

Extreme Groups Designs

Kristopher J. Preacher

Vanderbilt University

It is common for psychological researchers to examine the relationship between two variables X and Y by first selecting cases on the basis of extreme X scores (e.g., X scores situated in the outer tertiles or quartiles of the score distribution) and then using X to predict Y . For example, a researcher may administer a depression scale to a participant pool, randomly select participants from the lower and upper 25% of the distribution of depression scores, and then use depression as a predictor of alcohol consumption for the sampled cases. This sampling procedure has been termed the *extreme groups approach* (EGA), and it results in an *extreme groups design*. The extreme groups design can be seen as a special case of oversampling, in which the outer portions of a distribution are sampled to the complete exclusion of the central portion. The extreme groups design also represents a special case of restriction of range, in which the range of sampled X scores is restricted to the extremes of the distribution of X , with predictable consequences for subsequent statistical analyses. In addition, the extreme groups design is also a special case of observational or quasi-experimental design because participants are measured on X , not randomly assigned to treatment conditions.

Several modifications of EGA exist. For example, the extreme groups need not be equal in size, nor must they span the same range of scores in the lower and upper extremes of the score distribution. Most often, the extreme groups are chosen on the basis of sample quantiles, but also can be selected on the basis of conventional cutoff points or features of the score distribution. Often, the extreme groups, once selected, are assigned codes for "low"

(0) vs. "high" (1) prior to analysis, creating two groups, each of which is assumed to be homogeneous on X . EGA does not describe studies that limit attention to one extreme or the other (such as studies of academically gifted children). Nor does it describe designs in which data are collected across the full range of X and then discarded from the middle of the X distribution.

Past Investigations of Extreme Groups Designs

EGA has a long history in the literature on research design and applied statistics. Early uses in the 1930s through 1950s were motivated by attempts to enhance statistical power of subsequent tests (Kelley, 1939), simplify the task of fitting a line to data (Bartlett, 1949; Gibson & Jowett, 1957; Wald, 1940), and reduce the cost of data collection (Peters, 1941; Peters & Van Voorhis, 1940). These reasons likely number among the most popular reasons for employing EGA in more modern times as well.

Consequences of Using Extreme Groups Designs

Cost-Efficiency

Cost—in terms of time, money, or labor—is often a limiting factor in data collection. Assuming a fixed sample size and a desired level of power for detecting a relationship, EGA can improve cost-efficiency by reducing the sample size required to achieve the desired power. Improved cost-efficiency is one of the few clear benefits of EGA, but it is beneficial only if the researcher is interested in hypothesis testing and is willing to assume that the same model that applies to the extremes of the X distribution also applies to the omitted middle.

Statistical Power

Power for the test of the correlation of X and Y is generally enhanced after using EGA on X . All else being equal, power increases with the extremity of selection on X . For example, power for the test of r_{XY} will be greater if N cases are selected from the outer quartiles than if they are selected from the outer tertiles. This difference in power is more pronounced for smaller N . Many studies spanning the 1920s to 1970s (e.g., Feldt, 1961) showed that the power for a t test comparing the mean of Y across groups created on the basis of extreme X scores is maximized when the proportion of cases drawn from each extreme is approximately 27%, nearly a quartile split (this came to be known as the "twenty-seven percent rule"). However, higher power can be achieved by not grouping after EGA, instead preserving the individual differences within each extreme of X and using correlation or regression to gauge the relationship between X and Y (Alf & Abrahams, 1975).

Effect Size

EGA has no biasing effect on the covariance or regression slope relating X and Y . However, EGA usually increases measures of standardized effect size (e.g., r , R^2) relative to using data from the full range of X (Cortina and DeShon 1998). This inflated effect size leads to higher power for tests of linear association, but also gives a misleading picture of the size or practical significance of the effect. Thus, standardized effect sizes derived from extreme groups designs should not be generalized to the population or used in meta-analyses. This inflationary effect is most pronounced for correlations near .5 and is attenuated as r nears 0 and ± 1 (Preacher, Rucker, MacCallum, and Nicewander, 2005).

Reliability

EGA is often thought to improve reliability through selection of the X scores perceived to be the most reliable (the extremes). However, the opposite is generally true; it can be shown through application of item response theory

principles that scores in the *interior* of the X distribution tend to be those that are the most reliable (Preacher et al. 2005). In addition, scores in the extremes of the X distribution are those that are most vulnerable to regression toward the mean. That is, the extreme X scores are those that are the most liable to change by a noteworthy amount upon repeated measurement, and thus have the lowest test-retest reliability.

Model Misspecification

EGA removes from consideration the middle portion of the distribution of X . Thus, the (typically linear) relationship that is observed to hold between X and Y when using extreme groups must be assumed to also hold for cases in the omitted portion of the X distribution. Because this segment of the population has been omitted from consideration, it is not possible to explore a possibly nonlinear relationship. EGA typically increases the power of subsequent tests of linear relationships between X and Y , but at the expense of rendering the detection of nonlinear effects improbable.

Group Assignment

Sometimes the extreme scores of a continuous X distribution are recoded into a binary (0,1) variable to represent groups that are low and high on X , which then are used in subsequent analyses (e.g., a t test for group means on Y). There are multiple arguments against this practice. It assumes that (1) two underlying categories really exist and that observed individual differences reflect only measurement error rather than meaningful individual differences; (2) group assignment is accurate; (3) the proportions of cases in each group are equal in the population; and (4) group constituency is stable over samples, contexts, and occasions. Furthermore, true dichotomies are arguably rare in psychological research. The implication is that, if EGA is to be used, individual differences within each extreme should be preserved, and regression-based methods should be used rather than a t test. If the existence of unobserved groups is plausible, more suitable

methods exist for determining whether scores on X reflect a discrete vs. continuous variable.

Interaction Effects

Sometimes EGA is used to create subgroups for use in subgroup analysis, a common tool for investigating interaction effects. In subgroup analysis, groups are created on the basis of extreme scores on a moderator variable Z , and the relationship between X and Y is assessed in each group. If the relationship of X and Y is found to differ between Z groups, an interaction is said to exist. Whereas this procedure is intuitive and simple to implement, the creation of subgroups on Z has many of the limitations mentioned above for group creation. Methodologists have suggested more appropriate and powerful methods for detecting interactions that do not require the creation of arbitrary groups, such as moderated multiple regression (MMR) with product terms. If EGA is used in conjunction with MMR—that is, if extreme cases are sampled on both X and Z , and these extreme scores are treated as continuous predictors in a regression analysis predicting Y from X , Z , and the product XZ —the power for detecting interaction effects can be greatly enhanced (McClelland and Judd, 1993).

Potential Applications of Extreme Groups Designs

On balance, the usefulness of EGA as a methodological tool is somewhat limited. Several costs have been described above. However, there are valid reasons researchers may want to employ extreme groups designs in some situations. As indicated, costs associated with data collection may prohibit the examination of data from the full range of X . In such cases, EGA can be used to increase the power for detecting a linear association given a fixed sample size. EGA can be useful for enhancing power not only for detecting bivariate relationships, but also interactions. It should be borne in mind, however, that using EGA to enhance power for a fixed N carries with it several dangers and assumptions. Typically, only linear associations

may be explored after EGA, reliability is almost always reduced relative to using data from the full range of X , and the researcher is obligated to avoid interpreting standardized effect size. Due to its power-enhancement properties, EGA may find its best use in pilot studies, where the goal is to detect promising trends in small samples that may warrant further investigation. When EGA is used, its use should be thoughtfully considered and explicitly justified. Researchers who are unaware of the limitations of extreme groups designs are in danger of overstating results.

A method closely related to EGA is *post hoc subgrouping*. In contrast to EGA, which involves collecting N data points from a restricted range of X scores, post hoc subgrouping involves deleting from a data set those cases having X scores in the interior of the X distribution, and sometimes dichotomizing the remaining cases into low and high groups. This practice shares the disadvantages associated with EGA and lacks its advantages. Unlike EGA, subgrouping does not improve cost-efficiency (the data have already been collected). Second, because data are discarded in subgrouping, the power to detect linear relationships relative to analyses using the full range of X scores is diminished.

SEE ALSO: Construct Validity; Correlational Designs; Distinguishing Categories from Continua; Effect Size; Item Response Theory; Quasi-Experiments; Psychopathology Research, Selection of Subjects in; Statistical Power

References

- Alf, Jr, E. F., & Abrahams, N. M. (1975). The use of extreme groups in assessing relationships. *Psychometrika*, *40*, 563–572.
- Bartlett, M. S. (1949). Fitting a straight line when both variables are subject to error. *Biometrics*, *5*, 207–212.
- Cortina, J. M., & DeShon, R. P. (1988). Determining relative importance of predictors with the observational design. *Journal of Applied Psychology*, *83*, 798–804.
- Feldt, L. S. (1961). The use of extreme groups to test for the presence of a relationship. *Psychometrika*, *26*, 307–316.

- Gibson, W. M., & Jowett, G. H. (1957). 'Three-group' regression analysis: Pt. I. Simple regression analysis. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 6, 114–122.
- Kelley, T. L. (1939). The selection of upper and lower groups for the validation of test items. *Journal of Educational Psychology*, 30, 17–24.
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, 114, 376–390.
- Peters, C. C. (1941). A technique for correlating measurable traits with freely observed social behaviors. *Psychometrika*, 6, 209–219.
- Peters, C. C., & Van Voorhis, W. R. (1940). *Statistical procedures and their mathematical bases*. New York: McGraw-Hill.
- Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods*, 10, 178–192.
- Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *Annals of Mathematical Statistics*, 11, 284.

Further Reading

- Abrahams, N. M., & Alf, Jr, E. F. (1978). Relative costs and statistical power in the extreme groups approach. *Psychometrika*, 43, 11–17.
- Borich, G. D., & Godbout, R. C. (1974). Extreme groups designs and the calculation of statistical power. *Educational and Psychological Measurement*, 34, 663–675.