

# Statistical Analysis for Quality Assurance

Some methods used by FAEIS to check for  
inconsistency and irregularities

1

## Quality Assurance Projects

- ▶ Check for consistency of the report builder
- ▶ Check the data that is sent by the universities
  - Search for irregular values
  - Search for missing values

2

## Checking Consistency

- ▶ Objective: Reproduce results of Report Builder
- ▶ Steps
  - Create data set
  - Use SAS to reproduce Report Builder analyses
  - Compare results

3

## Example: SAS Output identical to Report Builder

Baccalaureate Enrollment in Food Science and Technology

<i>Institution</i>	2004	2005	2006	2007	2008	2009
<i>Alabama Agricultural and Mechanical University</i>	39	51	50	49	55	54
<i>California State Polytechnic University, Pomona</i>	.	.	42	54	63	79
<i>California State University - Fresno</i>	90	100	99	109	116	151
<i>California State University - Long Beach</i>	16	23	27	29	28	37
<i>Clemson University</i>	97	113	145	167	209	238
<i>Cornell University</i>	40	127	152	167	156	200
<i>Delaware Valley College</i>	.	.	.	.	.	41
<i>Florida A and M University</i>	.	.	21	.	24	.
<i>Framingham State College</i>	.	.	.	.	.	14
<i>Iowa State University</i>	67	72	74	47	35	81
<i>Kansas State University</i>	90	109	137	138	145	188
<i>Louisiana State University</i>	22	28	25	27	19	.
<i>Michigan State University</i>	37	57	73	77	92	103
<i>Mississippi State University</i>	16	63	87	82	90	102
<i>North Carolina State University at Raleigh</i>	60	55	50	39	61	.
<i>North Dakota State University</i>	16	17	17	27	33	35
<i>Oklahoma State University</i>	.	.	.	28	31	23
<i>Oregon State University</i>	67	67	90	86	86	100

4

## Check for Irregularities

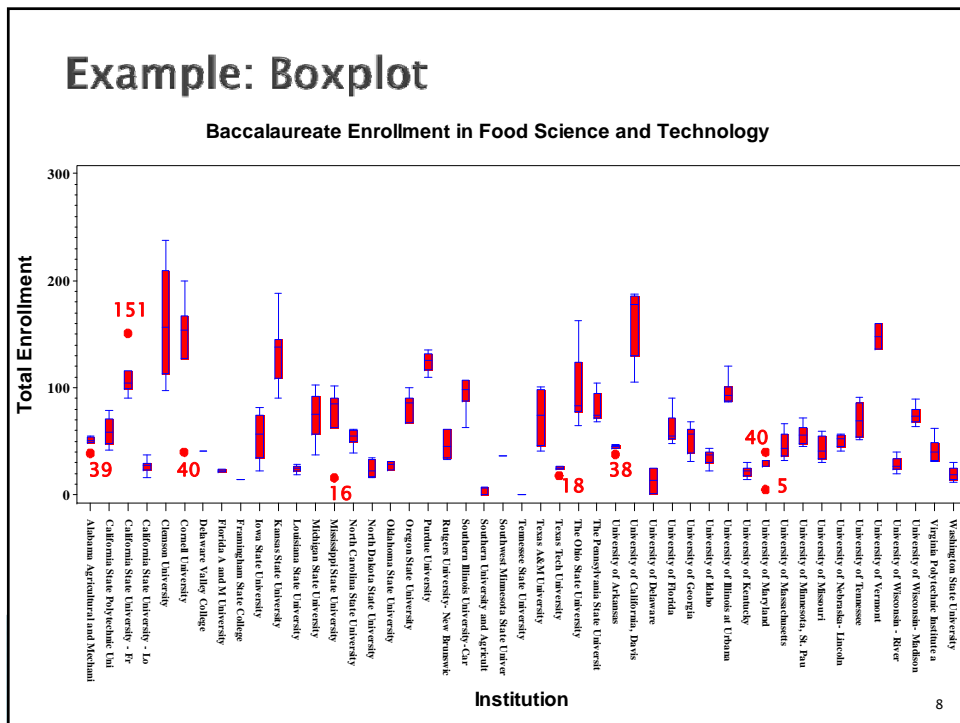
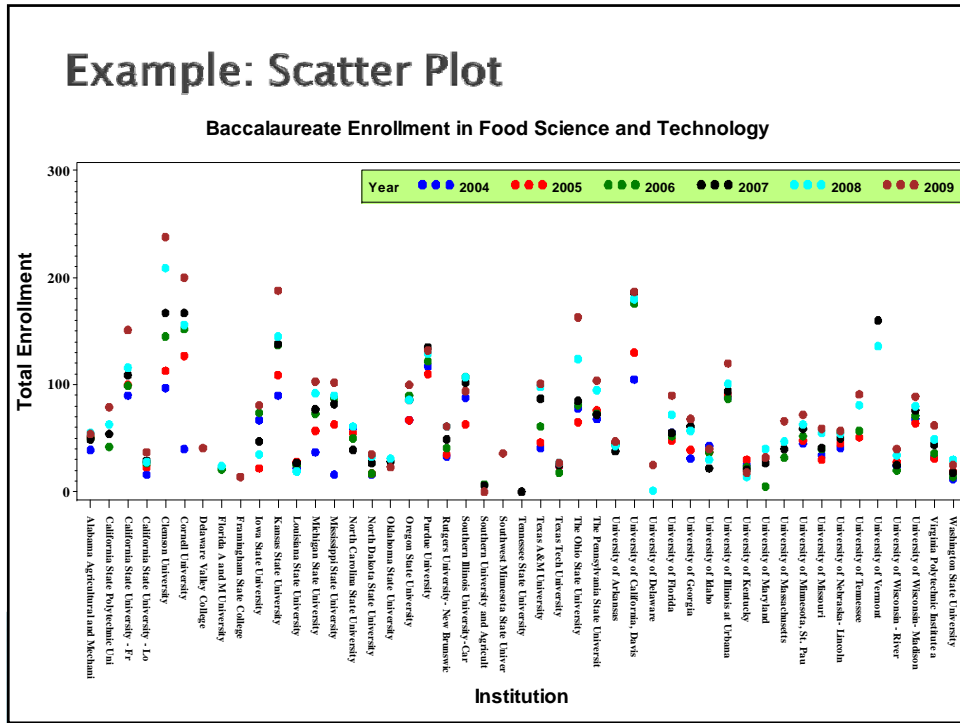
- ▶ Outliers/irregular values
- ▶ Search for two types
  - Change in a program that was relatively stable
    - Observations inconsistent with average observations
  - Programs that change following consistent pattern
    - Observations that change relative to previous observations (or pattern)

5

## Methods

- ▶ Inconsistent observations
  - Evaluate data for each university and CIP
  - Scatter plots to visualize the data
    - identify reporting frequency
    - Identify odd observations
  - Flag observations based on boxplots
  - Outlier checks
    - Extreme observations
    - Extreme differences

6

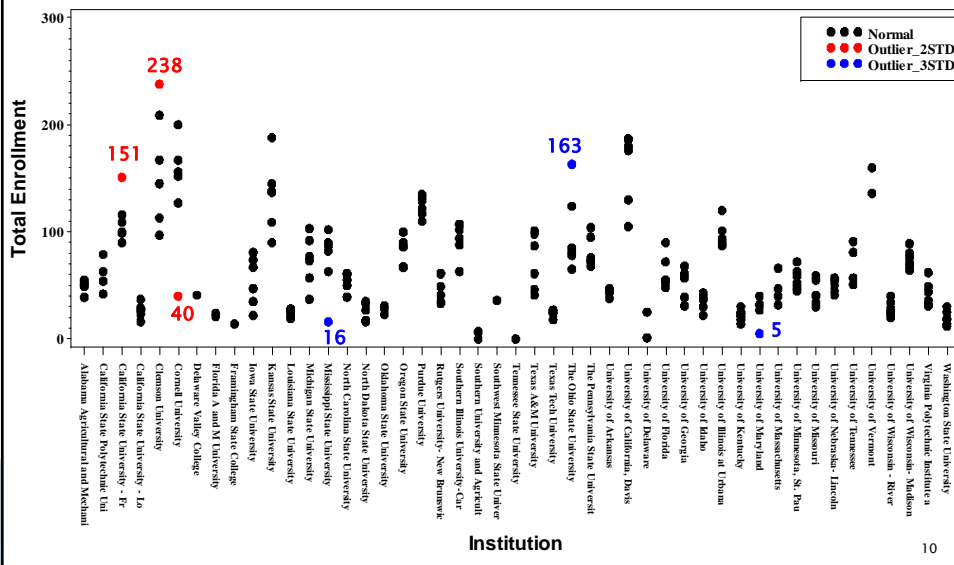


# Methods: outlier checks

- ▶ Inconsistent observations
  - Evaluate data for each university and CIP
  - Flag observations that are 2 and 3 standard deviations from the mean
  - Calculation of standard deviation:
    - Create groups based on size (small, medium, large, extra large)
    - Calculate standard deviation for each group
    - Also calculate pseudo standard deviation (based on interquartile range)

# Example: Regular Standard Deviation

Baccalaureate Enrollment in Food Science and Technology



## Example: Regular Standard Deviation

<i>Institution</i>	2004	2005	2006	2007	2008	2009
California State University - Fresno	90	100	99	109	116	51
Clemson University	97	113	145	167	209	238
Cornell University	40	127	152	167	156	200
Mississippi State University	16	63	87	82	90	102
The Ohio State University	78	65	81	85	124	163
University of Maryland	.	32	5	27	40	32

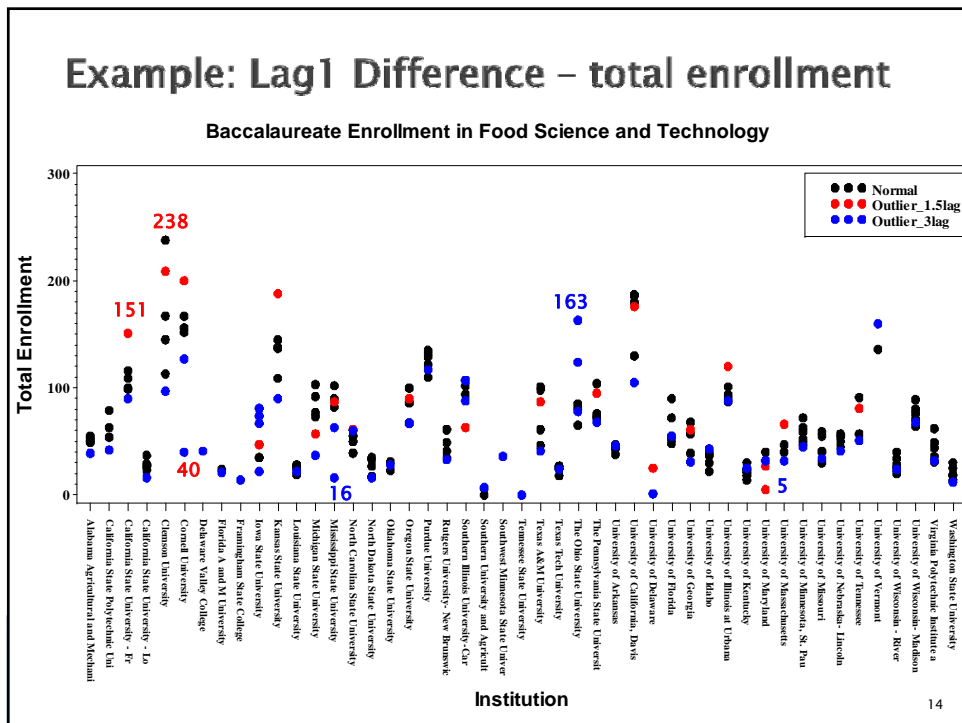
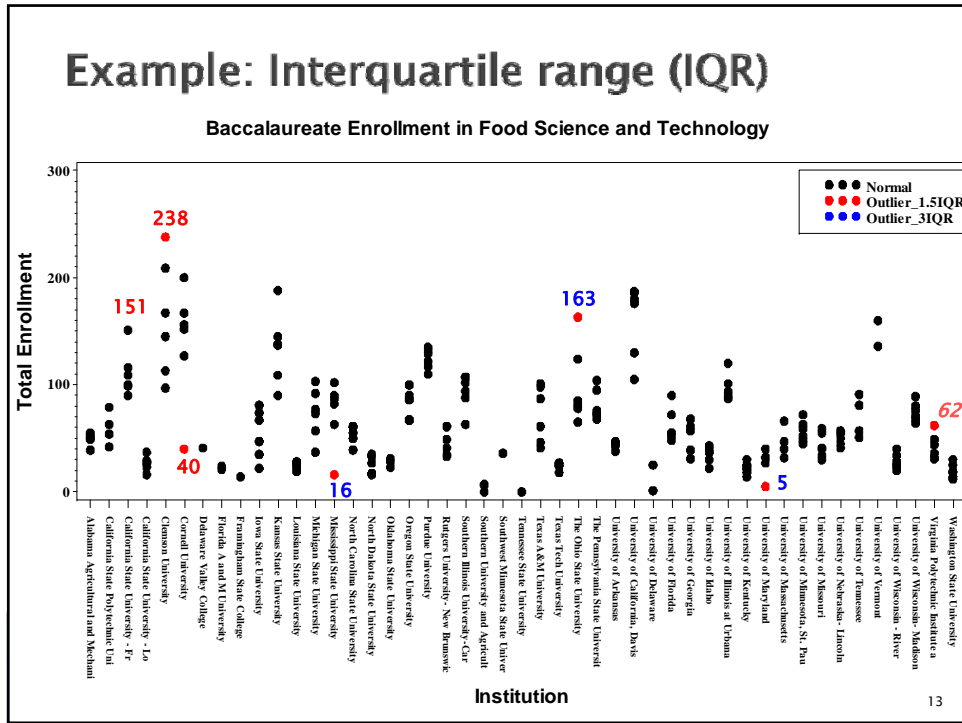
checked and validated  
 changed to 134  
 still checking  
 checked and validated  
 changed to 17 from IR report  
 removed data due to user breaking down into different CIP codes that follow previous reporting

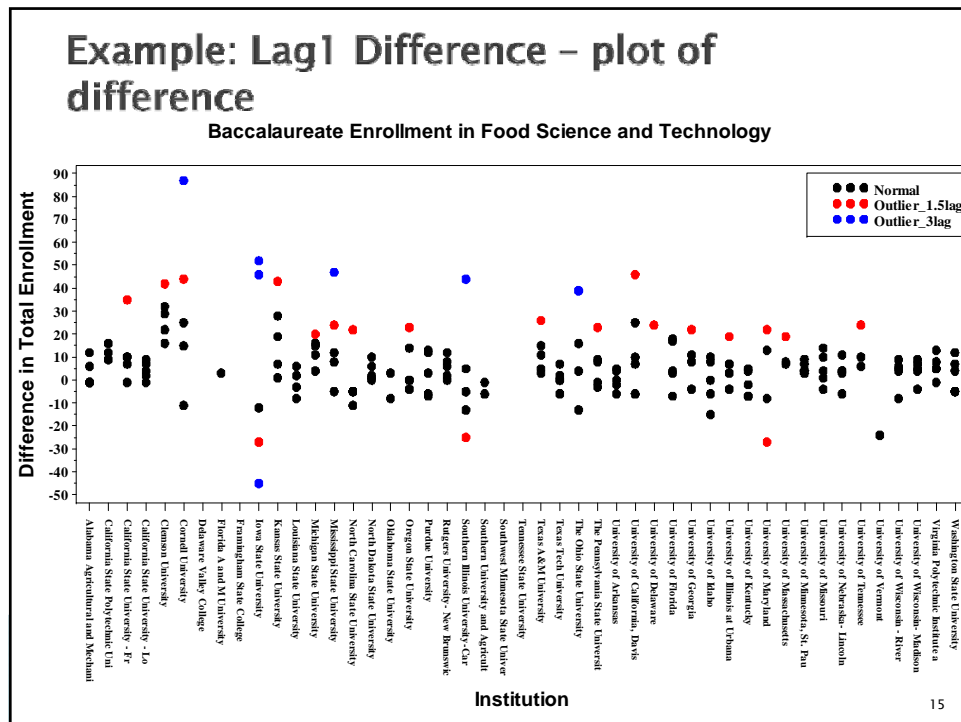
11

## Methods: difference oddities

- ▶ Detecting an odd change
  - Example: increasing enrollment followed by sharp drop in enrollment may suggest a change in program or an entry error
  - Look for a change in the pattern
  - Calculate change from previous observation
  - look at changes that are greater than 2–3 standard deviations

12





## Methods

- ▶ Detecting missing data
  - List the institutions which has missing records
  - Identify the “holes” in the records
  - Identify the abrupt ending of records
  - Compare with similar CIP code within the same institution to identify possible misplacement
- ▶ Following detection, check records and/or contact institution for information



## Example: Missing Data

Baccalaureate Enrollment in Food Science and Technology  
Missing Data Entry

Institution	2004	2005	2006	2007	2008	2009
California State Polytechnic University, Pomona	.	.	42	54	63	79
Delaware Valley College	.	.	.	.	.	41
Florida A and M University	.	.	21		24	
Framingham State College	.	.	.	.	.	14
Louisiana State University	22	28	25	27	19	
North Carolina State University at Raleigh	60	55	50	39	61	
Oklahoma State University	.	.	.	28	31	23
Southern University and Agricultural and Mechanical College	.	.	7	6		
Southwest Minnesota State University	.	.	.	.	.	36
University of Delaware	.	.	.	.	1	25
University of Georgia	31	39		61	57	68
University of Maryland	.	32	5	27	40	32
University of Massachusetts	.	.	32	40	47	66
University of Tennessee	.	51	57		81	91
University of Vermont	.	.	.	160	136	

17

## Summary

- ▶ FAEIS help desk team and PIs provide quality assurance through checks on data
  - Missing values analysis
  - Outlier analysis
  - CIP code analysis
- ▶ Updates are made to ensure quality
  - Contacting university representatives
  - Contact IR
  - Use other sources

18