Teaching Statistics---NOT Probability

Frank H. Dietrich II, Northern Kentucky University

James T. Snodgrass III, Xavier University

INTRODUCTION:

We will compare two ways to present an introductory statistics course, where we define an introductory statistics course to be one for which the only prerequisite is an understanding of high school algebra. There are three broad content areas that generally appear in all such courses: probability, descriptive statistics, and inferential statistics. The difference from course to course is how much time is spent on each of these content areas, and the specific topics that are treated. Both the time spent and the topics can vary greatly from course to course. In this presentation we specifically address the coverage of probability included in an introductory statistics course. We begin by describing two general types of these courses. The first is a course that includes enough coverage of probability topics to merit the title: *An Introduction to Probability and Statistics*. An alternative is a course that has a minimal coverage of probability; we would call this course: *An Introduction to Statistics*. We will conclude with our rationale as to when and why the second course is preferable to the first.

COURSE I: AN INTRODUCTION TO PROBABILITY AND STATISTICS

Courses such as this typically include the following topics from discrete probability: Experiments; sample spaces; events; unions, intersections, and complements of events; mutually exclusive events; conditional probability; independence of events; and the multiplicative and additive rules for finding probabilities. Counting rules with permutations and combinations may also be included, and discrete probability distributions are often covered. Coverage of discrete probability distributions usually includes the binomial distribution, and may include others such as the hypergeometric and Poisson distributions. In addition to these discrete probability topics, most such courses include extensive discussion of the normal distribution, followed by coverage of sampling distributions for statistics like the sample mean and the sample proportion. It is easy to see how instructors might spend as much as, or perhaps even more than, half of the course on these topics.

The descriptive statistics segment of this course will cover both numerical and graphical descriptive methods. The specific methods covered depend on the instructor, but include at minimum the mean, median, variance, standard deviation, and histograms. Some instructors like to discuss the mode, measures of relative standing, bar or pie charts, box-and-whisker plots, stem-and-leaf plots, scatterplots, and correlation.

Finally some basic inferential techniques are discussed. Generally, tests of hypothesis and confidence intervals are both introduced. Depending on the amount of time devoted to probability and descriptive statistics, the inferential statistics segment of the course may be limited to inferences about a population mean, and perhaps a population proportion, using the normal distribution. If time is available, inferences about a population mean using Student's t-distribution may also be covered. There is usually not enough time to consider two sample procedures, or other more advanced statistical methods.

COURSE II: AN INTRODUCTION TO STATISTICS COURSE

This course covers none of the discrete probability topics mentioned in the description of the previous course, beginning instead with a discussion of descriptive statistical methods. As with the first course, exactly which topics are included is at the discretion of the instructor. The normal distribution is then introduced as a way to describe a population with a very specific shape. The notion of a probability arises by considering the percentage or fraction of measurements under some portion of a normal curve.

Sampling distributions are introduced next in the context of describing possible values of sample quantities, namely statistics. Certainly, the sampling distribution of the sample mean is covered, and a discussion of the sampling distribution of the sample proportion would also be an appropriate topic.

Unless descriptive statistics is covered in great detail, it generally takes about 6 weeks of a 15-week semester (40% of the course) to cover this material. This means that the majority of the semester can be devoted to inferential statistics.

We consider the following inferential topics to be essential: Inferences[1] about a population mean using both the z-distribution and the t-distribution, inferences about a population proportion using the z-distribution, and inferences to compare two population means using the t-distribution. The rest of the course can include whatever statistical inferential procedures the instructor finds most valuable or appropriate. Possible candidates include: Inferences to compare two proportions, analysis of variance, chi-squared tests for contingency tables, regression and correlation, and nonparametric tests. It is our experience that there is time to discuss most (all but one in fact!) of these topics.

WHICH COURSE SHOULD I TEACH?

The distinction between these two courses is obvious. The first devotes a significant amount of time to discrete probability, while the second includes none of this material and instead covers many more inferential statistical methods. We believe that you should choose the type of course that best meets your educational or curricular objectives. If you want to introduce your students to probability, descriptive statistics, and inferential statistics, then you should teach the first type of course. However, if your primary goal is to help your students understand the science of statistics, then we think you should teach the second course.

Having taught both types of courses over several years, we have reached the following conclusion: Students do not need to study discrete probability in order to understand inferential statistics. Statistical reasoning hinges on understanding when an observation is very likely or very unlikely, and we find that most students already have an intuitive understanding of these ideas. In fact, for a student taking a single introductory statistics course, time spent on discrete probability only detracts from the student's study of the science of statistics.

If we want our students to have a meaningful understanding of inferential statistics, then they must spend more than three or four weeks studying the topic. In the first type of course we think there are insufficient opportunities to discuss the concepts that are essential to understanding the inference-making process: Type I and Type II errors, the meaning of confidence, p-values, and the rationale behind the testing of hypotheses. It has been our experience that even the brightest students do not really understand these

[1] The word "*inferences*" in this context is meant to include both tests of hypothesis and confidence intervals.

concepts without repeatedly discussing them in multiple settings. By studying a large variety of statistical techniques, students can continually revisit all of the essential inferential concepts. We believe that this repetition leads to real understanding.

CONCLUSION:

There are many different course designs for an introductory statistics course. We strongly recommend that instructors carefully and thoughtfully consider their objectives for such a course. What do you want your students to learn? If the answer is inferential statistics, then you do not need to teach discrete probability. In this case, we strongly recommend that you try teaching *An Introduction to Statistics Course*, rather than *An Introduction to Probability and Statistics*. By eliminating the discrete probability coverage, there is time for students to gain a deeper understanding of statistical methods. In fact we like to think of this course as a course in applied statistics, which is accessible to even students with a modest mathematical background. *An Introduction to Statistics* gives students the tools to succeed in discipline-specific research courses, promotes critical thinking skills for liberal arts or general education students, and prepares students for responsible citizenship in a data-rich culture.

We would be happy to discuss these ideas with our colleagues. Feel free to contact us via e-mail at:

dietrichf@nku.edu
snodgrassjt@xavier.edu