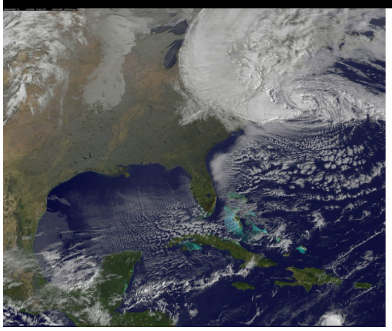
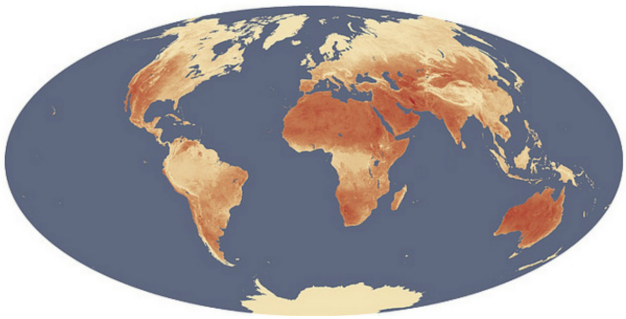
Engaging Mathematics: Teaching Manual for Elementary Statistics: Society and Environment

Engaging Mathematics Teaching Manual





Dr. Mangala Kothari

Mathematics, Engineering and Computer Science Department
LaGuardia Community College, City University of New York

Contributors: ***LaGuardia Community College Engaging Mathematics Team: Alioune Khoule, Marina Nechayeva, Nana Osei Bonsu, Steven Cosares, Milena Cuellar, Brian Johnston (Quantitative Reasoning Fellow, CUNY)***

Contents

License Information and Credit

This teaching manual was created for Engaging Mathematics with support from the National Science Foundation.

An initiative of the National Center for Science and Civic Engagement, Engaging Mathematics applies the well-established SENCER method to college level mathematics courses, with the goal of using civic issues to make math more relevant to students.

Engaging Mathematics will: (1) develop and deliver enhanced and new mathematics courses and course modules that engage students through meaningful civic applications, (2) draw upon state-of-the-art curriculum in mathematics, already developed through federal and other support programs, to complement and broaden the impact of the SENCER approach to course design, (3) create a wider community of mathematics scholars within SENCER capable of implementing and sustaining curricular reforms, (4) broaden project impacts beyond SENCER by offering national dissemination through workshops, online webinars, publications, presentations at local, regional, and national venues, and catalytic site visits, and (5) develop assessment tools to monitor students' perceptions of the usefulness of mathematics, interest and confidence in doing mathematics, growth in knowledge content, and ability to apply mathematics to better understand complex civic issues.

Updates and resources developed throughout the initiative will be available online at `http://www.engagingmathematics.net`. Follow the initiative on Twitter: @MathEngaging.

Support for this work was provided by the National Science Foundation under grant DUE-1613211, formerly DUE-1322883, to the National Center for Science and Civic Engagement. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Cover Photo Credits

Top: M. Bob. Licensed under (CC BY-NC-ND 2.0). Link to original.
Bottom left: NASA Goddard Space Flight Center. Licensed under CC BY 2.0. Link to original.
Bottom right: NASA's Earth Observatory. Licensed under CC BY 2.0. Link to original.

0.1 Introduction

Engaging Mathematics seeks to create a set of courses that can make students' learning meaningful and relevant to their lives. At LaGuardia Community College, we developed an alternative version of our Elementary Statistics course, which we call *Elementary Statistics: Society and Environment*, that incorporates quantitative reasoning projects or "modules" to increase students' awareness of current events and connect their learning of mathematics to the outside world. A module contains exercises that help a student explore a theme or an issue, e.g., demographics, healthcare, or climate change, to help him or her understand and apply fundamental concepts in statistics in a realistic setting. A module also includes writing assignments that help the students to reflect on their activities and to present reasonable conclusions. A module-driven approach based on practical contexts is particularly suitable for students taking *Elementary Statistics* at LaGuardia because they hail from a wide variety of majors—most of them outside the natural sciences and technology. Core requirements state that every student must take at least one college-level course in mathematics and/or quantitative reasoning. Since the statistics course is often more relevant to many of these students than the college algebra/pre-calculus courses, it is a requirement for majors in, say, nursing, elementary education, and the social sciences. It is also a preferred alternative for students that have little more than high-school level (elementary algebra) or developmental mathematics experience.

The modules are designed to be portable units that can be implemented within an elementary statistics curriculum as in-class activities, as group-based collaborative work, and/or as independent inquiry projects. Each module in our implementation of the course is relevant to a theme in society or the environment that gives rise to a particular set of course concepts in statistics. Collectively, the modules are intended to support the course objectives to familiarize students with the nature of statistics and its technical underpinnings. They illustrate the key role that probability, uncertainty, and statistics play in students' lives and their ubiquity in everyday situations. The *Elementary Statistics* course covers topics such as the design of experiments, the data collection processes, potential sources of bias in samples, and techniques for converting data into information. A software package is used in the course so students may calculate sample statistics like means, medians, percentiles, standard deviations, confidence intervals, and critical values for hypothesis testing. They help demonstrate fundamental theorems and allow the student to work with larger, more realistic data sets and create tables and charts. The course also introduces the concepts of causation and correlation, how to use these to study multi-variate data, and how to use regression equations to make predictions. It also covers the basic concepts of probability and distributions of random variables. Finally the course presents real-life processes and events, to allow the student to apply course concepts and develop meaningful and well-presented conclusions and decisions.

0.2 Core Concepts Covered by Course Modules

As with many versions of elementary statistics offered throughout the country, the content covered in *Elementary Statistics: Society and Environment* at LaGuardia Community College can be partitioned into four major sets of concepts:

- Concept Set 1: Data Collection, Sampling Methods, and Descriptive Statistics

This part of the course introduces basic statistical concepts such as populations, variables, data collection, sampling techniques, sources of bias in samples, demographics, graphical representation and the organization and processing of data, measures of central tendency, and dispersion. Statistics is shown to be a means to generate meaningful information from a data set.

- Concept Set 2: Describing Relationships between Variables: Causation, Correlation, and Regression

  Students learn about controlled experiments and observational studies so that they can understand the distinctions between causation and correlation between variables. This part of the course focuses on (linear) correlations between two variables, describing data sets using scatterplots or two-way tables, the linear correlation coefficient and the concept of independence. In addition, they learn to identify which meaningful relationships are demonstrated when analyzing data from real studies.

- Concept Set 3: Probability and Probability Distributions

  Fundamental concepts of probability and discrete, continuous and multi-variate probability distributions are presented in this part of the course. The students learn where probability values might come from and which probability models are commonly used to describe realistic phenomena.

- Concept Set 4: Inferential Statistics and Hypothesis Testing

  Students are shown how descriptive sample statistics are used to infer population parameter values, calculate confidence intervals, and apply the fundamentals of hypothesis testing. For this course, most of the inferences are about the mean from a population.

  The modules that the LaGuardia Community College Engaging Mathematics team developed for *Elementary Statistics: Society and Environment* are briefly described below. Each is mapped to the concept sets it supports. We note that a course module that covers some themes in personal health, society, or the environment may have activities that are focused on one of these concepts, while another may have activities that span multiple concept sets. Thus it is important for the instructor to carefully plan when in the course to introduce a module and when to set the due dates for the activities. A module can be presented early, i.e., to help introduce a concept, or later, to reinforce one or more concepts by providing students an opportunity to practice what they had learned. A module that spans multiple concept sets might be better suited as a "capstone" experience. For each module, a website is provided that allows instructors to download the assignment text, a set of instructor recommendations, and electronic copies of any associated data sets.

Module I: My Class, My College and My City
https://lagcc-cuny.digication.com/project_data/Demographics

Students are presented with the fundamental question: how can we describe an individual so that he or she can be studied as part of some group? The module covers Concept Set #1, introducing the nomenclature of statistics and practices such as data collection, sampling methods, and descriptive statistics. Students are asked to design a sample framework to collect data for variables about themselves and their peers, like the country of origin, age, gender, commute time, and NYC borough of residence to perform an observational study. The assignment covers topics on sampling

techniques, sources of bias, sample and population data, and concludes with a demographic comparative study of the college. Students are also prompted to critically reflect on the implications of the differences at several geographical scales. The assignment can be extended to asking students to perform a demographic comparative study of the neighborhood or the city.

Module II: Intensity of Hurricanes
https://lagcc-cuny.digication.com/project_data/Hurricanes

The objective of this module is to have students examine data to identify whether a relationship exists between wind speeds and pressure within a hurricane and to develop a (linear) model that describes such a relationship. Thus it covers topics from Concept Sets #1 and #2. Data on maximum sustained wind speeds (in miles per hour) and the pressure (in millibars) within hurricanes from year 2000 to 2005 is provided. The students are asked to either use the entire data set or to take a sample (by year) to perform statistical analyses using SPSS or Excel. Through this activity, students explore and learn how to identify linear correlations between two data sets using scatterplots and the linear correlation coefficient. At the end of the assignment students reflect on their findings and relate it to their research on the underlying topic of hurricanes.

Module III: Making Connections, "Does This Ad Make Me Fat?"
https://lagcc-cuny.digication.com/project_data/Connections

This activity presents students with an opportunity to learn key concepts of causation and correlation as described in Concept Set #2. Students will read an article from the *New York Times*, "Does This Ad Make Me Fat?" and examine the underlying statistical study that attempts to draw a relationship between aggressive food advertising and the level of obesity in a community. Students learn the distinction between a study that is "observational" in nature, so can only be used to establish a correlation, and a study that includes an "experiment" that aims to establish a causal relationship between a pair of variables. Additionally, students review introductory concepts such as explanatory variables, response variables, lurking variables, and their significance in a statistical study. Students are given the opportunity to explore the original study report to reflect on whether the article's criticism is fair.

Module IV: Temperature Trends
https://lagcc-cuny.digication.com/project_data/Temperature_Trends

In this activity, students examine historical data about average temperatures in New York and in Sydney, Australia. Students will use a statistical tool to calculate statistics, plot time-series graphs and test hypotheses to determine whether there is a significant trend to this data. As an extra credit exercise, they are asked to find data for another city and then consider whether any similarities discovered about these diverse locations might allow them to make more broad conclusions, e.g., about Global Warming. This module spans all of the concept sets in the course, so would be suitable for a semester-long, multi-phased project or as a capstone activity.

Module V: What is my calorie budget?
https://lagcc-cuny.digication.com/project_data/Metabolic_Rate

Students read an article about Basal Metabolic Rate (BMR) to gather basic information, including the underlying mathematical formulas involved. They are expected to collect data from their peers, create graphs and charts, and calculate descriptive statistics. Thus the module covers topics from Concept Set #1. They will be asked to reach some conclusions and write about their

experiences.

Module VI: Presidential Heights and Weights
https://lagcc-cuny.digication.com/project_data/Presidents

In this module, students are presented with a two-way table compiled from data on the recorded history of presidential elections, giving the number of times that the winning candidate was taller or shorter, and heavier or lighter, than his or her opponent. The project covers topics from Concept Sets #2 and #3. Students compute statistics and estimate probabilities to help determine whether height and weight play a role in political success.

Module VII: Is This Generation Taller?
https://lagcc-cuny.digication.com/project_data/Generations

In this module, students test hypotheses about whether there are significant differences in height between their generation and their parents' generation, in support of Concept Set #4. In the preparatory stage, students draw informal conclusions based on their family members, friends, and acquaintances to decide whether there is any indication that a difference exists. They are prompted to write a paragraph describing their observations and develop appropriate hypotheses. In the subsequent stages, students may gather data on their own or use sample data provided by the instructor to test their hypotheses. The different data sources provided give rise to a variety of potential class discussions, student speculations, and reflective writing activities based on the topic.

Module VIII: Coffee and Sleep
https://lagcc-cuny.digication.com/project_data/Coffee_Sleep

This is an in-class activity based on data collected in the classroom. Students are asked two questions: "How many hours of sleep did you get two nights ago?" and, "How much coffee did you drink yesterday?" The instructor gathers the data and uses it to review topics in Concept Set #1 and #2. Students are asked to summarize the results in writing. The data set can also be used to illustrate visualization techniques such as five-number summaries and box-plots. Class discussions regarding the correlation of coffee consumption and sleep, based on the sample data, can further lead to discussion about extending results to the population, a future topic in inferential statistics.

0.3 Module Implementation and Assessment

The activities described in each module can be used either to introduce a course topic or to review the course material later in the semester. Based on our experiences, we recommend using about two or three of these modules to partially cover the concept sets for a typical semester. Where appropriate, Instructor Notes are included with the Student Handouts in the module to provide detailed guidelines and suggestions. Some of the activities, like those in Modules I and IV, could be assigned as a semester-long project. Other activities, like those in Modules VI and VIII, could be used as classwork exercises, while activities like those in Modules II, III, and V could be assigned as homework. Moreover, the design of each activity allows the instructor to use them to either present or scaffold course material as well as develop quantitative reasoning and writing skills.

Students' work can be assessed based on their demonstrated abilities to comprehend the underlying topic, to perform appropriate statistical analysis, and to make decisions that are consistent with their analyses. Each activity introduced in the modules involves exercises based on three components: reading, statistical analysis and reflection (through writing exercises). This allows students to develop skills in inquiry-based learning and to be able to draw meaningful conclusions based on their quantitative reasoning and analysis skills. Instructors could create rubrics for the Module activities that address these competencies as well as those based on the Statistics concept sets and could possibly be included in the course syllabus as grading criteria.

0.3.1 MODULE I - Demographics: My Class, My College & My City

Mangala Kothari and Milena Cuellar

LaGuardia Community College, Long Island City, NY

Part_1: Demographics of NYC

New York City is extremely diverse. Immigration in New York City has helped to make the city grow to a total of 8.34 million people as of 2012. Three million of these New Yorkers were born outside of the United States, and 49% of New Yorkers speak a non-English language at home. These New Yorkers come from all around the world and have had a huge impact on the racial and ethnic diversity of New York City.

The above information is summarized from "The Newest New Yorkers," a report on NYC immigration that is based on Census and American Community Survey data. Please read Chapter 1:

http://www1.nyc.gov/assets/planning/download/pdf/data-maps/nyc-population/nny2013/chapter1.pdf

You may also visit

http://www1.nyc.gov/site/planning/data-maps/nyc-population/newest-new-yorkers-2013.page

to view the full report.

Based on your reading please answer the following questions:

1. How does this report of New York City immigration relate to your experience as a New Yorker?
2. What is the specific group of people (i.e., the population) that the report focuses on?
3. This report makes claims, such as 49% of New Yorkers speak a non-English language at home. How do you think the writer made this claim? What is their evidence?

Part-2: Demographics of the College

Is LaGuardia Community College similar to the rest of New York in terms of its population? Suppose you are hired to complete a statistical study for LaGuardia Community College Students Affairs Division. The goal of this study is to learn about the country of origin, gender, NYC

borough of residence, and age of LaGuardia students. This information could help us identify ethnic group majorities and minorities within the college, estimate commute times to offer extracurricular support programs for students, suggest students' societies and clubs according to student's country of origin, etc.

Please answer the following questions:

1. Identify the population and variables in your study.
2. Use the following chart to list the name of each variable, state if the variable is qualitative or quantitative, and specify the possible values it could take. For example, a "Country of Origin" is the name of a qualitative variable that can take values, such as "United States," "Mexico," "India," "South Africa," etc.

Variable Name	Variable Type	Possible Values

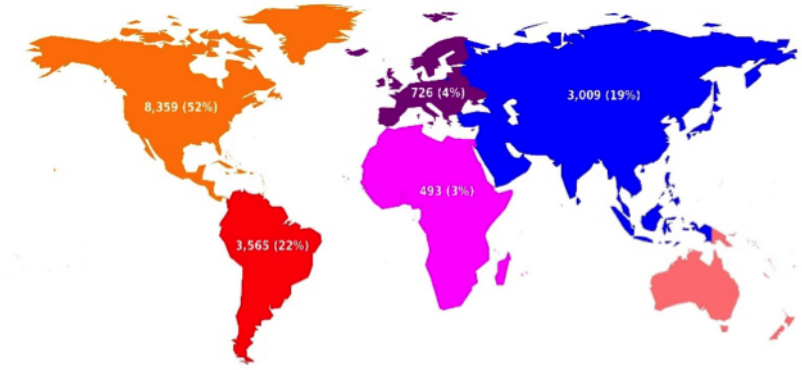
3. Due to time and budget constraints, it is impractical to conduct a population census (a survey of all students in the college), so you can only sample 30 students in order to estimate the population characteristics of interest in your study. Describe how you will go about selecting an unbiased sample and gathering data for the four variables of interest? Be sure to justify your choice of sampling method. Why do you think the method you chose is appropriate?
4. Find a sample of 30 LaGuardia Community College students using the method you described in the last exercise. Include a table to display your data using one column per variable.
5. For each variable, create an appropriate graphical representation, such as a histogram, pie chart, etc. for the data collected. Note that the data values for the variable "country of origin" can be grouped by the continents Asia, Africa, North America, South America, and Europe, and summarized using a frequency distribution.
6. Consider the data you collected. To what extent do you think the sample you obtained is representative of the student population at LaGuardia? Did you use a "convenience sample?" What are some sources of bias that could possibly have occurred in your sample?
7. What could you do to determine whether your sample is representative of LaGuardia's student population?

Part-3: Comparison of Sample and College Demographics

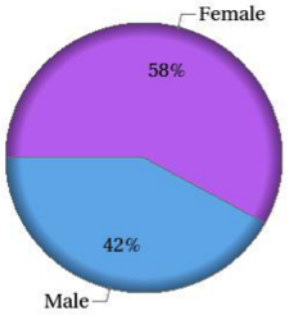
Figure 1 contains a graphical summary of the four variables from Part-2 for the student population at LaGuardia Community College obtained from the Institutional Research office (http://www.lagcc.cuny.edu/IR/IR-facts), which effectively maintains and reports a census of LaGuardia's student population. Use your sample data from Part-2, the information in Figure 1, and your responses from Part-1 to support your answers to the following questions.

1. How do your sample results compare with the entire student population at LaGuardia? What similarities and differences do you notice?
2. Based on your summary of similarities and differences, do you think you succeeded in obtaining a sample that is representative of the entire student population? Why or why not? Carefully examine the sampling process you used in Part-2 to support your answer.

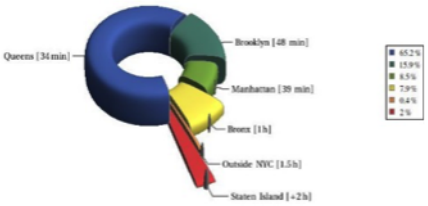
3. Is it reasonable to generalize the results of your sample to the entire student population at LaGuardia?
4. If you have identified sources of potential bias in your sample, what are possible remedies to minimize it? Do you think the sampling method or sample size has any influence on the results?
5. Pick a partner and discuss each other's sampling process. Can you identify sources of potential bias in your partner's sample? If so, explain it to your partner.
6. As a group, come up with a best strategy and report it to the class.



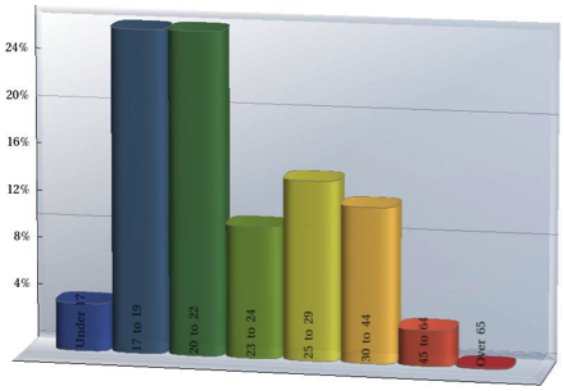
a) Country of origin



b) Gender



c) NYC City Borough of Residence



d) Age

Figure 1: LaGuardia CC institutional profile 2013 for demographic variables

Module I: Demographics of my Class and College & NYC - Guideline for Instructors

1. How can this topic be introduced in a statistics class?

   As an introduction to the defining components of statistics and its practice, the instructor may initiate the discussion with students by asking questions such as, "How can we describe a group of people?" or, "Who are we or where do we live and work?" thereby allowing students to come up with some variables. This conversation may then help the instructor to discuss the demographics of the class or the institution.

   The reading assignment in Part-1 of the module introduces the topic of demographics in New York. It allows instructors to help students build an inquiry and lay out the theme background information about the model. It could be made optional to students. Instead, the instructor may just briefly discuss the summary of the article and assign students only Parts 2 and 3 of the module. Instructors living in other communities will need to seek out demographic information relevant to their community. (A list of online demographic resources is included with module materials at https://lagcc-cuny.digication.com/project_data/Demographics).

2. When could this project be assigned in a Statistics course?

   We assigned the module after introducing students to the basic process of statistics and data collection; in particular, after discussing sampling methods and bias. Students are then prepared to conduct an observational study on demographics, collecting data for variables like

country of origin, age, gender, city of residence, and commute time. We assign the second part of the activity after discussing approaches for summarizing and visualizing qualitative and quantitative data; in particular, after introducing frequency and relative frequency distributions, bar charts, pie charts, histograms, stem-and-leaf plots, dot plots, shapes of distributions, and graphical representation of data.

3. What are objectives of this module?

   This module introduces the fundamental statistical process of data collection. Students learn through experience the key issues and limitations of the data collection process in real world settings. Students will be able to:

   - Classify quantitative and qualitative variables used in the study
   - Determine and create appropriate graphical summaries, including frequency distributions, for different types of variables
   - Demonstrate understanding of sampling techniques and bias
   - Make meaningful comparisons between the characteristics of a sample and the population from which the sample was drawn

4. What is the time commitment for this module?

   - After 20-30 minutes of introduction in class, students are given about a week for data collection, working individually or in a group.
   - Allow students 1-2 weeks to complete the full project report.

5. How is student work on this module assessed?

   An Instructor can choose to use his/her own rubric to assess the students' work. One can look at the students' ability to apply the knowledge of course contents consistently to make the inferences about their findings (questions in Part-2). A summative assessment of the conceptual understanding and its applications would be helpful in evaluating the students' work.

6. Additional suggestions:

   After collecting the project report, have a discussion in class about the students' findings. Discuss with students the limitations of the data collection process and, if possible, discuss examples of multi-stage sampling techniques.

   - The instructor may choose some of the students' samples to create graphical displays and use them in class to demonstrate the comparison between summaries of different sample sizes. This could help instructors to facilitate the discussion with students on the accuracy and shortcomings of their sampling methods. The instructors could focus on clarifying students' common misconceptions, such as the sample results must exactly match with the population, and the use of some sampling method always yields a random unbiased sample.
   - Once the lessons on Measures of Central Tendency and Dispersion are covered, the instructor can create one data file with the data collected by all students (by variable), which he or she can share with students and work with them at the Computer Lab hour.

- The instructor may also collect and use the class demographic information as an example to illustrate how representative the class is of the institutional profile and help students understand the similarities and differences in the results.

0.3.2 MODULE II - Intensity of Hurricanes

Nana Osei Bonsu

LaGuardia Community College, Long Island City, NY

Introduction

Hurricanes are natural events that bring destruction in many different ways. They are tropical storms with high wind speeds that can unleash gallons of rain. The high winds may spawn tornadoes and the torrential rains may cause floods and landslides. The destructive nature of hurricanes after making landfall is very devastating, to the extent that houses are wiped off the map and the floods deposit debris of damaged homes. Sometimes human lives are lost through this destructive natural event. As residents in our various communities, it is very important to have some level of knowledge about hurricanes and familiarize ourselves with the best practices and safe measures in the event of a hurricane.

A hurricane is an intense cyclonic storm that develops over the warm oceans of the tropics. It usually begins as a tropical disturbance and turns into a tropical depression when the speed of the wind attains 61 kilometers per hour (km/h) or equivalently 38 miles per hour (mph) at the storm center. When the sustained wind speed attains 63 km/h (or 39 mph) the tropical depression becomes a tropical storm. The tropical storm is classified a hurricane when the sustained wind speed reaches 119 km/h (or 74 mph). Using the Saffir-Simpson scale the hurricane is categorized a rating of 1 to 5. In the northern Indian Ocean, the tropical storm is known as a cyclone and in the western Pacific Ocean it is referred to as a typhoon [1].

We wish to determine how sustained wind speed in a hurricane is related to the surface pressure of the storm.

Part_1: Learning about Hurricanes and Tropical Storms

Reading: Tropical weather and hurricanes. (Source: Pidwirny, M., 2012).

http://www.physicalgeography.net/fundamentals/7u.html

Based on the reading please answer the following questions. You may also use additional websites to obtain more information about hurricanes, but you must provide the name of the site and the web address.

1. How are hurricanes developed?
2. What weather-related conditions will lead a meteorologist to predict a hurricane?
3. How are the categories of hurricanes defined? What do we mean by the strength of a hurricane?
4. What is a tropical storm? How are they different than hurricanes?

Part_2 Relating Surface Pressure and Wind Speed

The goal of this part of the module is to consider the relationship between wind speed and pressure within a hurricane and, if possible, develop a model that describes this relationship. Please refer to the dataset on maximum sustained wind speeds measured (mph) and pressures (mb) within hurricanes for the period 2000-2005 at https://lagcc-cuny.digication.com/project_data/Hurricanes. Using either SPSS or Excel for statistical analysis of this data set, please answer the following questions.

1. Identify the response and explanatory variables. Construct a scatter plot for this data set. What kind of relationship appears to exist, if any, between the two variables?
2. Determine the linear correlation coefficient. Does the value support your observation in exercise 1?
3. Is the linear correlation coefficient statistically significant at the 5% level? Explain. What does this tell you about the existence of a linear relationship between these two variables?
4. Develop a least squares regression model for the two variables. Graph it along with the scatter plot.
5. Interpret the slope of the least squares regression model in the context of wind speed and atmospheric pressure.
6. Determine the coefficient of determination and interpret its meaning in the context of wind speed and atmospheric pressure.
7. Use the least squares regression model to estimate the maximum sustained wind speed in a hurricane when the pressure reading is 950 mb.
8. In 2000 and 2004, hurricanes DEBBY and JEANNE recorded maximum wind speeds of 75 mph and 127 mph, respectively. However, their corresponding pressure readings remain unknown. Use your regression model to predict these pressure readings. Be sure to include appropriate units!

Part_3

1. Write a paragraph on hurricanes reflecting your work on this module. In particular, please address the following questions:
   - What impacts do hurricanes have on the weather?
   - How do hurricanes affect communities along their path?
   - When a hurricane is forecasted and its surface pressure is provided, how will you use this information to educate and advise people living in that area?
2. Gather information about the recent hurricane SANDY and write a paragraph on its impact on the communities affected. Include information about the wind speed and pressure of this hurricane.

References

1. Pidwirny, M. (2012). Tropical weather and hurricanes.

2. Aguado, E. and James E. Burt. 2010. Understanding Weather and Climate. Fifth Edition. Prentice Hall, Upper Saddle River, New Jersey.

3. Ahrens, C. D. 2006. Meteorology Today. An Introduction to Weather, Climate, and the Environment. Eighth Edition. Thompson, Brooks/Cole. USA.

4. Lutgens, F.K. and E.J. Tarbuck. 2004. The Atmosphere: An Introduction to Meteorology. Ninth Edition. Prentice Hall, Upper Saddle River, New Jersey.

5. Oliver, J.E. and J.J. Hidore. 2002. Climatology: An Atmospheric Science. Second Edition. Prentice Hall, Upper Saddle River, New Jersey.

6. Tsokos, Chris P. and Wooten Rebecca D. 2013. The Joy of Statistics: Learning with Real World Data. First Edition. Kendall Hunt, USA.

Module II: Intensity of Hurricanes - Guidelines for Instructors

1. How do I introduce this topic in my class?

   Use information from the reading assignment or any recent information on hurricanes to introduce the topic. The dataset for the module can be found at https://lagcc-cuny.digication.com/project_data/Hurricanes.

2. When do I assign this project?

   This project should be assigned after the lesson on correlation and simple linear regression.

3. What are the objectives of the module?

   This module addresses the use of statistical software for obtaining a least square regression model for maximum sustained wind speed and pressure in a hurricane. By the end of this project, students will be able to:

   - Identify explanatory and response variables and represent their relationship visually using a scatterplot.
   - Measure the strength of a linear relationship between the two variables and test its significance at the 5% level.
   - Interpret the slope of a least squares regression model in terms of the two variables.
   - Use the least squares regression model for estimation.
   - Interpret the coefficient of determination.

4. What is the timeline for the module?

   The Instructor may use 40-45 minutes to introduce the module, to perform illustrative analyses using dummy data in SPSS or Excel and to identify keys measures or information from the output. Allow 15-20 minutes in a lab for students to run their analysis in SPSS or Excel and copy the output to MS-Word to complete the write up. Allow a week or two for the completion of the full project report.

5. What are other instructions, besides students' handout, would you give to students before starting to work on the project?

The data should be made available to students through blackboard or any student accessible learning platform.

6. Any other suggestions?

   The data contains two missing pressure values so the instructor needs to clean up the data before posting it on blackboard/learning platform or the instructor can guide students to individually clean up the data before using it. The project prompts were self-explanatory to the students and there were no observable confusions. Overall, the project went very well.

0.3.3 MODULE III - Making Connections: "Does This Ad Make Me Fat?"

Steven Cosares

LaGuardia Community College, Long Island City, NY

Since obesity has become a growing problem in the United States, researchers have been trying to identify what factors may cause an increased level of obesity and what policies they could recommend in order to curtail the level of obesity, especially in urban areas. In a New York Times Op-Ed piece, "Does This Ad Make Me Fat?" (http://www.nytimes.com/2013/03/10/opinion/sunday/does-this-ad-make-me-fat.html?emc=eta1) the authors describe a study that attempts to find a relationship between outdoor advertising for food, like those on billboards, and the level of obesity within some sample urban regions near Los Angeles and New Orleans called "census tracts." Each randomly selected census tract represents a single subject in the sample.

Read the article and answer the following questions:

1. To get a value for the *variable* that represents the amount of outdoor advertising in the tract, the researchers went along every street and counted the number of billboards that advertise some food or restaurant. How do you suppose the researchers obtained a value for the variable representing the overall level of obesity in a census tract? (Hint: "Body Mass Index," or "BMI," is a variable that represents the total fat of an individual by measuring both height and weight. It is discussed in: http://www.nhlbi.nih.gov/health/educational/lose_wt/BMI/bmicalc.htm.)

2. The study finds that the two variables are *correlated*. Do you think that the associated correlation coefficient is positive or negative? Why?

3. The study indicates that, in addition, there may be a *causal relationship* between these variables. If this is so, which is supposed to be the *explanatory* variable and which is the *response* variable? The Op-Ed piece claims that the roles of the two variables could be reversed. Do you agree? Why or why not?

4. A *lurking variable* is some factor that is not included in the analysis of a situation, but can affect how other variables are related. Can you identify some other variables, beside the number of billboards, which might be reasons for one neighborhood (tract) having a higher level of obesity than another?

5. The Op-Ed piece describes some ways in which an experiment might be designed to determine whether a causal relationship exists. Describe one of these and show how it would more clearly demonstrate cause and effect.

Extra Credit:

The article gives the reference to the original study report, http://www.biomedcentral.com/1471-2458/13/20. The authors say that "...the researchers themselves went out on a limb that their data did not fully support" by claiming a causal relationship between the variables. Do you think that is a fair assessment of the original study? Why or why not?

Module III: Making Connections - Guidelines for Instructors

Materials for this module may be found at https://lagcc-cuny.digication.com/project_data/Connections.

In this lesson, students are asked to read a *New York Times* Op-Ed piece, "Does This Ad Make Me Fat?" (http://www.nytimes.com/2013/03/10/opinion/sunday/does-this-ad-make-me-fat.html?emc=eta1). In the piece, the authors present their opinions about a study that draws a connection between the level of obesity of the residents in some metropolitan areas and their level of exposure to food-related outdoor advertisements. After completing the assignment, students would understand the distinctions between a study that is "observational" in nature, so can only be used to establish a correlation, and a study that includes an "experiment" that aims to establish a causal relationship between a pair of variables.

If you wish, you can ask students to refer to the original study report, "Outdoor advertising, obesity, and soda consumption: a cross-sectional study," (http://bmcpublichealth.biomedcentral.com/articles/10.1186/1471-2458-13-20) so that they may evaluate the opinions and conclusions of the Op-Ed's authors.

0.3.4 MODULE IV - Historical Temperature Trends

Marina Nechayeva and Steven Cosares

LaGuardia Community College, Long Island City, NY

Has the average temperature in cities throughout the world changed over time? Are there any recent trends that we should be aware of? In this module we will look at historical temperature records for selected cities and use statistical methods to determine whether the data shows any significant patterns over time.

You will be provided historical data for the annual average temperature records for New York City, USA, and Sydney, Australia. Data for other cities throughout the world may also be provided; the cities should represent a variety of regions having different climates. For each year, the data value is calculated by taking the average of the daily high temperature readings from every day of that year, observed in a specific location, e.g., Central Park for New York and Observatory Hill for Sydney. The data are sorted by year in ascending order. For your convenience, temperatures have been converted to the Fahrenheit scale. You can find the data and the original data sources at https://lagcc-cuny.digication.com/project_data/Temperature_Trends.

Exercise 1

a) Use a statistical software package to produce a scatter plot of the temperatures for New York City using the "year" as x-axis.

b) Provide a paragraph describing whether you see any trends or patterns in the data from your graph.

c) Produce a similar plot for Sydney, (perhaps using the same axes as the NYC plot).

d) In a paragraph, describe what you see for that city.

e) Describe any similarities and/or differences you observe between Sydney and New York City.

Exercise 2

a) Use a software package to obtain summary statistics for the annual average temperatures in New York City and Sydney over the entire recorded period: min, max, mean, median, and standard deviation. Identify the hottest year and the coldest year for each city.

b) Make a Box Plot for each data set.

c) Write a paragraph describing how the two cities differ in temperature. Hint: Base your answer on any notable differences you observe in the two Box Plots.

Exercise 3

a) Use software to obtain a frequency distribution histogram of the annual average temperatures in New York City. Describe the "shape" of distribution.

b) Repeat the process for Sydney, using the same class limits you had for New York City.

c) In a paragraph, describe any similarities and differences you observe between the shapes of the two distributions and interpret the meaning of these differences.

Exercise 4

Limit your attention to the most "recent" temperature data, i.e., the last 25 years for New York City. Answer the following questions:

a) Is the mean temperature for the last 25 years higher or lower than that of the entire period? Calculate the z-score for the recent mean temperature based on the mean and standard deviation for the entire period. What does the value of the z-score tell you about the difference between the two means?

b) For what percentage of the last 25 years was the annual temperature above the historical mean temperature from 1869-present? Comment on whether you think this average higher or lower than you expected.

c) Of the 10 warmest years in recorded history, how many occurred during the last 25 years? Of the 10 coldest years in recorded history, how many occurred during the last 25 years? Do you find this result surprising?

d) Create a box-plot representing the temperatures for the last 25 years.

e) Write a paragraph summarizing your findings. Are NYC annual temperatures notably different in recent times from what they were in the earlier part of the observation period? In what way?

Extra Credit: Repeat the exercise for the Sydney data.

Exercise 6

a) Based on your analysis of the historical temperature data for these cities, make a hypothesis about the nature of temperature trends throughout the world. What would you need to do to "test" your hypothesis?

b) Global warming is a much publicized and controversial issue. Based on your work in this module, what role can statistics play in shaping our discussions concerning this important topic?

Module IV: Historical Temperature Trends - Guidelines for Instructors

Data for this module can be found at https://lagcc-cuny.digication.com/project_data/Temperature_Trends.

Learning Goals: Students gain insight into the very important, much publicized, and controversial issue of Global Warming using a data-driven, scientific approach. Students become aware of the power of statistical methods learned in an elementary statistics course to guide an inquiry, evaluate the sufficiency of evidence, and provide a data-informed discussion around this important issue. The topic allows for students to engage in evidence-based, persuasive writing exercises.

Topics covered: The activity spans a variety of topics from descriptive statistics such as graphical representation (e.g. histograms, box plots) and numerical summaries (e.g. measures of center and variability) for comparison of large data sets, as well as correlation and regression analyses. Inferential statistics topics, i.e. basics of hypothesis testing, are informally introduced. Students gain relevant technical skills working with a statistical software package such as SPSS.

Data: Students access online databases to get average annual temperature records (going back at least 100 years) for selected cities in different parts of the world (New York, Sydney, and others). For each city, students detect and analyze temperature trends.

Timeline: Different parts of the project can be staged throughout the semester. The first exercises can be used to help students become familiar with larger data-sets and the need to employ some statistical software. The last three exercises can be part of a capstone exercise that contributes to the final grade.

Outline:

(Exercises 1-2) Global Temperature Trends activity is introduced. Students are provided with link to annual average temperature records of two cities, New York, USA, and Sydney, Australia, each going back over 120 years. The first activity involves summarizing average annual temperature distribution for both cities.

The instructor may choose to perform all necessary steps (making a histogram, describing the shape, finding and interpreting center and spread) using data records of one city, as a way to introduce the statistical package and relevant functions/procedures as well as to review key statistical ideas. Students would then be asked to perform all the steps in small groups, using data from the second city.

Students should get familiar with their variable, average annual temperature, and get an idea about typical average annual temperatures in the selected cities and typical variations in the data. They might observe that average annual temperatures don't seem to fluctuate by much (the standard deviation is "small"), they should also become aware (from assigned readings and/or discussion) that even "small" changes in average annual temperature over time can signal catastrophic consequences.

(Exercise 3) At this point students should have been introduced to measures of relative standing and box-plots as means for comparing different data sets. The project itself can serve as a means for introducing/reinforcing these topics. The emphasis here is on comparing two sets of data for each city: observations over the entire recorded period and observations limited to the last 25 years. As before the instructor can choose to do one city with the class, ask student to work in groups on the second city, and have them work individually, at home, on the third city of their choice. Students should observe that the average temperature over the recent period is higher for each city. There may also be notable differences in range and spread.

(Exercise 4-5) At this point students should have been introduced to time series, scatter plots, and the correlation coefficient as a measure of strength and direction of the association between variables, and also have a good intuitive idea of the notion of significance as applied to the correlation coefficient and an understanding of the concept/purpose of the least square regression line. The project itself can serve as a means for introducing/ reinforcing these topics. Students should be able to spot and make sense of the upward temperature trend in all three cities, measure the strength of the association (statistical significance) and look into effective significance of the temperature change (interpret the slope of the regression line and compare it to the value of the standard deviation).

(Exercise 6) Students are asked to reflect on the results they have obtained so far and write a short essay regarding the valid conclusions they could draw from the exercises. The instructor may allow students to discuss their ideas in groups but each essay should be written individually.

As a potential **Extra Credit** exercise, each student could be asked to choose a third city (they could use databases provided by the instructor or look for their own relevant records online) and repeat the steps using this data, describing any differences with NYC and with Sydney. Students should justify their choice of the city and realize that temperature records must go back at least 100 years and that their selection should differ in climate from NYC and Sydney.

0.3.5 MODULE V - What is my calorie budget?

Alioune Khoule

LaGuardia Community College, Long Island City, NY

Knowing how many calories you consume in the foods you eat each day is an important part of maintaining a healthy weight. Besides providing the energy you need to move about, calories are also used by your body to maintain the basic functions that keep you alive. (Think about the total energy needed just to keep your heart beating about 60 times every minute!) Basal Metabolic Rate (BMR) is the average rate at which energy is used by a person at complete rest, when he or she does not expend energy for any activity except for the basic functions for survival. BMR can be

expressed as the number of calories released, e.g., through body heat, per kilogram of body weight per hour. A person's BMR value might depend on his or her gender, age, weight, and height. Once we know our BMR, we can calculate how many calories from food we would need per day.

For more information read the following article: http://www.dbbe.fcen.uba.ar/old/materias/fac/PaperMetabolismo(2).pdf

Key definitions:

- Your *Activity Level* is a category variable that is used to represent your daily physical activity.
- Your *Energy Factor* represents your increased energy expenditure due to your activity level. The following chart shows energy factors associated with five different activity levels.

Activity Level	Energy Factor
Sedentary	1.2
Lightly Active	1.375
Moderately Active	1.55
Very Active	1.725
Extremely Active	1.9

For example, the 1.9 suggests that an extremely active person's total energy expenditure is 90% greater than his or her BMR. A sedentary person has an energy factor of 1.2 indicating that his or her energy expenditure is 20% larger than their BMR.

1. Find 40 New Yorkers (about 20 males and 20 females) and record their ages, heights, weights, and activity levels in the form of a chart as below.

Age	Height	Weight	Gender	Activity Level

2. Use the sample data you have collected to find the means and standard deviations for the weights, heights, and ages in your sample.

3. The BMR formulas for men and women are as follows:

   Women: BMR = 655 + (4.35 x weight in pounds) + (4.7 x height in inches) - (4.7 x age in years)

   Men: BMR = 66 + (6.23 x weight in pounds) + (12.7 x height in inches) - (6.8 x age in year)

   (http://www.bmi-calculator.net/bmr-calculator/bmr-formula.php)

   a. Use these two formulas to calculate the BMR for each of the sample subjects in your chart, and then record the BMR values in two separate lists, one for men and one for women.

   b. Calculate the mean and standard deviation for the BMR values for men.

   c. Calculate the mean and standard deviation for the BMR values for women.

   d. What, if any, significant differences do you notice between the BMR for the men and the BMR for the women in your sample? Why do you suppose this is so?

4. Calculate the mean BMR value for each category of activity level in your sample. Does there appear to be any significant pattern that would suggest a relationship between BMR value and activity level?

5. Construct a frequency distribution for BMR values using a class width of size 10. Display the distribution on a histogram. Describe the shape of the distribution.

6. Use the following formula and the energy factor chart given above to calculate total energy expenditure (TEE) for each subject in your chart.

   TEE = BMR x Energy Factor

   Construct a frequency distribution for the TEE values and graph the histogram using a class width of 10. Does this distribution have a different shape than the BMR distribution? Why do you suppose this is so?

7. Individuals with higher energy expenditures require larger calorie diets to fuel their higher activity level. Given this, explain what diet recommendations nutritionists should give to patients, based on their TEE. Based on the statistics from your sample, for example the spread in the TEE values, should nutritionists supply many different dietary recommendations, based on their TEE, or would a couple of overall standard recommendations for people be sufficient? Why?

8. Determine your personal BMR and TEE. Your calculated TEE value is an estimate of your personal energy expenditure in calories per day. Given this, about how many calories should you consume daily?

Module V: What is my calorie budget? - Guidelines for Instructors

1. **Objective:** The main purpose of this module is to conceptualize the idea of descriptive statistics and potential relationships between variables using a realistic context. Students gain experience with collecting, organizing, analyzing, and summarizing data in order to draw a conclusion. This module focuses on how many calories we should consume in order to maintain a healthy weight.

2. **Topics covered:** This module focuses mainly on descriptive statistics: frequency distribution, measures of central tendencies, and measures of dispersion.

3. **Data Collection:** Students will be asked to collect data from about 40 people (20 male and 20 female) outside or inside school. This will be input to a software system like Excel.

4. **Project timeline:** The activity could be either assigned early in the course, or after a discussion of potential bias in samples.

Step 1: Instructor will hand out the reading assignment, to develop students' baseline knowledge on BMR and its relevance.

Reading assignment: http://www.dbbe.fcen.uba.ar/old/materias/fac/PaperMetabolismo(2).pdf.

Instructor should spend about 30 minutes discussing the article just to make sure that students grasp the concept of BMR and Total Energy Expenditure.

Step 2: Students will be asked to collect data for 20 males and 20 females. These data include age, height and weight, and activity level. Students should be given about two weeks to collect the data.

Step 3: Instructor will use lab time, if available, to help students become familiar with the technologies needed, for example SPSS or Excel. Students will calculate the BMR and TEE for all individuals in their samples. The instructor can help students generate the BMR and TEE formulas in Excel, so they don't have to do the calculation for each subject.

Step 4: Students will calculate mean, median, and standard deviation for each. They may use the systems to help generate histograms.

Step 5: Students are expected to answer the questions and write up their results.

0.3.6 MODULE VI - The Heights & Weights of the Presidents

Brian Johnston

Quantitative Reasoning Fellow, CUNY, NY

Introduction

Do the heights and weights of presidential candidates affect whether they win the presidency? For example, in the 2008 presidential election, Obama was taller and heavier than McCain, and Obama won that election. Is there a pattern of taller and heavier candidates winning? The following op-ed piece from the *New York Times*, (http://www.nytimes.com/interactive/2008/10/06/opinion/06opchart.html?_r=0) summarizes presidential election data from 1896-2008. We could use the information to determine the number of times the winning candidate was taller/shorter and the number of times the winner was heavier/lighter. Please use the data from the article to answer the following questions.

1. How many times was the winning candidate taller? How many times was the winning candidate heavier?
2. Use the information to fill out the following table:

# times	Taller	Shorter	Total
Heavier			
Lighter			
Total			

3. What is a good estimate for the probability that the winning candidate is taller? Heavier?
4. What is the probability that the winner is taller and heavier?
5. If a candidate is taller, what is the probability the candidate is heavier?
6. If a candidate is shorter, what is the probability the candidate is heavier?
7. Is there a pattern of these winning candidates being taller and/or heavier? Please explain your answer.

Module VI: The Heights & Weights of the Presidents - Guidelines for Instructors

This assignment uses historical data on U.S. presidential elections to introduce probability concepts. The goal is to first get students thinking about how the heights and weights of presidential candidates could influence who wins elections. Then, students develop a summary of the data in a two-way table and respond to the assignment questions. By doing so, students will be introduced

to reading two-way tables, as well as calculating and interpreting probability (including marginal, joint, and conditional probabilities).

The raw data and the references can be found at https://lagcc-cuny.digication.com/project_data/Presidents.

0.3.7 MODULE VII - Is this generation taller than the last one?

Marina Nechayeva

LaGuardia Community College, Long Island City, NY

Part 1: Preparation/Reflection

I. Consider your family, friends, and acquaintances and decide whether there is any indication that the current generation of college-age students, (i.e., people born in the late 1990s) is different in average height (taller/shorter) from their parents' generation, (i.e., people born in the late 1960's). Write a paragraph stating your informal "claim" about this, providing whatever evidence you need to support it.

II. Consult the article: "Why have humans evolved to be taller over the last three hundred years?" by Dr. Christopher S. Baird here: http://sciencequestionswithsurprisinganswers.org/mobile/2015/08/19/why-have-humans-evolved-to-be-taller-over-the-last-three-hundred-years/.

Watch the video: "Will humans keep getting taller?" on *USA Today* site: http://www.usatoday.com/videos/news/2015/10/26/74652970.

Feel free to explore the Internet for additional answers to the question, "Are humans getting taller, and if so, then why?" Select a source that you find the most interesting/illuminating. If you use a different source from the two suggested here, please include a full citation for it along with your answers to the questions below.

What does your chosen source suggest?

1. Are humans still getting taller or has the growth stalled?
2. Have people "grown" taller all over the world, or is it happening more in some areas and less/not at all in others?
3. What are the reasons researchers provide to explain "growth" or "the lack of growth?"
4. What statistical evidence, if any, do the sources provide to support their claim?
5. What projections does your source make for future human height?

III. Now you have the chance to see for yourself! Write a short description of a statistical study that would provide insight into the question, "Is your generation taller than the previous one?" Address the following:

1. What population would you like to focus on for your study?
2. What kind of data will you need to collect?

3. How would you go about selecting your sample?
4. What sample statistics would you obtain?
5. How would you analyze your sample data?

Part 2: Descriptive Statistics

I. To obtain meaningful data that would answer the question we posed, it would make sense to narrow your population down to a particular nation (ethnicity or country) and gender. Why do you think this is the case?

II. You have been provided with a table listing average (mean) heights of males born around 1960 by nation (country). We can consider these values as population means for males born 1946-1975. To obtain an estimate for the mean female height in any particular country, just take the value listed for males and divide by 1.08.

Pick a nation you wish to study. We suggest you choose a nation you are connected to personally or through your circle of family/friends, as you will be asked to obtain a random sample of young people descendant from that country. Do you think people descendant from this country are getting taller/shorter? Let your instructor know which group you have chosen to study.

III. Collect and bring to class data on height from about 20 college-age adults (either all male or all female) from the population you have selected.

IV. Calculate (using SPSS or Excel) the sample mean and sample standard deviation for your data set.

V. Construct a relative frequency distribution for your sample. What is the shape of the distribution?

Part 3: Hypothesis Testing Activity

1. Compare the sample mean you obtained in Part 2 with the population mean of the previous generation you have obtained from the reference table. What does the data suggest? State your claim in words.
2. We are ready for formal hypothesis testing now. Are conditions for 1-sample t-test satisfied?
3. Set up null and alternative hypotheses at significance level 0.05.
4. Make a sketch of the relevant distribution and shade critical region(s).
5. Proceed using Critical Value or P-value method below:

   (Critical Value Approach)

   (a) Find and label critical value(s) on your sketch.

   (b) Calculate and label the test statistic on your sketch.

   (P-Value Approach)

   (a) Calculate the value of the test statistic.

   (b) Calculate the value of the test statistic.

6. Determine whether you have sufficient evidence to reject the null hypothesis at a 5% significance level. Discuss the conclusions of your statistical analysis in practical terms. What is your work saying about any differences in heights between generations within your population? If your work suggests that there is a height difference, is this difference significant in practical terms?

Final Reflection

I. Write a short essay describing your statistical study and its findings. Address the following:

1. What population did you use?
2. How did you select your sample/collect the data?
3. What conclusion have you arrived at?
4. Do your findings agree with expert claims you have found on the Internet? If not, what do you think are the possible reasons?
5. Did you find statistically significant evidence to support your claim?
6. Is the difference between heights of generations significant in the everyday sense of the word?

II. Choose one of the three topics below (or suggest your own) related to our theme and write a one page response. Please cite all sources that you use to support your response.

1. Many researchers point out that while Europeans (especially the Dutch) continue to grow taller, the heights of Americans have pretty much stalled for the past 50 years or so. What factors account for this observed difference between the two populations?
2. Many researchers point out that, in most regions of the world, a person's height is positively correlated with their socio-economic status. Moreover, highly developed and wealthier nations tend to have taller populations and vice versa. Is there a cause and effect relationship here, and if so, what is the cause and what is the effect?
3. Nature or Nurture? Are differences in average heights and growth trends between nations caused mainly by genetics or by socio-economic conditions within the nations (GDP, income inequality, healthcare system)?

Module VII: Is this generation taller than the last one? - Guidelines for Instructor

A variety of relevant datasets are available at https://lagcc-cuny.digication.com/project_data/Generations

1. When do I assign this project?

   Part 1 of this project can be assigned as soon as basic statistical concepts of study design, sampling, etc. are introduced (usually by week 2). Part 2 can be assigned after the lesson on measures of center and spread (by week 3 or 4). Part 3 deals with hypothesis testing and should be assigned at the same time as covering this topic in class (this is typically the last topic on the syllabus). Alternatively, the entire project may be assigned at the end of the course with Parts 1 and 2 serving as a review.

2. What are objectives of the lesson?

- Students will develop deeper understanding of statistical study design and learn to see data as evidence for inferential statistics.
- Students will focus on all aspects of hypothesis testing, from setting up the claim to interpreting the conclusion.

3. What is the timeline for the lesson? (How much time will the instructor spend in class? How much time will students get to complete the project?)

   The instructor may use 1-2 hours (staged throughout the semester) to discuss the most interesting/challenging aspects of the activity and to let students share the results, debate, and formulate common conclusions. The amount of time students spend will depend on the format chosen by the instructor. Students can spend 10-15 minutes calculating sample statistics using technology and another 10-30 minutes on hypothesis testing. Additional time outside of the classroom will be needed if students are asked to collect their own sample data. This activity is writing intensive. Students will need 1-2 hours to work with relevant information sources, and to write a preliminary and a final reflection.

4. What are other instructions, besides student handouts, would you give to students before starting to work on the project?

   Some "mock up" sample data may be provided for Part 2 of the activity, to save time and ensure viable samples. The data on average heights of the previous generation (by nation) should not be made available until part 2 is completed, to avoid introducing a bias.

5. Any other suggestions?

   The activity lends itself well to group work. Students may team up to collect data and to write a reflection piece. Students working with different data samples may be instructed to read and comment on each other's drafts before the final submission. Online discussion on Piazza or other educational forums/platforms can be used to supplement the class discussion.

0.3.8 MODULE VIII - Coffee & Sleep Data

Brian Johnston

Quantitative Reasoning Fellow, CUNY, NY

The objective of this assignment is to become familiar with a variety of topics in statistics, including:

1. Setting up data files (e.g., Excel, SPSS)
2. Scales of measurement for survey questions
3. How mean, median, and mode apply to data (reflect how everyone in a sample is similar)
4. How variance and standard deviation apply to data (reflect how everyone in a sample is different)
5. Sampling

First, think about how much you sleep and drink coffee. Then, please answer the following two questions:

1. How many hours of *sleep* did you get *two nights ago*?
2. How much *coffee* did you drink *yesterday*?

Next, the instructor will go over how to put your sleep and coffee data into a statistical program, along with the sleep and coffee data for the entire class.

Please answer the following questions:

1. What are the scales of measurement for the sleep and coffee variables? (nominal, ordinal, interval, ratio)
2. Doing calculations by hand, please find the mean, median, and mode for the sleep variable.
   - Please describe what these tell us in words.
3. Doing calculations with the computer program, please find the mean, median, and mode for the coffee variable.
   - Please describe what these tell us in words.
4. Doing calculations by hand, please find the variance and standard deviation for the sleep variable.
   - Please describe what these tell us in words.
5. Doing calculations with the computer program, please find the variance and standard deviation for the coffee variable.
   - Please describe what these tell us in words.
6. What sampling technique did we use?
7. What can we say about the sleeping and coffee drinking behavior in the population of students in this classroom?
8. What can we say about the sleeping and coffee drinking behavior in the population of students at your school?

0.4 Acknowledgements

We are grateful to the Engaging Mathematics project staff, especially Christine DeCarlo, and the project's PI Wm. David Burns and Co-PIs for their constant support, encouragement, and assistance throughout the work of this manual, without which this manual would not have been in existence. The author of Module VI would like to express his deep sense of gratitude to Dr. Nicholas Christenfeld from the Department of Psychology, University of California, San Diego, for providing crucial feedback on the presidential heights and weights assignment. We are also thankful to Roslyn Orgel from the Center of Teaching and Learning at LaGuardia Community College CUNY, Dr. Howard Wach, the VP for Academic Affairs, and the Quantitative Reasoning (QR) Fellows

program of the CUNY Graduate Center for the assignment of QR fellow Brian Johnston to assist our efforts on the development of the course in the context of the Engaging Mathematics project. We are grateful to Dr. Milena Cuellar for collaborating on the work on this manual during the initial stages of the project during the year 2014.

We would also like to express our deep sense of gratitude to the Provost, Dr. Paul Arcario, the dean of academic affairs, Dr. Bret Eynon, and the department chairperson, Dr. Abderrazak Belkharraz, for their guidance and constant support and encouragement throughout the work of this project. Finally, we would like to thank our colleagues, Dr. Marina Nechayeva and the faculty project advisor Dr. Prabha Betne, for assessing the elements of this curriculum in their classrooms, and our students for providing us their crucial feedback on the activities. We would like to extend our special thanks to Dr. Steven Cosares for providing critical feedback and assisting us in the revision of this work.

0.5 Course Map at a Glance

Suggested modules	Concepts sets	Skills used/presented	Comments
Demographics of my class, my college, my neighborhood and my city (Module I)	Introduction to practice of statistics, sampling methods, sources of bias, organizing qualitative and quantitative data, inference-estimating the population mean	Practical use of population, sample, type of variable, level of measurement, sampling methods, 4-steps of statistics process	The module activity can be assigned to either introduce or review the underlying topics. The module could be assigned as a semester-long project
Intensity of Hurricanes: Hurricane wind speeds and pressure (Module II)	Scatter diagram and correlation, regression and equation of the line coefficient of determination	Construct scatter diagrams and calculate correlation coefficients, identifying response and explanatory variables and relationship between them, calculate and graph regression equations and use it for predictions, interpretation of slope and y-intercept, calculating the coefficient of determination and its interpretation in context to the variables such as wind speed and pressure in the hurricane	The writing component of the assignment allows students to explore more about the similar data and use their findings to comprehend and analyze the information available online. The module could be assigned as a homework activity.
Making Connections, "Does This Ad Make Me Fat?" (Module III)	Correlation and causation	Difference between observational and experimental statistical study, students will learn how to identify causation and its limits when analyzing real data	The module could be assigned as a homework activity.
Historical Temperature Trends (Module IV)	Descriptive statistics (histograms, box plots), numeric summaries measures of center and variability), correlation and regression analyses, basics of hypothesis testing	Technical skills using statistical software such as SPSS or Excel, research and data analyses, writing and quantitative reasoning skills	The module can be staged in parts throughout the semester or the entire assignment can be assigned as a semester-long project.
What is my calorie budget? (Module V)	Data collection, sampling techniques, descriptive statistics, (measures of center, frequency distribution, histograms, spread)	Use of statistical software SPSS or Excel to learn descriptive statistics, writing up the results interpreting potential relationships between variables using a realistic context	This module could be assigned as a homework activity.
Presidential heights and weights (Module VI)	Probability rules, two way tables, independence conditional and joint and marginal probability	Calculating single event probabilities, interpretation of probabilities	The module could be used as a class exercise.
Is this generation taller than the last one? (Module VII)	The language of hypothesis testing, hypothesis testing for a population mean	Small-scaled research skills, data analyses, applications of hypothesis testing to real-world question	The module could be used either as a class exercises or homework activity.
Coffee & Sleep (Module VIII)	Measures of Central Tendency and Dispersion (raw/grouped data) Measures of Position, Outliers and 5 Five number summary and box-plot	Sampling (convenience); designing survey questions; scales of measurement for survey questions; how mean, median, and mode apply to data (reflect how everyone in sample is similar); how variance and standard deviation apply to data (reflect how everyone in sample is different); analyzing data by hand in lecture; setting up data files (e.g., Excel, SPSS) and analyzing data with software.	The module could be used as class activity to introduce the topics. The data files obtained can be used later for visualization of the concepts such as five- number summary outliers, or regression, correlation etc.