How Do Actuaries Use Data Containing Errors?: Models of Error Detection and Error Correction

BARBARA D. KLEIN
University of Michigan-Dearborn

Information systems provide data for business processes and decision making. There is strong evidence that data items stored in organizational databases have a significant rate of errors. If undetected in use, errors in data may significantly affect business outcomes. The question examined in this paper is the extent to which business professionals are able to evaluate the quality of data in the information systems they use and the impact of their evaluations on decision-making behavior. Models of error detection and error correction processes are developed. The validity of the models is then examined through an analysis of interviews with ten actuaries. The findings show that actuaries detect errors in data using three general methods and that actuaries consider feasibility and costs when deciding whether to correct data errors.
found to believe that the usefulness of their organization’s data is limited because it is only 95 percent accurate (Nayar, 1993). Knight (1992) reports the findings of a survey conducted by Computerworld in which two-thirds of the surveyed organizations acknowledged problems stemming from inaccurate or incomplete data.

This paper begins with a review of the literature related to the detection of errors in data and the impact of data errors on decision making. Next, the paper draws on theories of decision making to examine the research question. Models of error detection and error correction processes are then developed, and the validity of the models is examined through an analysis of interviews with ten practicing actuaries.

**Literature Review**

In a broad sense, this study falls in the literature on data quality. Several general conclusions can be drawn from the existing research on data quality. First, while no single definition of data quality has been accepted by researchers working in this area, there is agreement that data accuracy, currency, and completeness are important areas of concern (Agmon and Ahituv, 1987; Davis and Olson, 1985; Fox et al., 1993; Huh et al., 1990; Madnick and Wang, 1992; Wand and Wang, 1994; Zmud, 1978). Second, while it is difficult to compare error rates across studies, rates substantially greater than zero have been found in all of the studies addressing the extent to which data errors exist in databases (Ham et al., 1985; Johnson et al., 1981; Knight, 1992; Laudon, 1986; Stone and Bublitz, 1984). Third, there is disagreement about the extent to which efforts to purge all errors from databases should be attempted. Some researchers propose methods designed to completely rid databases of errors (Janson, 1988; Misvanks, 1988; Naus, 1975; Parsaye and Chignell, 1993), while others propose tools for determining how to best allocate limited resources to controlling the level of data errors (Ballou and Pazer, 1987; Ballou and Tayi, 1989; Ballou et al., 1987; Bowen, 1992; Paradise and Fuerst, 1991). Fourth, many researchers argue that users need not discard data containing errors. A variety of approaches for using imperfect data have been suggested (Ballou and Pazer, 1985; Ballou and Pazer, 1987; Ballou and Pazer, 1995; Bansal, 1993; Gaba and Winkler, 1992; Garfinkel et al., 1986; O’Leary, 1993; O’Neill and Vizine-Goetz, 1988).

This study fits into the literature on data quality that examines the detection of data errors and the impact of data errors on the quality of managerial decisions.

**The Detection of Errors in Data**

In general, the data quality literature argues that users are not very capable of finding errors in data. There is evidence in this literature that errors in data are not always detected by users (e.g., Davis et al., 1967; Laudon, 1986; Ricketts, 1990). Davis et al. (1967) conducted a field experiment in which individuals were mailed banking confirmation statements with embedded errors. The individuals were asked to verify their account information, and approximately half failed to detect important errors. Laudon (1986) found that users of criminal information systems rarely detected errors in these records even though information provided to police departments by the FBI is accompanied by a warning stating that the user should verify that the information is accurate. Ricketts (1990) conducted a laboratory experiment in which over ninety percent of the subjects failed to detect a substantial data error in production planning reports. The failure of humans to detect errors in data is also assumed in much of the literature on data quality in which it is argued that resources should be devoted to the improvement of the quality of the data stored in all organizational databases (e.g., Redman, 1992; Redman 1995).

The findings of these three studies suggest that users may be poor detectors of errors in data. However, this interpretation may underestimate the extent to which users can detect errors. For example, subjects in Ricketts’ study may not have understood that error detection was part of the experimental task. Subjects who failed to detect the error in the Davis et al. study may simply not have accepted responsibility for carefully reviewing the banking statement.

**The Impact of Data Errors**

Ballou and Pazer (1985) proposed a methodology for analyzing the impact of data quality on computer-generated outputs. Ballou et al. (1987) applied this model to an analysis of the impact of data quality in the context of the use of spreadsheets by end users. The analysis of spreadsheet data assumes that errors are normally distributed around a true value.

The idea underlying the Ballou and Pazer (1985) methodology is that if a user has some knowledge of the degree of random errors in the inputs to a spreadsheet, then he or she can calculate the degree of error in outputs calculated using these inputs. The framework is not proposed as a model of how people actually use data containing errors. It is, rather, a prescriptive methodology for understanding the sensitivity of computer-generated outputs to errors in data inputs. The framework is a way of analyzing the robustness of outputs to potential data errors. According to Ballou et al. (1987), “the purpose of our analysis is not to give precise information concerning the quality of spreadsheet results but rather to provide insight as to possible error magnitudes (bull park estimates) which a manager would incorporate perhaps at an intuitive level, into the decision-making process” (page 19). The question remains as to how managers do or could incorporate insights into possible errors and the magnitudes of these possible errors into decision-making processes.
Related Content

A Comparative Analysis of Online and Traditional Undergraduate Business Law Classes
[www.igi-global.com/chapter/comparative-analysis-online-traditional-undergraduate/22889?camid=4v1a](www.igi-global.com/chapter/comparative-analysis-online-traditional-undergraduate/22889?camid=4v1a)

Enterprise-wide Strategic Information Systems Planning for Shanghai Bell Corporation
[www.igi-global.com/chapter/enterprise-wide-strategic-information-systems/44557?camid=4v1a](www.igi-global.com/chapter/enterprise-wide-strategic-information-systems/44557?camid=4v1a)

[www.igi-global.com/chapter/technology-trust-b2b-electronic-commerce/6113?camid=4v1a](www.igi-global.com/chapter/technology-trust-b2b-electronic-commerce/6113?camid=4v1a)

End-User’s Disconfirmed Expectations and the Success of Information Systems
[www.igi-global.com/article/end-user-disconfirmed-expectations-success/51000?camid=4v1a](www.igi-global.com/article/end-user-disconfirmed-expectations-success/51000?camid=4v1a)