

Integrating Statistics into College Algebra to Meet the Needs of Biology Students[†]

Sheldon P. Gordon¹, Department of Mathematics, Farmingdale State College
Florence S. Gordon, Department of Mathematics, New York Institute of Technology

Size	6500
Institution Type	Medium-sized state college with focus on technology
Student Demographic	Diverse student body with many first-generation and minority students
Department Structure	Mathematics and Biology are individual departments in the College of Arts and Sciences

ABSTRACT

Most of the mathematics used by students in introductory biology and other laboratory science courses arises in the laboratory when students are faced with experimental data to analyze and to use as the basis for answering predictive questions. In this article, we build a case for integrating, in natural ways, significant amounts of statistical ideas and methods into courses in college algebra and precalculus to provide students with the mathematical skills they use in biology and other science courses. We provide examples to illustrate how statistics can be incorporated in ways that are natural extensions and applications of mathematical topics that are usually included in mathematics courses.

COURSE STRUCTURE

- Weeks per term: 15
- Classes per week/type/length: two 75-minute and one 50-minute class each week
- Labs per week/length: none
- Average class size: 30 students per section
- Enrollment requirements: For high school students and freshmen, though many postpone taking the mathematics course as long as possible.
- Faculty/dept per class, TAs: Taught by one mathematics instructor (usually a full-time instructor).
- Next course: The purpose of the course is to teach the mathematical ideas and methods needed by students who plan to major in biology or who want a mathematical approach that reflects the needs of biology and other laboratory sciences. It is also taken by some business majors. The biology majors will subsequently take at least one semester of calculus and one semester of statistics.
- Website: <http://www.farmingdale.edu/faculty/sheldon-gordon/>

Introduction

[†] supported by NSF grants DUE-0089400, DUE-0310123, and DUE-0442160

¹ gordonsp@farmingdale.edu

Each year, some one million students (Lutzer et al. 2007) take college algebra and related courses. Most of the courses were designed to prepare students for the mainstream calculus sequence by focusing almost entirely on developing the algebraic skills needed for calculus.

However, only about 10% of those who complete college algebra successfully ever go on to take Calculus I and virtually none have gone on to take Calculus III (Dunbar 2006). This suggests that offering college algebra courses to prepare students for calculus is doing a disservice to the vast majority of the students.

The challenge we face is to change the focus in these courses to make them meaningful to the departments, especially the biology department, that require them, so as to better meet the needs of the students and the disciplines that are our clients and partners.

Why do so many students take the courses?

The overwhelming majority of students take college algebra and related courses to fulfill general education requirements (particularly at large state universities and state colleges) or to fulfill the mathematics requirements of other departments. One of the partner disciplines that sends us a large proportion of our students at this level is biology. But, according to leading educators from biology and from other partner disciplines (Ganter and Barker 2004), their students do not need or ever use the algebraic manipulation that is a hallmark of most of these courses. Corporate and government leaders say we need a quantitatively literate workforce and public to compete in today's technological world and function effectively as citizens (Steen and Madison 2003). The primary mathematics needed – by the other disciplines, for today's workplace, and for effective citizenship – is more knowledge of statistics, not computational skill in algebra.

In many states, a course in college algebra is the primary mathematics requirement for students planning to become elementary and middle school teachers. The traditional college algebra course with its predominant focus on algebraic skills is not particularly appropriate for these students, especially if they will be teaching in schools that use NCTM Standards-based curricula.

Across the country, traditional college algebra courses have been identified as the primary barrier preventing students from gaining access to careers in quantitative fields. For instance, the Economic Development Council of San Antonio has identified the college algebra courses offered in the city's colleges as an impediment to the city not having the quantitatively-

trained workforce needed for the high-tech economy that they envision as the future of San Antonio. In response, the mayor has appointed a taskforce, including representatives of all eleven public and private colleges in San Antonio and representatives of business, industry, and government to consider the problem and change the nature of the college algebra experience.

Most students who take college algebra are poorly served by the traditional courses. There is a need to change the focus to better serve the students who take them, to meet the increasingly quantitative needs of most other disciplines, and to meet the needs for a well-educated workforce and citizenry that can function effectively in a quantitative environment. In all of these instances, college algebra has little to do with moving on to calculus.

These issues are discussed in considerably greater detail in (Gordon 2009) and in the articles in the volume edited by Hastings (Hastings 2006).

What do the students need from these courses?

What do other disciplines, particularly the biological sciences, want their students to bring to their courses from the mathematics courses they require? The MAA's committee on Curriculum Renewal Across the First Two Years (CRAFTY) recently brought together leading educators from seventeen quantitative disciplines in workshops to discuss the current mathematical needs of their disciplines and to make recommendations about what should be in the mathematics offerings to satisfy their needs. The seventeen discipline reports and a summary of all Curriculum Foundations project recommendations appear in (Ganter and Barker 2004); a discussion of their implications for courses at the college algebra and precalculus level appears in (Gordon 2009). The reports provided background for the recommendations on the undergraduate mathematics curriculum in the MAA's *Curriculum Guidelines 2004* (Lutzer et al. 2004) developed by CUPM (the Committee on the Undergraduate Program in Mathematics).

For almost all of the disciplines involved in the project, the focus was on courses below calculus, particularly college algebra. As mentioned, virtually all the disciplines see a need for a different focus in the mathematical training of their students, one that stresses conceptual understanding, mathematical modeling and problem solving, and a heavy emphasis on statistical reasoning and interpretation of real-world data. Among the points that the biologists made were

1. The collection and analysis of data that is central to biological investigations inevitably leads to the use of mathematics.

2. Mathematics provides a language for the development and expression of biological concepts and theories. It allows biologists to summarize data, to describe it in logical terms, to draw inferences, and to make predictions.
3. Statistics, modeling, and graphical representation should take priority over calculus.

The biologists said:

Biology students need to understand the meaning and use of variables, parameters, functions, and relations. They need to know how to formulate linear, exponential, and logarithmic functions from data or from general principles. They must also understand the basic periodic nature of the sine and cosine functions. It is fundamentally important that students are familiar with the graphical representation of data in a variety of formats (histograms, scatterplots, pie charts, log-log and semi-log graphs).

Perhaps most telling is the biologists' comment that "The current mathematics curriculum for biology majors does not provide biology students with appropriate quantitative skills".

One of the themes in virtually every discipline report in the Curriculum Foundations document is the need for more statistical training. In discussions with faculty in biology and the other laboratory sciences (such as chemistry and earth and space science), it has become clear that the mathematical limitations of their students appear most dramatically in the laboratory when they are asked to analyze and interpret experimental data. In most of the courses in these fields, little if any mathematics arises in the classroom; it is in the labs that students need to apply mathematics and that mathematics is almost always statistical. Typically, they need to find trends (usually linear) in a data set and to answer predictive questions (that is, solve the resulting equation). These are the primary connections to topics in college algebra courses. Many of the social sciences, especially business, focus on using data to produce models that can be used to answer predictive questions. Their students also need more exposure to statistics.

At many schools, college algebra is the mathematical prerequisite for the first course in statistics. The introductory statistics course is already crammed with too many topics, and it is usually not possible to cover everything that students should know. One solution is to require a second statistics course, but crowded curricula make that an unrealistic option. A better solution is to provide an introduction to standard statistical topics in the prerequisite course, so that the full treatment in statistics courses can go more quickly and smoothly.

Repeating the statistics topics may seem to be a waste of time. However, it is not. Our students see the equation of a line in pre-algebra classes, again in elementary algebra, in intermediate algebra, in college algebra, and in precalculus. Yet, sadly, many students still seem not to have fully mastered the concepts or the ability to find the equation of a line in Calculus I, despite the repetition. And this is one of the simplest things that they see in their mathematical training. The concepts and methods of an introductory statistics course are less intuitive and are much broader in variety. Despite this, we expect the students who take a one semester introductory statistics course to understand and be able to apply them based on a single exposure in one semester. Also the students are mathematically weaker than the ones who go through the traditional mathematics curriculum toward calculus. So, it is unrealistic to assume that one exposure to the ideas and methods of statistics is sufficient. Students need to see statistical ideas repeatedly, just as calculus-bound students need to see many techniques and ideas repeatedly.

Accordingly, we feel that there are compelling reasons to try to integrate a substantial amount of statistical reasoning and methods into college algebra and related courses.

Statistical Analysis and Reasoning

The challenge we face is finding ways in which statistical ideas and data analysis can be integrated into a college algebra course so as to support and reinforce the concepts and methods of college algebra. Data analysis, in the sense of fitting functions to data, has become a common topic in most textbooks as a way in which interesting and realistic applications of families of functions (linear, exponential, power, logarithmic, polynomial, and even sinusoidal) covered in the course can be applied. The extent to which this material is actually used by instructors is, unfortunately, uncertain.

Several reform college algebra texts include a chapter that looks at some simple statistical ideas such as finding the center and spread in a data set and displaying data graphically. However, the texts have been written so that the course will satisfy quantitative literacy requirements for students who will take no more mathematics. They do not meet the needs of students in biology and the other laboratory sciences, nor do they provide a broad introduction to statistical ideas. Also, the statistical ideas arise only in the free-standing chapter, so they are neither extensive nor integrated into the entire course.

How then could we incorporate statistical ideas and methods in a natural way throughout an entire college algebra course? The question is complicated by the wide variety of audiences for the course, including those who have not seen statistics previously, and who would be best served by a good introduction, and those who have previously taken a statistics course, and who would be best served by seeing many of the same ideas in a new and more mathematical context.

Many students in college algebra courses lack a sense of what mathematical notation is all about—the symbols that stand for variables, the types of variables that arise in connecting mathematics to the real world where everything is not mindlessly x and y , and the notion of the scales in reading and interpreting (let alone creating) graphs and charts. All the concepts can be reinforced by looking at real-world data and creating tables and graphs. This gives a wonderful opportunity to stress the difference between the dependent and the independent variables and the practical meaning of the domain and range of functions. These are some of the key notions that biologists called for in their report in the Curriculum Foundations project (Ganter and Barker 2004).

In introducing different types of behavior for functions (increasing versus decreasing, concave up versus concave down, turning points, and inflection points), we can look at a normal distribution as an example. It provides an effective way of reinforcing the notion of the mean (the center of a data set) and the standard deviation (the spread in a data set). Subsequently, the idea of the z -value associated with a measurement x can be introduced as nothing more than a linear function relating the variables.

In discussing the regression line to fit a data set, particularly if it is laboratory data, it is natural to point out that there could be many different sets of data for the same experiment, each leading to a different regression line. A computer graphics simulation provides visual support to make the different lines come to life and to investigate the effects of different sample sizes. We have developed an effective version of such a simulation in Excel that is available to any interested reader (Gordon cited 2012). We illustrate the possible results in Figures 1 and 2, which show, respectively, the outcomes associated with samples of size $n = 4$ and $n = 20$. In the first case, we see that many of the sample regression lines have slopes that vary dramatically from that of the population's regression line (the heavy line in the figure); in the second case, almost all of the sample regression lines have slopes that are close to that of the population regression line.

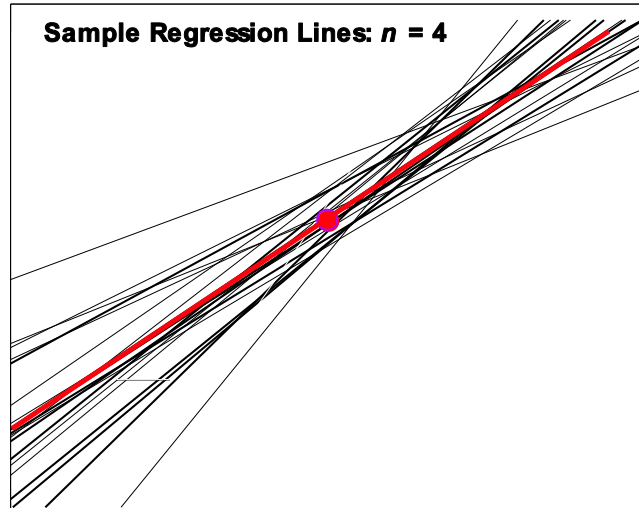


Figure 1: Sample size $n = 4$.

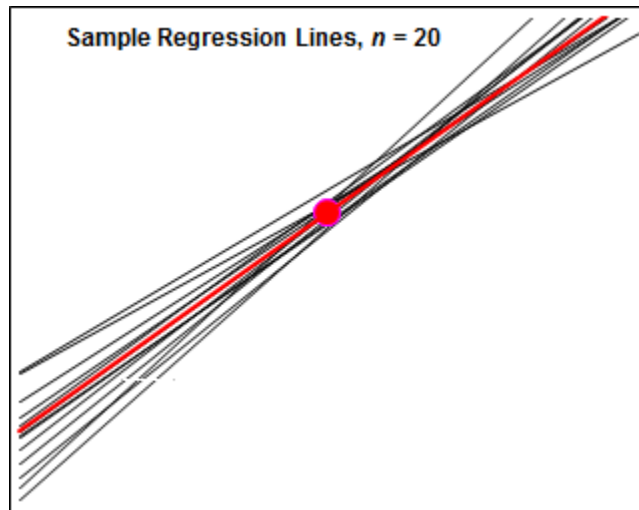


Figure 2: Sample size $n = 20$.

Such an investigation provides a wonderful opportunity to stress two of the key themes in statistics—the effect of sample size on the outcome and the variation that occurs within a sample and between different samples. These notions are critical for anyone who will be working with laboratory data.

Later stage in the course, in discussing shifting and stretching of functions, we can return to the normal distribution function

$$N(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2 / 2\sigma^2}$$

that is centered at the mean μ and has standard deviation σ . We can emphasize the fact that the standard normal distribution curve has been stretched or squeezed horizontally by the effect of σ^2

as a divisor in the exponent, shifted horizontally by an amount equal to μ , because of the presence of the $(x - \mu)$ term, and stretched or squeezed vertically by the effect of σ as a multiple in the denominator of the coefficient.

We can also look at the normal distribution function as an example of a composite function. Unlike the artificial functions we typically use to illustrate the idea of a function of a function, the normal distribution provides a meaningful example that the students will see again. This makes the concepts more meaningful to the students and provides motivation that is often not present in the standard treatment of composite functions.

Subsequently, we can introduce the notion of the distribution of sample means. It consists of the means of all possible samples of a size n drawn from an underlying population having mean μ and standard deviation σ . The Central Limit Theorem, which is probably the single most important result in inferential statistics, provides information on the characteristics of the population of sample means. It tell us

1. The mean of the distribution is $\mu_{\bar{x}} = \mu$.
2. The standard deviation of the distribution is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

3. If the underlying population is roughly normally distributed, then the sample means are also normally distributed.
4. If the sample size n is sufficiently large, then the distribution of sample means is roughly normally distributed whatever the underlying population is. Typically, samples of size $n > 30$ are sufficiently large to assure approximate normality.

The distribution of sample means provides a wonderful opportunity to reinforce the notions of shifting and stretching functions. A computer graphics simulation that draws repeated random samples and displays their means in a histogram can provide visual and numerical support. We would start with the case of large samples means, so that the distribution of sample means will be roughly normal. The distribution of sample means is centered at the mean of the underlying population, μ , so the center of the histogram representing the sample means is typically close to the center of the underlying population, and the numerical value for the mean of all the sample means typically comes out close to the population mean μ . Therefore,

we have the same horizontal shift in the distribution of sample means as in the underlying population.

The spread in the distribution of sample means clearly depends on the sample size n . Using the simulation with large numbers of samples of size $n = 4, 9, 16$, and 25 , it quickly becomes clear to the students, both visually from the graphs and numerically from the calculated values of the mean of the standard deviations of the samples, that the spread is, respectively, about $1/2$, $1/3$, $1/4$, and $1/5$ of σ . Since we are dealing with fractional multiples of σ , the distribution of sample means is clearly equivalent to a horizontal squeeze of a normally distributed population and the formula that quickly materializes is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Once the Central Limit Theorem has been introduced, it is a simple matter to develop the idea of a confidence interval for the mean of an unknown population and tie it to the standard problem that arises in many laboratory experiments where the student is expected to estimate the mean of a population based on a single experimental sample.

In discussions with some biology faculty, it has become clear that there is also a need to introduce hypothesis testing. It is a fundamental statistical tool, particularly in an introductory biology course that addresses genetics. We have not been able to devise a natural way to integrate that into the flow of topics that would normally be considered part of a course at the college algebra level. The alternatives appear to be either to incorporate hypothesis testing as part of an additional unit on probability (and there simply does not appear to be adequate time in a four-hour course to do that) or to introduce it on its own without it being a natural application of a mathematical topic being addressed.

Classroom Experiences

To implement the ideas outlined above, we have developed a full set of classroom materials, including text materials and accompanying Excel spreadsheets. Much of it has been tried in a new four-credit, four-hour per week semester-long precalculus/college algebra course designed for bioscience and business students. Not all of the statistical ideas and methods have been tried yet; in particular, the ideas relating to the Central Limit Theorem and confidence intervals have not been covered, but most of the others have been implemented.

The fact that the course is intended for both the bioscience students and business students make for some interesting compromises, but also provide the opportunity to see how different kinds of students reacted to the materials. For the bioscience students, the course serves as a precalculus experience. For most of the business students, the course is a terminal one, although some of the specific majors within our business school require a subsequent business calculus course. While we view the course as something that the students in both programs should take early in their academic careers, both groups tended to put it off as long as possible. So, most of the business students previously had taken a required introductory statistics course offered through the business school, while virtually none of the bioscience students had had any formal experience with statistics, other than what they may have done in lab reports in some of their biology courses.

With both groups, the reaction to including the statistical ideas was positive. Rather than being bored by the repetition, the business students seemed happy with the reminders about what they had seen in their first statistics course and it seemed to help them make new connections between the mathematical concepts and the statistical ones. The bioscience students seemed to find that the statistical ideas made the course more valuable because they had direct connection to what they had been doing in their lab courses. This was facilitated by the gentle introduction of statistical ideas with little formal emphasis on the computation; that was left for a subsequent statistics course, which we hope the students might consider taking.

Although the statistical ideas were not tested on exams, the students generally did better on test problems on topics that were enhanced by some of the statistical ideas than students in comparable courses have done in recent years. For instance, the students had no trouble with problems on shifting and stretching functions, which was reflected in the examples on the normal distribution curve; in previous years, this kind of problem invariably elicited numerous “solutions” that displayed absolutely no idea of what was going on, while this year, virtually every student solved them easily.

Some Concluding Thoughts

At most schools, college algebra is the prerequisite for introductory statistics, though virtually none of the ideas and algebraic techniques in traditional college algebra courses are relevant to introductory statistics. Introductory statistics courses are typically much too crowded

and rushed and many students are poorly served by them. If a significant amount of statistical thinking can appear in the prerequisite college algebra course, as we have done, we believe that it can make the subsequent statistics course more reasonable and more successful for many of the students.

Integrating statistics into college algebra and other math offerings will become increasingly imperative as the Common Core in Mathematics program begins to be implemented. Already approved by at least 45 states, this curriculum calls for introducing statistical ideas starting in sixth grade and anticipates that *every* high school graduate will have the equivalent of a solid introductory statistics course. Statistics will be a thread throughout the curriculum but will not be integrated into the other mathematics topics. Students coming out of the Common Core will need reminders about statistical ideas, as discussed here. Eventually, the introductory statistics course as we know it today will become a new developmental offering, while those students who really need statistics in college will start with what we currently consider a second course.

References

- Dunbar, S., 2006: Enrollment Flow to and from Courses below Calculus, in *A Fresh Start for Collegiate Mathematics: Rethinking the Courses below Calculus*, N. B. Hastings, Ed., MAA Notes, Mathematical Association of America.
- Ganter, S. and W. Barker, editors, 2004: *The Curriculum Foundations Project: Voices of the Partner Disciplines*, Mathematical Association of America.
- Gordon, S. P., 2009: What's Wrong with College Algebra?, *PRIMUS*, **18**, 516-541.
- Gordon, S. P., cited in 2012: DIGMath Software Packages [Available online at <http://www.farmingdale.edu/faculty/sheldon-gordon/>].
- Hastings, N. B. Ed., 2006: *A Fresh Start for Collegiate Mathematics: Rethinking the Courses below Calculus*, MAA Notes #69, Mathematical Association of America.
- Lutzer, D. J., S. B. Rodi, E. E. Kirkman, and J.W. Maxwell, 2007: *CBMS 2005: Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States*, American Mathematical Society.
- Pollatsek, H., W. Barker, D. Bressoud, S. Epp, S. Ganter, and B. Haver, 2004: *Undergraduate Programs and Courses in the Mathematical Sciences: CUPM Curriculum Guide 2004*, Mathematical Association of America.
- Steen, L. and B. Madison, 2003: *Quantitative Literacy: Why Numeracy Matters for Schools and Colleges*, Woodrow Wilson National Foundation Press, Princeton, NJ.

