

# Calculus and Statistics

Daniel Kaplan, DeWitt Wallace Professor at Macalester College

Danny Kaplan is DeWitt Wallace professor at Macalester College, where he directs the applied math and statistics major. He's the author of *Statistical Modeling: A Fresh Approach* and *Start R in Calculus*.

Statistics can be introduced best when students have a solid grounding in calculus. This statement would have been mainstream 50 years ago. Nowadays it is controversial and provocative, even goading. I write it, and believe it, even though there is considerable experience to the contrary.

For at least 20 years, there has been a lively reform movement in statistics education, an improvement in pedagogy based on understanding how students perceive statistical concepts. One reform strategy is to strip away mathematical formalism that's not strictly needed, including integration and differentiation, the hallmarks of calculus. This has made statistical thinking more accessible.

A visible sign of success is the rapid growth of Advanced Placement statistics. There are important and legitimate criticisms of the AP curriculum and how it connects with more advanced statistics, but the success of the AP program is inspiring and a model to be emulated. AP statistics bootstrapped itself into high schools by providing training opportunities for high-school teachers who often had little or no statistics education themselves. Many students find the AP statistics course an attractive alternative to calculus because they see statistics as useful.

A decade ago, the Mathematical Association of America Committee on the Undergraduate Program in Mathematics worked with many

partner disciplines to see how the mathematics curriculum can better serve them. The findings, published in the CRAFTY reports, include a recommendation that students broadly be taught statistics without a calculus prerequisite.

The no-calculus form of statistics is also a pragmatic choice; that's where the students are. Nationally, the most heavily enrolled mathematics course at the college level is "college algebra," a pre-calculus course designed in almost all cases to lead to calculus, but with a success percentage in the single digits.

Even among those reaching the calculus level, attrition is high. The half-life of a student in the university-level mathematics curriculum is one course. Calculus is a filter that has become a choke-point as the economy becomes more and more technical.

This might make sense if the material learned in the traditional calculus sequence were more directly connected to success in technical careers. But, for many students, the calculus path leads to a destination of uncertain value. A student who spends a year learning techniques for symbolic differentiation and integration of functions of a single variable, along with definitions of limit and techniques for the analysis of sequences and series, by and large learns techniques that will rarely, if ever, be used by instructors in the partner disciplines and even less in their eventual careers. Statistics is often much more relevant to a student's ongoing and future work.

In thinking about the relationship between calculus and statistics, many people think first about integration and differentiation, cumulatives and



**a website geared toward new statistics professionals**



### STATtr@k offers tips on:

- how to apply for a job
- how to be a successful graduate student
- how to make the transition from coursework to research

STATtr@k also offers information about career and mentorship sites, upcoming conferences, and awards and competitions. New articles will appear monthly.

<http://stattrak.amstat.org>

densities, areas and slopes. One quickly realizes that calculus doesn't provide much insight. "Area" and "slope" are intuitive, elementary concepts. Indeed, much calculus pedagogy relies on areas and slopes to motivate derivatives and integrals. Beyond that, the algebraic techniques of calculus courses (e.g.,  $x^2 \rightarrow 2x$ ) don't get traction against the common distributions of statistics (e.g., the normal and  $t$  distributions).

Put aside for a moment the methods for differentiation and integration and think about the tools and language one needs to describe relationships among variables. Calculus and statistics both center on models of relationships: constructing them, analyzing them, evaluating them. In calculus, the choice to add a term to a model reflects some knowledge or hypothesis about mechanism. In statistics, choices are based on evidence provided by data. These are complementary perspectives with a shared foundation in mathematical modeling.

Traditionally, calculus instruction has emphasized functions of one variable,  $y = f(x)$ . In algebra and pre-calculus, students take on linear forms ( $mx + b$ ), then quadratics ( $ax^2 + bx + c$ ) and factoring. In calculus, they learn that  $mx + b$  can be used as a local approximation to many forms of functions, while  $ax^2 + bx + c$  gives a better approximation.

For statistical thinking, what's needed instead of quadratics and factoring is the incorporation of covariates. This can be as simple as the linear function with two inputs,  $z = a + bx + cy$ . This general-purpose form—extended often to more than two variables but remaining linear—is the workhorse of statistical modeling. It's a first representation of what might be called complexity that more than one variable can play a role.

A powerful way of thinking about functions like

$z = a + bx + cy$  is to ask how the output changes when either of the inputs,  $x$  or  $y$ , is changed. An important strategy from calculus is the partial derivative—examining the change in outcome as one input is changed while others are held constant. This aligns with experimental method in science; examining partial change and developing a formal language for describing it helps students understand that there are different ways for change to happen. In my view, understanding what's a partial change and what's not is fundamental to thinking about covariates and causation and therefore to the most compelling issues for applying statistics.

Ideally, statistical notions of fitting functions to data are taught hand in hand with the introduction of functions and their parameters in calculus. With this, and with the idea of partial change, students are better able to make mathematical sense of statistical ideas such as adjustment and how the relationship between two quantities,  $z$  and  $x$ , is informed by the participation of additional quantities.

At Macalester College, drawing on resources from NSF-sponsored Project MOSAIC, we redesigned our introduction to calculus to incorporate the CRAFTY recommendations: modeling, multiple variables, using only essential algebra, using data to inform models. Although it serves needs of all the disciplines, the primary orientation is toward statistics (even using R software for teaching calculus). In one semester of calculus, students gain experience building and interpreting models in multiple variables. They understand why it's important to consider relationships among multiple variables and learn the language to express such relationships. Then, when they move on to statistics, they can connect their models to data and

## For More Information

Blair, R., E.E. Kirkman, and J.W. Maxwell. 2013. *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 survey*. American Mathematical Society.

Garfield, J., and D. Ben-Zvi. 2008. *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer.

AP Statistics discussed in *Amstat News*, Sept./Oct./Nov. 2009, <http://magazine.amstat.org/blog/2009/11/01/statviewnov09>

The CRAFTY reports, [www.maa.org/cupm/crafty](http://www.maa.org/cupm/crafty)

Project MOSAIC, [www.mosaic-web.org](http://www.mosaic-web.org)  
NSF DUE-0920350

examine and evaluate the extent to which the data provide evidence for their models.

The success in making elementary statistics accessible without calculus is remarkable. But it's not clear how students can move forward along this path to the sorts of statistics needed in contemporary work involving complex, real-world systems. The intellectual skills students need to advance in statistics can be supported by engaging calculus, remodeling the calculus curriculum as needed to support the description, analysis, and judgment needed in statistical work. ■