



Data Integrity and APPA's 2010 FPI

By Ernest Hunter and Maggie Kinnaman

Every year, APPA's Facilities Performance Indicators (FPI) Survey and Report increases in the number of participants, and in the quality of data gathered. For FY2010 we continued our emphasis on the number of participants, but also implemented a significant enhancement that vastly improved quality and integrity of the data being collected.

In August 2010 a team of four started the process of creating a design that would support a data scrubbing team and end users throughout the survey cycle. The team included:

- Christina Hills, APPA FPI Administrator
- Maggie Kinnaman, APPA Fellow and Past APPA President
- Ernest Hunter, Hunter Consulting and Training
- Heather Lukes, Digital Wise, Inc.

Previously, all data scrubbing occurred after the survey cycle closed. At

that point, many participants had completed their survey and had moved onto other initiatives. This made it difficult to address data integrity concerns in January/February. Our data scrubbing team believed that the time to create a participant safety net – and to ensure data integrity was in the moment – was as participants completed a module, not after the entire survey was closed. We wanted to create a design that helped to ensure participants were completing their survey as part of a team, all focused on achieving maximum data integrity, and the most accurate picture of their campus reality.

Scrubbing team members each brought to the table their particular area of expertise. Ernest was focused on statistical analysis of the data, and automating the process of detecting data anomaly based on standard deviation triggers. Maggie was focused on survey participant interaction including mentoring new participants.

Heather was focused on creating the supporting technological platform and tools to support the design. And, Christina provided overall leadership for the project. The goal was to help participants by identifying data that fell outside of upper and lower trigger limits and then working with participant to either validate the correctness of the data, or helping them correct the data. The scrubbing team asked targeted questions to ensure that the participant understood both the definitions for the data points but also how to apply the definition to their institution. This scrubbing process was divided into three phases.

PHASE ONE

The first phase focused on actual data input module by module. The new data scrubbing survey design created a survey home page that enabled the participant to request a data review once they had completed a module. This review team

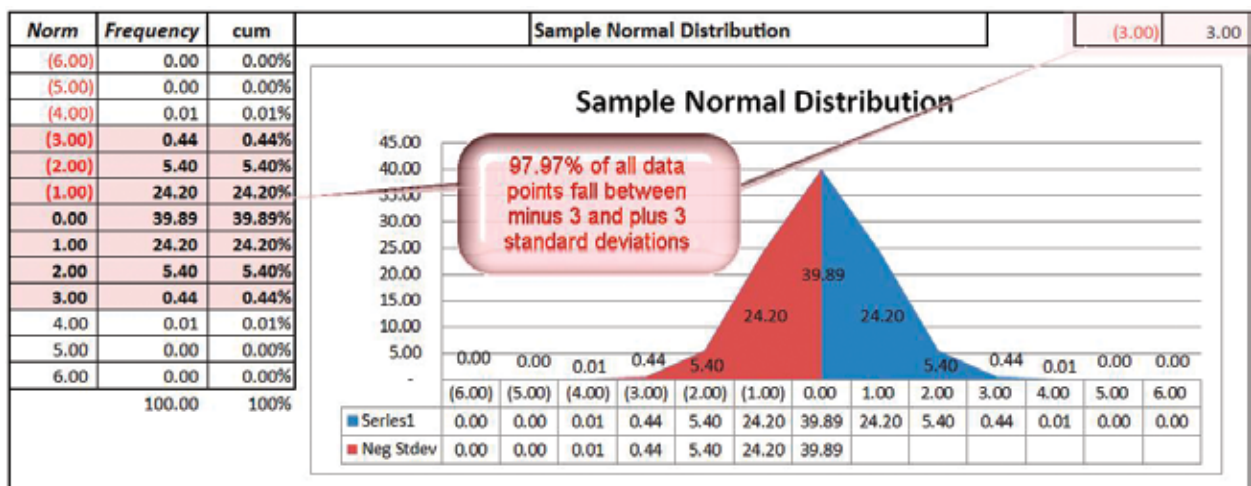


Figure 1 Normal Distribution

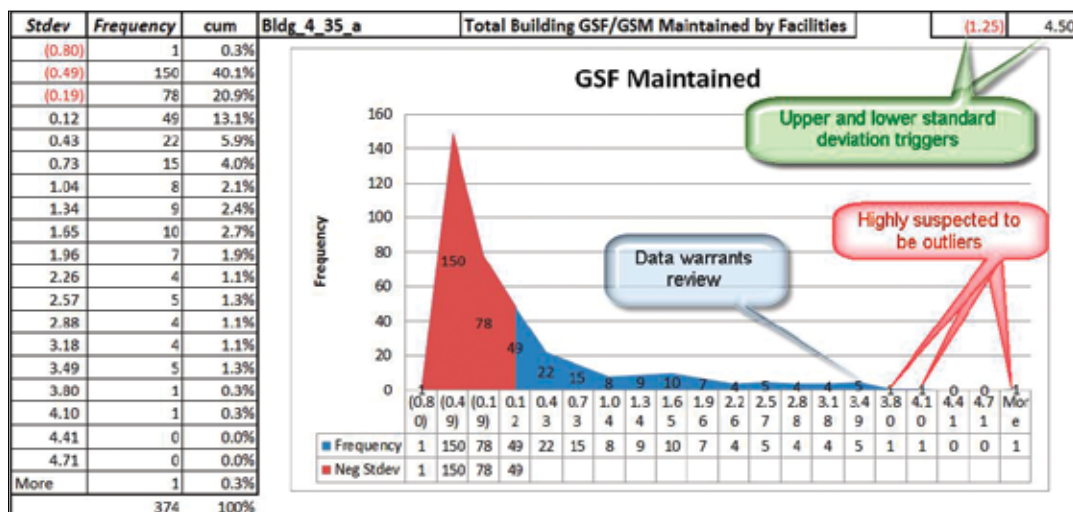


Figure 2 Data Distribution for GSF Maintained

ran two sets of statistics for every data point. They first looked at the data in relation to the prior year input and if there was a variance of more than 20 percent, the data point was flagged as a potential outlier. A second filter was also applied that looked at where the data fell in respect to our upper and lower standard deviation triggers. If no outliers were identified, the participant received a green light on their survey page. If outliers were identified, the home page showed a red light for that module, indicating that future communication would be needed.

Once an outlier was identified, institutions were contacted via e-mail, advised of the possible outlier, and asked to review their data. If the scrubber suspected possible misinterpretation of definitions, then that was pointed out in the e-mail. Participants were asked to change data where necessary and to let us know if they felt that, even after their review, the data point was valid. In most cases data issues were resolved by e-mail.

However, Maggie also spent many hours on the phone with participants resolving data integrity issues and providing one-on-one survey training. Once communication from the end user was completed, the data point was either released as a valid point, or con-

tinued to be flagged as a data point or possible outlier. Once all data points for an institution were resolved, the radio button on the institution's survey home page changed from red to green. With

As an example, a data point such as *Annual Facilities Operating Expenditure* and *Gross Institutional Expenditures*, could fall within normal limits as a stand-alone data points, but once used

388 participants in the 2009-2010 survey, over 10,000 data points were scrubbed using this approach.

PHASE TWO

The second phase of data scrubbing occurred once the survey closed in December 2010. As most of our members are aware, the FPI program takes data input and produces reports of the actual data, as well as using that data to create ratios and measures. These ratios and measures were the focus of phase two data scrubbing.

Amish Country Gazebos

FREE Shipping... Nationwide!

Call 1-800-700-1777 for your FREE catalog, DVD & price guide!

Call Now!

www.AmishGazebos.com

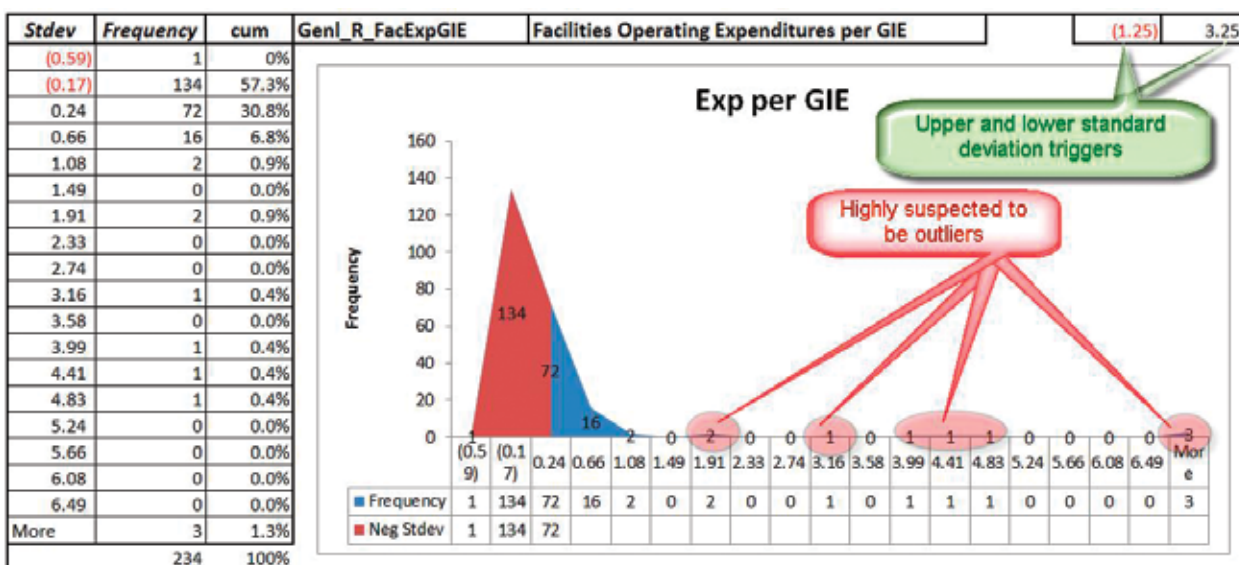


Figure 3 Data Distribution for GIE Index

in a ratio like the GIE index (*Annual Facilities Operating Expenditure* divided by *Gross Institutional Expenditures*), the GIE index might be well outside of our upper and lower standard deviation triggers.

The team developed tools to automate detection of ratios and measures that fell outside of our standard deviation triggers, and focused on them during the second phase. End users were again contacted about the potential outliers and asked to review their data and to let the data scrubber know if a data point used in an outlying ratio required a

change. Often it was necessary to set up a telephone call to resolve the issue as one of two numbers could be the culprit in causing the ratio to be identified as an outlier. At this stage of the data scrubbing process, the participant could not change data themselves, but had to communicate clearly to the data scrubber so that they could make the appropriate changes to the data input.

The new survey design allowed participants to view their survey data and also the reports that were being generated. This helped to facilitate communication with the data scrubber and also helped to achieve a better understanding of how their data input would be used in the published report. In phase two, over 3,000 outlier ratios were resolved ending the cycle, with only five suspect ratios remaining redlined in the published report. This redlined data, although shown in the report, was not utilized in creating any of the overall averages.

In the second phase it was important for the data scrubbing team to understand the natural data distribution profile for the FPI data. They began their pursuit of this understanding by a review of Statistics101, which teaches that the normal distribution is considered the most prominent probability distribution in statistics for sets of data. As can be seen in Figure 1, 97.97 percent of all data points

fall between -3 and +3 standard deviations for a normal distribution.

While the data scrubbing team realized that the FPI data is not a random data set, and that it would not come close to approaching a normal distribution profile, distribution profile analysis with visual presentation such as Figure 2 and Figure 3 would serve as a good tool for systematically identifying potential data outliers.

It turns out that the data distribution for GSF Maintained as shown in Figure 2 is fairly typical for many of the FPI data elements and ratios and measures. Figure 2 is an example of how the scrubbing team reviewed 60 of the most prominent data fields and ratios and measures, and thus easily identified three highly suspected data entries and five data points that warranted review. The participants were contacted and the concerns resolved. With this picture of the distribution profile, the scrubbing team was able to set the upper and lower standard deviation triggers to help them identify other potential data outliers.

Lastly, in further explanation of the second phase and the use of data distribution profile analysis, Figure 3 illustrates the use of data distribution profile analysis to identify ratio and measure outliers – even in instances where the individual stand-alone data fields that make up the ratio and measures were in expected

The 2010-11 FPI Survey Now Open!

- APPA member institutions that participate in the FPI Survey receive the FPI Report at no charge.
- APPA Members who do not participate in the FPI survey pay \$500 for the FPI Report.
- Non-member institutions wishing to participate in the survey pay \$895 for the FPI Report.

Start the survey at:

<http://www.appa.org/Research/FPI/index.cfm>

How Much Could You Save Not Lighting Empty Spaces?

Verve Living Systems' products offer incredibly simple, cost effective solutions for reducing energy usage in student residence rooms, hallways and common areas. Our products are incredibly simple to install, are wireless and battery-less, and require less time and money for on-going maintenance.

Contact us today for more information on how we can help you save energy without sacrificing the safety and comfort of your students, faculty, or visitors.

Toll free 1.877.874.8774 • verve@vervelivingsystems.com

ranges. As can be seen from Figure 3, the scrubbing team identified nine potential outliers. For some of these potential outliers one or both of their companion raw data fields (*Annual Facilities Operating Expenditure* and *Gross Institutional Expenditures*) were also out of range, but for some others, the raw data fields were perfectly in an acceptable range, but mismatched to each other causing the ratio to become an outlier.

PHASE THREE

Finally, phase three of the data scrubbing occurred during the Beta report review period. This phase continued to resolve outlying ratios but also focused on ensuring that data collected in one module of the report was consistent with data collected in other modules. During this phase of scrubbing, hundreds of inconsistencies between modules were resolved, and participants were better educated about the survey and

how each data point is an integral piece in telling an important story.

Feedback from survey participants regarding the data scrubbing initiative was overall very favorable. They indicated that it was comforting to know that someone, other than them, was looking at the data and asking questions to help ascertain the accuracy of the information. Additionally they seemed to appreciate the ongoing nature of the review as they completed each section. They truly felt that they had a partner in helping to ensure that their data was the most accurate representation of their campus facilities realities.

Reflecting on the data scrubbing process for the 2009-2010 FPI survey, the data scrubbing team can safely say that the newly designed process was effective in enhancing the quality of our data. The new process created an environment of enhanced communication and teamwork with our participants. It also resulted in a better

understanding of how future survey design enhancements can reduce the number of data integrity issues.

With new automated statistical data integrity tools now in place, next year's data scrubbing team will be able to further improve data integrity and carve out additional time to better assist our first time FPI participants. Lastly, APPA and the data scrubbing team is well positioned for the future to help the facilities management profession better understand the FPI survey, and how to use the survey report to educate a campus community about their facilities realities. ☺

Ernest Hunter is president of Hunter Consulting and Training in Austin, TX. He can be reached at ernesthunter@gmail.com. Maggie Kinnaman is past APPA President and can be reached at maggiiekinnaman@comcast.net.

We know you've never seen anything like this before

Visual Workloading Software

The future of cleaning management

Smart software for smart people

INFORMED

www.InformedCorp.com



“Switching to an online
work order solution has taught
this old dog new tricks!”

Easy, Efficient Work Order Management.

Loyola Marymount University saved time and money
by moving their work order system to the web.

Learn how you can too!

Go to www.schooldude.com/saveworkorders

