İstanbul Üniversitesi İşletme Fakültesi Dergisi
Istanbul University Journal of the School of Business Administration
Cilt/Vol:39, Sayı/No:1, 2010, 169-174
ISSN: 1303-1732 – www.ifdergisi.org © 2010

Invited Paper

The central role of the unit of analysis concept in research design

Serkan Dolma[1]
Department of Organizational Behavior,
School of Business Administration
Istanbul University, Istanbul, Turkey

Abstract

In this paper the importance of the unit of analysis in research design is emphasized. The relations between the unit of analysis and other related concepts are described, and possible analysis alternatives in case of violation of independence assumption are briefly discussed.

Keywords: *Unit of analysis, observational unit, the assumption of independence of observations, hierarchical linear modeling.*

Analiz birimi kavramının araştırma tasarımındaki merkezi rolü

Özet

Bu makelede analiz birimi olgusunun, araştırma tasarımı açısından önemi vurgulanmıştır. Analiz biriminin ilgili diğer kavramlarla ilişkisi açıklanmış ve bağımsızlık varsayımının sağlanamadığı durumlar için geçerli olan analiz yöntemi alternatiflerinden kısaca bahsedilmiştir.

Anahtar Sözcükler: *Analiz birimi, gözlem birimi, gözlemlerin bağımsızlığı varsayımı, hiyerarşik doğrusal modelleme.*

1. Introduction

The term 'unit of analysis' can be simply defined as "the entity that is being analyzed in a scientific research". Determining or being cognizant of the unit of analysis of the research has a pivotal role in any research endeavor. This may not seem such a serious problem at first since most of the time the appropriate unit of analysis in a study is pretty obvious. For instance, it is for sure that when the relationship between employees' job satisfaction and their performance is investigated, the employee is the unit of analysis as both job satisfaction and job performance are attributes of employees.

But other times, detecting the unit of analysis of a given study may not be that simple. For example, the unit of analysis of a study, in which the possible effect of organizational climate on subordinates' loyalty to their supervisors is examined, cannot be identified without scrutinizing the details of the research design. Unit of analysis may be "organization" in this study since climate is an attribute of organizations or it may be "subordinate" in that, loyalty is an attribute of subordinates or it may even be "supervisor" as it is the entity that is being loyal to. There is no pat answer for determining the unit of analysis in complex studies and only a closer look into the details

[1] *dolma@istanbul.edu.tr (S. Dolma)*

of the research can unravel this problem. The research problem, the hypotheses, variables, measurement instruments, sampling method and population to which the findings are intended to generalize, data analysis techniques and all other elements of a scientific research affect[2] the unit of analysis of a given study.

2. Level of Analysis

Almost anything can be the unit of analysis in a social research, yet the branch of social science mostly delimit, albeit not severely, the possibilities of the unit of analysis of a scientific study. In organizational behavior and managerial sciences for example, typical unit of analysis are employees, supervisors, top managers, customers, work teams, departments, business corporations to name some. In educational research, the most common units of analysis are students, parents, teachers, classes, schools or school districts. Although some units are more popular in a particular discipline, one can specify any type of social entity as the unit of analysis of her study so long as she finds it interesting to investigate.

The units of analysis of studies may be classified into fewer categories or levels. It may not seem very significant to do so, but it, in fact, may help to see the hierarchical relations between the units of analysis possibilities that one can choose of, for her study.

2.1. Individual Level

Individuals are the most common units of analysis in social sciences. Students, employees, union members, registered voters, citizens, political party members, managers, teachers, faculty members, officers, customers, sales representatives can be given as examples of individual level units of analysis.

2.2. Group Level

Sometimes, groups, which consist of multiple individuals, are the main focus of a study such as; study groups, work teams, departments, families, divisions, project teams, residents of an apartment building/a block/a neighborhood. The groups which consist of only two individuals who have a defined relationship (usually called dyads) may also be of interest in a study (e.g. subordinate-supervisor or mother-son dyads or married couples). In studies where the unit of analysis is specified as groups, instead of the individual attributes of the members of the group, the attributes of the group as a whole are of interest (such as group size) although they might be operationalized as the sum or mean of individuals' scores (e.g. the success of a class can be defined as the average score of the students in that class. These types of data are called aggregated data).

2.3. Organizational Level

In sociology, managerial sciences and other social science disciplines, investigating units that are wider than groups (and usually these wider units involve multiple groups within themselves) are not rare at all. Studies that analyze business corporations, not-for-profit organizations, unions, army divisions, schools and universities are some of them. Even a wider social entity may be the unit of analysis of scientific studies such as those in which societies, cities, nations are investigated.

---

[2] All these elements of research do not only affect but also are affected by the unit of analysis (and other related concepts) of the study and by each other due to the fact that scientific research is not a linear but rather interactive and circular process.

2.4. Social Artifacts and Social Interaction Level

Other than individuals and social entities that comprise of individuals, sometimes products of social beings or interactions between social beings may be the unit of analysis of a social science study. Research in which the buildings, books, songs, jokes, tales, scientific discoveries, weddings, wars, strikes, laws, constitutions and meetings are investigated, are examples of studies with these kinds of unit of analysis.

3. Related but Distinct Concepts

There are other concepts in research methodology which are very similar and highly related to the unit of analysis. A researcher should be aware of the difference among these concepts in order not to misspecify the unit of analysis of her study.

3.1. Observational Unit

Observational unit is the concept that is most frequently confused with the unit of analysis even by seasoned researchers. Because observational unit is defined as the entity on which measurements are obtained, the unit of analysis and observational unit are the same for most of the research, but not always. Consider a study inquiring the effect of team size on work team cohesiveness. The unit of analysis of this study will probably be team although team cohesiveness would be measured by applying a questionnaire to team members individually.

Since 'variables' are defined as characteristics of observational units, there would be as many observational units as there are variables in a study. Unit of analysis, on the other hand, is more closely related to the data analysis method. Suppose that the correlation between the heights of fathers and their sons is analyzed. There would be two observational units in this study which are father and son even though it is the father-son dyad which is investigated, thus would be the unit of analysis.

3.2. Sampling Unit and Unit of Generalization[3]

Sampling unit and unit of generalization concepts are relevant for any study which employs inferential statistical analyses. Both of these units, in principle, are advised to be at the same level with the unit of analysis of the study [1]. Sampling unit is the one by which the observations are selected to enter the study, and unit of generalization is the level at which the researcher wants to make generalizations. Normally, one can only make generalizations to the larger group (population) that the samples are drawn from and legitimate generalizations can only be made if a type of random sampling procedure is applied.[4] This concludes that in theory, the sampling unit and unit of generalization (therefore, unit of analysis) of a study will always be at the same level. However, there might be some instances where these units, for practical reasons, are specified at different levels without violating the generalization rule seriously. For example, in a study where the average amount of allowance of the pupils in a city is investigated, in order to

[3] Another related concept is 'unit of assignment' (or experimental unit), which is only relevant to experimental studies and can be described as the unit by which observations are assigned to levels of the independent variable (experimental conditions). It is also advised to perform analyses at the level of assignment, in other words to keep the unit of assignment with the same level of unit of analysis.

[4] Unfortunately random sampling is seldom used in social sciences inspite of the fact that most of the statistical analysis methods are based on the assumption of that. It is true that random sampling is very hard to achieve in most of the real life situations but it is also true that researchers in general do not give enough effort to employ a sampling procedure that at least approximate random sampling.

be truly able to generalize the results to all pupils in that city, one must first have the list of all the pupils and then draw random sample from that list. It may be very hard, if not impossible, to acquire such a list and instead, the list of households which would be a lot easier to attain, can be used to draw from. In this case, the unit of analysis and generalization would be the pupils while sampling unit would be the household.[5]

One common mistake that is done by inexperienced researchers while conducting a research is assuming the unit of sampling to be always the same as the observational unit. For instance, depending on this false assumption, it is a widespread -but not very appropriate- practice to calculate the adequate sample size for the study in terms of individuals (e.g. employees) although the attributes in the study belong to wider units (e.g. teams, departments, organizations). High number of employees (from a few number of organizations) who participate in the study would not be very impressive if the inquiry is about the relationship between organizational culture and performance.

Another error, which is so frequent that it has a name of its own (i.e. *ecological fallacy* [2]), is about making generalization to lower levels of units (e.g. individuals) although the unit of analysis in the study is at a higher level (e.g. organizations, schools, cities). Suppose that it is investigated whether there is a correlation between crime rates and the unemployment rates of the cities. It is obvious that the unit of analysis is the city in this study. Therefore, if such a correlation is found, it will only be appropriate to generalize findings to geographical unit (cities) but not to individuals. A so-called significant correlation does not actually imply that crimes in that city are committed by unemployed people of that city.

4. Data Analysis Alternatives

Specifying the unit of analysis and conducting the appropriate analysis is not a challenging task as long as all the variables in the study are attributes of the same social entity. On the other hand, when the variables (i.e. dependent, independent, mediator, moderator variables) of the inquiry are operationalizations of the attributes of entities with different levels, it may not be that simple to determine the appropriate unit of analysis and method of analysis. Consider, in its simplest form, a study in which the possible effect of the size of a student's school class on her academic success is investigated. Since size (number of students) is an attribute of the class and success is an individual attribute, it may confuse researchers in determining the correct unit of analysis. Actually, if simple random sampling procedure is utilized, there is no harm in conceptually accepting the class size as an individual attribute and specify students (individual) as the unit of analysis. It is safe to claim that although size is an attribute that belongs to classes (to keep it simple let us say it is a dichotomous variable with "small" and "large" categories), it would also be legitimate to conceptualize it as an individual attribute since one can categorize students as those who are enrolled in a small size classes and those who are enrolled in a large (rather crowded) classes. However, in determining the unit of analysis, the most important thing is to ensure that observations are independent of one another as most of the analyses assume independence of observations. Drawing a sample which involves more than one student from the same school class would violate this assumption because the values of independent variable for these classmates cannot be different from one another, in other words, are fixed to be the same. It is not very probable to have such a sample if simple random sampling is employed, though. But in practice, simple random sampling is rarely

[5] For this example, the independence assumption would be slightly violated. Since to have more than one pupil from the same household in a randomly drawn sample is not very likely, it is hoped that this violation will not affect the validity of the findings.

preferred or considered feasible and cluster (multistage) sampling is more widely applied. Typically, classes (or even schools) would be the sampling unit of such a study due to budget and time constraints which means the assumption of independence of observations would be violated at an unacceptable level. Researchers have several options in these situations, three of which are:

1. *Naïve analysis*[6] which involves ignoring the violation of the assumption or unrealistically hoping that independence assumption is not violated which is unfortunately the least legitimate yet the most preferred approach. It should be reminded that the independence of observations assumption is considered, by far, the most important assumption and for even a small violation of it produces a substantial effect on both the level of significance and the power of the test statistic [3].
2. *Aggregation method* which involves aggregation (usually by taking the average or sum or proportion) of the data of the variable which has a lower level and then conduct the analysis at the higher level. For the example above, it would involve calculating the average within every class and then correlate these class averages with class sizes. Two of the most crucial drawbacks of this approach are that (i) the unit of analysis will change to class level which means it would not be justifiable to generalize results to students (in order not to commit an ecological fallacy), and that (ii) the individual variability will be discarded from the analysis which may have provided valuable information in explaining the causes/predictors of the outcome/criterion/dependent variable.
3. *Hierarchical Linear Modeling*[7] (HLM), a more advanced analysis in which the independence of observations assumptions is circumvented by taking the hierarchical structure of the data into consideration. It is done so by allowing the components (parameters) of the regression equation to vary among clusters (i.e. at higher levels) and estimating the variation of these so-called "random" parameters. The most challenging caveat of this approach is that it requires much larger sample sizes. While it is quite a popular advice to employ HLM if the research design involves multiple levels, one should bear in mind that it is not having observational units of different levels that force researcher to implement HLM but the method of the sampling that is preferred. Provided that simple random sampling is employed, it would be totally justifiable to use single level models (e.g Anova, regression) for most of the research ventures.[8] The reason that HLM is the best method for most of the research (which involves different levels of observational) is due to the fact that simple random sampling is not feasible (or at least not cost-efficient) for most of the situations, thus cluster/multistage sampling is preferred, which means violation of the independence assumption is inescapable.

To conclude, it would be a fair advice to say that, the elements of the research design should be considered and determined at the very beginning of any research endeavor, at least before collecting data, and it should be kept in mind that there is a highly interactive structure among these research elements. Realizing that the unit of analysis is different than the intended one may have irreversible adverse effects on the research process or may even cause the researcher to terminate the project. Researchers should

---

[6] Sometimes it is called 'disaggregation method'.

[7] This approach has some many different names that I am not sure anyone knows all of them. Multilevel Analysis (Modeling), Mixed Effect Analysis (Modeling), Random Effect Analysis (modeling) and Random Coefficient Analysis (Modeling) are some of its synonyms.

[8] Only occasions where random sampling may not be enough to warrant single level analysis are those in which the number of cases of the higher unit (e.g. organization) in the population is so few that having a sample that includes multiple observations from the same higher unit is inevitable.

also bear in mind that preferences on sampling methods or level of analysis would dramatically affect the applicability of analysis methods, the validity and the generalizability of the research findings.

5. Suggestions for Further Reading

Although there are not any books which covers the unit of analysis topic in details, Babbie's [4], Singleton and Straits' [5] and Neuman's [6] books can be referred for further examples and alternative definitions. Kenny's web page [1], albeit very short, is one the most influential sources on this topic.

There are several books on HLM or Multilevel Modeling but most of them require high level of mathematical background. Twisk's monograph [7], Field's related chapter in his book [8] and the chapter written by Natasha Beretvas, in Stevens' book [9], by contrast, are less demanding. Osborne's short article [10] compares the three analysis methods mentioned above.

Researchers may also refer to Kenny and Judd's article [11] for more information on consequences of violating the independence assumption in Anova, and consult Hopkins' paper [12] to be informed about an alternative way to overcome the dependent observations problem.

References

[1] D.A. Kenny, "Unit of Analysis", Retrieved on September 2009, from http://davidakenny.net/u_o_a.htm.

[2] W.S. Robinson, Ecological Correlations and the Behavior of Individuals. *American Sociological Review*. 15, 351-357 (1950).

[3] J.P. Stevens, *Intermediate Statistics*. Lawrence Earlbaum, New York, 2007.

[4] E. Babbie, *The Basics of Social Research*. Thomson Wadsworth, Canada, 2005.

[5] R.A. Singleton, Jr., B.C. Straits, *Approaches to Social Research*. Oxford University Press, New York, 1988.

[6] W.L. Neuman, *Basics of Social Research: Qualitative and Quantitative Approaches*. Pearson Education, 2006.

[7] J.O.R. Twisk, *Applied Multilevel Analysis*. Cambridge University Press, Cambridge, 2006.

[8] A. Field, *Discovering Statistics Using SPSS*. Sage Publication, London, 2009.

[9] J.P. Stevens, *Applied Multivariate Statistics for the Social Sciences*. Routledge, USA, 2009.

[10] J.W. Osborne, Advantages of Hierarchical Linear Modeling. Practical Assessment, Research and Evaluation, 7, Article 1, Retrieved on January 2010 from htpp://pareonline.net/getvn.asp?v=7&n=1.

[11] D.A. Kenny, C.M. Judd, Consequences of Violating the Independence Assumption in Analysis of Variance. *Psychological Bulletin*. 99, 422-431 (1986).

[12] K.D. Hopkins, The Unit of Analysis: Group Means Versus Individual Observations. *American Educational Research Journal*. 19, 5-18 (1982).