

Investigations for Introducing Mathematically Inclined Students to Statistics

Allan Rossman and Beth Chance, Cal Poly - San Luis Obispo

Motivation: Key features of the statistics education reform movement include the use of data, activities, and technology to help students understand fundamental concepts and the nature of statistical thinking. Many innovative and effective materials have been developed, but the vast majority have been aimed at “Stat 101,” the algebra-based service course. Our goal in this project is to support an introductory course at the post-calculus level around the best features of statistics education reform: data, activities, concepts, and technology. We emphasize issues of data collection and data analysis as well as statistical inference. We ask students to investigate the applications of statistics and also its mathematical underpinnings.

Guiding Principles:

- *Experience the entire statistical process over and over.* From the first chapter we ask students to consider issues of data collection, produce graphical and numerical summaries, consider whether inference procedures apply to the situation, apply inference procedures when appropriate, and communicate their findings in the context of the original research question. This pattern is then repeated over and over as students encounter new situations. We hope that this frequent repetition helps students to appreciate the “big picture” of the statistical process, and to develop their confidence for “doing statistics.”
- *Motivate with real studies and genuine data.* Almost all of these investigations center on genuine data from real studies. The contexts come from a variety of scientific disciplines and also from popular media. Some of these studies include classic studies on smoking and lung cancer, as well as recent cases of pulmonary diseases in popcorn manufacturing plants, relationship between night lights and myopia in children, use of shared armrests on airplanes, and lasting effects of sleep deprivation on cognitive ability.
- *Put students in the role of active investigator.* Our materials emphasize active learning and ask students to construct their own knowledge of statistical concepts and methods. They consist primarily of investigations that present a directed series of questions that lead students to explore and apply statistical ideas, often asking them to make conjectures before carrying out the analysis about a statistical concept or a research question.
- *Emphasize connections among study design, inference technique, and scope of conclusion.* Issues of study design come up early and recur throughout our materials. From the opening chapter we lead students to examine when causal conclusions can be drawn, the distinction between randomization and random sampling, and the concept of statistical significance, motivated first through simulation and then through probability models.
- *Use variety of computational tools.* As appropriate, students are asked to use different computational tools, both to analyze data and to explore statistical concepts. Many of the concepts are visualized through interactive java applets and much of the statistical analysis is done with Minitab.
- *Investigate mathematical underpinnings.* We often ask students to use their mathematical training to investigate some of the underpinnings behind statistical procedures (e.g., principles of least squares, effects of transformations).
- *Provide frequent feedback.* Through a series of practice problems and an extensive collection of exercises, students can be asked follow-up questions to immediately test their

understanding of basic terminology and applications, as well as more involved data analyses and mathematical derivations. Instructors can utilize the practice problems between classes to informally assess which topics merit more discussion.

Outline:

	Chapter 1	Chapter 2	Chapter 3	Chapter 4	Chapter 5	Chapter 6
Data Collection	Observation vs. experiment, confounding, randomization		Random sampling, bias, precision, nonsampling errors	Paired data	Independent random samples	Bivariate
Descriptive Statistics	Conditional proportions, segmented bar graphs, odds ratio	Quantitative summaries, transformations, z-scores, resistance	Bar graph	Models, Probability plots, trimmed mean		Scatterplots, correlation, simple linear regression
Probability	counting, random variable, expected value	empirical rule	Bermoulli processes, rules for variances, expected value	Normal , Central Limit Theorem		
Sampling/ Randomization Distribution	Randomization distribution of $\hat{p}_1 - \hat{p}_2$	Randomization distribution of $\bar{x}_1 - \bar{x}_2$	Sampling distribution of X, \hat{p}	Large sample sampling distributions for \bar{x}, \hat{p}	Sampling distribution of $\hat{p}_1 - \hat{p}_2, \bar{x}_1 - \bar{x}_2$, OR	Chi-square statistic, F statistic, regression coefficients
Model	Hypergeometric		Binomial	Normal, t	Normal, t , log-normal	Chi-square, F, t
Statistical Inference	p-value, significance, Fisher's Exact Test	p-value, significance, effect of variability	Binomial tests and intervals, two-sided p-values, type I/II errors	z-procedures for proportions t -procedures, robustness, bootstrapping	Two-sample z - and t -procedures, bootstrap, CI for OR	Chi-square for homogeneity, independence, ANOVA, regression

Faculty Development Workshop: As part of the MAA's PREP program, we will offer a professional development workshop on these materials in San Luis Obispo on July 18-22. See www.rossmanchance.com/prep/workshop.html for more information and to apply on-line.

Acknowledgements: We thank the National Science Foundation (DUE/CCLI #9950476, 0321973) for supporting the development of these materials, and Duxbury Press, publishers of *Investigating Statistical Concepts, Applications, and Methods*, by B. Chance and A. Rossman. The preliminary version is available now and the first edition will be available this summer. Additional information and resources are available at: www.rossmanchance.com/iscam/.