Application of Statistical techniques in Data Analysis

Dr.Jiteshkumar K Talati (Principal)

Shree R.P.Anada College of Education Borsad, Dist: Anand

Introduction:

New technologies offer new opportunities, although pervasive use of computers in research has not come about without problems. Some of these problems are technological, some financial. Underlying many of them are complex institutional and behavioral constraints.

Nearly five decades ago, the first programmable, electronic, digital computer was switched on. The amount of data that can be analyzed has expanded, as has the complexity of analyses. And researchers can collaborate more widely and efficiently.

Different scientific disciplines use information technology differently. Uses vary according to the phenomena the discipline studies and the rate at which the discipline obtains information. In such disciplines as high energy physics, neurobiology, chemistry, or materials science, experiments generate millions of observations per second, and these must be screened and recorded as they happen. For these disciplines, computers that can handle large amounts of information quickly are essential and have made possible research that was previously impractical.

Data Analysis

Data Analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate, condense and recap, and evaluate data. According to Shamoo and Resnik (2003) various analytic procedures "provide a way of drawing inductive inferences from data and distinguishing the signal (the phenomenon of interest) from the noise (statistical fluctuations) present in the data".

While data analysis in qualitative research can include statistical procedures, many times analysis becomes an ongoing iterative process where data is continuously collected and analyzed almost simultaneously. Indeed, researchers generally analyze for patterns in observations through the entire data collection phase (Savenye, Robinson, 2004). The form of the analysis is determined by the specific qualitative approach taken (field study, ethnography content analysis, oral history, biography, unobtrusive research) and the form of the data (field notes, documents, audiotape, videotape).

An essential component of ensuring data integrity is the accurate and appropriate analysis of research findings. Improper statistical analyses distort scientific findings, mislead casual readers (Shepard, 2002), and may negatively influence the public perception of research. Integrity issues are just as relevant to analysis of non-statistical data as well.

Considerations/issues in data analysis

There are a number of issues that researchers should be cognizant of with respect to data analysis. These include:

- ✓ Having the necessary skills to analyze
- ✓ Concurrently selecting data collection methods and appropriate analysis
- ✓ Drawing unbiased inference
- ✓ Inappropriate subgroup analysis
- ✓ Following acceptable norms for disciplines

- ✓ Determining statistical significance
- ✓ Lack of clearly defined and objective outcome measurements
- ✓ Providing honest and accurate analysis
- ✓ Manner of presenting data
- ✓ Environmental/contextual issues
- ✓ Data recording method
- ✓ Partitioning 'text' when analyzing qualitative data
- ✓ Training of staff conducting analyses
- ✓ Reliability and Validity
- ✓ Extent of analysis

Having necessary skills to analyse

A common practice of investigators is to defer the selection of analytic procedure to a research team 'statistician'. Ideally, investigators should have substantially more than a basic understanding of the rationale for selecting one method of analysis over another. This can allow investigators to better supervise staff who conduct the data analyses process and make informed decisions.

Concurrently selecting data collection methods and appropriate analysis

While methods of analysis may differ by scientific discipline, the optimal stage for determining appropriate analytic procedures occurs early in the research process and should not be an afterthought. According to Smeeton and Goda (2003), "Statistical advice should be obtained at the stage of initial planning of an investigation so that, for example, the method of sampling and design are appropriate.

Drawing unbiased inference

The chief aim of analysis is to distinguish between an event occurring as either reflecting a true effect versus a false one. Any bias occurring in the collection of the data, or selection of method of analysis, will increase the likelihood of drawing a biased inference. Bias can occur when recruitment of study participant's falls below minimum number required to demonstrate statistical power or failure to maintain a sufficient follow-up period needed to demonstrate an effect.

Inappropriate subgroup analysis

When failing to demonstrate statistically different levels between treatment groups, investigators may resort to breaking down the analysis to smaller and smaller subgroups in order to find a difference. Although this practice may not inherently be unethical, these analyses should be proposed before beginning the study even if the intent is exploratory in nature. If it the study is exploratory in nature, the investigator should make this explicit so that readers understand that the research is more of a hunting expedition rather than being primarily theory driven. Although a researcher may not have a theory-based hypothesis for testing relationships between previously untested variables, a theory will have to be developed to explain an unanticipated finding. Indeed, in exploratory science, there are no a priori hypotheses therefore there are no hypothetical tests. Although theories can often drive the processes used in the investigation of qualitative studies, many times patterns of behavior or occurrences derived from analyzed data can result in developing new theoretical frameworks rather than determined **a priori.** It is conceivable that multiple statistical tests could yield a significant finding by chance alone rather than reflecting a true effect. Integrity is compromised if the investigator only reports tests with significant findings, and neglects to mention a large number of tests failing to reach significance. While access to computer-based statistical packages can facilitate application of increasingly complex analytic procedures, inappropriate uses of these packages can result in abuses.

Following acceptable norms for disciplines

Every field of study has developed its accepted practices for data analysis. Resnik (2000) states that it is prudent for investigators to follow these accepted norms. Resnik further states that the norms are '...based on two factors:

(1) The nature of the variables used (i.e., quantitative, comparative, or qualitative),

(2) Assumptions about the population from which the data are drawn (i.e., random distribution, independence, sample size, etc.). If one uses unconventional norms, it is crucial to clearly state this is being done, and to show how this new and possibly unaccepted method of analysis is being used, as well as how it differs from other more traditional methods.

Determining significance

While the conventional practice is to establish a standard of acceptability for statistical significance, with certain disciplines, it may also be appropriate to discuss whether attaining statistical significance has a true practical meaning, i.e., **'clinical significance'**. Jeans (1992) defines 'clinical significance' as "the potential for research findings to make a real and important difference to clients or clinical practice, to health status or to any other problem identified as a relevant priority for the discipline".

Kendall and Grove (1988) define clinical significance in terms of what happens when "… troubled and disordered clients are now, after treatment, not distinguishable from a meaningful and representative non-disturbed reference group". Thompson and Noferi (2002) suggest that readers of counselling literature should expect authors to report either practical or clinical significance indices, or both, within their research reports. Shepard (2003) questions why some authors fail to point out that the magnitude of observed changes may too small to have any clinical or practical significance, "sometimes, a supposed change may be described in some detail, but the investigator fails to disclose that the trend is not statistically significant".

Provide honest and accurate analysis

The basis for this issue is the urgency of reducing the likelihood of statistical error. Common challenges include the exclusion of outliers, filling in missing data, altering or otherwise changing data, data mining, and developing graphical representations of the data (Shamo0, Resnik2003).

Manner of presenting data

At times investigators may enhance the impression of a significant finding by determining how to present derived data (as opposed to data in its raw form), which portion of the data is shown, why, how and to whom (Shamoo, Resnik, 2003). Nowak (1994) notes that even experts do not agree in distinguishing between analyzing and massaging data. Shamoo (1989) recommends that investigators maintain a sufficient and accurate paper trail of how data was manipulated for future review.

Data recording method

Analyses could also be influenced by the method in which data was recorded. For example, research events could be documented by:

a. recording audio and/or video and transcribing later
b. either a researcher or self-administered survey
c. either closed ended survey or open-ended survey
d. preparing ethnographic field notes from a participant/observer
e. requesting that participants themselves take notes, compile and submit them to researchers.

While each methodology employed has rationale and advantages, issues of objectivity and subjectivity may be raised when data is analyzed.

Reliability and Validity

Researchers performing analysis on either quantitative or qualitative analyses should be aware of challenges to reliability and validity. For example, in the area of content analysis, Gottschalk (1995) identifies three factors that can affect the reliability of analyzed data:

- ✓ **stability,** or the tendency for coders to consistently re-code the same data in the same way over a period of time
- ✓ **reproducibility,** or the tendency for a group of coders to classify categories membership in the same way
- ✓ **accuracy,** or the extent to which the classification of a text corresponds to a standard or norm statistically

The potential for compromising data integrity arises when researchers cannot consistently demonstrate stability, reproducibility, or accuracy of data analysis.

According Gottschalk, (1995), the validity of a content analysis study refers to the correspondence of the categories (the classification that raters' assigned to text content) to the conclusions, and the generalizability of results to a theory (did the categories support the study's conclusion, and was the finding adequately robust to support or be applied to a selected theoretical rationale?).

Extent of analysis

Up on coding text material for content analysis. Rates must classify each code into an appropriate category of a cross-reference matrix. Relying on computer software to determine a frequency or word count can lead to inaccuracies. "One may obtain an accurate count of that word's occurrence and frequency, but not have an accurate accounting of the meaning inherent in each particular usage (Gottschalk, 1995). Further analyses might be appropriate to discover the dimensionality of the data set or identity new meaningful underlying variables.

Whether statistical or non-statistical methods of analyses are used, researchers should be aware of the potential for compromising data integrity. While statistical analysis is typically performed on quantitative data, there are numerous analytic procedures specifically designed for qualitative material including content, thematic, and ethnographic analysis. Regardless of whether one studies quantitative or qualitative phenomena, researchers use a variety of tools to analyze data in order to test hypotheses, discern patterns of behavior, Conclusion:

It is very efficient and need to develop new research data techniques to evaluate new researches done as qualitative and quantitative research both. In present paper there was a minor try to mention some research techniques for analysis of data and prior care should be taken while handling data in research.

References:

- Best, J.W & Kahn, J.V. (1989). Research in education. (6Ih ed.) New Delhi: Prentice Hall of India.
- Bowman-Alden, S. (n.d.). "The Role Technology Can Play in Preparing Our Children For the 21st Century." Computer Learning Foundation. Retrieved March 30, 2003.
- Daccord, T. Teaching History with Technology. Retrieved on January 14, 2004.
- Garrett, H.E. (1981). Statistics in Psychology and Education. Bombay: Vakils fifer and Simons Ltd.
- P.Bhatnagar, Poonam Rani Bhatnagar (2002), Readings in methodology of research in Education and Psychology, Rajhans Publications, Meerut.