Utilizing Open-Source Government Data Sets in Health Information Management Teaching: An Application in Statistics and Data Analytics

by Zahraa M. Alakrawi, MS; Valerie Watzlaf, PhD, RHIA, FAHIMA; Jon Aborde, RHIA; Carey Johnson, RHIA; and Timothy Johnson, RHIA

Abstract

The American Health Information Management Association (AHIMA) Council for Excellence in Education acknowledges the importance of improving the data analytics skills of health information management (HIM) students at the baccalaureate level. This importance is especially reflected in the Informatics, Analytics, and Data Use domain of the HIM curricular competencies defined by AHIMA. Open-source health data represent a valuable, yet unrecognized tool for teaching applied statistics and big data analytics in HIM. This article describes the adoption of a new method to teach the application of statistical concepts in HIM by utilizing open-source government health data sets. These data sets were assigned to students in class projects in which the students had to develop a hypothesis and specific aims based on the data set and then use statistical and data analytic techniques to prove or disprove the hypothesis. A major goal of this work is to integrate research methods, data analytics, and statistics to develop an innovative yet applied method for teaching statistics at the baccalaureate level. This method can be adopted to enhance many curricular competencies related to information and data use.

Keywords: data analytics; health data; health information management (HIM) teaching; health statistics; open-source data

Introduction

The American Health Information Management Association (AHIMA) Council for Excellence in Education (CEE) was created to support the health information management (HIM) profession by developing leading education strategies that would pave the way for future changes and guide HIM educators through the ever-changing healthcare industry.[1–4] Therefore, maintaining high-quality education standards as well as curricular competencies for the different levels of HIM education constitutes one of the highest priorities of the AHIMA CEE.[5–8] However, meeting the curricular competencies can be a daunting task for educators, especially in the area of statistics, data analytics, and research methods.[9–14] Therefore, educators are always looking for creative and innovative ways to teach these concepts so that students can "hit the ground running" when they become employed in the HIM field.[15]

Background and Significance

This project emphasizes the importance of leveraging HIM education to meet the ever-increasing challenges in the healthcare industry. Major trends that have contributed substantially to these changes are the fast growth in the health information technology sector and the evolution of big data analytics in healthcare.[16–19] Both of these areas require critical thinking skills, sound knowledge of research methods and statistics, and practical experience in the use of information technology to help healthcare stakeholders make informed business decisions regarding healthcare operations.[20, 21]

Furthermore, the importance of improving data analytics skills of HIM students at the baccalaureate level is especially reflected in Domain III, Informatics, Analytics, and Data Use, in the HIM curricular competencies identified by AHIMA.[22, 23] Open-source health data may be valuable in the teaching of applied statistics[24] and big data analytics[25] in HIM. Therefore, one of the major aims of this study is to integrate data analytics and research methods with statistics to develop an innovative yet applied method for teaching statistics at the baccalaureate level. This method can be adopted to enhance many curricular competencies related to information and data use, such as the following:[26]

Subdomain III.A. Health Information Technologies
Subdomain III.C. Analytics and Decision Support

a. Apply analytical results to facilitate decision-making
b. Apply data extraction methodologies
c. Recommend organizational action based on knowledge obtained from data exploration and mining
d. Analyze clinical data to identify trends that demonstrate quality, safety, and effectiveness of healthcare

Subdomain III.D. Health Care Statistics

a. Interpret inferential statistics
b. Analyze statistical data for decision making

Subdomain III.E. Research Methods

Apply principles of research and clinical literature evaluation to improve outcomes

Subdomain III.F. Consumer Informatics

Educate consumers on patient-centered health information technologies[27]

Objectives

The objectives of this project included the following:

1. Utilize open-source government data sets (health-related) in the classroom with hands-on statistics projects to integrate principles of research methodology and biostatistics with big data analytics in healthcare.
2. Demonstrate how open-source data can be incorporated into analytical projects in HIM courses such as statistics, epidemiology, research methods, and quality management at the baccalaureate level.
3. Prepare students with critical and creative thinking skills in the statistical analysis of large data sets for use in the healthcare industry.

Methods

Students

A total of 30 students participated in the Application of Statistical Concepts in HIM course. This course is an applied statistics course for HIM students at the baccalaureate level. It is designed to provide students with a practical understanding of the use of statistics in HIM. Descriptive statistics, inferential statistics, and nonparametric methods were discussed and utilized with HIM data or concepts. It is offered once per year and is taken during the spring semester of the HIM student's junior year for three credits. The course requirements include completion of a statistics project using an open-source data set.

Project Description and Evaluation

Each student was required to use a randomly assigned open-source data set to develop a statistics project. Students were guided to generate hypotheses and specific aims related to the data set. Evaluation criteria and a proposed timeline are presented in Table 1. Students' progress was assessed frequently throughout the course on the basis of the proposed timeline. Also, the instructor and teaching assistant arranged to meet weekly during and after class time with any student who might have questions. An individualized record was developed for each student, including the student's progress throughout the semester, using e-mails and qualitative notes to record items such as technical problems, as well as outstanding performance.

Selection of Data Sets

A total of 37 data sets were initially selected to be assigned to 30 students. These data sets were mainly obtained from the US government's open data website (www.data.gov). Alternatively, some data sets were obtained from the US Department of Health and Human Services open data website (www.healthdata.gov/dataset), which can generate comparable results after relevant filters are applied. The major focus was on data sets that could be relevant to HIM. Examples are provided in the following sections to demonstrate the search results.

Figure 1 depicts the data set selection process. First, data sets were retrieved from www.data.gov. The website has a wide variety of topics for which data sets can be selected. For the purpose of this course, health, public safety, and science and research data sets were included. Second, filters were applied to narrow the results. Examples of filters include topic category, data type, format, organization type, and publisher. Browsing data sets by format was very important because raw data in Microsoft Excel was the preferred format. A similar search was performed at www.healthdata.gov/dataset. Finally, each data set was manually inspected by the teaching assistant, and a list of 37 potential data sets was chosen.

Selection Criteria

Each data set was examined and manually selected to satisfy predetermined criteria, which included the following:

1. Health topic or health-related issues that are relevant to HIM as a field of study;
2. Microsoft Excel or compatible format;
3. Adequate sample size with respect to the unit of analysis;
4. Type of variables (quantitative and qualitative); and
5. Comparable data sets with respect to the level of complexity.

The selected data sets represent a wide spectrum of current health-related issues and were collected by different healthcare entities at either the national or state level. Examples of the healthcare organizations include the Department of Health and Human Services, the Centers for Medicare and Medicaid Services, state health departments, acute care and short-stay hospitals, and long-term healthcare settings.

In addition to the topic, it was necessary to select data sets in Microsoft Excel format or other formats that can be read by Excel, such as comma-separated values, so that the data would be easy to transport

and use with statistical software. In addition, students were exposed to and encouraged to use advanced statistical analysis software besides Microsoft Excel, including SPSS, SAS, and Python programming.

Sample size was a determinant in the search because it was important to incorporate the concept of big data analytics into this course to help students understand the proper methods for analyzing big data in different healthcare settings. Therefore, large data sets were preferred.

Alignment of the data sets with the course objectives, in terms of potential statistical analysis and the types of variables required for such analysis, was also included in the search. These data sets contain both quantitative and qualitative variables, as well as various types of numerical variables, including nominal, ordinal, interval, and ratio. The goal was to have students work with data that could be used to generate descriptive statistics and at least one higher-level inferential statistic (parametric or nonparametric).

Lastly, to maintain equality and fairness in classroom teaching, comparable data sets were selected with relatively the same level of difficulty in terms of the potential statistical analysis, data aggregation, and sample size. In the case of a very high sample size, students were encouraged to take a random sample of the data for statistical analysis.

Selected Topics

The data sets that satisfied the selection criteria reflect a wide variety of health-related topics. Examples of topics include the following:

- Electronic health records
- Medicare hospital utilization
- Hospital-acquired infections
- HIV ambulatory care quality of care performance
- Hospital inpatient discharges by diagnosis-related group (DRG)
- Student weight status categories
- Health prevention indicators (health status and health disparities; chronic diseases; HIV/STDs, vaccine-preventable diseases, healthcare-associated infections; healthy and safe environment; healthy women, infants, and children; and mental health and substance abuse)
- Community health status indicators on obesity, heart disease, and cancer
- Inpatient psychiatric facility quality measure data by state

See Table 2 for a full list of the selected data sets.

Assignment of Data Sets

Each student picked a number from a basket to randomly choose a data set. A list of the data sets, numbered from 1 to 37, was provided to the students so that they could determine their chosen data set. Providing a unique data set to each student helps to develop an early sense of ownership and stimulates students' potential and creativity in developing their work.[28, 29]

Results

All students successfully completed their class project requirements. Most students perceived their experience as a positive learning opportunity regardless of the amount of work and the challenges that they might have had with their data sets.

Examples of Student Projects and Student Comments

The following are examples of student projects. All students whose projects are included here are listed as authors of this article and gave permission for the discussion of their projects.

"An Investigation of Nursing Home Quality Based on Location"

In this project, the data set titled "Nursing Home Weekly Bed Census," retrieved from www.data.gov, was used. An external data set from the Medicare Nursing Home Compare website (www.medicare.gov/nursinghomecompare/About/What-Is-NHC.html) was also included. The objective of this project was to discover if there was a significant difference in the quality of care provided by nursing home facilities between urban and rural counties in New York.

The author of this project (T.J.) describes his experience utilizing the data set as "overwhelming at first; however, the answer came from the skills learned in both the statistics class as well as the database management class. In fact, this project included a great deal of hard work and hours of research techniques. The research techniques and statistical tests included . . . were learned in the statistics class."

Microsoft Excel and SPSS were both utilized for the purpose of data analysis, which included filtering data, hiding unwanted data, and formatting the data to see trends. In addition to descriptive statistics, the author of the project performed two separate independent-sample *t*-tests to investigate the following:

1. whether there was a statistically significant difference in the quality score of nursing homes between urban and rural areas in New York; and
2. whether there was a statistically significant difference in the overall rating of nursing homes between urban and rural areas in New York.

The statistical analysis for this project revealed a significant difference in quality of care between urban and rural nursing homes.

This project demonstrated that the data sets can be used to enhance students' knowledge and experience of statistics by giving them the experience firsthand of how to formulate a hypothesis, collect data, enter it into the SPSS software, analyze the data, create graphs and charts from the data, and draw a conclusion on the significance of the data.

Furthermore, such experience can reinforce students' research skills by providing examples of real-life situations in which research can be put into practice. Using both research methods and statistics in this way could help implement change in our country's healthcare system and encourage further research in the area of quality of care in nursing homes.

"The Implications of Chronic Disease on Hospital Spending per Patient"

In this example of the use of a data set in the statistics project, the author of the project (C.J.) was assigned to analyze a data set that broke down the average Medicare spending per patient in hospitals across the United States. The following data sets were used: Medicare Spending per Patient—Hospital Data Sets, retrieved from www.data.gov, and Medicare Chronic Conditions Dashboard, which can be found at www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/CCDashboard.html.

The specific aims of this project were to determine the following:

1. How much money do hospitals in areas with a particularly high prevalence of a disease with many associated comorbidities, such as chronic kidney disease, spend on their patients?
2. Is their spending, on average, significantly greater than another hospital in an area with less chronic disease prevalence?

To study the research questions, the second data set was used to obtain chronic kidney disease prevalence rates for 702 counties nationwide. In each of the selected counties, the original data set

(Medicare Spending per Patient) was used to identify hospitals in counties with high or low prevalence of chronic kidney disease. Then, the differences between the spending per patient at these hospitals were analyzed with Microsoft Excel and SPSS. In addition to basic descriptive statistics, two separate statistics, including the independent-sample *t*-test, were generated.

Results showed that the difference between the scores in counties with high versus low prevalence was not significant: that is, there was no statistically significant difference in the amount of money hospitals spent if they were situated in a county with a high or low prevalence of chronic kidney disease. However, when the sample size was modified to include only the hospitals from the 25 counties with the highest and lowest prevalence of chronic kidney disease, results were significant. Even though not all results were significant, this project experience led to several discoveries, including, first, that data can lead to discoveries and conclusions about numerous topics and, second, that blending data sources can lead to additional discoveries that would not have been made with just one data source.

This project extended far beyond the classroom and provided an opportunity for the student to embrace ambiguity by creating an entire project, an experience that was immensely helpful with internships and other school projects. An applied approach that blends principles of research and statistics can help increase students' confidence with respect to their research and analytic skills.

"Health Coalitions and the Community; The Effect of Anti–Teen Pregnancy Coalitions in New York"

This project used a data set regarding health coalitions. After realizing that the original data set consisted of only addresses and telephone numbers of different health coalitions, the author of the project (J.A.) made the decision to sort them by coalition type, which revealed a trend that led to determining that a significant percentage of the coalitions focused on teen pregnancy.

This data set was also supplemented with other data sources. External data were obtained from the New York State Department of Health's list of funding programs, which can be found at https://www.health.ny.gov/funding/, and from County Health Rankings and Roadmaps, available at http://www.countyhealthrankings.org/. After collecting teen birth numbers in New York from multiple sources, the author of the project had enough data to begin the analysis.

The primary goal of the project was to explain why the original data set showed more coalitions that focused on teen pregnancy than any on other health issue. The student concluded that there was in fact an issue regarding teen pregnancies in New York, which generally showed rates of teen pregnancy higher than those in a majority of states in the country.

Microsoft Excel was used to create and organize a new data set to discover any potential correlations between the teen birth rates and the number of anti–teen pregnancy coalitions established in New York. Charts and formulas were used to illustrate the data.

The data analysis revealed that the presence of anti–teen pregnancy coalitions is a strong factor in improving this particular health disparity. However, this conclusion was based on statistical correlation rather than the establishment of causal relationships: while the decline in teen birth rates cannot be directly attributed to the presence of these coalitions on the basis of these findings, it can be stated that these coalitions may have a positive effect on their communities, supporting the consideration of future establishment of similar coalitions.

Course Feedback

Students completed questionnaires regarding their opinions of the teaching. When students were asked what aspects of the course were most beneficial to them, the results included the following:

- "All of the Excel and SPSS usage really helped me refresh my memory of both of these applications. I know I will definitely be using Excel in the future and learning how to compute certain statistical tests really enhanced my knowledge."
- "Becoming exposed to new statistical software such as SPSS and SAS."
- "Learning Python and analyzing big data; SPSS."
- "Learning how to do statistical problems and tests in SPSS. I really enjoyed being able to learn how to use this program."
- "In all, this experience enhanced my knowledge of the SPSS software."

Discussion

This study demonstrated how open-source health data sets can be used to provide hands-on teaching of statistics in HIM. It also demonstrated how open data could be incorporated into analytical projects in HIM courses at the baccalaureate level, not only in statistics, but also in epidemiology and quality management. A total of 37 data sets were selected on the basis of predetermined criteria. These data sets were randomly assigned to 30 students to develop unique class projects. The selection process was somewhat challenging at first because each data set needed to be manually inspected to ensure that the data were of reasonable quality, were in the right format, and had adequate variables to satisfy the course requirements. However, the instructor can save these data sets by creating and routinely updating a list for future use.

In general, all students successfully completed the course requirements with respect to the statistics project. Student evaluation was performed every two weeks in class, and grades and notes were taken and recorded in a spreadsheet. In addition, time was allocated in class each week for students to discuss any issues or concerns if they did not have an opportunity to arrange an individual meeting. Some projects were more challenging than others, and therefore the evaluation was performed on a case-by-case basis.

It is strongly recommended that educators use open-source material and incorporate it into their courses. This project represents a very good exercise in data analysis and is comparable to what HIM students would encounter in real-life settings.

Limitations

Limitations of this study include issues with data integrity and time constraints. The data that were utilized in this course include secondary data collected by other sources. Therefore, issues encountered were mostly related to data integrity (missing data, missing variables, inappropriate coding) because the data were collected by the original source. To address these issues, each data set was examined to ensure that the data could be used to support the intended analysis and answer research questions. Students were instructed to inspect their data sets and perform relevant variable transformation to resolve issues related to inappropriate coding of variables. In addition, students were given alternative data sets if data collection errors could not be fixed. However, in a real-life setting, data sets are messy and include missing variables. This experience provided students with real-life data sets, and students were able to work through the problems associated with the data.

Some students could not locate the assigned data sets on the websites because of broken links. We believed this problem was due to regular maintenance of the websites or changes to the website maps. This problem was overcome by providing these students with copies of the data sets that were previously downloaded by the teaching assistant. Also, another limitation of this study is the time frame allocated for the project. This course is provided in the spring semester, and extending the project into the summer semester could possibly be an effective solution that would provide students with more time to complete the project.

Conclusion

Open-source health data—which comes from a variety of sources—represents a valuable tool for teaching applied statistics and data analytics in HIM. In this course, the instructors adopted a new method to teach the application of statistical concepts in HIM by assigning government health data sets to students for use in class projects. Each student was randomly assigned a unique data set to develop a final project throughout the course. All students successfully completed their class project requirements and had a positive attitude toward the experience in general. However, the initial process of selecting the data sets, as well as monitoring student progress, was time consuming and somewhat labor intensive. Furthermore, some projects were more challenging than others, which requires that special consideration to be given to each case to ensure a fair assessment. Data quality varies greatly across these data sets, and therefore it is recommended that data be manually inspected to verify that each data set satisfies AHIMA's Data Quality Management guidelines.[30] Adopting this approach will contribute to quality education and teaching of HIM at all levels and help students attain the CEE competencies in all relevant domains concerning statistics, data analytics, and research methods.

Zahraa M. Alakrawi, MS, is a PhD candidate in the Department of Health Information Management at the University of Pittsburgh School of Health and Rehabilitation Sciences in Pittsburgh, PA.

Valerie Watzlaf, PhD, RHIA, FAHIMA, is an associate professor in the Department of Health Information Management at the University of Pittsburgh School of Health and Rehabilitation Sciences in Pittsburgh, PA.

Jon Aborde, RHIA, is an adoption services intern at M*Modal in Easton, PA.

Carey Johnson, RHIA, is a Consulting Analyst in Cerner Corporation, MO

Timothy Johnson, RHIA, is a Data Analyst Intern at St. Clair Hospital, PA.

Notes

1. Gordon, Leslie, Valerie J. M. Watzlaf, and Desla Mancilla. "Mapping the Future of HIM Education." *Journal of AHIMA* 84, no. 8 (2013): 54–57.
2. Cortelyou-Ward, Kendall, Summerpal Kahlon, and Alice Noblin. "Competencies for Global Health Informatics Education: Leveraging the US Experience." *Educational Perspectives in Health Informatics and Information Management* (Winter 2013): 1–8. Available at http://eduperspectives.ahima.org/competencies-for-global-health-informatics-education-leveraging-the-us-experience/.
3. Joshi, Ashish, and Douglas Marcel Perin. "Gaps in the Existing Public Health Informatics Training Programs: A Challenge to the Development of a Skilled Global Workforce." *Perspectives in Health Information Management* (Fall 2012). Available at http://bok.ahima.org/doc?oid=105943.
4. Butler, Mary. "Mind the Gap: HIM Rushes to Bridge Educational and Professional Gaps Caused by a Quickly Advancing Industry." *Journal of AHIMA* 86, no. 2 (2015): 20–24. Available at http://bok.ahima.org/doc?oid=107552.
5. Council for Excellence in Education. "Baccalaureate Level HIM Curriculum Map." AHIMA Foundation. 2014. Available at https://ahimafoundation.org/downloads/pdfs/2014%20Final%20Baccalaureate_Level_Curriculum_Map.pdf. Accessed November, 2015.
6. Cortelyou-Ward, Kendall, Summerpal Kahlon, and Alice Noblin. "Competencies for Global Health Informatics Education: Leveraging the US Experience."
7. Joshi, Ashish, and Douglas Marcel Perin. "Gaps in the Existing Public Health Informatics Training Programs: A Challenge to the Development of a Skilled Global Workforce."
8. Butler, Mary. "Mind the Gap: HIM Rushes to Bridge Educational and Professional Gaps Caused by a Quickly Advancing Industry."
9. Gordon, Leslie, Valerie J. M. Watzlaf, and Desla Mancilla. "Mapping the Future of HIM Education."
10. Raghupathi, Wullianallur, and Viju Raghupathi. "Big Data Analytics in Healthcare: Promise and Potential." *Health Information Science and Systems* 2 (2014): 3. Available at http://hissjournal.biomedcentral.com/articles/10.1186/2047-2501-2-3.
11. Joshi, Ashish, and Douglas Marcel Perin. "Gaps in the Existing Public Health Informatics Training Programs: A Challenge to the Development of a Skilled Global Workforce."
12. Butler, Mary. "Mind the Gap: HIM Rushes to Bridge Educational and Professional Gaps Caused by a Quickly Advancing Industry."
13. Sandefer, Ryan, and Ellen Shakespeare Karl. "Ready or Not, HIM Is Changing: Results of the New HIM Competencies Survey Show Skill Gaps Between Education Levels, Students, and Working Professionals." *Journal of AHIMA* 86, no. 3 (2015): 24–27. Available at http://bok.ahima.org/doc?oid=107578.
14. Kozak, Marcin. "Teaching Statistics = Teaching Thinking Statistically." *Model Assisted Statistics and Applications* 4, no. 4 (2009): 275–79.
15. Sandefer, Ryan, and Ellen Shakespeare Karl. "Ready or Not, HIM Is Changing: Results of the New HIM Competencies Survey Show Skill Gaps Between Education Levels, Students, and Working Professionals."
16. Gordon, Leslie, Valerie J. M. Watzlaf, and Desla Mancilla. "Mapping the Future of HIM Education."
17. Raghupathi, Wullianallur, and Viju Raghupathi. "Big Data Analytics in Healthcare: Promise and Potential."

18. Sandefer, Ryan, and Ellen Shakespeare Karl. “Ready or Not, HIM Is Changing: Results of the New HIM Competencies Survey Show Skill Gaps Between Education Levels, Students, and Working Professionals.”
19. Fenton, Susan H., Elizabeth Joost, Jimena Gongora, Davis G. Patterson, C. Holly A. Andrilla, and Susan M. Skillman. “Health Information Technology Employer Needs Survey: An Assessment Instrument for HIT Workforce Planning.” *Educational Perspectives in Health Informatics and Information Management* (Winter 2013): 1–36. Available at http://eduperspectives.ahima.org/health-information-technology-employer-needs-survey-an-assessment-instrument-for-workforce-planning/.
20. Raghupathi, Wullianallur, and Viju Raghupathi. “Big Data Analytics in Healthcare: Promise and Potential.”
21. Fenton, Susan H., Elizabeth Joost, Jimena Gongora, Davis G. Patterson, C. Holly A. Andrilla, and Susan M. Skillman. “Health Information Technology Employer Needs Survey: An Assessment Instrument for HIT Workforce Planning.”
22. Gordon, Leslie, Valerie J. M. Watzlaf, and Desla Mancilla. “Mapping the Future of HIM Education.”
23. Cortelyou-Ward, Kendall, Summerpal Kahlon, and Alice Noblin. “Competencies for Global Health Informatics Education: Leveraging the US Experience.”
24. Molloy, Jennifer C. “The Open Knowledge Foundation: Open Data Means Better Science.” *PLoS Biology* 9, no. 12 (2011): e1001195. Available at http://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001195.
25. Raghupathi, Wullianallur, and Viju Raghupathi. “Big Data Analytics in Healthcare: Promise and Potential.”
26. Cortelyou-Ward, Kendall, Summerpal Kahlon, and Alice Noblin. “Competencies for Global Health Informatics Education: Leveraging the US Experience.”
27. Gordon, Leslie, Valerie J. M. Watzlaf, and Desla Mancilla. “Mapping the Future of HIM Education.”
28. Bland, J. Martin. “Teaching Statistics to Medical Students Using Problem-Based Learning: The Australian Experience.” *BMC Medical Education* 4, no. 1 (2004): 31. Available at http://bmcmededuc.biomedcentral.com/articles/10.1186/1472-6920-4-31.
29. Gelman, Andrew. “Going beyond the Book: Towards Critical Reading in Statistics Teaching.” *Teaching Statistics* 34, no. 3 (2012): 82–86. Available at http://www.stat.columbia.edu/~gelman/research/published/dontbelieve3.pdf.
30. Davoudi, Sion, Julie A. Dooling, Barbara Glondys, Theresa D. Jones, Lesley Kadlec, Shauna M. Overgaard, Kerry Ruben, and Annemarie Wendicke. “Data Quality Management Model (2015 Update).” *Journal of AHIMA* 86, no. 10 (2015). Available at http://bok.ahima.org/doc?oid=107773.

Table 1

Evaluation Criteria and Proposed Timeline

Statistical Analysis Project: ***The statistical analysis project requires that each student complete the following:***		
1	A health-related data set will be provided to you. Based on your review of the topic and analysis of the data, you will complete the following steps:	Week 1
2	Develop a hypothesis or statement of the problem.	Week 3
3	Develop specific aims or goals related to hypothesis.	Week 3
4	Develop a database or spreadsheet to collect your data.	Week 5
5	Design tables and graphs to analyze data.	Week 7
6	Conduct appropriate statistical analysis to examine the area. You must conduct at least one higher-level statistical test (such as a *t*-test, ANOVA, correlation, regression) to prove your hypothesis.	Week 8
Spring Break		
7	Provide an overall analysis of the data from tables and statistical tests, etc.	Week 11
8	Based on the analysis of the data, make conclusions that you will use to continue to improve.	Week 12
9	Hand in a written report of the project discussing each of the areas listed above.	Week 14
10	Report the results to the class.	Week 14

Table 2

Full List of the Selected Data Sets

No.	Title of Data Set
1	Electronic Health Record Vendors Reported by Health Care Providers Participating in Federal EHR Incentive Programs
2	Health Insurance Marketplace: Summary Enrollment Data for the Initial Annual Open Enrollment Period
3	Medicare Short Stay Hospital Utilization
4	Hospital-Acquired Infections: Beginning 2008
5	Environmental Radiation Surveillance Nine Mile Point and James A Fitzpatrick Readings: Beginning 2009
6	Hospital Discharges by Facility (SPARCS): Beginning 2010
7	HIV Ambulatory Care Quality of Care Performance Results: Beginning 2011
8	Nursing Home Weekly Bed Census: Last Submission
9	Hospital Inpatient Discharges by DRG, U.S., FY2011
10	Medicare Hospital Spending by Claim
11	Student Weight Status Category Reporting Results: Beginning 2010
12	Percutaneous Coronary Interventions by Cardiologist: Beginning 2008
13	Medicaid Program Enrollment by Month: Beginning 2009
14	Managed Care Plan Utilization Data: Beginning 2009
15	Quality Assurance Reporting Requirements: Beginning 2008
16	Prevention Agenda 2013-2017 Tracking Indicators: State Baselines and Targets
17	Health Coalitions: A Resource for Community Collaboration
18	Professional Medical Conduct Board Actions: Beginning 1990
19	CMS Medicare and Medicaid EHR Incentive Program, Electronic Health Record Products Used for Attestation
20	Medicare Hospital Spending Per Patient – Hospital
21	Community Health Status Indicators (CHSI) to Combat Obesity, Heart Disease, and Cancer
22	State Averages (Quality Measures)
23	Healthcare Associated Infections – Hospital
24	Readmissions Complications and Deaths – Hospital
25	Timely and Effective Care – Hospital
26	HCAHPS – Hospital
27	BRFSS Prevalence And Trends Data: Tobacco Use – Adults Who Are Current Smokers for 1995-2010
28	Inpatient Psychiatric Facility Quality Measure Data – by State
29	Hospital ACS Measures
30	Veterans Health Administration 2008 Hospital Report Card – Availability of Services
31	Hospital Readmission Reduction
32	Veterans Health Administration 2008 Hospital Report Card – Quality of Care – Populations
33	2010 Chronic Conditions PUF
34	2008 Institutional Provider & Beneficiary Summary PUF
35	2008 Basic Stand Alone Prescription Drug Events PUF
36	2010 Basic Stand Alone Home Health Agency Beneficiary PUF
37	Cross Federal Workgroup on Telehealth (FedTel) Inventory of Activities

Figure 1

Data Set Selection Process