faeis Food and Agricultural Education Information System

400 Turner Street, Suite 102
Blacksburg, VA 24061

May 11, 2011

Dr. Shane Ball
USDA-National Institute of Food and Agriculture

Dear Shane:

Greetings from the FAEIS team! Following this letter you will find the second statistical quarterly report, as required by NIFA's RFA which states that FAEIS will "Produce quarterly reports on the progress in addressing transcription errors, outliers and missing values." (Appendix B). Major accomplishments since the first report (January 2011), include a presentation to NIFA staff on April 7, 2011 and a program review by the FAEIS Statistical Expert Panel on April 8th in Washington, DC. The FAEIS team prepared extensive notebooks for panelists and NIFA representatives, distributed prior to the Panel meeting. Please consider both the FAEIS Statistical Expert Panel notebook and this subsequent report as our second quarterly progress report.

In this report, we present three statistical methods to identify outliers and compare these to the boxplot method presented in the first quarterly report. Results show that these three methods—natural standard deviation, pseudo standard deviation, and Lag1 difference—are more sensitive than the boxplot method for identifying outliers. Also, we hired one additional statistical graduate student to assist in our statistical efforts (Appendix A).

This report also compares FAEIS to the Integrated Postsecondary Education Data System (IPEDS). A PowerPoint presentation given by Dr. Eric Vance at the Statistical Expert Panel meeting is included as Appendix C. Results show that overall, FAEIS data provide users greater specificity by collecting all data using Classification of Instructional Programs (CIP) codes. In contrast, IPEDS only provides CIP code breakdown in its degrees awarded survey. This is important since data analyses can be disaggregated at the degree/discipline level using CIP codes with FAEIS. Here is a brief summary that addresses each of the items mentioned in the RFA:

RFA Items	Timeline for Deliverables		
3	√	Creation of SAS dataset and report verification	10/2010
2,5	√	Identification of outliers and missing data	04/2010
9	√	Statistical update quarterly report	01/2011
2,5	√	Improvement of Identification of outliers	04/2010
2	√	Comparisons of IPEDS and FAEIS	04/2011
1,6	√	Statistical Expert Panel meeting	04/2011
9	√	Statistical update quarterly report	05/2011
2		Identification of redundant entries and miscoded CIP codes	06/2011
2		Automated identification of problematic data	06/2011
		Non-Universal database problem	07/2011

Thank you and please contact us if you have any questions on this second quarterly report.

Sincerely,

Mary A. Marchant, Ph.D.
FAEIS Principal Director

Progress on the Statistical Analysis of FAEIS Data Quality —Second Quarterly Report—

May 11, 2011

Food and Agricultural Education Information System (FAEIS)

http://faeis.usda.gov
mailto:faeis@vt.edu

540-231-4941

- Mary Marchant, Ph.D., FAEIS Principal Investigator
  Agriculture and Applied Economics Department, Virginia Tech
- Timothy Mack, Ph.D., FAEIS Co-Principal Investigator
  School of Graduate Studies and Research, Indiana University of Pennsylvania
- Eric Smith, Ph.D., FAEIS Co-Principal Investigator
  Statistics Department, Virginia Tech
- Bill Richardson, FAEIS Project Manager
  Agriculture, Human and Natural Resources Information Technology (AHNR-IT), Virginia Tech
- Eric Vance, Ph.D., FAEIS Statistical Project Manager and LISA Director
  LISA (Laboratory for Interdisciplinary Statistical Analysis) and Statistics Department, Virginia Tech
- Albert Shen, Ph.D., FAEIS Statistical Graduate Research Assistant (GRA)
  Statistics Department, Virginia Tech
- Katie Griffin, FAEIS Graduate Research Assistant
  Statistics Department, Virginia Tech
- Ashley Bell, FAEIS Graduate Research Assistant
  Dairy Science Department, Virginia Tech
- Lisa Hightower, FAEIS Graduate Research Assistant
  Agricultural and Extension Education Department, Virginia Tech

Progress on the Statistical Analysis of FAEIS Data Quality —Second Quarterly Report—

Introduction

This is the second in a series of quarterly reports from the Food and Agricultural Education System (FAEIS) to the U.S. Department of Agriculture-National Institute of Food and Agriculture (USDA-NIFA), in response to item #9 in the FAEIS RFA (see Appendix B) which states: “Produce quarterly reports on the progress in addressing transcription errors, outliers and missing values. Include statistical procedures used to correct and process FAEIS data.”

Summary

In the first progress report, the FAEIS Team explored the use of the boxplot method to identify outliers. In this second report, we explore three additional methods and compare them to the earlier boxplot method. These three methods include: natural standard deviation (NSD), pseudo standard deviation (PSD), and Lag1 difference. Results show that these three methods are more sensitive than the boxplot method for identifying outliers.

On April 8, 2011, the FAEIS Statistical Expert Panel met in Washington, DC in response to item #1 of the FAEIS RFA. Prior to the meeting FAEIS staff prepared and sent notebooks to panel members and USDA-NIFA representatives. This progress report includes both the report that follows, as well as the FAEIS Statistical Expert Panel notebook, previously sent to USDA-NIFA.

In this report we compare FAEIS to IPEDS (see section 3). Overall, FAEIS data are collected with greater specificity by collecting all data using Classification of Instructional Programs (CIP) code classifications. In contrast, IPEDS only provides CIP code breakdown in its degrees awarded survey. Some of the additional major differences found between FAEIS and IPEDS are as follows:

- FAEIS enrollment data have greater granularity, since data are collected annually by CIP code and thus is far more useful to make comparisons at the degree/discipline level. In contrast IPEDS collects enrollment data every 2 years and for only 6 aggregated fields of study at the undergraduate level and 9 aggregated fields of study at the graduate level; none of the aggregated fields are related to Agriculture.
- FAEIS data are released ~10 months before IPEDS data.
- FAEIS collects finer data on faculty, allowing average salary comparisons by discipline, rank, tenure status, contract length, gender and ethnicity. IPEDS only collects faculty data aggregated at the institution level.
- FAEIS reporting features are faster, more powerful and easier to use than IPEDS.

1. FAEIS has added additional statistical expertise

(Refers to RFA item 2; see Appendix B)

In January, FAEIS added a second Graduate Research Assistant (GRA) from the Statistics department, Katie Griffin, to the Help Desk team. As a graduate student pursuing a Masters degree in statistics, Katie brings in statistical expertise and part of her duty is to work on the comparison of FAEIS data with the Integrated Postsecondary Education Data System (IPEDS) and other data sources. Some of the comparison results were presented at the FAEIS Statistical Expert Panel on April 8, 2011, and received positive feedback. She joins FAEIS Team members Eric Smith, Eric Vance, and Albert Shen from the Virginia Tech Statistics Department, also described on the first page of this report and in Appendix A.

2. SAS algorithms have been improved to identify outliers

(Refers to RFA items 2 & 5; see Appendix B)

In the first quarterly progress report, we reported the development of boxplot and strip plot methods to identify outliers in the FAEIS data. Boxplots and strip plots provide limited information when the sample size is small. In the FAEIS data, the boxplots are based on only up to six data points. Therefore, in this quarter we developed three new methods to improve the identification of outliers. We use Bachelor Enrollment in Food Science and Technology as an example to illustrate each method and to make comparisons.

In the boxplot (Figure 1; top), we should pay attention to two features of the display. The first feature is the outliers, which are labeled as red circles with the values (enrollment) to the right. The second feature is the tall boxes, which indicate large variation of the data between 2004 and 2009. To take a closer look at the variations in the data, the strip plot is very helpful.

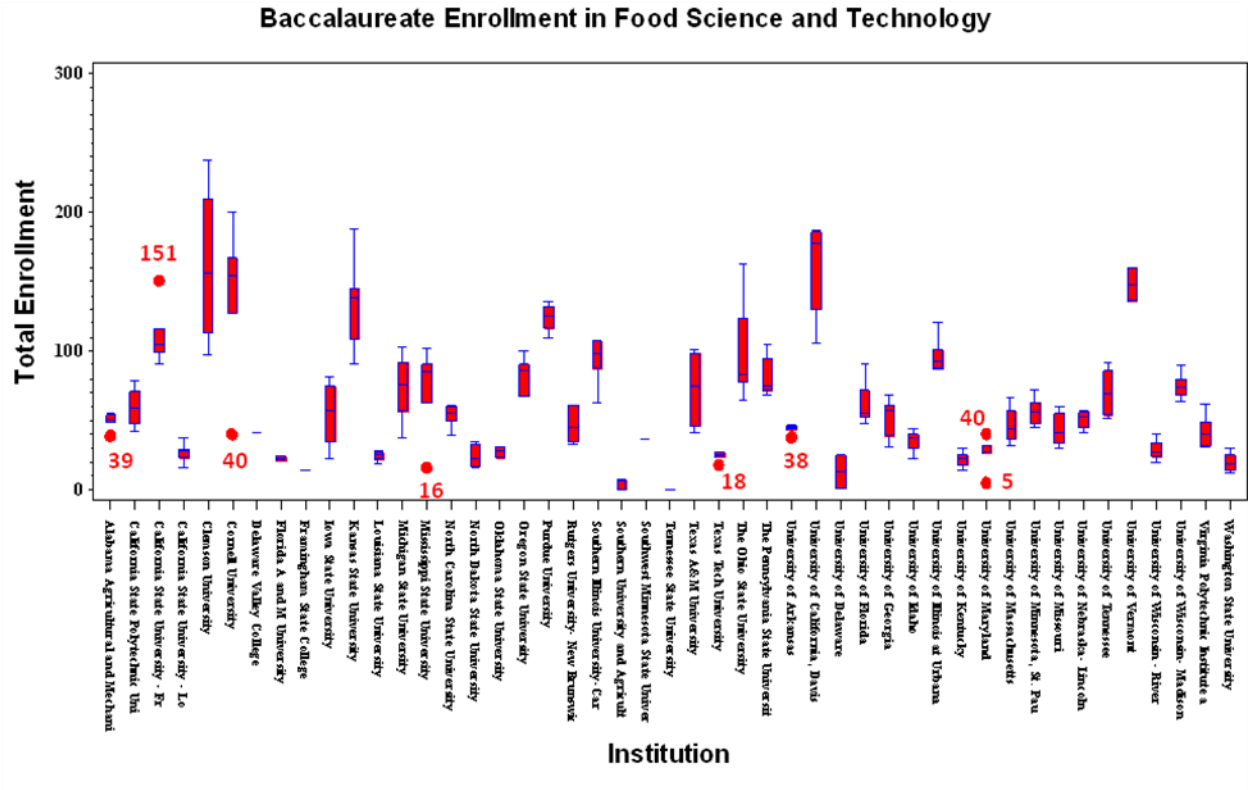
In the strip plot (Figure 1; bottom), the enrollment for each year is plotted as a dot. Clusters of dots will reflect small variation and a small box in the boxplot. On the other hand, if the dots are widely scattered, there is large variation and a large box is formed in the boxplot. Another issue to explore from the strip plot is the “trend” of enrollment when the variation is large. If there is an obvious trend (increase or decrease) of enrollment with year, the data is more reliable. On the other hand, if there is no obvious trend for the large variation, the data may be questionable and would be flagged for further investigation.

The first new method uses the ordinary or **natural standard deviation** (NSD) derived from the FAEIS data. First, we create groups (small, medium, large, and extra large) based on the enrollment size. We then calculate the standard deviation for each group in the same academic area. The outliers are flagged for the observations that are outside two or three standard deviation from the mean. An example using this method is shown below (Figure 2; top). Six observations are identified as outliers, three of them (151, 238, 40) are outside two

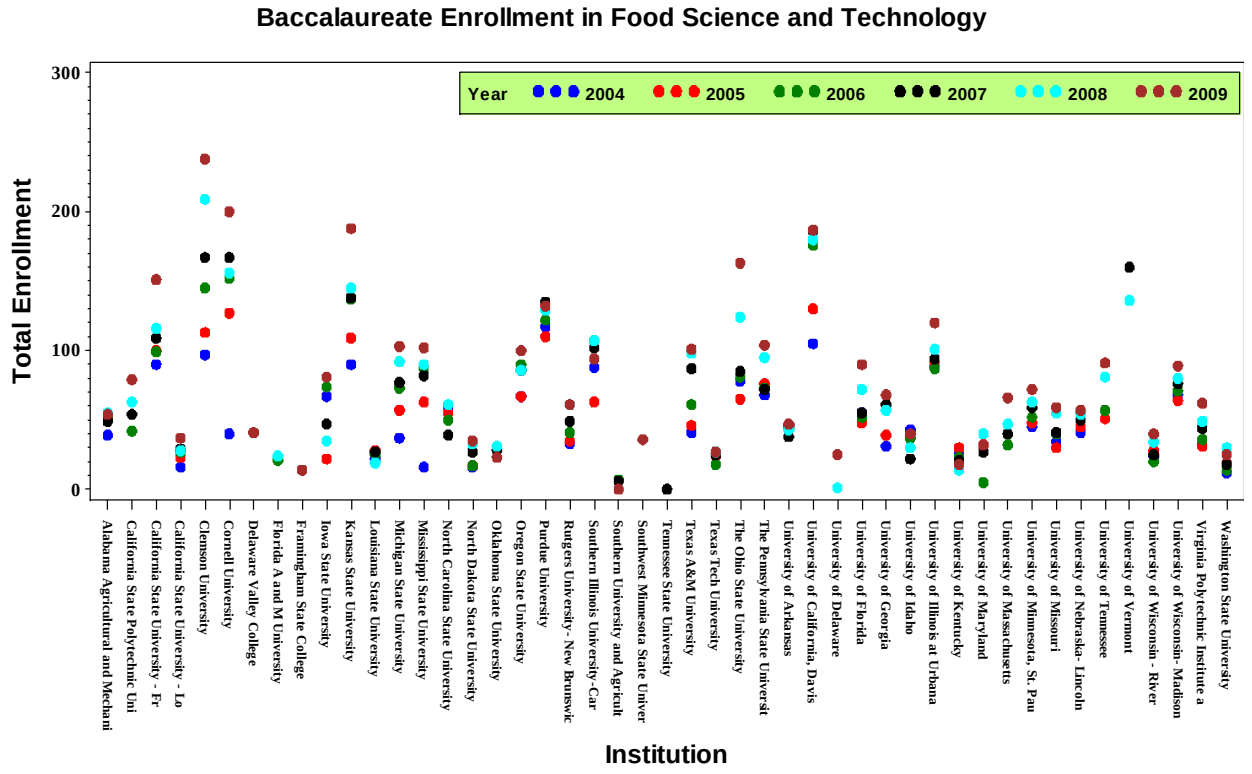
standard deviations from the mean and the other three (16, 163, 5) outside three standard deviations. When compared to the boxplot method, the NSD method identified fewer outliers (8 from boxplot *vs.* 6 from NSD). Four (151, 16, 40, 5) of the six observations identified by the NSD method are also identified by the boxplot method. The NSD method is more sensitive to the observations that stretch over a wide range (i.e. the long boxes in the boxplot) and less sensitive to the observations that stretch over a narrow range, i.e. the short boxes, than the boxplot method. Overall, the NSD method is better than the boxplot method.

The new second method uses the **pseudo standard deviation** (PSD) based on interquartile range (IQR). Again we first create groups based on the enrollment size. We then obtain the IQR, the distance between the 75^{th} percentile and the 25^{th} percentile, for each group in the same academic area. The pseudo standard deviation is calculated as IQR/1.35. The outliers are flagged for the observations that are outside two or four PSD from the median, which is equivalent to 1.5 or 3 IQR from the median. An example using this method is shown below (Figure 2; bottom). The identified outliers are almost identical to those using the NSD method. One observation (61) identified using the PSD method is not identified using the NSD method. Therefore, the PSD method can be used as the supplementary method to the NSD method.

The third new method looks for **an odd change in the data pattern (the Lag1 Difference**). The Lag1 difference plot is the plot of the difference between enrollments in adjacent years. The difference between one year and the next year is calculated. Then the standard deviations of the differences are calculated using the same approach as in the regular calculations. Again, an ordinary standard deviation may be calculated or one that is robust to outliers based on the interquartile range may be used. The odd values that are identified in the two plots are the same values. However, in the first plot we note that some values that are in the middle of the total enrollment plot (Figure 3; top) do not appear to be outliers. When displayed in the difference plot (Figure 3; bottom) we note that they are extreme. For example, compare Oregon State University in both plots. The middle value in the first plot appears extreme in the second plot.

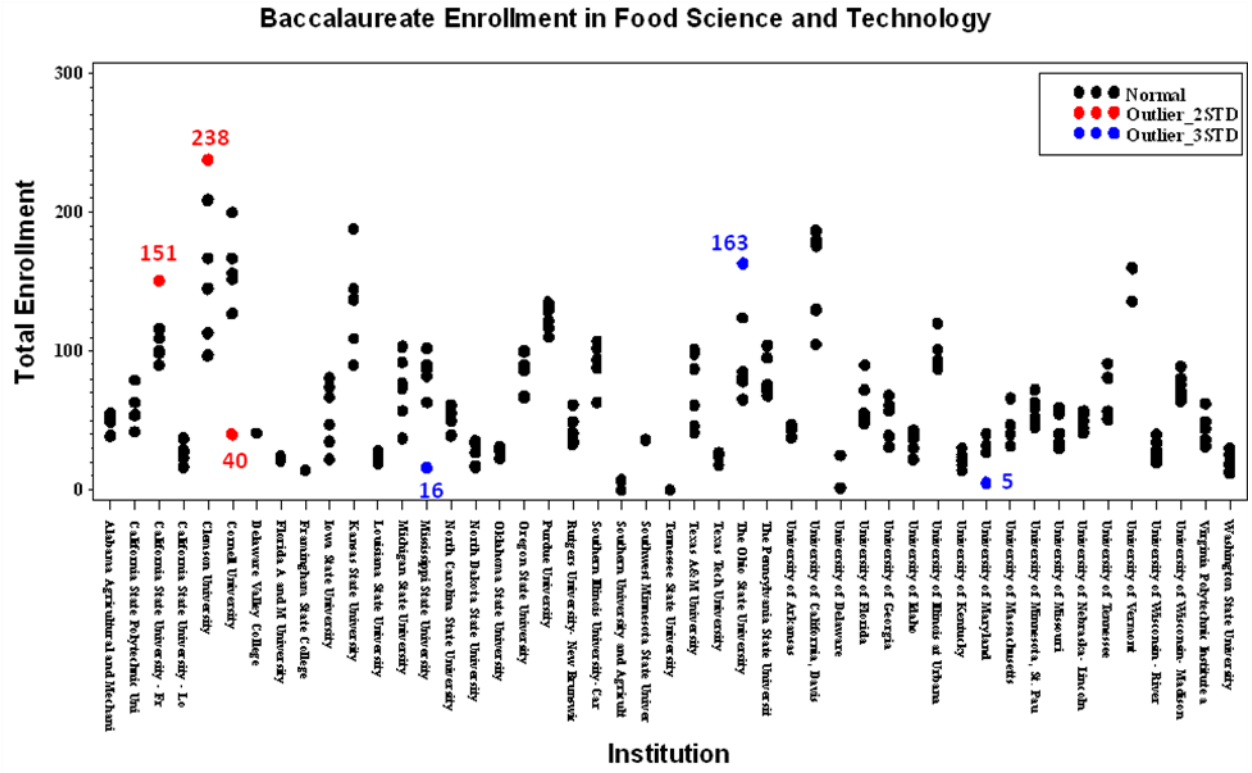


Boxplot of Enrollment by Institution

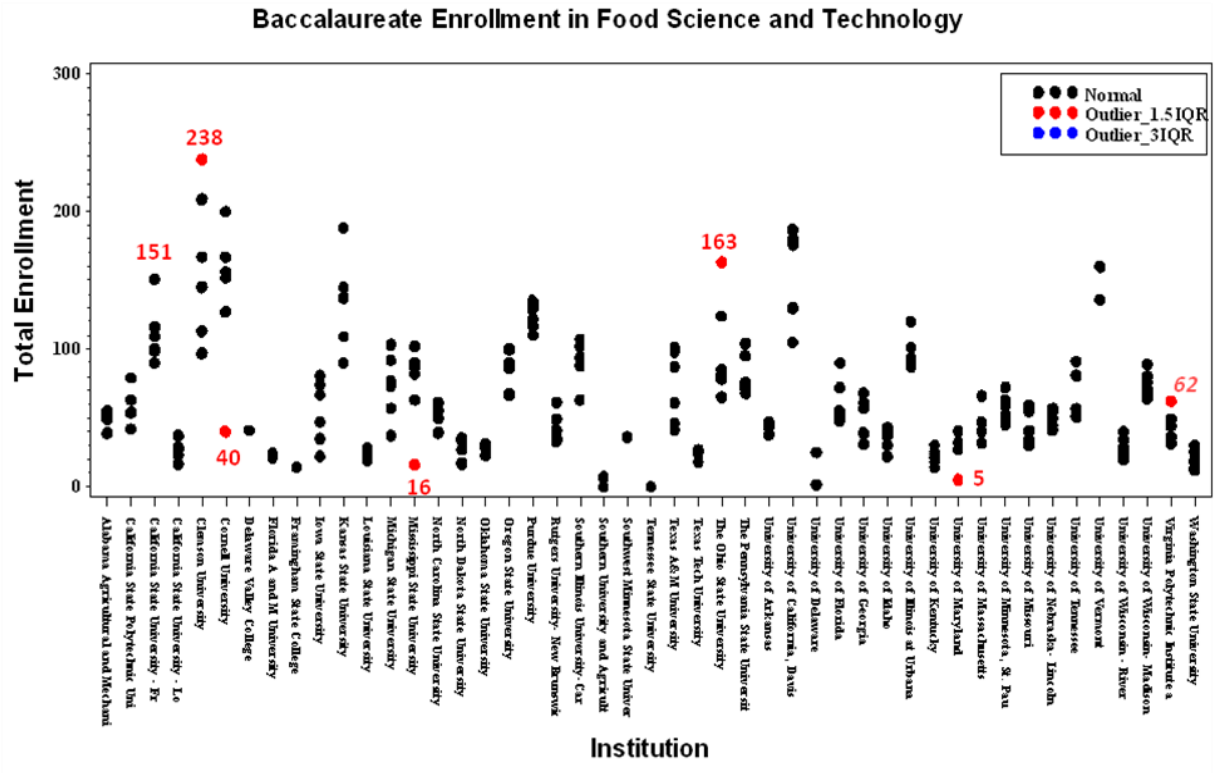


Strip Plot of Enrollment by Institution

Figure 1. Boxplot and Strip Plot for Identifying Data Quality

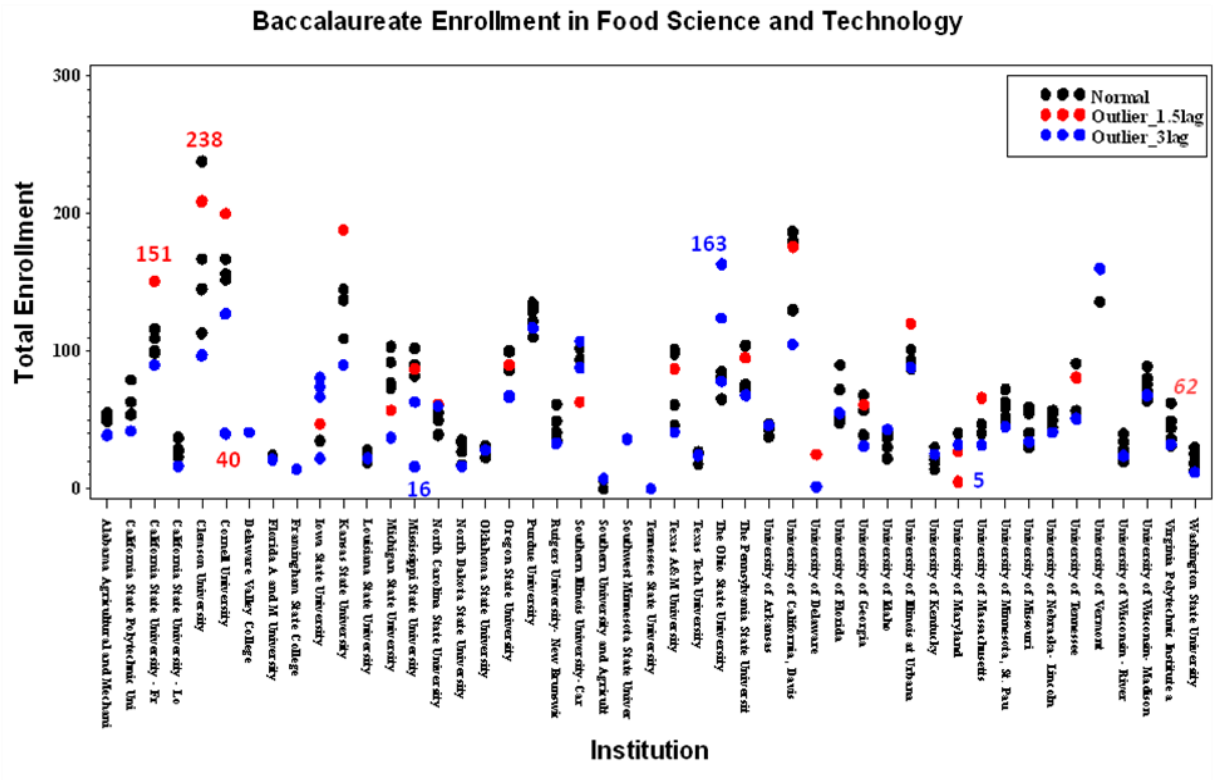


Outliers of Enrollment by Institution Using Natural Standard Deviation

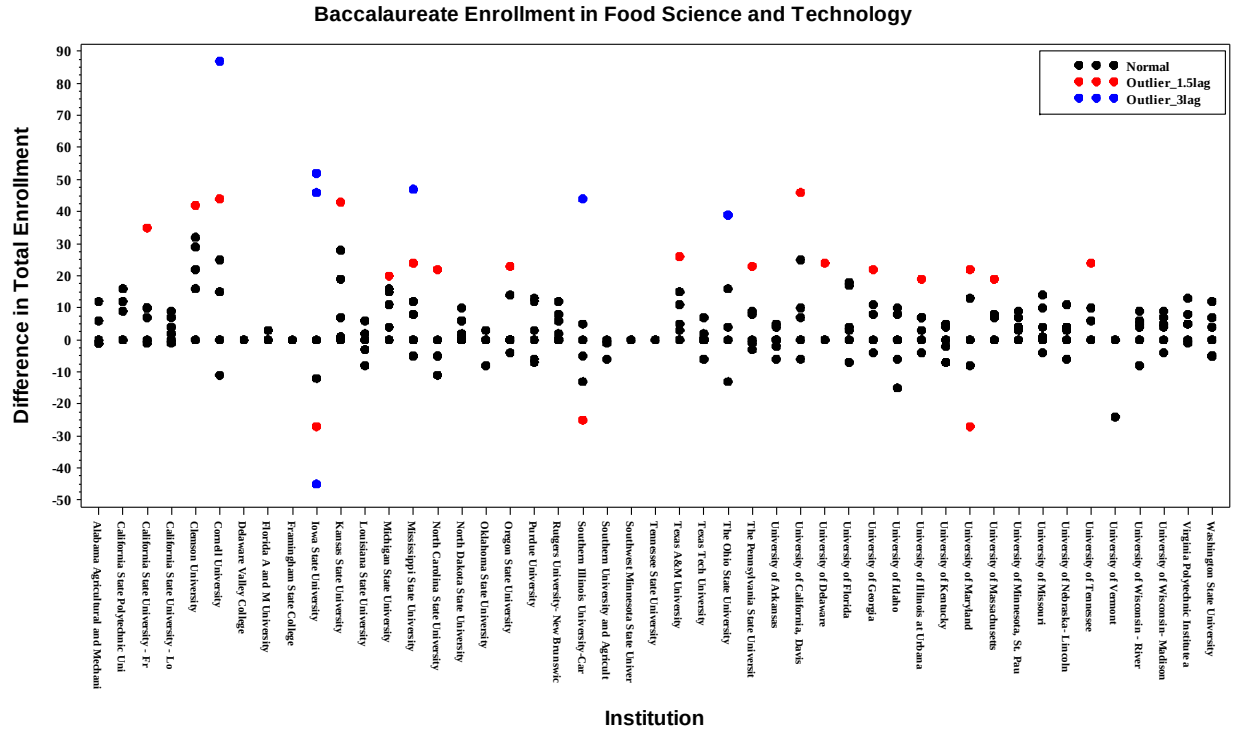


Outliers of Enrollment by Institution Using Inter Quartile Range

Figure 2. Plots for Identifying Data Quality



Outliers of Enrollment by Institution Using Lag1 Difference



Plot of Difference in Enrollment by Institution Using Lag1 Difference

Figure 3. Plots for Identifying Data Quality

3. Comparisons of FAEIS and IPEDS

(Refers to RFA item 2; see Appendix B)

In March 2011, FAEIS began a project to provide a detailed comparison of FAEIS to IPEDS and other comparable databases, such as the Survey of Earned Doctorates (SED) containing student and faculty information at the national level. A PowerPoint presentation on this subject was given by Dr. Eric Vance at the Statistical Expert Panel meeting on April 8, 2011 and is included in this report as Appendix C.

The FAEIS database system has been compared to the Integrated Postsecondary Education Data System (IPEDS). IPEDS collects data in seven areas that include institutional characteristics, institutional prices, enrollment, student financial aid, degrees and certificates conferred, student persistence and success, and institutional resources. FAEIS and IPEDS can be compared on three of these main areas: enrollment, degrees and certificates conferred, and institutional resources.

Table 1 compares FAEIS and IPEDS data sources. In making comparisons across all categories, it is important to note that **FAEIS always collects and reports all data by CIP codes, while IPEDS collects only degrees awarded by CIP codes.** This provides a greater level of specificity for FAEIS data. For example, FAEIS collects detailed annual data for student enrollment and faculty while IPEDS does not.

Table 1

	FAEIS	**IPEDS**
Degrees Awarded	Yes	Yes
- every year by demographic variables	Gender Race/Ethnicity Degree level	Gender Race/Ethnicity Degree level
- every year by CIP codes	Yes	Yes
Faculty	Yes	Yes (instructional only)
- every year by CIP codes	Yes	*No*
- every year by demographic variables	*Race/Ethnicity* *Tenure Status* *Age* Gender Academic rank Contract length	Gender Academic rank Contract length
- every year by average salary	Yes	Yes

Table 1-Continued:

Fall enrollment	Yes	Yes*
- every year by demographic variables	Gender Ethnicity	Gender Ethnicity
- every year by CIP codes	Yes	*No*
- every year by degree level (AA, BS, MS, PHD)	Yes	*No (graduate & undergraduate totals only)*
Academic year enrollment	No	Yes
- every year by demographic variables	--	Yes
- every year by CIP codes	--	No

For **enrollment,** both databases collect data every year by the demographic variables of race/ethnicity and gender. IPEDS also collects fall enrollment data by level of study, which indicates whether a student is an undergraduate or graduate. FAEIS collects these data in far more detail, subdividing enrollment by Associates, Bachelors, Masters, and Doctorate programs. But the biggest difference between the two databases' fall enrollment data is that **FAEIS collects fall undergraduate and graduate enrollment data by CIP code every year, whereas IPEDS collects these data every two years aggregated by six fields of study at the undergraduate level and nine fields of study at the graduate level.**

IPEDS and FAEIS contain comparable data for **degrees awarded,** referred to in IPEDS as degrees and certificates conferred or completions. FAEIS and IPEDS both use the demographic variables of race/ethnicity and gender for this survey. Also, both are collected at the degree level. For these data IPEDS collects completions by CIP codes, as does FAEIS. Essentially, the types of data included in these surveys appear to be identical between FAEIS and IPEDS. However, upon further investigation when producing reports from both systems, more often than not there are inconsistencies among the number of degrees awarded. This could be due to the fact that data come from different sources for the two database systems. The data in FAEIS are reported by individuals within the appropriate college or department where the degree was completed within the institution; thus closer to the source. Whereas, IPEDS data come from a central source for each institution, usually institutional research (IR) or comparable office. Also important to note is that reporting to IPEDS is required by all institutions participating in student financial aid programs from the government, whereas reporting to FAEIS is voluntary.

Finally, for **faculty numbers and salaries,** within the IPEDS survey of institutional resources there is a component called human resources, which covers headcount and salary of all institutional staff members, including faculty; thus it can be compared to the faculty survey in FAEIS. As with other IPEDS data, faculty data are not collected by CIP code, but rather at the institution level. Both systems collect academic rank, contract length, tenure status, gender, and race/ethnicity for faculty. Another difference is that IPEDS can only report faculty salary data by race/ethnicity for salary ranges, not by average salary. However, FAEIS can report average faculty salary by race/ethnicity. Furthermore, FAEIS collects all faculty data by CIP codes, whereas IPEDS collects faculty data by entire institution. **Thus, FAEIS can specify faculty salaries for specific disciplines by gender, ethnicity and rank, while IPEDS reports faculty salary at the institution level by gender and rank with ethnicity reported in salary ranges.**

Summary

- FAEIS always collects and reports all data by CIP codes, while IPEDS collects only degrees awarded by CIP codes.

- IPEDS does not collect fall enrollment data for individual CIPs or degree areas. In odd-numbered years, IPEDS collects fall enrollment on six major fields of study for undergraduates and three major fields of study for graduates. None of the fields of study are related to agriculture.

- FAEIS salary data are collected with greater specificity. FAEIS data provides information on average salaries by gender and race/ethnicity by CIP codes.

- FAEIS reporting is both easier for users and more powerful than IPEDS. IPEDS requires more steps than FAEIS to generate even the simplest reports. IPEDS can not report across years, so multiple reports must be generated and combined externally. IPEDS does not generate multi-dimensional reports, and does not generate finished reports, only extracted data, which needs editing to eliminate unwanted columns.

4. Statistical Expert Panel Meeting

(Refers to RFA item 1; see Appendix B)

The FAEIS Statistical Expert Panel meeting was held on April 8th, 2011 in Washington, DC, consisting of the following panel members:

Dr. Ali I Mohamed, **Panel Chair**
Director, Division of Environmental Systems
U.S. Department of Agriculture
National Institute of Food and Agriculture

Mr. Jim Alessio
Director, State Council of Higher Education for Virginia

Dr. Ken Esbenshade
Associate Dean and Director of Academic Programs
College of Agriculture and Life Sciences
North Carolina State University

Dr. Nagaraj Neerchal
Chair, Math and Stat Department
Department of Mathematics & Statistics
University of Maryland, Baltimore County

Ms. Sabrina Ratchford
Statistician, Postsecondary, Adult and Career Education
U.S. Department of Education, NCES

Dr. Nicole Smith
Research Professor and Economist
Center on Education and the Workforce
Georgetown University

The meeting began at 9:00am with a brief introduction from Dr. Frank Boteler, USDA-NIFA. The report and recommendations from the Statistics Expert panel are forthcoming and will be addressed in the next edition of this report.

5. Future Work

(Refers to RFA items 1, 2B, 6; see Appendix B)

In the subsequent quarters, the following will be enacted:

1. SAS algorithms to identify redundant/repeated data entries and miscoded CIP codes –

   Redundant or repeated data entries have been found in the FAEIS data, as well as miscoded CIP codes. These types of data errors are not easily identified manually. Redundant data often occur when the same information was entered multiple times using different FAEIS accounts. Often these data appear to be outliers when compared to other years. A SAS algorithm is being developed to identify the redundant data by searching for multiple accounts and for outliers. Misplaced CIP codes often have the feature of missing data for a certain CIP code in certain years when the data is placed in another similar CIP code. A SAS algorithm is being developed to identify the misplaced CIP codes by matching the CIP codes with missing data.

2. Automated identification of invalid/problematic data –

   Once the identification of erroneous data with SAS is fully developed and tested, it will be automated with Microsoft Task Manager on a daily or semi-weekly basis. The detected erroneous data will be sent by e-mail to the graduate research assistants at the FAEIS help desk for further investigation. The FAEIS GRAs will follow up with the corresponding institution representatives. This will be a routine data quality assurance procedure. We have achieved some success in the preliminary tests and are in the process of improving the formats and results.

3. Responses to suggestions from the Statistical Expert Panel.

4. Further comparisons of FAEIS with IPEDS and other national databases.

5. Expanded use of data mining and direct contacts with Institutional Research offices for obtaining data for institutions with missing data.

RFA Items	Timeline for Deliverables		
3	√	Creation of SAS dataset and report verification	10/2010
2,5	√	Identification of outliers and missing data	04/2010
9	√	Statistical update quarterly report	01/2011
2,5	√	Improvement of Identification of outliers	04/2010
2	√	Comparisons of IPEDS and FAEIS	04/2011
1,6	√	Statistical Expert Panel meeting	04/2011
9	√	Statistical update quarterly report	05/2011
2		Identification of redundant entries & miscoded CIP codes	06/2011
2		Automated identification of problematic data	06/2011
		Non-Universal database problem	07/2011

Appendix A—FAEIS TEAM MEMBERS

FAEIS Help Desk Staff

Bill Richardson: FAEIS Project Manager

Bill Richardson received a Bachelor of Science in Forestry at Virginia Tech in 1976. He began working at Virginia Tech in 1983 and in 1993 in Agriculture, Human and Natural Resources Information Technology in the College of Agriculture and Life Sciences. He has been with the FAEIS project since it came to Virginia Tech in 2001, starting as the lead programmer and later adding the dual role of project manager.

Dr. Jolene Hamm: FAEIS Consultant / Former Help Desk Manager

Jolene Hamm completed her PhD in Agricultural Education and Extension at Virginia Tech in December 2010. She worked as the FAEIS Help Desk Manager and a graduate research assistant for nearly three years prior to graduation. Dr. Hamm has authored a series of refereed journal articles, including one on FAEIS. Dr. Hamm is currently working at the Office of Institutional Research and Effectiveness at Ferrum College in Virginia.

Albert Shen: FAEIS Statistics Graduate Research Assistant

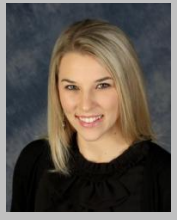
Albert Shen received a Bachelor of Science in physics at National Tsing-Hua University in Taiwan. He received a Masters degree in statistics from Columbia University. He completed a doctorate in biophysics from the University of Virginia. He is currently working toward a doctorate in the Statistics Department at Virginia Tech.

Katie Griffin: FAEIS Statistics Graduate Research Assistant

Katie Griffin received a Bachelor of Science in Mathematical Sciences from Loyola University Maryland. She is currently completing her Masters degree in the Statistics Department at Virginia Tech.

Lisa Hightower: FAEIS Help Desk Graduate Research Assistant

Lisa Hightower received a Bachelor of Science in Journalism and minored in video production and a Masters degree in agricultural communication at the University of Florida. She is currently completing a doctorate degree in the Agricultural Education and Extension Department at Virginia Tech.



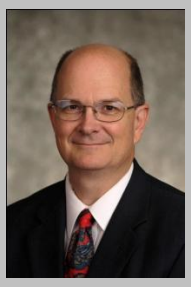
Ashley Bell: FAEIS Help Desk Graduate Research Assistant

Ashley Bell received a Bachelor of Science in Animal and Poultry Sciences and minored in Biology at Virginia Tech. She is currently working on a Masters degree in the Dairy Science Department at Virginia Tech.

FAEIS Principal Investigators

Dr. Mary Marchant: Principal Investigator

Dr. Mary Marchant obtained all of her advanced degrees at the University of California Davis. Upon graduating with a PhD in agricultural economics, she joined the University of Kentucky faculty, where she worked for 17 years. Dr. Marchant joined Virginia Tech (VT) as Associate Dean and Director of Academic Programs for the College of Agriculture and Life Sciences in April 2006 and recently joined the VT faculty in the Department of Agriculture and Applied Economics.

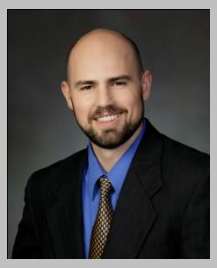


Dr. Tim Mack: Co-Principal Investigator

Dr. Tim Mack is the dean of the School of Graduate Studies and Research at Indiana University of Pennsylvania (IUP). Mack came to IUP from Georgia Southern University, where he was the Dean of the Jack N. Averitt College of Graduate Studies. Previous to that position, he worked at Virginia Tech, serving for three years as Associate Dean for Information Technology and Distance Education in the College of Agriculture and Life Sciences. Dr. Mack was instrumental in bringing FAIES to VT and served as the original principal investigator.

Dr. Eric Smith: Co-Principal Investigator

Dr. Eric P. Smith has been a member of the Statistics Department at Virginia Tech faculty since 1982 and chair of the department since 2006. His research focuses on the development and application of statistical methods to help understand and solve environmental and ecological problems. He was the director of the Statistical Consulting Center 1995-2004.



Dr. Eric Vance: Statistical Project Manager

Dr. Eric Vance is an Assistant Research Professor in the Department of Statistics at Virginia Tech. He received his MS in statistics and decision sciences from the Institute of Statistics and Decision Sciences at Duke University and his PhD in Statistical Science from the Department of Statistical Science at Duke University. He has more than 7 years of experience contributing statistical expertise to interdisciplinary research projects. Since 2008, he has been the director of the Laboratory for Interdisciplinary Statistical Analysis (LISA).

Appendix B--USDA-NIFA 2010-2011 RFA--Final Year of Contract

NIFA calls for significant advanced expertise, detailed reporting and communications:

1. Because FAEIS is a national database, it is expected that data management and analyses must be reviewed by an external expert panel. The expert panel will determine the limitations of FAEIS data, proper interpretation and analyses of data from a voluntary data submission process, which is an unreliable data collection source.

2. Data and data analyses must be the products of significant statistical expertise that reflects standards for survey data management, analyses, and interpretation. Transcription errors will be corrected by implementing quality control procedures before the statistical analyses are performed. Methods to accomplish this include:
   a. Proofing data visually in column by row format (Excel or SAS file) by FAEIS employees. In addition, use exploratory data analyses as an additional quality check and test the assumptions.
   b. Systematic testing of data to determine its data accuracy to a "gold-standard" database – which is the IPEDs database.
   c. Outlier tests to highlight abnormal values and eliminate them before other statistics are calculated. Also, remove any redundancy and "orphan records" from the database. Applicant should use appropriate statistical outlier tests to determine if data are wrong and can be removed. Examples include the Shapiro-Wilks test, non-parametric methods and robust statistics (e.g., median and median absolute deviation). Generally, census/survey data are messy and often require multiple imputation methods.

3. SAS software is to be used to conduct data management (Excel and SAS files as output) to ensure ease of data transfer.

4. Data must correctly reflect the real world. All tests and procedures correcting the data must be completed before the FAEIS clienteles are given access to the data.

5. Missing data must be addressed. Examples of analysis techniques to perform this include the (1) Casewise deletion, (2) Pairwise deletion, (3) Mean substitution, (4) Hot-deck imputations, (5) sample weight imputation, and (6) Proxy pattern-mixture analysis or a combination of others.

6. The "no-universe database problem" must be addressed. Each year, the numbers in FAEIS have increased because FAEIS has captured more data – not because the number of students has necessarily increased. Statisticians might call this a trend in the mean/count. In addition, degrees within Classification of Instructional Program (CIP) codes have changed dramatically over the years – including degrees that were not part of the original CIP (1981). For example, the 01 CIP (agriculture) has been changed 35 times since 1978 (Survey of Earned Doctorates – SED, NSF) according to the NSF. This process has by definition increased the numbers by adding new degrees to the CIP. This is also a trend in the mean/count. To adjust for the effect of the population (universe) increasing, the total number of science/engineering majors must be added to the FAEIS dataset.

7. Prepare a complete and up-to-date list of all sources, that is, FAEIS contacts, and include in the final report.

8. Develop a final report using a similar format as the National Science Foundation (use SED, as example) that shows the improvement of the database and tables with summarized results compared to IPEDS data.

9. Produce quarterly reports (due January 1, April 1, July 1 and October 1) on the progress in addressing transcription errors, outliers, and missing values. Include statistical procedures used to correct and process FAEIS data.

10. Conduct a statistically valid, random survey to collect FAEIS clientele data. This survey must clearly define the target population and the random sample must match the target population. The sample size must be large enough and the response rate must exceed 70% (90-99% far better) (NSF=93%, IPEDS >99+ %). The survey must use various methods: mail, Internet, telephone surveys, etc. The survey must be well written, tested and contain no leading questions. The survey personnel selected must be professionals well trained to conduct surveys.

Appendix C—COMPARISON OF FAEIS TO IPEDS

Please see the file ***faeis_ipeds_review_stat_report_appendix_c.*****pdf,** sent with this document.