K is usually held to be 3 or 4). Another analysis approach offers the opportunity to preserve this variation.

Random Effects Models

Thus far, we have assumed that data have been collected using the classic individual difference design in which multiple individuals are measured at a single point in time. To fully understand individual differences in psychological processes, however, a better approach is to collect repeated-measures data. The logic behind this design shift is intuitive. Ordinarily, we collect one observation of X and Y per person and pool the data over persons to permit estimation of the regression line. Now, however, we collect multiple observations on X and Y for each person to enable estimation of an individual-specific regression line, with an intercept and slope that is unique to the person. We thus allow for the full range individual differences in the effect of X on Y.

A great advantage of obtaining repeated measures is that we can separate variability that resides within the person (intraindividual variability) from variability across persons (interindividual variability), and thereby isolate the level at which effects operate. More typically, these two sources of variability are confounded. For instance, suppose we conduct a conventional study in which we measure negative affect and alcohol use once for a sample of N individuals. We wish to evaluate the self-medication hypothesis: Does the experience of negative affect positively predict drinking behavior? We conduct a standard regression analysis and find a statistically significant effect. This effect tells us that individuals reporting higher negative affect tend to report higher alcohol use. But we do not know whether, when a given individual experiences greater negative affect than usual, this individual will also tend to drink more. The latter effect, the *within-person* effect, is more pertinent to our hypothesis, but our data do not provide sufficient information for us to estimate this effect. In the simple regression analysis, it is confounded with the *between-person effect*: Individuals with higher average levels of alcohol use may also have higher average levels of negative affect. By collecting multiple observations per person over time, however, we can parse intraindividual and interindividual variability, permitting us to separate within-person and between-person effects, respectively. In particular, we are interested in these questions: Do episodic increases in negative affect predict increased consumption of alcohol, as supposed by the self-medication hypothesis? And might this relationship be stronger for some people than for others?

To answer these questions, we can posit the following model

$$Y_{ti} = \beta_{0i} + \beta_{1i}\dot{X}_{ti} + e_{ti}$$

where $t = 1, 2, \ldots, T_i$ indexes the time point at which the observation was taken. The presence of repeated measures allows us to deviate from the standard regression model in several ways. First, the notation $\dot{X}$ indicates that the predictor has been person-mean centered. In person-mean centering, the mean value of X is calculated separately for each person and then subtracted from that person's scores—that is, $\dot{X}_{ti} = X_{ti} - \bar{X}_{.i}$, where the person mean is calculated as $\bar{X}_{.i} = \sum_{t=1}^{T_i} X_{ti}/T_i$. The purpose of this procedure is to produce a version of the predictor that contains only within-person variability so that β_{1i} unambiguously represents a within-person effect.[3] Thus, when person i experiences a one-point increase in X, we expect that person i will also experience a β_{1i}-unit change in Y. Second, note the introduction of the i subscript on the regression coefficients, to designate that the intercept and slope have unique values for each person. The same amount of change in X may thus produce different changes in Y for different people. This difference from the standard regression model is particularly exciting, as it allows us to consider individual differences in the way that variables are related to one another within persons, providing insight into individual differences in psychological processes.

With sufficiently many repeated measures per person, we could literally conduct a separate linear regression analysis for each person to obtain estimates of β_{0i} and β_{1i}. But we would ideally like to make conclusions about individual differences in the population as a whole, not just persons $1, 2, \ldots, N$ of the sample. The random effects model (also called the mixed model, multilevel model, or hierarchical linear model) allows us to do just this (Raudenbush & Bryk, 2002). Rather than estimate separate values of β_{0i} and β_{1i} for each person in the *sample*, we estimate characteristics of the distribution of β_{0i} and β_{1i} that holds in the *population*. We can thereby determine how much the relationship between X and Y differs across persons. We can also include predictors of slope variation in the model. In our investigation of the effect of negative affect on alcohol use, for instance, we could profitably use these models to first establish whether the relationship between negative affect and alcohol use is the same for everyone (i.e., is the effect fixed or constant?) or whether there are individual differences in this relationship (i.e., is the effect random, varying across persons?). Then, provided there are individual differences in the effect of negative affect on alcohol use, we could attempt to explain these differences on the basis of known characteristics of the individual, such as gender.

Random effects models are by no means new to psychology, but applications of random effects models have generally been restricted to one of two types. The first type of application is to clustered data—that is, single-time-point data on individuals clustered within groups (e.g., students within schools). The second type of application is in growth modeling: Y is an outcome of interest, and X measures the passage of time, so that β_{0i}

and β_{1i} characterize individual trajectories of change over time. Only recently have psychologists begun to collect intensive repeated-measures data over short time intervals in order to look at individual differences in relationships between variables. This new and promising application of random effects models to intensive short-term longitudinal data has been greatly aided by the development of new technologies (e.g., online data entry, personal digital assistants) and new data collection designs (e.g., experience sampling techniques; Stone, Shiffman, Atienza, & Nebeling, 2007). As the trend toward collecting repeated measures increases, the creative application of random effects models to this data holds much promise for individual differences research (Curran & Bauer, 2011).

Conclusions

The analysis approaches I have discussed offer us the opportunity to more fully evaluate individual differences in psychological processes. They take us beyond the design and analysis approaches that have characterized individual difference research for more than a century. Each approach is based on well-established statistical theory. But methodological innovations continue and must be appraised by the scientific community for their potential to improve psychological research. In particular, Molenaar (2004) has argued in favor of a more ideographic approach to evaluating individual differences in within-person processes. Consideration of such innovations is critical if we are to avoid establishing new habits of design and methodology that ultimately again fail to keep pace with theoretical developments. Nevertheless, in the short term, fuller use of the analysis approaches reviewed here, and particularly the random effects modeling approach with repeated measures data, has the potential to greatly advance research on individual differences.

Notes

1. Of course, the standard linear regression model is still useful for addressing other research questions.
2. The multiple-groups structural equation model is another vehicle for conducting similar moderation analyses.
3. We could expand this equation to also include the between-person effect—that is, how individual differences in average levels of X (person means) are predictive of individual differences in average levels of Y, but this is seldom of as much interest as the within-person effect (see Kreft, de Leeuw, & Aiken, 1995; Raudenbush & Bryk, 2002).

Declaration of Conflicting Interests

The author declared that he had no conflicts of interest with respect to his authorship or the publication of this article.

Recommended Reading

Aiken, L.S., & West, S.G. (1991). (See References). Provides an excellent overview of how interactions are specified and evaluated in linear regression models.

DeSarbo, W.S., & Cron, W.L. (1988). (See References). Offers a useful, though slightly technical, introduction to finite mixture regression models.

Everitt, B.S., & Hand, D.J. (1981). *Finite mixture distributions.* London, England: Chapman & Hall. A well-written and generally accessible overview of finite mixture models.

Schwartz, J.E., & Stone, A.A. (1998). Strategies for analyzing ecological momentary assessment data. *Health Psychology*, *17*, 616. Provides a concise introduction to the use of random effects models with intensive longitudinal data.

References

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bauer, D.J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, *42*, 757–786.

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Curran, P.J., & Bauer, D.J. (2011). The disaggregation of within-person and between-person effects in longitudinal models of change. *Annual Review of Psychology*, *62*, 583–619.

DeSarbo, W.S., & Cron, W.L. (1988). A maximum likelihood methodology for clusterwise linear regression. *Journal of Classification*, *5*, 249–282.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression* (2nd ed.). Thousand Oaks, CA: Sage.

Kreft, I.G., de Leeuw, J., & Aiken, L. (1995). The effect of different forms of centering in hierarchical linear models. *Multivariate Behavioral Research*, *30*, 1–21.

Molenaar, P.C.M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspective*, *2*, 201–218.

Muthén, B.O. (2001). Second-generation structural equation modeling with a combination of categorical and continuous latent variables: New opportunities for latent class/latent growth modeling. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 291–322). Washington, DC: American Psychological Association.

Nagin, D.S. (1999). Analyzing developmental trajectories: A semi-parametric, group-based approach. *Psychological Methods*, *4*, 139–157.

Nagin, D.S. (2005). *Group-based modeling of development*. Cambridge, MA: Harvard University Press.

Raudenbush, S.W., & Bryk, A.S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Stone, A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). *The science of real-time data capture: Self-reports in health research*. New York, NY: Oxford University Press.