# The First Year of Calculus and Statistics at Macalester College

**Dan Flath**[1]
**Tom Halverson**[2]
**Danny Kaplan**[3]
**Karen Saxe**[4]

**Department of Mathematics, Statistics, and Computer Science, Macalester College**

| | |
|---|---|
| Name of Institution | Macalester College |
| Size | about 2000 students |
| Institution Type | selective 4-year undergraduate liberal arts college |
| Student Demographic | first semester calculus students; required course sequence for all math, biology, economics students; often taken by others fulfilling distribution requirement. |
| Department Structure | Mathematics and Statistics are housed in same department, together with Computer Science and Applied Mathematics. |

## Abstract

At Macalester College, we have redesigned our introductory calculus course to make a more useful mathematics sequence for life science students, a change that also works better for students from the physical sciences, economics, and other social sciences. This redesigned curriculum consists of a two semester course sequence: Applied Calculus (AC) and Introduction to Statistical Modeling (ISM).

## COURSE STRUCTURE

- Weeks per term: 14-week semester
- Classes per week/type/length: three 1-hour lecture periods each week
- Labs per week/length: none required, but AC meets frequently in lab; ISM meets every day in lab
- Average class size: AC has 30-40 students in 5-6 sections per year; ISM has 30 in 3-4 sections per year

[1]flath@macalester.edu
[2]halverson@macalester.edu
[3]kaplan@macalester.edu
[4]saxe@macalester.edu

- Enrollment requirements: AC – Macalester admission; ISM – either AC, or Multivariable Calculus, or Linear Algebra
- Faculty/dept per class, TAs: One department faculty member per section, plus one undergraduate assigned to help with grading, and to help in lab, as instructor sees fit
- Next course: After AC either ISM, or second semester calculus; After ISM, Applied Multivariate Statistics
- Websites:

  http://www.causeweb.org/wiki/mosaic/index.php/Macalester_Math_135

  http://www.causeweb.org/wiki/mosaic/index.php/Macalester_Stat_155

## Introduction

Macalester College is a selective liberal arts college in Saint Paul, MN with about 2000 students. The median SAT scores for current students are 670 in mathematics and 700 in critical reading; the ACT composite median is 31. A vast majority of the students who take our introductory courses fall into one of two categories: (1) those who take quite a bit of mathematics, including Linear Algebra and Multivariable Calculus, and (2) those who are interested in another discipline and are required (or limited by time) to take two mathematics courses. Our former Calculus I was designed as the first semester of a 2 or 3 course calculus sequence, and we found that it was no longer meeting the needs of the type (2) students and that very few of the type (1) students were enrolling in it (one or two per year), since they placed into Calculus II or higher. This motivated us to change the role of Calculus I and to think of it as part of a two-course sequence in Calculus and Statistics. During this same time (November 2000), our department was host to a Curriculum Foundations Workshop in Biology (Dilts and Salem, 2004) and Chemistry (Craig, 2004), where members of these partner disciplines gathered to discuss the ways that the mathematics curriculum supports their students. The themes that arose from these workshops — especially multivariable relationships, scale and estimation, modeling, and data analysis — were the foundation of our vision for our new courses. Note that Applied Calculus (AC) is our first calculus course; we do not offer precalculus courses. Mathematics majors typically enter Macalester with calculus credit; since they all take Introduction to Statistical Modeling (ISM) and some multivariable background is needed for this course, most take ISM following their Multivariable Calculus or Linear Algebra course. Students from client departments for whom the AC/ISM sequence is required can instead take Multivariable Calculus and ISM (and many economics students, for example, do exactly this).

## The first semester – Applied Calculus

Applied Calculus (AC) is a one semester course on mathematics for modeling. It is distinctive for the range of topics covered, the active approach taken in the classroom, and the significant computer component using the software R[5]. This is professional-level software and is used for the entire year in AC and its follow-up course (discussed below). Algebraic techniques are present but deemphasized; there is no calculus prerequisite. Most students go from this course to the statistics class designed to follow it and described in the next section of this paper.

The content of the course is broken into four broad sections. The first section (roughly 9 classes) focuses on families of functions and modeling. Single and multivariable phenomena are discussed. The emphasis is on building intuition about basic families of functions of one variable — linear, exponential, and sinusoidal — and how they are used in modeling. Later in the course, the logistic family is added. The graphical significance of parameters is explored through curve fitting, in the case of the exponential function with the aid of semi-log graphs. After an initial computer lab on defining and graphing functions with R, students model real experimental data by a graphical curve fit – exponential decay is represented by automatically recorded heights of successive bounces of a ball, and a sinusoidal model is fit to a digital recording of a human whistle. Along the way, the importance of keeping track of units of measurement is emphasized. Functions of two and more variables are introduced with a Cobb-Douglas model and the connection between rectangular tables of values and their contour diagram representations are made with a fun class activity to create a contour diagram from a large table of values by drawing the diagram right on the table.

The second section (roughly 14 classes) is on differentiation of functions of one and several variables. The derivative is introduced as a sensitivity parameter that can be approximated by difference quotients. This is reinforced with a computer lab on numerical differentiation that features experimentation with the size of $\Delta x$. Symbolic differentiation, including the chain rule and the product rule, are treated but not dwelt on. Partial and directional derivatives are introduced as natural extensions of derivatives to functions of two or more variables and students learn how to estimate these from contour plots and tables of data. Second order derivatives are used to make local approximation of 1- and 2-variable functions with quadratic Taylor polynomials. Gradient vectors are developed in a class activity with a contour map. A lab lets students explore how to use gradients in the steepest descent method to find the minimal potential energy of a configuration of two masses and two springs. Constrained optimization including Lagrange multipliers is taught graphically and the significance of

[5] R is a language and environment designed for statistical computing and graphics. One of R's strengths is the ease with which well-designed publication-quality plots can be produced, including mathematical symbols and formulae where needed. It is a numerical computation environment, like Matlab, and is popular with biologists. R is free software, available at www.r-project.org/

the multiplier $\lambda$ is explained.

The third section (roughly 8 classes) is on modeling with differential equations. The focus is on interpretation of the equations, and this section of the course includes a variety of extended examples. The concept of a differential equation is introduced as the mathematical description of simple models of population and pollution that are based on rate of change assumptions. Presentation as a slope field permits graphical solution. The exact solution of the linear differential equation $y' = ky$ is taught, but solutions are produced and graphed by computer for all other differential equations. A common sense explanation of Euler's method suggests how numerical solutions may be generated. First order models studied include (exponential and logistic models) population growth, predator-prey dynamics, susceptible-infected-removed (SIR) model of the spread of disease, passage of a cold medication through the body, and a general first order constant coefficient linear system of two coupled functions of time. In a computer lab, students graphically explore the variety of solutions of the coupled system that arise by changing the signs and magnitudes of the coefficients in the equation. As a last example, the second order harmonic oscillator equation is studied, and students learn to interpret a curve in the $y$-$y'$ phase plane.

The course ends with a section (roughly 7 classes) on linear algebra. The driving force behind including basic linear algebra is to understand least squares geometrically. This material is critical for, and has proven a good segue to, the next semester on statistical modeling. Students come to the course understanding two simultaneous linear equations as seeking the point of intersection of two lines in the plane. They are taught the dual interpretation, that of finding a linear combination, if any, of given vectors that "hits" a target vector. In the computer lab, students play a video game in which they attempt to hit a target on the screen by visually changing the lengths of given vectors that are added. Since the vectors are in 3-space, they must view the system from different angles, leading to some challenge and fun. Consideration of existence and uniqueness of solutions in examples leads naturally to the concepts of span and linear independence. In the applied setting the target vector, given by data, belongs to a high dimensional space and is not in the span of the few vectors available. What to do? The students are told to find the linear combination that gets closest to the target. Students see that finding a least squares fit can be interpreted in just this way, the parameters sought being the coefficients in a linear combination of vectors determined by the model. Using low dimensional examples, lengths, angles, dot products, and ultimately orthogonal projections are introduced that are used to solve the least squares problem.

Many locally written materials, including daily syllabus, class activities and a module on linear algebra, are available at the course website given above. They supplement the basic textbook (Hughes-Hallett et al., 2009).

## The second semester – Introduction to Statistical Modeling

The development of our Introduction to Statistical Modeling course (ISM) stemmed from a complete rethinking of the goals of introducing college-level statistics. ISM is typically taken right after AC. To accommodate students who have taken calculus in high school, students can also enter ISM after taking Multivariable Calculus or Linear Algebra; students in ISM must have some exposure to calculus of several variables and basic vector techniques. Note that most of our students take Multivariable Calculus before Linear Algebra, though the former is not required for the latter. Students who have taken AP Statistics in high school do not place out of ISM, as ISM is a multivariate modeling course.

ISM students should be able to describe realistic systems in meaningful ways. Descriptive statistics in ISM emphasizes multivariate modeling: how two variables are related in the context set by other variables. The idea of a partial derivative is important in describing relationships. To illustrate, consider the analysis of data from a trial of a cancer drug. Our previous course involved a $t$-test comparing patients who received the drug and those who got a placebo. In ISM, the effect of the drug can be analyzed in terms of the dose, adjusting for the sex and condition of the patients.

ISM students see the central concepts of statistical inference, and we teach inferential statistics as built on a core framework that unifies the various tests from the $t$-test to analysis of covariance. The central idea is that models partition variability into "deterministic" and "random" components (or "explained" vs "unexplained" or "modeled" vs "unmodeled") and that inference involves comparing the sizes of the two components.

Students should be equipped to build and interpret statistical models that can be of real use in their work in client disciplines. Computational statistics in ISM includes the organization of multivariate data, simulation and bootstrapping, and of course the interpretation of standard reports such as the regression and ANOVA reports.

ISM starts with building and interpreting linear models in a deterministic framework. The emphasis is on how to choose explanatory model terms that can capture important aspects of the variability in a response variable, and how to interpret the coefficients found by fitting. The central decision that a modeler makes is which terms (the columns of the model matrix $A$, in algebraic language) to include. Since the modeling is multivariate, we can introduce both "main effects" and "interactions" early on.

To describe the process of fitting a model, we introduce some mathematical abstractions: data as a point in an $N$-dimensional space and fitting as a projection onto a subspace. We emphasize the geometry of the situation. A central concept is the "model triangle," a right triangle whose hypotenuse is the response variable $b$ and whose legs are the fitted values ($Ax$) and the residuals ($b - Ax$). Correlation coefficients are cosines of angles, variances are square lengths, $R^2$ is a ratio of square lengths.

We move out of the deterministic framework half-way through the course. Confidence intervals are introduced through resampling. The importance of $\sqrt{N}$ is highlighted and reinforced by teaching about the nature of random walks.

The central inferential paradigm of hypothesis testing is also presented geometrically. The null hypothesis is that the explanatory vectors (the columns of $A$) point in random directions. In this framework, the one-sample $t$-test can be done with a protractor. To teach ANOVA and ANCOVA, we build on the idea of a sequence of models and how adding a new model term moves the fitted values ($Ax$) closer to the response variable ($b$). The $F$ statistic compares how far our new term took us to what would be expected for a randomly pointing term.

The overall result is a course that is very mathematical and perceived by the students as useful. They emerge with a set of concepts and skills in statistical reasoning that are a match for their native reasoning skills. Statistics becomes a way to describe and understand relationships of some complexity.

ISM is an ambitious course. There are a lot of rich and powerful ideas to cover and it's important not to spend time on technicalities. For example, we do not spend time on the difference between $z = 1.96$ and $t = 2.09$. We do not talk about the unequal variance $t$-test. Non-parametrics are covered concisely: take the rank before modeling. As mentioned above, we use professional-level software (the R package), presenting carefully selected aspects to the students. Since this is the second semester of using R, we can expect students to learn those aspects fluently.

Students emerge from ISM with a sense for the power of statistics and why so many fields rely on statistical methodology. The course is challenging, but not inaccessible; it is taken by about one-quarter of all students at Macalester. A typical course section includes a mix of students heading toward majors in biology (the biology major requires the ACM/ISM sequence), economics, and several other disciplines. It is also taken by all math majors, who particularly benefit from seeing practical applications of mathematics such as linear algebra.

The basic textbook (Kaplan, 2011) is available at the mosaic website given above.

## Results and Challenges

We are very happy with the way the AC/ISM sequence is working at Macalester. Students enjoy the courses, find them useful in further courses and in jobs, and faculty members in other departments (most notably, in biology and economics) appreciate the topics we teach. The 'big picture' view and wide variety of topics that students are exposed to early in college serves students majoring in these client disciplines well. The biggest challenge we face is the integration of these courses into our own majors' plans, in mathematics, applied mathematics, and statistics. On the math side, we still need to work on the transition from AC to second semester calculus for the very few students who come to us with no calculus at all, start in AC, and then want to consider a math major. For these students, AC lacks some of the algebraic formulations of a more traditional first calculus course, that they might need to go on in theoretical and/or applied mathematics. That said, they too get a very good feel for calculus and are often better prepared for Multivariable Calculus and Linear Algebra when they take them. On the statistics end, there is perhaps too much overlap between ISM and the next Applied Multivariate Statistics course, which is taken by students who decide to complete a statistics major or minor. In short, the courses work very well as terminal courses in our department, which was our intent. Our challenge is that their success has attracted more students with a greater diversity of mathematical backgrounds to want to go on in our department than we saw with a traditional calculus sequence in place. We are happy to continue dealing with this challenge.

## References

Craig, N.C., 2004: Chemistry. *Curriculum Foundations Project: Voices of the Partner Disciplines*, S. Ganter and W. Barker, Eds., MAA, 27-35.

Dilts, J. and A. Salem, 2004: Biology. *Curriculum Foundations Project: Voices of the Partner Disciplines*, S. Ganter and W. Barker, Eds., MAA, 15-18.

Hughes-Hallett, D., P. F. Lock, A. M. Gleason, D. E. Flath, S. P. Gordon, D. O. Lomen, D. Lovelock, W. G. McCallum, B. G. Osgood, A. Pasquale, J. Tecosky-Feldman, J. Thrash, K. R. Rhea, Thomas W. Tucker, 2009: *Applied Calculus*. 4th ed., Wiley.

Kaplan, D.T., 2011: *Statistical Modeling: A Fresh Approach*. 2nd ed., Project MOSAIC.