David A. Kenny
March =1, 2003

Unit of Analysis

This page provides the practicing researcher with =uidance concerning the unit in the statistical analysis. I thank Charles =udd for helping me with many of the ideas on this page. Any feedback, =ither technical or pedagogical, would be most appreciated.

Outline =BR> **Statement of the Problem**
Independence of =nits
=/FONT>The Measurement of Nonindependence
Unit of =eneralization
=/FONT>Unit of =easurement
=/FONT>Unit of Assignment or Sampling
How Do I Conduct the Analysis?
References

Statement =f the Problem

In statistical analysis, =t is sometimes not clear what is the appropriate level of analysis. For =nstance, persons are in groups (e.g., children in classrooms), and either person =r group could be the unit of analysis. (The group would be the unit of =nalysis by computing a mean of those persons who are members of the group.) =ometimes the two units are crossed instead of nested; for example, 30 judges rate =0 targets. Either target, rater, or even observation could be the =nit of analysis. Because nesting (e.g., children in classrooms) is much =ore common than crossing, that case is generally assumed in the following discussion.

Independence of Units

At the heart of statistical =nalysis is replication or the repeated observation of a phenomenon. For a replication to be a true replication, there must be independence of observations. (For example, duplicating your data is not replication!) Independence of observations is presumed in standard =measures of variability. For there to be

independence, two =bservations are no more likely to be similar (or different) than any other two observations. There are several factors that make units =onindependent (Kenny & Judd, 1986). Observations can be nonindependent =ecause of compositional effects, common fate, and social interaction:

- Compositional effects refer to the fact that sets =f observations are already similar before the study even begins.
- Common fate refers to the fact that sets of =bservations may have common causes.
- Social interaction refers to direct and indirect =nfluence between pairs of observations.

Using =ath analysis notation, a compositional effect is a curved line between a =air of observations, common fate is spuriousness (the observation caused by =ommon variable), and social interaction is a direct effect. The =onindependence would be positive if the nonindependent observations were more similar =han independent observations; the nonindependence would be negative if the nonindependent observations were more different than independent observations. The degree of nonindependence can be viewed as a =orrelation coefficient, though it is not usually measured by an ordinary Pearson product-moment correlation.

The =easurement of Nonindependence

To determine the =nit of analysis, an assessment of whether observations are independent is often =helpful. That is, the observations that are thought to be =onindependent, may in fact be independent. The measurement of nonindependence can be =omplicated, but in many cases an intraclass correlation can be used to measure the =egree of nonindependence. (Read =bout this measure for dyads.) This measure is appropriate when groups of =observations are all linked to one another. Kenny and Judd (1996) =iscuss a wide variety of measures of nonindependence.

Unit of Generalization

Another factor in =eciding the unit of analysis is the level of generalization that the researcher =eeks to make. Consider a researcher who measures 10 children in 10 classrooms =rom 10 different schools, or 1000 children in all. There are three =ossible levels of generalizations: the student, the classroom, and the school. =ne simple rule is to conduct the analysis at the level at which one wants =o make generalizations. So if one wants to draw conclusions about =ersons, person should be the unit of analysis. However, as will be seen, this =imple rule cannot always be followed.

The researcher should be aware of the *ecological =allacy* (Robinson, 1950). The conclusions drawn from an analysis conducted at a =roup level may not apply at the individual level. Conversely, analyses =t the individual level may not apply to the group level. In principal, =he analysis should be conducted at the level at which generalizations =hould be made. However, there are exceptions to this rule.

Unit of Measurement

Another consideration =s the unit of measurement. Again returning to the example of children, =lassroom, and school, some variables may be measured on children (e.g., achievement), =ome on the classroom (e.g., teacher's gender), and some on the school (e.g., =chool size). Just because one measures a variable at a certain level =oes not imply that the variable operates at that level. Consider the =ariable group size. Presumably this variable operates at the group =evel. However, if a researcher changed the unit of measurement of the variable =nd asked persons how big the group was, the variable will still likely =perates at the group level, not at the individual level.

A related issue is that sometimes a researcher =ggregates across units (i.e., averages) and so changes the unit of =easurement. For example, to measure organizational climate, the mean of individual =easures might be used. Just because the mean is at the level of the =rganization, does not mean that it, in fact, operates at that level.

Unit of =ssignment or Sampling

A final consideration in =he decision about the unit of analysis is design factors. It is =ecessary to consider the unit by which observations are selected to enter the study =r are assigned to levels of the independent variable. A good idea is to =erform the statistical analysis at the level of the selection or assignment. =o, for instance, if floors in a dormitory are assigned to experimental =onditions, dormitory floor, not person, should be the unit of analysis. This is not = "hard-and-fast rule," just a helpful guideline. For instance, =ndividuals may be the unit of assignment, but if individuals interact with one another, =hen it may not be possible to use individual as the unit of analysis.

How Do I =onduct the Analysis?

There are three major =pproaches to the unit of analysis question when persons are nested within groups =or observations are nested within persons) and is based on discussion in =enny (1996):

- **Aggregation**: Determine the lowest level at which observations are independent and =hen average scores of both the causal and outcome measures at that =evel. For instance, if children are nested in classrooms, which are nested =n schools, make school the unit of analysis. Child is the lowest =evel and school is the highest level. If there are no school or classroom =ffect, then child would be the unit of analysis. If there were no school effects, =ut there were classroom effects, then classroom would be the unit of =nalysis. If there were school effect, then school would be the unit of analysis. =his strategy is advisable when the causal variables are measured at the =evel of aggregation or when most of the variation in the causal variable is at =hat level. Thus, when scores are aggregated on the causal variable, they =re essentially the same.
- **Within =nalysis**: Determine the lowest level at which observations are not independent =nd conduct the analysis within each of these units and within each =chool. Save the estimates from these separate analyses and then test if the =ean of the

estimates is different from zero. This strategy is advisable =f the causal variable varies considerably within the nonindependent =nits. So for instance, if classrooms were not independent and gender of student =as an independent variable, then one would compute the mean difference =etween boys and girls for each classroom.

- **Combined or pooled analysis**: Multilevel or hierarchical linear modeling essentially combines the two above strategies. In essence, it =olves the unit of analysis question by making it a pseudo question. All =he observations are analyzed, and the degree of nonindependence is =mpirically estimated. (This strategy is virtually required when units are crossed.)

There are then two key =uestions in determining the unit of analysis. First, a determination must be =ade about the lowest level of units that are independent. Often =tatistical analysis is necessary to determine the extent to which units are =ndependent (though this can be tricky: see Kenny, Kashy, & Bolger's (1998) =oncept of "consequential nonindependence"). Second, a determination must be =ade about the degree of variation in the causal variable. If most of =ts variation is between the nonindependent units, then aggregation or =veraging should be used. If not, then the within analysis should be =sed.

Sometimes, rules about the unit of assignment and the =nit of generalization will be violated. For instance, classrooms may be =he unit of assignment, but if there is no evidence of nonindependence due to =lassroom, person can be the unit of analysis. Alternatively, if there is =vidence that classrooms are nonindependent, then person should not be the unit =f analysis, even if person is the unit of generalization. Because =ll of the variation of treatment is between classrooms (recall that classroom is =he unit of assignment), then the treatment's effect will be seen in between =lassroom variation, not within classroom.

References

Kenny, D. A. (1996). The design and =nalysis of social-interaction research. *Annual Review of Psychology, =7,* 59-86.

Kenny, D. A., & Judd, C. M. (1986). Consequences =f violating the independence assumption in analysis of variance. =I>Psychological Bulletin, 99, 422-431.

Kenny, D. A., & Judd, C. M. (1996). A =eneral procedure for the estimation of interdependence. *Psychological Bulletin, 119,* 138-148.

Kenny, D. A., Kashy, D. A., & Bolger, N. (1998). Data analysis in social psychology. In D. Gilbert, =. Fiske, & G. Lindzey (Eds.), Handbook of social psychology (4th ed., Vol. 1,

=p. 233-265). Boston, MA: McGraw-Hill.

Robinson, W. S. (1950). Ecological =orrelations and the behavior of individuals. *American Sociological Review, =5,* 351-357.

Go =ack to the homepage.