Data entry methods: Is double entry the way to go?

Jenna R. Scott, Ashleigh R. Thompson, Dionne Wright-Thomas, Xiaoyan Xu, & Kimberly A. Barchard
University of Nevada, Las Vegas

Reference: Scott, J.R., Thompson, A.R., Wright-Thomas, D., Xu, X. & Barchard, K.A. (April, 2008) *Data entry methods: Is double entry the way to go?* Poster presented at the Western Psychological Association Annual Convention, Irvine, CA.

Contact Information: Kim Barchard, Department of Psychology, University of Nevada, Las Vegas, 4505 S. Maryland Parkway, P.O. Box 455030, Las Vegas, NV, 89154-5030, USA, barchard@unlv.nevada.edu

Abstract

Data entry errors can be disastrous. They can alter your results and invalidate your conclusions (Kruskal, 1960; Wilcox, 1998). This study examined the accuracy of three different data entry methods: Single Entry, Single Entry with Visual Checking, and Double Entry. Previous research has shown that entering data twice (Double Entry) is more effective in reducing data entry errors than entering the data just once. Therefore, Double Entry is expected to be more accurate than the other methods studied.

A total of 31 participants have completed this study so far. Each participant was randomly assigned to one of the three data entry methods. First, the participants watched two separate videos teaching them basic Microsoft Excel skills and how to enter data according to their assigned data entry method. Afterwards, they were given two different sets of data entry sheets to enter using the method they were assigned. Contrary to prior research, the preliminary results showed no significant differences between the three data entry methods in terms of their average accuracy. However, Double Entry was the most accurate of the methods examined, and with a larger sample size, the differences will probably become statistically significant. On the other hand, even with this small sample size, the methods did differ in their variability: the Single Entry with Visual Checking method had a significantly higher standard deviation than the other two methods. Sometimes Single Entry with Visual Checking resulted in very low accuracy.

These results demonstrate the importance of good data entry methods. Data entry errors can wreck havoc on the results and conclusions of a research experiment and thus accurate data entry methods must be found and adopted by all researchers.

Introduction

Accuracy is vital for any study, especially in the critical step of data entry. There are various unfortunate and costly consequences to inaccurately entering data. According to Barchard and Pace (in press), inaccurate data entry can introduce random error to experiments, which consequently reduces reliability. Entering a few numbers incorrectly can even invalidate the results of a statistical analysis (Kruskal, 1960; Wilcox, 1998). Researchers have developed methods to fix data entry errors after the fact (see e.g., Winkler, 2004) but none seem to be as efficient as simply entering the data correctly at the outset.

There are many factors that could lead to data entry errors. Fatigue, for example, greatly impacts accuracy. As fatigue increases, accuracy decreases (Healy, Kole, Buck-Gengler & Bourne, 2004). In addition, when speed is introduced as an important aspect of data entry, accuracy decreases (Norr, 2000; Galinsky, Schleifer & Pan, 1995). Age differences and differing levels of experience also influence performance on data entry tasks. Czaja and Sharit (1998) found that older people actually enter data more accurately but younger people enter data at a quicker pace. They also noted that as experience increases on the data entry tasks, the number of errors decreases. According to Glaser, Tatum, Nebeker, Sorenson and Aiello, (1999), workload and stress during data entry tasks can decrease performance. Sleep deprivation also has an adverse effect on performance of tasks in general and could cause data entry errors (Broughton, 1991). Because there are so many possible causes for data entry errors, it is inevitable that they will occur. Researchers therefore need a method of detecting and correcting these errors.

The purpose of this study is to compare three data entry methods, to determine which method results in the most accurate data entry. The first method is the Single Entry method. This involves the person simply entering the data and not checking their work. The second method is Single Entry with Visual Checking, in which the participant enters the data and then visually checks the entry for errors. The third method is the Double Entry method, in which the participant enters the data twice and is alerted if the data are out of range or if it does not match. This allows the data entry person to correct any mistakes immediately.

We hypothesize that the Double Entry method will be more accurate than the other methods studied. Healy et al. (2004) argues that feedback on errors increases accuracy. As well, Beaty (1999) showed that PowerChecker, an earlier double-entry system, is more accurate than entering the data once. In order to test our hypothesis, we will use ANOVA to compare the accuracy of the three data entry methods: Single Entry, Single Entry with Visual Checking, and Double Entry.

Method

Participants

There were 31 undergraduate students (22 female, 9 male) who participated in this study in return for course credit at a large western university. They ranged in age from 18 to 34 (mean 20.71, standard deviation 3.32). In terms of ethnicity, 42% identified themselves as Caucasians, 10% African American, 16% Asian, 13% Hispanic, 16% Pacific Islander, and 3% other.

Procedures

This study involves a single 90-minute appointment. The study is completed on the computer during individual appointments with a study administrator.

The participants are randomly assigned to one of three methods of data entry: Single Entry, Single Entry with Visual Checking, and Double Entry. The participants assigned to the Single Entry method are instructed to enter the data as accurately as possible and to simply type the data once into their Microsoft Excel spreadsheet. The participants assigned to the Single Entry with Visual Checking are given the same instructions, but in addition to entering the data once, they are asked to visually check for errors after each data sheet is entered.

The participants assigned to Double Entry are instructed to enter all their data twice. The spreadsheet is set up to automatically compare the two sets of data for mismatches and check for entries that are outside the allowable range for those variables. The participants then have two columns that give them the number of mismatches and out-of-range errors. Also, the cells where the errors occurred are then highlighted, making errors visually noticeable. Barchard and Pace (in press) claim that double entry, comparing for mismatches, and checking for out-of-range errors may fend against most data entry errors.

During the testing session, participants watch two videos. The first video teaches them basic skills on how to operate a Microsoft Excel spreadsheet. The second video instructs them on how to enter data using one of the three methods (Single Entry, Single Entry with Visual Checking, or Double Entry). Thus, there are three versions of the second video and the participant views the video that corresponds to the data entry method to which they are assigned.

After the participant watches the video on their assigned data entry method, the participant enters two sets of data entry sheets into two different spreadsheets. First the participant opens an Excel spreadsheet labeled "Part 1" and enters a set of five data entry sheets. Once they complete the first set, they save and close the current spreadsheet. Then they open the second spreadsheet, labeled "Part 2", and enter the primary data for this study, a set of thirty data entry sheets.

The data entry sheets that make up Part 1 and Part 2 consist of six different types of data that the participant must enter: ID number, Sex, and four different types of scales that each consist of ten entries. The four difference scales are labeled "Family Background," "School Experiences," "Extraversion," and "Social Skills Test."

When entering the data sheets, the participant is asked to convert letter responses into corresponding numbers. For "Sex", the participant must convert the letters "M" and "F" into the corresponding numbers of "1" and "2". For the "Family Background" scale, the participant must convert "SD" into "1", "D" into "2", "N" into "3", "A" into "4" and "SA" into "5". Finally, for the "School Experiences" scale, the participant must convert "D" into "1", "N" into "2", and "A" into "3." For the scales of "Extraversion" and "Social Skills Test", the responses are already in numbers and require no conversions.

Measures

For each participant, accuracy was calculated as the percentage of correct entries across all data sheets.

Statistical Analysis

We performed a one-way ANOVA to compare the three data entry methods in terms of their accuracy during Part 2 of the study, where each participant entered 30 data entry sheets. In addition, we used Levene's Statistic to compare the methods in terms of the variability of their accuracies.

Results

The descriptive analyses given in Table 1 show that Double Entry and Single Entry both resulted in very high accuracy rates. Single Entry with Visual Checking resulted in lower accuracy. To determine if these differences were statistically significant, a one-way ANOVA was conducted. There were no statistically significant differences among the three groups (see Table 2).

Accuracy was much more variable for Single Entry with Visual Checking than for Double Entry or Single Entry. The standard deviation for Single Entry with Visual Checking was much larger than the standard deviations for the other two techniques (Levene's Statistic (2, 28) = 7.57, p = .002). From examining Table 1, the reader will note that the primary difference between the techniques was the minimum accuracy score. For both Double Entry and Single Entry, the very worst data enterer was still relatively accurate. However, for Single Entry with Visual Checking, one participant had an accuracy score of only 61%. Thus, both Single Entry with Visual Checking and Double Entry seem to result in consistently high accuracy, but Single Entry with Visual Checking can sometimes be disastrous.

Table 1

Accuracy of the Three Data Entry Methods

Source	Single Entry	Visual Checking	Double Entry
Mean	.9967	.9437	.9997
Standard Deviation	.0041	.1355	.0005
Minimum	.9849	.6111	.9984
Maximum	1.0000	.9992	1.0000

Table 2

ANOVA Comparing Accuracy between the Three Data Entry Methods

Source	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	.02	2	.01	1.93	.164
Within Groups	.13	28	.01		
Total	.15	30			

Conclusions

The purpose of this research was to discover which of the three data entry methods yields the most accurate data. We hypothesized that the Double Entry method would be the most accurate. Upon testing our theory however, the preliminary results showed no significant differences between the three data entry methods in terms of their average accuracy. However, Double Entry was the most accurate of the methods examined, and with a larger sample size, the differences will probably become statistically significant. Sample size is one of the primary determinants of statistical power but, even with this small sample size, methods did differ in their variability. The Single Entry with Visual Checking method had a significantly higher standard deviation than the other two methods: sometimes, Single Entry with Visual Checking resulted in very low accuracy. The relative accuracy of different data entry methods is important because of the negative effects data entry errors can have on the results of a research experiment. Prevention of data entry errors will result in more accurate and reliable conclusions in research experiments.

References

Barchard, K. A., & Pace, L. A. (in press). Meeting the Challenge of High Quality Data Entry: A Free Double-Entry System. *International Journal of Services and Standards*. Available from Kim Barchard, Department of Psychology, University of Nevada, Las Vegas 4505 S. Maryland Parkway, PO Box 455030 Las Vegas, NV 89154-5030, barchard@unlv.nevada.edu.

Beaty, J. C. (1999). The PowerChecker: A visual basic program for ensuring data integrity. *Behavior, Research Methods, Innovation, & Computers, 31*, 737-740.

Broughton, R. J. (1991). Field studies of sleep/wake patterns and performance: A laboratory experience. *Canadian Journal of Psychology, 45*, 240-253.

Czaja, S. J. & Sharit, J. (1998). Ability-performance relationships as a function of age and task experience for a data entry task. *Journal of Experimental Psychology: Applied, 4*, 332-351.

Galinsky, T. L., Schleifer, L. M., & Pan, C. S. (1995). The influence of performance standards and feedback on speed and accuracy in an electronically monitored data-entry task. *International Journal of Human-Computer Interaction, 17*, 25-36.

Glaser, D. N., Tatum, B. C., Nebeker, D. M., Sorenson, R. C., & Aiello, J. R. (1999). Workload and social support: Effects on performance and stress. *Human Performance, 12*, 155-176.

Healy, A. F., Kole, J.A., Buck-Gengler, C. J., & Bourne, L. E. (2004). Effects of prolonged work on data entry speed and accuracy. *Journal of Experimental Psychology: Applied, 10*, 188-199.

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics*, 2. Available at http://www.tufts.edu/~gdallal/out.htm Accessed September 11, 2007.

Norr, K. B. (2000). An investigation of the relationship between speed and accuracy utilizing a data entry task. *Dissertation Abstracts International: Section B: The Sciences and Engineering, 60*, 3606.

Wilcox, R. R. (1998). How many discoveries have been lost by ignoring modern statistical methods. *American Psychologist, 53*, 300-314.

Winkler, W. E. (2004). Methods for evaluating and creating quality data. *Information Systems, 29*, 531-550.