Math 135/155 at Macalester College

Project Mosaic Kick-Off Workshop

Daniel Kaplan

July 1, 2010

Background

- Math 135/155 is designed as a *first-year sequence* of courses
  1. 135: Applied Calculus
  2. 155: Introduction to Statistical Modeling
- It is often actually taken as a first-year sequence, but about half of the students take it in some other way.
  - Applying high-school calculus instead of 135.
  - Math majors take 155 after linear algebra.
- The 135 syllabus varies somewhat from instructor to instructor.
  Founders in first three years: DTK, Tom Halverson, Karen Saxe, Dan Flath. Also taught by David Bressoud, Andrew Beveridge, Chad Topaz.
  I'll describe my planned arrangement for next year.

Background and Motivation

Origin: We made a deal with the biology department. They would require two specific math courses, designed by us, and we would make sure that their students got “twice as much statistics.”

- Biologists also take physics, chemistry, and statistics, so the sequence had to satisfy the needs of these disciplines.
- Economics is a large calculus-requiring department.

Fortunately ... Outside of math, chemistry, physics, no one actually relies on the calculus taught in calculus courses.

Design Goals and Constraints for 135

- Needed to make sense as a stand-alone mathematics course for science & social science majors. Give them a sensible mathematical background for their work in the major.
- Support the multivariable modeling planned for Math 155
- Accessible to students who enter with no calculus, and also those who enter with substantial high-school calculus.
- Cover calculus as needed for physics, chemistry, and economics. Be interesting and attractive to biology students. Make a plausible calculus course for pre-meds.
- Be able to replace Calc I as an entry point to the calculus sequence.

Design Goals for 155

Background:

- The conventional course introduced t- and related tests. Students learned what p-values and confidence intervals are. Description topics: center and spread, outliers, correlation as a linear trend.
- In their work in biology and economics, they were doing analysis of covariance and multiple regression. The p-value provides the connection between t-tests and these.

No surprise, then, that the p-value was being emphasized.

Provide students with reasonable skills for dealing with **Large, Complex Data**

- Shift emphasis from “tests” to covariation and modeling.
- Make **adjustment** and **analysis of covariance** accessible and understandable.
- Introduce logic of statistical inference rather than formulas. George Cobb describes the logic as the 3 Rs: Randomize, Repeat, Reject.

Lessons from CRAFTY

- Cover **multivariate** topics in a meaningful way.
- Provide **modeling** skills.
- Use **computation** in a serious way.
- De-emphasize **algebraic** manipulation.

Relate to Other Courses

- Feature **dynamics**: change, equilibrium, stability.
- Relate to **statistics**:
  - the multivariate linear model
  - functions as approximations to data
  - linear algebra/geometry to support statistical theory.
- Constrained optimization for economics.
- Include case studies that relate to a variety of fields, e.g. damped harmonic oscillator for **physics**.

Math 135 Topics I

1. Modeling Basics and Functions
   1.1 Important function types: constant, linear (affine), exponential, power-law, logarithm, sine, sigmoidal.
   1.2 Parameters
   1.3 Root-finding and inversion: graphical, numerical, table lookup, analytic.
   1.4 Setting parameters to match data
   1.5 Functions of two variables
   1.6 Polynomials (up to quadratic in two variables)
   1.7 Finite-difference models: exponential, "logistic"
2. Units and Dimensions
   2.1 Basic dimensions: Length, Time, Mass
   2.2 Conversion of units
   2.3 Commensurability
   2.4 Dimensionless constants.
3. Linear Algebra/Geometry
   3.1 Vectors
   3.2 Subspaces

Math 135 Topics II

3.3 Linear Combinations
3.4 Projection
3.5 Solving $Ax = b$ (emphasis on least squares)
3.6 Singularity (“redundancy”)

4. Derivatives
   4.1 ordinary, partial, gradients, directional
   4.2 differential equations
   4.3 the phase plane
   4.4 solutions through anzätze, substitution, intuitive accumulation
   4.5 contrast Taylor with Least Squares
5. Optimization
   5.1 Fitting
   5.2 Constraints
6. Accumulation and Integration
   6.1 Sums
   6.2 Euler
   6.3 Averages
   6.4 Relationship to derivatives. Definite and indefinite integrals.

Software

R — the statistics package.

- A calculator and graphing engine.
- Functional notation including operators on functions: composition, differentiation.
- Reading in data and fitting.
- Differential equation solving.
- Several interactive “apps.”

Why R Works for Calculus

- Students know the same package will be used in statistics, providing additional motivation.
- High quality graphics are easy.
- The editing/command/revision cycle is easy. Even more so now with RStudio interface.
- The "functional" nature of the language supports calculus. Example: A derivative function:

  ```
  D = function(f, h=0.00001) {
    function(x) { (f(x+h) - f(x))/h }
  }
  ```

  This takes a function as an input and returns a function as an output.
- Has **little or no symbolic capabilities**. Benefits: syntax is much easier; reduces *drift* back to the traditional Calc I; encourages instructors to emphasize numerics/graphics/data.

Problem: Not so powerful for GUIs/Apps.

Some Examples of the Style Used in 135

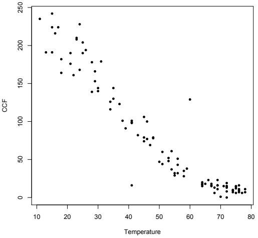
1. A model from data. You can ask interesting questions about linear relationships!
2. The bivariate quadratic: a guide for modelers.
3. Euler integration.
4. The phase plane.
5. Solutions of linear systems.

A Model from Data

Problem: How much energy can I save by reducing my electricity use?
Issue: Energy used for lighting, appliances, etc. may also contribute to household heating, so reducing electricity use might increase use of natural gas for heating. (Similarly, reducing electricity for lighting and appliances use might give double-savings in summer, by reducing air conditioning demand.)
Approach: Collect data on **natural gas** use vs **electricity** use during heating months to find the relationship. Potentially include another important determinants of natural gas use: **temperature**.

A Model from Data

Problem: How much energy can I save by reducing my electricity use?
Issue: Energy used for lighting, appliances, etc. may also contribute to household heating, so reducing electricity use might increase use of natural gas for heating. (Similarly, reducing electricity for lighting and appliances use might give double-savings in summer, by reducing air conditioning demand.)
Approach: Collect data on **natural gas** use vs **electricity** use during heating months to find the relationship. Potentially include another important determinants of natural gas use: **temperature**.



A Model from Data (cont.) I

Consider two models:

$$\begin{aligned} \text{thermsPerDay} &= 3.9215 + 0.8977 \text{ kwh} \\ \text{thermsPerDay} &= 10.513 - 0.001462 \text{ kwh} - 0.1484 \text{ temp} \end{aligned}$$

1. Interpret each of the models in terms of the basic question: Does electricity use offset natural gas use during heating months?
2. Pick one of the models and show that the coefficients are optimal, that is, that using the given numbers gives a smaller set of residuals than other numbers. First, extract a subset of the data for heating months (and that avoids the two outlier points that resulted from a misreading of the meter):

```
> u = ISMdata("utilities.csv")
> h = subset(u, temp < 60 & ccf > 20)
```

A Model from Data (cont.) II

Make a graph showing the size of the residuals (summarized in some appropriate way) versus the value of one of the coefficients, holding the others constant, and confirm that the minimum is near the given coefficients.

3. Are these models conceivably describing the same data? Plug in some typical values of kwh and temp and see if they give similar results for thermsPerDay.
4. What are the units of the coefficients? Keep in mind that kwh is actually kilowatt-hours per month. The thermsPerDay variable is what it says: therms per day of natural gas.
5. Which of the two models is better in the sense of producing the smallest residuals?

A Model from Data (cont.) III

6. A physicist proposes a sanity check for the models. Her reasoning is that a kilowatt-hour is a unit of energy and a therm is also a unit of energy. For heating a house, it doesn't matter whether the energy comes from a furnace burning natural gas or a light bulb "burning" electricity. So the amount of extra natural gas that will be needed when you save a kilowatt-hour of electricity is simply the equivalent to a kilowatt-hour in therms.
   Use the Internet to find the conversion factor between kwh and therms and compare the result to the appropriate coefficient from each model. Remember that the data are in thermsPerDay and kilowatt-hours per month, so you will have to convert the ratio of therms/kwh to a conversion factor of therms-per-day/kwh-per-month.
   Which, if either, of the two models is consistent with the physicist's sanity check?

A Model from Data (cont.) IV

7. Use the Internet or some other source to find reasonable typical prices for a therm of natural gas and for a kilowatt-hour of electricity. Using the prices, calculate how much money, if any, would be saved by reducing electricity use by one kwh taking into account any additional cost or saving in the bill for natural gas.

Some Modeling Exercises

1. Construct a model of how fast a bicycle goes as a function of hill steepness and gear. Assume that the rider is expending a constant effort.
2. Construct a model of how production depends on the amount of capital and labor available.
3. Construct a model of daylight length as a function of time of year and latitude.

Some background mathematics

Students often know about

- Line models: $y = f(x) = mx + b$
- Polynomials, e.g.: $y = f(x) = ax^2 + bx + c$

Students don't usually know about functions of two variables:

$$z = f(x, y) = a + bx + cy + dxy + ex^2 + fy^2$$

- How to evaluate them (given a, b, c, d, e, f and x, y)
- How to interpret them

Interpreting polynomials in two variables

$$z = f(x, y) = a + bx + cy + dxy + ex^2 + fy^2$$

When to include the various components.

$a + bx + cy$	almost always included: a simple roof shape
dxy	the bilinear or interaction term. Needed when effect of x changes with y.
ex^2 or fy^2	curvature (or extrema) in those variables.

Example: A polynomial in two variables

How fast does a bicycle go depending on hill steepness and gear? $s = f(h, g)$

- s bicycle speed
- h hill steepness
- g gear

- Is there an optimium hill slope: No. Don't need h^2 term.
- Is there an optimum gear? Yes: include g^2 term.
- Does the optimum gear depend on steepness? Yes: include hg interaction term.

Example: Economic Production

Factory output depends on the amount of capital and labor:
$P = f(C, L)$

- P Production
- C Capital
- L Labor

- Output increases with C and L.
- More capital increase productivity of labor: Interaction term CL

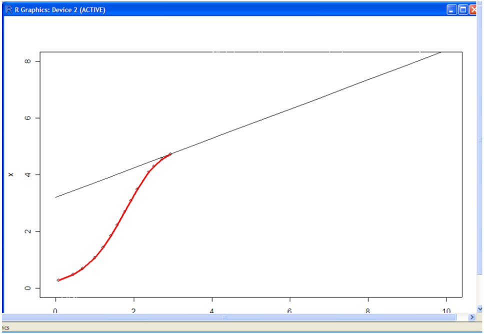
Example: Duration of daylight

Duration depends on latitude and month.

- D Duration
- L Latitude
- M Month

- Changes with both latitude and month.
- Change over the months depends on the latitude: Interaction term LM.
- Saturates with month: M^2 (or higher)

A graphical approach to integration



The logistic-growth system:

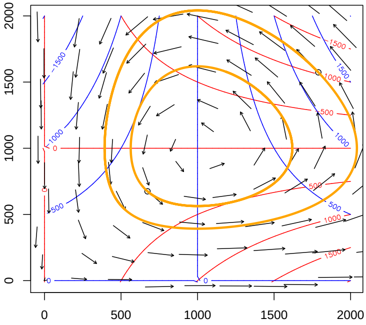
$\dot{x} = rx(1 - x/K)$

- The differential equation describes **local** dynamics.
- Growth rate changes with x.
- Accumulate small increments.

Dynamics on the Phase Plane

Predator-Prey system. Emphasize formation of the model itself ...

- Exponential growth of prey (in absence of predator)
- Exponential decay of predator (in absence of prey)
- Bilinear term: “Interaction” of predator and prey.



Software draws:

1. Flow field.
2. Null-clines. More generally: contour plot of dynamical functions (blue and red)
3. Trajectories (click on starting point)

Calculus does not need to focus on Analytic Solutions

It's also calculus to teach ...

- The phenomenology of differential equations: equilibrium and stability, oscillation
- What is a function of what? Phase plane $\dot{x} = f(x, y)$ versus solution $x(t)$ versus t.
- The analysis of trajectories (e.g., they don't cross one another, they cross null-clines in a particular way)
- Identifying consequences of modeling assumptions: What happens if prey growth is sigmoidal rather than exponential?

Computers can solve the DEs, so solution techniques are no longer central.

Solving Simultaneous Equations

$$\begin{array}{rcrcrcr} 3 & x & + & 2 & y & = & 5 \\ -1 & x & + & 4 & y & = & 3 \end{array}$$

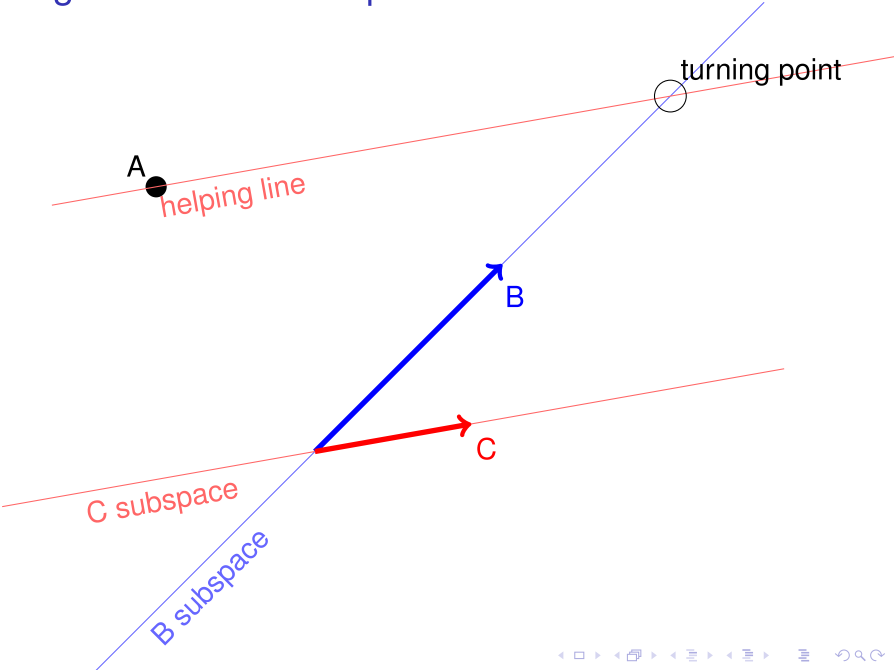
- Students learn how to solve such systems by substitution.
- Systems like this are central to science and technology, but few students know why.
- They often involve hundreds to millions of variables and equations.
- In statistics applications, typically there are many more equations than variables.

Simultaneous Equations

What should students learn about such systems?

- Where they come from. (Linear approximations of systems.)
- When they don't have solutions.
- How to find the *best* solution when no exact solution exists. (This is much of statistics.)
- When the solutions are sensitive to the quantities in the equations (which themselves are usually known only approximately or with error).

Solving Simultaneous Equations



Math 135 Impact at Macalester

- Courses are liked by students. They typically report this is the first time the see the *uses* of mathematics.
- One-third of all students at Mac take the calculus course, one-quarter take the statistics course.
- Courses are well liked by faculty. Senior faculty teach Applied Calculus. Client faculty see the connection to their disciplines.
- Advanced placement students are pleased with the courses — they are learning a lot that is new; the courses aren't too easy for them.
- Non-advanced placement students do well in the course.

Enrollments

Compare enrollments for the class of 2005 and the 2009-2010 school year. Note: entering class in 2009 was approx. 15% bigger than usual. Longitudinal data not yet available.

	Calc I/Applied	Calc II/Single	Calc III	Lin. Alg.	Diff. Eq.
2005	112	123	78	57	26
	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$	$\Downarrow$
2009	171	111	109	78	46

Observations:

- Substantial increases in all courses *except* Calc II.
- We see students skipping from Applied Calculus to Calc III. We think these are the students who took BC-level calculus in high school and who were motivated to re-enter math.
- More common to have math majors starting in Applied Calc. compared to the old Calc I. *Doing well by doing good!*

Math 155: Introduction to Statistical Modeling

- Follow-up course to Math 135.
  But ... about half of students in Math 155 are coming from a high-school calculus course. Physics and chemistry students tend **not** to take 155.
- Largest groups of clients: economists, biologists, math majors.
- Builds on modeling skills from Math 135: multivariate modeling.

Topics in Introduction to Statistical Modeling

- Organization and (simple) descriptions of data.
- Construction of (linear) statistical models. This includes multiple variables and nonlinear terms, esp. interactions.
- Adjustment for covariation. The idea of "partial change."
- Inference:
  - Confidence intervals and the effects of collinearity.
  - Analysis of Covariance. Central question: Does this variable contribute to the explanation.
- Causation & Experimental design: Randomization, blocking, and orthogonality.
- Logistic regression and non-parametrics.

For the preface and outline, see
`http://www.macalester.edu/~kaplan/ISM`

CS 121: Scientific Programming

- A programming course designed for science students.
- Language: MATLAB
- Programming topics: functions, flow control, scope, data structures, GUIs, trees.
- Everything introduced in the context of science/statistics/mathematics questions.
- Sound and image processing featured.
- Scientific graphics from week 1.
- Has been the largest enrollment intro CS course.