Learning Statistics through Excel simulations and Hands-on Surveys

Chintamani Jog, Jeremy Oller
Department of Economics, University of Central Oklahoma

Learning Objectives:

- Understand why more data is better
- Accomplish simulations to see Central Limit Theorem at work
- Give a feel for primary data collection and analysis.
- Learn about practical aspects of the process of acquiring data.

Abstract:

This presentation proposes two pedagogical innovations in teaching statistics at the college level. The first involves using interactive, MS-Excel based simulations to illustrate the Central Limit Theorem. The second is a team-based, semester-long activity, wherein students create questionnaires on the topic of their choice, collect primary data, analyze and present basic results. We expect these methods to facilitate self-motivated spirit of inquiry among students, be a part of their e-Portfolios, as well as reinforce the practice of transformative learning. The generality and relative ease of implementing these methods across different disciplines will be an added bonus.

Integrated Project Plan:

The simulations and hands-on project described are connected in several ways. Here are some potential ways in which the integration of these ideas will help students in a typical undergraduate introductory statistics course:

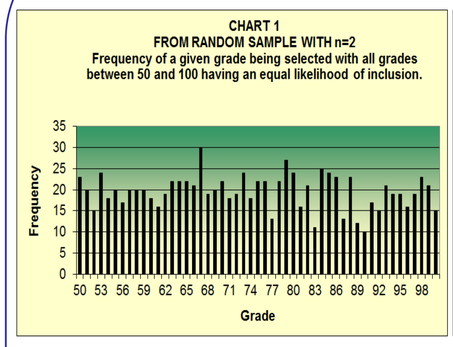
- Applicability across disciplines: Introductory statistics courses are ubiquitous in nature at undergraduate level.
  - Central Limit Theorem is often the common denominator that is part and parcel of every such course - hence, there is a dire need to understand the basic concept (i.e. the CLT) on which the rest of the topics (Inference, Hypothesis Testing, Regression etc.) are based
  - Standard errors are a common computation utilized in all statistics courses. These projects allow instructors to differentiate a *sample standard deviation* from a *standard deviation of a sampling mean* in a practical setting. Students frequently mesh these concepts; however, these practical applications and simulations reinforce the similarities and distinctions between the two variables.
  - Hands-on projects could be customized to each discipline based on the field of inquiry - for instance, psychology majors can collect primary data related to various trait tests.
- Use of MS Excel: Microsoft Excel availability is more widespread than specialized statistical software (i.e. Stata, SPSS that require licenses and cost on the part of colleges). Use of MS Excel prepares students for skills that employers are looking for, especially rudimentary data management and basic presentation.
- Segue to experiential learning: Based on the response of students, more ideas could be developed for projects across campus/community. Typically, they may involve data sharing across different disciplines and/or institutions in the community. The goal will be to provide students a chance to analyze real life data strengthen their understanding.

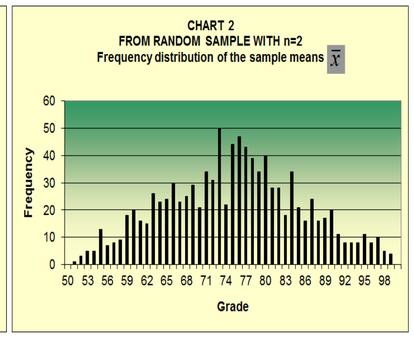
MS-Excel Simulations

Motivation and Assignment:
Teaching sampling distributions can be challenging because students have difficulty visualizing the fact that they are only selecting one sample out of many potential samples that they could have selected. Mills' work specifically discussed how simulations help to clarify statistical theorems, and this MS Excel assignment requires students to help create the simulation process. Student created simulations allow the learner to visually examine the implications of the Central Limit Theorem, while enhancing their MS Excel skills in the process.

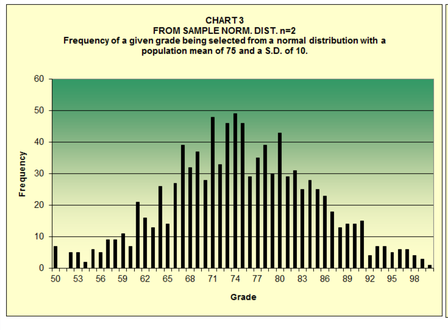
- Students are provided an excel sheet that allows them to enter random sample data with the mean and standard deviation provided. Students generate 1000 samples and calculate the mean for each of these samples. Charts are automatically created as they compute the sample means.
- Students do this procedure for a uniform distribution with n=2, a normal distribution with n=2, and a normal distribution with n=5.
- Students then answer questions about the graphs and data. Upon completion of the exercise, students generally have a better understanding of the Central Limit Theorem and the interpretation of a standard error.

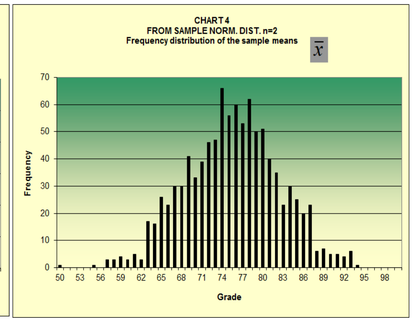
Randomly Selected Data Following a Uniform Probability Distribution





Randomly Selected Data Following a Normal Probability Distribution





Hands-on Projects

Backdrop and the Nature of Assignment:
Finding an optimal amount of intrinsic and external motivation for student engagement is often a quest for any teacher. In the spirit of Edward Deci's work, this assignment tries to blend autonomy and freedom to provide a unique experience for students by means of a hands-on project. While most of t he college level statistics courses begin with description of data collection processes, there is not enough time to demonstrate the idea that *there is no such thing as free data*. This assignment offers an alternative to work with real data in the hope of making it both instructive and rewarding for students. Students can relate more easily to the data collected and insights gained through this process.

Timeline of the Project:

The following timeline represents the course of a hands-on project over a typical 16-week class. It could be modified for shorter/longer classes as needed :

Weeks 1 - 6

- Students sort themselves into groups of two/three members each and pick a topic of their interest
- The enquiries are expected to be simple, everyday yet interesting for students. For instance,
  - Eating habits of UCO students
  - Workout habits of UCO students
  - Sleeping habits of UCO students
- After finding a topic of interest, students create a brief *questionnaire* built around the theme of their topic.
- The number of questions range between 10 – 15
- Questionnaire needs to be approved by the instructor.

Weeks 7 - 11

- Once approved, the questionnaire is circulated around the class to get data.
- *Quid-pro-quo* applies as students help each other in answering questionnaires.
- Students are encouraged to use software such as *Survey Monkey* or *Qualtrics* (if accessible for free of charge)
- At this point, most of them realize the delays/other difficulties encountered in acquiring data in a real-world setting
- Students are advised to collect *at least* 30 observations to help the overall validity of results
- After gathering the data, the students report it in an Excel spreadsheet

Weeks 12 - 16

- Once all the data are in, categorical variables are coded in, if required.
- At this point, students get to see the raw data before they begin analyzing it.
- Students are asked to *interpret* the data by applying appropriate statistical tools
- Depending on the nature of variables, basic descriptive statistics such as mean/median/mode are computed
- Graphical exposition of the data using histograms, bar charts, pie charts etc. is employed to make *'sense'* out of data
- Scatter plots may be used if students expect some *a-priori* relationship between the variables of interest.
- Finally, students make a brief presentation to the class incorporating the details of their findings
- Each presentation is required to have the following three components:
  - Introduction/Backdrop: Outlining motivation/idea behind the topic
  - Questionnaire/dataset: Describing the details about questions and responses
  - Results/Analysis: A brief summary of main results and the group's take on it
- A final report based on the presentation is turned in bearing the group member's names which concludes the activity.

Select References and Acknowledgments:

- Carnell, Lisa J., "The Effect of a Student-Designed Data Collection Project on Attitudes toward Statistics", *Journal of Statistics Education*,16.1, 2008.
- Deci, Edward L. and Ryan, Richard M., "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions", *Contemporary Educational Psychology*, 25, 54-67, 2000.
- Dinov, Ivo D., Christou, Nicolas, and Sanchez, Juana, "Central Limit Theorem: New SOCR Applet and Demonstration Activity", *Journal of Statistics Education*, 16.2, 2008.
- Doane, David P., "Using Simulation to Teach Distributions", *Journal of Statistics Education*, 12.1, 2004.
- Mills, Jamie D., "Using Computer Simulation Methods to Teach Statistics: A Review of the Literature", *Journal of Statistics Education*,10.1, 2002.

- The authors would like to thank Dr. Jeff King at the Center for Excellence in Transformative Teaching and Learning (CETTL) at the University of Central Oklahoma for his helpful comments.