Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 189 (2008) 583–593
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Testing the accuracy of employee-reported data: An inexpensive alternative approach to traditional methods

Douglas N. Hales [a,*], V. Sridharan [b,1], Abirami Radhakrishnan [c], Satya S. Chakravorty [d], Samia M. Siha [d]

[a] College of Business Administration, University of Rhode Island, Ballentine Hall, 7 Lippitt Road, Kingston, RI 02881, USA
[b] Department of Management, College of Business and Behavioral Science, 117-F Sirrine Hall, Clemson University, Clemson, SC 29634-1305, USA
[c] College of Business Administration, Morgan State University, Baltimore, MD, USA
[d] Department of Management and Entrepreneurship, Michael J. Coles College of Business, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144, USA

Received 31 December 2005; accepted 29 September 2006
Available online 19 January 2007

Abstract

Although Information Technology (IT) solutions improve the collection and validation of operational data, Operations Managers must often rely on self-reported data from workers to make decisions. The problem with this data is that they are subject to intentional manipulation, thus reducing their suitability for decision-making. A method of identifying manipulated data, *digital analysis*, addresses this problem at low cost. In this paper, we demonstrate how one uses this method in real-world companies to validate self-reported data from line workers. The results of our study suggest that *digital analysis* estimates the accuracy of employee reported data in operations management, within limited contexts. These findings lead to improved operating performance by providing a tool for practitioners to exclude inaccurate information.

Keywords: Digital analysis; Benford's law; Plastics industry; Operations management; Fraud detection

1. Introduction

In 2001, we conducted a quality audit for a manufacturing firm to discover the source of an uncharacteristic increase in the number of customer returns. The firm employs Statistical Process Control (SPC) to track performance of its production process. After review of the SPC charts, we found no patterns that indicated the level of defects observed in the returned goods, thus suggesting a product design flaw or error in specifications, which were transmitted into the equipment setup and charts. However, further investigation indicated that an experienced line worker (machine operator) fabricated product weight data on the SPC chart,

* Corresponding author. Tel.: +1 401 874 7882; fax: +1 401 874 4312.
E-mail addresses: dhales@uri.edu (D.N. Hales), suhas@clemson.edu (V. Sridharan), aradhak@gmail.com (A. Radhakrishnan).
[1] Tel.: +864 656 2624.

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2006.09.092

which was discovered after 10 days production were shipped, rejected, and returned from the customer. The operator failed to perform his required weight checks on production runs and thus randomly assigned fraudulent weights within the SPC limits in a pattern that appeared to be genuine. This meant that SPC did not detect the product defects. This single incident cost the company an estimated $300,000 (US) of which only $40,000 was recoverable.

This issue presented an interesting problem to managers. Since managers rely on employee reported data to make decisions, how can they estimate the accuracy of the information without reinstituting traditional quality control inspection and sampling procedures for which they worked hard to replace? Additionally, since the company's managers prided themselves on trusting employees, how could they ensure data accuracy without instilling a sense of distrust among machine operators? After all, this problem occurred with only one of 21 operators employed at this facility. If managers, in this example, had an inexpensive tool to validate the data reported by the dishonest operator, the problem could have been identified earlier.

The purpose of this study is to provide operations managers with a tool to validate self-reported data from line workers, where the reports are the only source of information, or where secondary sources are difficult or expensive to obtain. The problem with employee reported data is that it provides the opportunity for individuals to manipulate the results, thus reducing its suitability for decision-making. Managers who do not have a secondary source for verifying the accuracy of employee reported data, may find decreased performance in activities and processes that rely on this information. Hence, we identified how other disciplines address this issue. We found that financial auditors commonly employ a method called *digital analysis* to identify suspect data. One purpose of this study is to apply this method in two companies and industries to evaluate its ability to detect fraudulent data in the context of operations. A second purpose is to extend the use of *digital analysis* to data types not considered in previous studies, i.e. to distributions previously considered inappropriate for digital analysis.

2. Literature review

The literature is replete with methods for gathering and evaluating data from manufacturing processes, most notably Statistical Process Control, statistical (acceptance) sampling, and post-process quality control inspection (Deming, 1986). While powerful, these techniques rely heavily on the integrity of the workers collecting the data – which at times prove to be problematic (Hales et al., 2004). To combat fraud in self-reported data, managers rely on duplicate measurements and post-process inspections to validate information. The main problem with these approaches is that they are expensive to execute and run counter to many contemporary business techniques of trusting workers and eliminating waste through duplicate efforts (Deming, 1986). In reviewing how other disciplines detect fraudulent data, we found *digital analysis.*

Digital analysis refers to a technique for estimating the distributions of certain digits, 0–9, in naturally occurring data (i.e. data that is not intentionally manipulated). The premise of the technique is that naturally occurring data have different distributions than manipulated data. One estimates probabilities for these occurrences and then compares them to actual process data. If there are differences between the actual and estimated distributions, then the data is said to have a probability of containing systematic error or statistical bias (Nigrini, 1996a, 1998, 1999, 2000; Nigrini and Mittermaier, 1997). Other applications include detecting fraud in areas such as declaration records (Browne, 1998), and the feasibility of outputs from computer simulations and logistics models (Hill, 1995, 1996, 1998). While the procedure is not applicable in all situations, it does provide an inexpensive alternative to other forms of validation procedures such as statistical sampling or duplicate measurement. In the only Operations application found, Becker (1982) used the method to estimate the degree to which machine failure rate lists, based on Mean-Time-To-Failure calculations had systematic error, indicating intentional manipulation or defects in measurement processes.

The literature suggests the type of data appropriate for digital analysis. The primary qualifications are that they be from large data sets with preferably large ranges, generated naturally without pre-set limits or breakpoints, and without intervention. Evidence suggests that the larger the range, the smaller the required data set. The distributions are Weibull-like in shape and hold true for populations and representative samples (Brown, 2005; Nigrini, 1996b; Hill, 1995). These assumptions require that a test for appropriateness be conducted using values known to be generated without intervention. If the observed probability distribution matches that pre-