

## Robust Statistical Analysis for MSW Characterization Studies

Roger Painter\*, Valetta Watson and Ahmad Kheder

Civil and Environmental Engineering, Tennessee State University, Nashville, TN 37209, USA

### Abstract

Tennessee State University conducted a pilot municipal solid waste study for Tennessee Department of Environment at two Tennessee municipal solid waste disposal facilities. A major goal for the pilot study was to develop and demonstrate statistical analysis methodologies to be used in a future statewide municipal solid waste study. In municipal solid waste studies the costs of sorting and weighing a sufficient number of samples to obtain reasonably precise estimates is prohibitive for some waste constituents. This issue regarding the number of municipal solid waste samples was addressed using a real-time iterative analysis that involved simultaneous tracking of the mean, trimmed mean, and median of the sample populations. Sampling was terminated for a particular waste category based on observation of dimensioning incremental improvement in 90 percent confidence intervals for the mean and median. This approach was adopted to take advantage of the robustness of efficiency of the mean for waste categories with near normal distribution and the robustness of validity of the median for grossly non-normally distributed categories. The trimmed mean was included in the analysis as an intermediate estimator to the mean and median with regard to loss of sample information. The convergence of the three estimators for nearly normal data and the trimmed mean intermediate relationship to the mean and median provided excellent field guidance regarding the tradeoff between precision and sampling cost. This approach also provides the option of making statistical inference on the median for grossly non-normal waste subcategories when additional sampling to designate the mean is not an option.

**Keywords:** Municipal Solid Waste; Statistical Analysis, Field Sampling; Solid Waste Characterization

### Introduction

Tennessee Department of Environment and Conservation (TDEC) retained Tennessee State University (TSU) to conduct an initial phase municipal solid waste (MSW) characterization study to better understand the composition of MSW being disposed in Tennessee. The field portion of the study was sampling of waste being disposed at Cedar Ridge Landfill in Lewisburg, TN and Bi-County Landfill in Montgomery County, TN. Statistical analysis was then conducted on the sampling results to determine the composition of targeted MSW streams from the rural areas served by the Cedar Ridge facility and urban areas (Clarksville) served by the Bi-County facility. TSU was directed to develop and refine methodologies for statewide MSW characterization. The resulting statistical analysis methodology will provide TDEC with characterization of Tennessee's MSW in a cost effective manner. Application of the methodology is presented in terms of the paper, metal, and glass waste categories sampled during the MSW pilot study.

### Materials and Methods

Since the main goal of the pilot study was to verify a real time statistical analysis methodology, every effort was made to ensure sampling was representative and random in nature. Each MSW load selected for sampling was tipped into an elongated pile on the ground or the floor of the disposal facility. An imaginary 16-cell grid was superimposed on the tipped load and a random number generator provided the grid number for sampling. The field crew supervisor directed the loader operator to remove and mix the waste from the randomly selected cell and a minimum of 200 pounds of material from the cell was staged near the sorting tables [1]. The number of cells in the sampling grid was adjusted downward for small loads. For example small loads were divided into fewer than 16 cells to ensure that at least 200 pounds per cell was captured for sampling. The sorting crew sorted the material by hand into the prescribed 64 material types. Plastic laundry baskets were used to contain the separated components. The entire sample was initially sorted into the nine major waste types shown in Table 1. The sorting crew members then specialized in types of materials and sorted the major waste types into 64 subcategories

according to the subcategories definitions. The supervisor of the sorting crew monitored the homogeneity of the component baskets as they accumulated, rejecting materials that were improperly classified. Open laundry baskets allowed the supervisor to see the material at all times. The supervisor also verified the purity of each component as it was weighed, before recording the weight on field sheets. The materials were sorted to the greatest reasonable level of detail by hand, until no more than a small amount of homogeneous fine material remained. The supervisor recorded composition weights on Sample Tally Sheets.

The statistical analysis methodology involved an iterative analysis that involved simultaneous tracking of the mean, trimmed mean, and median of the sample populations. Ninety percent confidence intervals were calculated after each waste sort beginning after the fifth sort.  $A_k = 1$  trimmed mean was used in all cases. This represented aggressive

WASTE TYPE	SUB-CATEGORIES
PAPER	11 (1 – 11)
GLASS	6 (12 – 17)
METAL	7 (18 – 24)
ELECTRONICS	4 (25 – 28)
PLASTIC	13 (29 – 41)
ORGANIC	9 (42 – 50)
CONSTRUCTION & DEMOLITION	8 (51 – 58)
SPECIAL WASTE	5 (59 – 63)
MIXED RESIDUE	1 (64)

Table 1: Major Waste Categories.

\*Corresponding author: Roger Painter, Civil and Environmental Engineering, Tennessee State University, Nashville, TN 37209, USA, E-mail: [rpainter@tnstate.edu](mailto:rpainter@tnstate.edu)

Received November 18, 2011; Accepted December 12, 2011; Published December 14, 2011

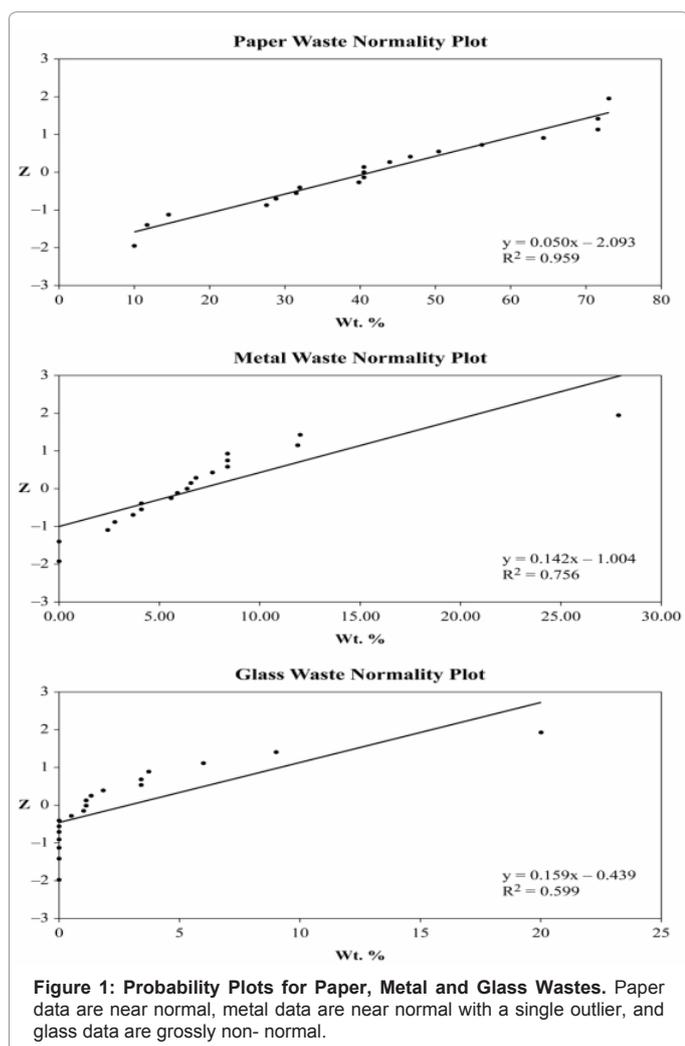
Citation: Painter R, Watson V, Kheder A (2011) Robust Statistical Analysis for MSW Characterization Studies. J Civil Environment Engg 1:102. doi:10.4172/2165-784X.1000102

Copyright: © 2011 Painter R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

trimming when the number of samples was small and became less aggressive and approaching the mean as the sampling continued.

## Results

The analysis for the paper, metal, and glass waste categories are presented below. These categories were chosen for demonstration because of their relative departure from normality as indicated by normality plots shown in Figure 1. The plots indicate that the paper waste data are near normal, the metal data are significantly non normal and the glass data are grossly non normal with multiple low outliers. The 90 percent confidence intervals for the means and medians were estimated incrementally during the study after the completion of the fifth solid waste sort. The confidence intervals are shown in Figure 2. During the pilot solid waste study, field sampling was terminated based on observation of dimensioning incremental improvement in 90 percent confidence intervals for the mean or median for the sample populations. The trimmed mean was not used in for inference of variance. Instead the trimmed mean was used as an indicator for real time sampling decisions. The intermediate nature of trimmed means relative to the mean and median was exploited to better visualize the trending of the confidence intervals for the mean and median. The visualization tool is shown in Figure 3 in for the paper, metal and glass data. The sample mean and  $k = 1$  trimmed mean are superimposed



on a box and whisker plot for the sample medians and the relative confidence ranges of all three estimators are also contrasted.

## Discussion

Statistical analysis of MSW constituents with small weight fractions often provides poor precision unless a very large number of samples are collected. In many studies the costs of sorting and weighing a sufficient number of samples to obtain reasonably precise estimates is prohibitive for all but the most common constituents. It is in this context that engineering judgment is employed to make best use of a usually limited budget by trading off precision and sampling costs. The number of samples needed to achieve a given level of statistical confidence in the results of a solid waste study is a function of the variation of the sampling results and the nature of the underlying distribution of the solid waste concentrations. These factors cannot be known in advance and often cannot be estimated based on the results of other studies. ASTM D 5231 recommends that the number of sorts be estimated based on the Student-t distribution [2, 3]. USEPA also recommends a computerized methodology entitled PROTOCOL for determining the number of samples required for statistical reliance. However PROTOCOL relies on out of date historical data in making statistical inference [4-6]. The Student's t-distribution arises from the estimation of the mean of a normally distributed population and is itself symmetric and bell shaped. Consequently approximations that rely on Student's t to estimate confidence intervals are based on an assumption of near normality distribution sample populations. Unfortunately MSW composition data are not normally distributed but moderately to severely skewed right with significant numbers of values many times greater than the mean. Furthermore, the most frequent value is invariably less than the mean and often zero. This results in poor estimates of confidence intervals for small samples by methods that rely on a normal approximation [7,8]. Approaches have also been developed that rely on real time statistical analysis of the field data and terminate sampling based on the dimensioning improvement in confidence intervals about the mean [9,10]. In these studies the sample means are estimated incrementally during sampling usually based on a Student-t distribution.

The mean as an estimator lacks robustness of validity if the underlying distribution is not normally distributed. A single outlier in the sample can dramatically impact the estimate of the confidence interval for the mean. The median as an estimator has robustness of validity but not robustness of efficiency. The estimation of the median is very tolerant of data outliers. Unfortunately has poor robustness of efficiency as the resulting confidence intervals tend to be relatively wider [11,12]. Trimmed means are a class of robust statistical estimators that attempt to balance the competing elements of robustness of the mean and median by trimming some (data) information to accommodate outliers and long tails but trimming less information than the median. Trimmed means form a continuum with the sample mean and the sample median at the extremes. In this regard the trimmed mean as an estimator of central tendency should coincide with the mean and median for normally distributed data and should serve as an intermediate location indicator between the mean and median of a skewed distribution [13].

The significance of the statistical analysis methodology is its capacity to allow statistical inference for non-normally distributed data in cases where it's not practical to continue field sampling for a desired precision for the mean. We demonstrate this based on the pilot study data for the paper, metal and glass wastes categories. Figure 3 shows that the mean, trimmed mean and median coincide for the near normal

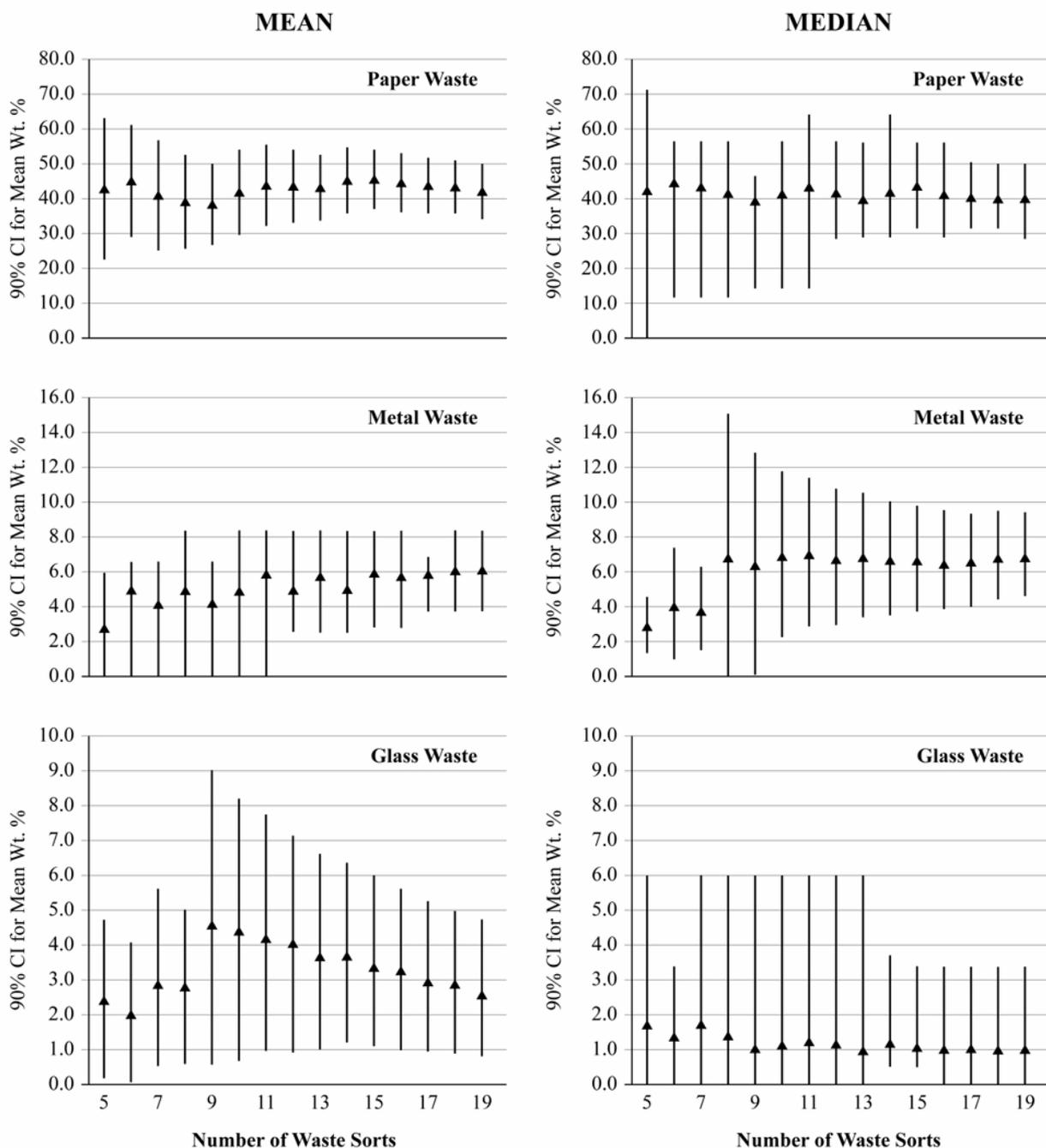


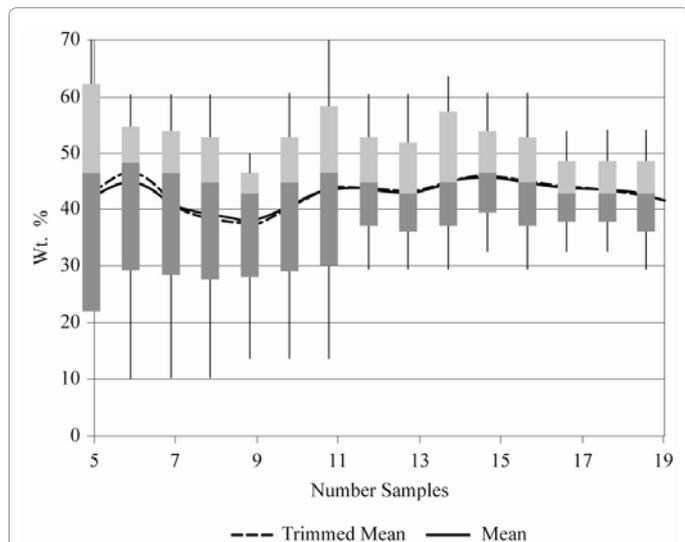
Figure 2: Ninety Percent Confidence Ranges for the Mean and Median for the Paper, Metal, and Glass Wastes.

paper waste data. This indicated that the assumption of normality was appropriate and sampling was terminated based on dimensioning improvement of the confidence interval for the mean. For the metal data, Figure 4 shows that an outlier at waste sort eight dramatically impacted the mean and that the mean and median were still divergent at waste sort nineteen. The trimmed mean was not impacted by the single outlier and the trimmed mean and median almost coincide at 19 samples. This scenario indicated that the data were likely near normal and the outlier was addressed in real time. Sampling was terminated based on the confidence range for the mean of the adjusted data. Finally,

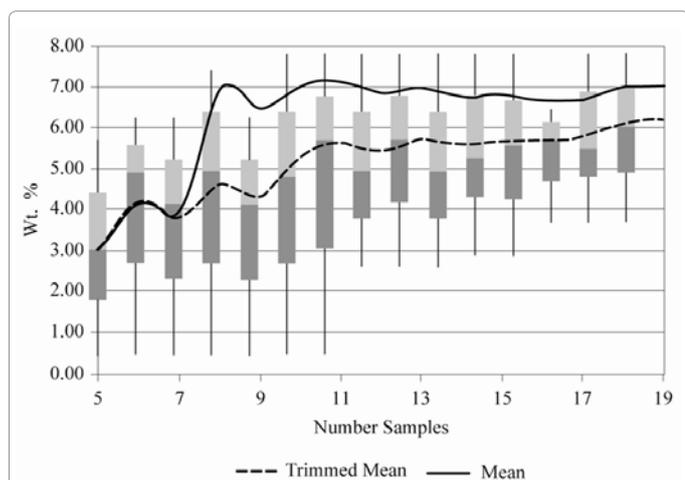
Figure 5 shows that for the grossly non normal glass data that the mean and median are divergent. The trimmed mean is also divergent of the median. This indicates that the normal approximation was not appropriate and sampling was terminated based on the confidence interval for the median.

### Conclusion

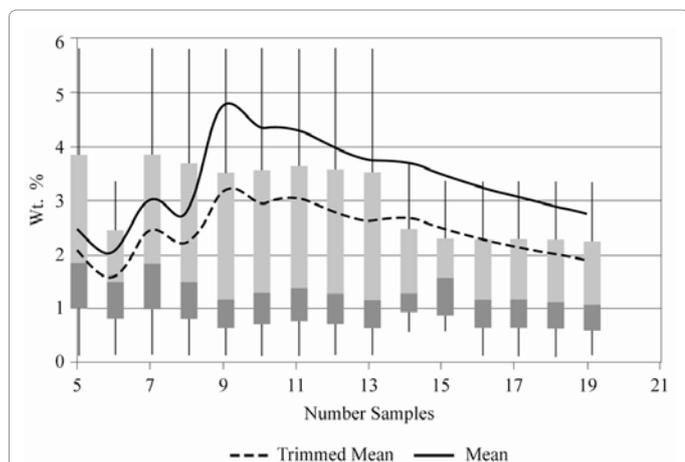
The statistical methodology developed for the pilot solid waste study facilitates field engineering judgment through a real-time iterative



**Figure 3: Paper Waste.** Ninety Percent Confidence Intervals for Mean and Trimmed Mean and Box and Whisker Plot for the Median.



**Figure 4: Metal Waste.** Ninety Percent Confidence Intervals for Mean and Trimmed Mean in Box and Whisker Plot for the Median.



**Figure 5: Glass Waste.** Ninety Percent Confidence Intervals for Mean and Trimmed Mean in Box and Whisker Plot for the Median.

analysis that involved simultaneous tracking of the mean, trimmed mean, and median of the sample populations. Sampling was terminated for a particular waste category based on observation of dimensioning incremental improvement in 90 percent confidence intervals for the mean and median. This approach was adopted to take advantage of the robustness of efficiency of the mean for waste categories with near normal distribution and the robustness of validity of the median for grossly non-normally distributed categories. The trimmed mean was included in the analysis as an intermediate estimator to the mean and median with regard to loss of sample information. The convergence of the three estimators for nearly normal data and the trimmed mean intermediate relationship to the mean and median provided excellent field guidance regarding the tradeoff between precision and sampling cost. This approach also gives the engineer an additional option of making statistical inference on the median for grossly non-normal waste subcategories when additional sampling to designate the mean is not an option.

### References

1. Musa E, Ho GE (1981) Optimum Sample Size in Refuse Analysis, ASCE J Environ Eng Div 107: 1247-1260.
2. American Society of Testing and Materials-ASTM (2008) Standard Test Method for Determination of the Composition of Unprocessed Municipal Solid Waste, ASTM Standard D 5231-5292.
3. California Code of Regulations - CCR (1990) General Guidelines for Sampling When Performing a Quantitative Field Analysis for a Solid Waste Generation Study.
4. Klee AJ (1991) PROTOCOL - A computerized Solid Waste Quantity and Composition Estimation System. EPA/600/2-91/005A RREL Cincinnati OH, February.
5. Klee AJ (1993) New Approaches to Estimation of Solid Waste Quantity and Composition. J Environ Eng 119: 248-261.
6. Reinhart DR, McCauley-Bell P (1996) Methodology for conducting composition study for discarded Solid Waste.
7. Tchobanoglous G, Theisen H, Vigil SA (1993) Integrated Solid Waste Management. McGraw-Hill, Inc New York.
8. Tchobanoglous G (2002) Handbook of Solid Waste Management second ed. McGraw-Hill, New York, USA 1139-1160.
9. Sharma M, McBean E (2007) A methodology for solid waste characterization based on diminishing marginal returns. Waste Manag 27: 337-344.
10. Sharma M, McBean E (2009) Strategy for use of Alternative Waste Sort Sizes for characterizing solid waste composition. Waste Manag Res 27: 38-45.
11. Hogg RV, Craig AT (1994) Introduction to mathematical statistics.
12. Mosteller F, Tukey JW (1977) Data Analysis and Regression. Addison-Wesley.
13. NIST/SEMATECH, 2011, e-Handbook of Statistical Methods. Div 898/handbook/ 2011l.