Classifying Courses: Standardized Courses, Clusters and Categories

(See appendix for decision tree useful for classifying a catalog entry.)

Table of Contents

Presenting the Data in a Table and Subtables

What campuses and departments are we surveying?

We are surveying campuses of 4-year, post-secondary, degree-granting institutions. Our campuses were chosen as what we hoped would be an interesting sample of institutions of various kinds. Interesting is more or less the same thing as being well-known, and we certainly do not claim our sample is random. Furthermore, within a type of institution (e.g., those that were women's colleges for most of the 20th century), we have only a few examples, so computing statistics with confidence intervals is presently out of the question.

Multiple departments.

We asked archivists at the campuses of our sample to send us photocopy pages from the school catalog corresponding to any undergraduate department with "mathematics" or "mathematical" in its title, but not Mathematics Education departments. Thus "Mathematical Sciences", "Mathematics and Computer Science", "Applied Mathematics" , would qualify[1]. The University of Texas at Austin, early in the 20th century, had departments of pure mathematics and applied mathematics. Each is to be surveyed in its own separate abbreviated catalog data workbook.

At campuses, such as Texas at Austin, which have multiple mathematics departments, we also present a "combination" ("combo" for short) workbook showing all courses in all departments at that campus.

There is more to this multiple department issue and a more extended discussion can be found in the file *DeptsCampusesWorkbks.doc*

[1] In this document, when we speak of a "mathematics department" we mean any and all such departments.

What courses are we surveying?

Not just courses for mathematics majors

We record all undergraduate courses, not just those for mathematics majors. There are a few exceptions noted below.

Graduate versus undergraduate

We are only interested in undergraduate courses. It may be hard to distinguish graduate from undergraduate courses. If there is a category entitled "Primarily for Graduate Students", we do not record these courses. If there is a category of courses which are said to be open to both graduate students and undergraduates (dual level courses), all these courses are to be recorded.

No doubt, any department with graduate courses will occasionally allow an exceptional undergraduate into a graduate course, no matter what it says in the catalog. In the strictest and most literal sense, there probably are no courses for graduate students only. But we are concerned about the rules and guidelines that apply to most students.

A course that might only be open to graduates at one department, might be available to undergraduates in another. For example, Stanford had a course in Elementary Functional Analysis. We do not take on the hopeless task of making an objective decision about the level of a course. We seek, instead to determine, as best we can, what the department in question thought about its courses.

What if we can't tell what a department is thinking? For example, at Stanford in 1955, there is no category of "primarily for grad students". The most advanced category is "COURSES PRIMARILY FOR UPPER DIVISION AND GRADUATE STUDENTS". This terse heading could perhaps mean that each course is suitable for both types of students. Or it might mean that each course is suitable for one or the other, but usually not both. Or maybe (most likely) it means nothing very precise except that students should be advised by faculty about course suitability. In the Stanford case cited, there are 100-level and 200 level courses in the ambiguously titled category. We have decided to take the 200 level courses as being graduate courses and not list them.

If an institution makes no distinction that we can detect in regard to undergraduate versus graduate courses, but specifies what courses are typically taken by a mathematics major in first year, second year, etc., this may perhaps be used as a guide to determine which are undergraduate courses. Course numbering might also give a clue.

Distinguishing graduate from undergraduate courses turned out to be a much smaller problem than you might guess from the ink expended on it here. And this comment might be generalized to much of the hairsplitting in this entire document.

Allied disciplines

Certain subjects have sometimes been taught in mathematics departments even though their relation to mathematics is, from the perspective of the early 21st century, somewhat peripheral. Examples we have in mind include: actuarial science, astronomy, computer science, engineering, military science, operations research and statistics. This raises the question of whether they ought to be regarded as mathematics courses and recorded in our tables merely because they are listed in the mathematics section of the catalog.

Various arguments can be given, pro and con, as to whether these courses should be recorded by our survey. Since our survey is, in large part, a recording of opinion[2] we are most interested in the opinions of the mathematics departments themselves at the various times in their history. There are two indications that allow us a glimpse, however imperfect, of departmental opinion.
1. The common practice of including an alphabetical prefix such as MATH (as in *MATH 123 Linear Algebra*) or CS or STAT in front of the course number. Our rule is that we go by these designations if they are available. We record courses with mathematics designations ("M", "MTH", "MATH", etc.) and only those.
2. If designations are not used, it may happen that the list of courses taught by the department is divided into named sublists. For example, at Samford in 1964 one sublist in the mathematics section is entitled "Mathematics" and another is entitled "Engineering". We consider this an indication that Samford did not regard the engineering courses as mathematics (except for administrative convenience) and we do not record them. On the other hand, if the sublists were (hypothetically) "Algebra", "Geometry", "Analysis" and "Statistics", we could not conclude that Statistics is outside mathematics in the opinion of this department.

If neither of these two phenomena is present, as is the case with some actuarial science courses, we will record all courses listed under the mathematics department.

It should be clear that the rules just described mean that this project is not a comprehensive tool for understanding these peripheral allied disciplines and their relation to mathematics.

Non-recording of "wandering" courses

[2] Given the huge volume of elementary mathematics, a department's curriculum is mostly its opinion about which subset of it is most valuable to put before students.

Occasionally some subjects are taught in mathematics departments but then in other years they are in other departments or a new department just for that subject. In the early part of the 20th century some astronomy courses were taught in the Appliued Mathematics Department at Stanford, but not each year. In some universities, computer science arose in the Mathematics Department, only to wind up in Computer Science departments. At Stanford, courses in foundations were taught sometimes in Mathematics, but sometimes in Philosophy. Probability and Statistics were sometimes in Mathematics and sometimes in Statistics at Stanford.

We record such courses when we see them in a mathematics section of the catalog and there is no designation or sublist that would indicate that they are not mathematics courses. No attempt will be made to record information about courses that are not described in the mathematics portion of the catalog simply because they used to be in the mathematics list and have (perhaps temporarily) wandered away. E.g., in years when astronomy is not listed under mathematics, we won't search through other departments to find and record astronomy simply because there were some years when it was a mathematics course.

Courses of a mathematical nature which are taught in a department without "mathematics" in its title but are nonetheless mentioned in a mathematics portion of the catalog will be not be recorded.

Cross-listed courses are generally not covered

A cross-listed course is one which is listed in a mathematics portion of a catalog but where there is also reference to another department. The main examples of such reference are:

1. In 1980, Stanford's mathematics department course list includes "136. Introduction to Computing – (Enroll in Computer Science 106.)". We do not record this course as it appears that it is taught by the Computer Science Dep't. – there is no course description in the mathematics section of the catalog.
2. We also do not record the course if the designation (abbreviation before the course number) is not a mathematics designation (e.g. Phy instead of Math.)

These rules still leave gray areas. For a time, Bowdoin's mathematics department listed a course in Mechanics with a physics designation, but the instructor was Prof. Christie, a member of the mathematics department. We recorded this as a mathematics course.

If there is a double designation such as *M360/P420 Philosophy Mathematics* , we would recorded it in our tables.

If there are no alphabetical prefixes designating department, and no clear indication of who teaches it, then we try to find some other clue about which department has primary responsibility, or simply use our best judgment. For example, at Samford in 1905,

Physics is listed in the mathematics department (with a reference to the same course in the Physics Dep't), but common sense indicates that we should not list the course in Cajori Two tables.

"Not given' courses

If a course is listed but there is a note that it is not given in the year covered by that catalog, we record it in the workbook nonetheless.

Interdisciplinary Programs

Programs that a mathematics department participates in but which may have significant numbers of courses in other departments, such as a business department or perhaps an operations research department, pose special problems. It would have required more resources than we had in order to do them justice. To see the issue, consider the Actuarial Science program at Purdue[3] in 2009. This program appears to be well-established and well developed, with a newsletter, an alumni awards program, an advisory council and significant space on the Mathematics Department website. But if one examines the program through the lens of courses with a mathematics designation (prefix) we find just two courses, *MA 373 Financial Modeling* and *MA/STAT 170 Introduction to Actuarial Science*. The rest of what is required are: statistics courses (STAT prefixes) – two of which are specifically dedicated to actuarial issues; four courses in management (MGMT); two in economics (ECON). Cajori Two only surveyed courses with mathematics designations (prefixes). Thus, one cannot gain any understanding whatsoever of Purdue's actuarial science program from the fact that we would list two mathematics courses in actuarial science in a Purdue workbook.

In short: Cajori Two is not a good means for studying interdisciplinary programs.

Mapping to Standardized Courses

Title and description variations, and standardized courses

Two courses in different colleges may be substantially the same even though they differ in title. For example *Linear Algebra* and *Theory of Vector Spaces* are probably very similar courses. Needless to say, the course descriptions, if available (not always the case in the early 20th century) need to be considered. But even if the course descriptions differ, if the differences do not seem great, we might consider the courses essentially the

[3] It is true that Purdue is not one of the schools in our sample, but we wanted our methods to be substantially sample independent, so from time to time we considered other schools known to us, such as Purdue.

same. In other language, the courses we find in the various catalogs fall into equivalence classes. Because Cajori Two seeks a birds-eye view of the century we are really interested in the equivalence classes. We implement this idea of equivalence classes by formulating *standardized courses*. For example we chose *Linear Algebra* as the standardized title for its equivalence class, and defined the class by a standardized course description (available in another document, entitled *Cajori Two Course Inventory* found in the file *ClusteredInventory.docx* in the folder *CajoriTwo 4.0\AbbreviatedCatalogData\HowTheAbbreviatedCatalogDataIsProduced\GeneralMethodology*.) Of course we do not require a course found in a catalog to have its catalog description to match up exactly with our standardized course description in order to map it to that standardized course. It is only necessary that the classifier determine that it fits tolerably well and better than anywhere else. As an example, we map *Introduction to Statistics* and *Introduction to Biostatistics* to the same standardized course *Elementary Statistics* on the grounds that introductory statistics courses necessarily have a great deal in common.

Categories and clusters of standardized courses

We have in the neighborhood of 250 standardized courses. For convenience in perusing our tables and for insightful data analysis, our standardized courses are grouped into categories: Elementary Courses in Algebra, Geometry and Trigonometry; Elementary Service Plus General Education; Mathematics Expressly for Teachers; Basic Calculus Sequences; Analysis Following Basic Calculus; Higher and Abstract Algebra, Linear Algebra and Number Theory; Advanced Geometry and Topology; Foundations; Advanced Applied; Discrete Mathematics; Advanced Probability and Statistics; Computer Science; Courses With Unspecified Content; Other.

Within each category, the courses are grouped into a small number of clusters to provide a level of "resolution" intermediate between the course and the category. See the document, entitled *Cajori Two Course Inventory* for the clusters. At present, the Excel-based Data Analysis Software takes no note of the clusters, so tabulation by cluster is not possible.

The Other category and Other courses.

If no standardized course seems like a good match for a course found in a catalog, the course should be mapped to the catchall course *Other*. Each category has a course entitled *Other*. Furthermore, there is even a category entitled Other (with courses such as *History of Mathematics*) for courses that do not fit gracefully in one of our explicitly named categories. (And, yes, for ultimate "otherness", the Other category has a course entitled *Other*.)

Standardized courses are one term long

All our standardized courses are courses that extend over one term (where a term is either a semester or quarter; our sample has no trimester schools.) In rare instances (mostly early in the century) a course may last for less than a term. When a catalog shows a course extending over a year, we break it into it constituent terms and look for a standardized course to match each term to. So course [101] extending over 3 quarters becomes 3 courses for us to classify and map to standardized courses: [101 1st term], [101 2nd term] and [101 3rd term].

In some cases, there may be a single title, number and entry in the catalog but there are different sections for different audiences. For example in 1911 at Reed College, courses 11 and 12 are precalculus courses that each have two sections for two different audiences, one more advanced than the other. The contents of the sections differ. We separate each of them into two courses, 11 section 1 and 11 section 2 and 12 section 1 and 12 section 2.

Calculus sequences

It is very common to think of basic calculus as being a sequence of courses. Accordingly, in our inventory we group basic calculus courses into sequences (also known in our inventory as clusters) whose titles reflect the nature of the sequence, or the audience (e.g. Mainstream Calculus Sequence, Accelerated/Honors Calculus Sequence, Briefer Calculus Sequence[4] .) The standardized courses themselves are named rather colorlessly, e.g., Mainstream Calculus Term 1.

Creating the standardized course list and course descriptions.

The first stage of Cajori Two was to make a list of undergraduate mathematics courses which might be found at 4-year colleges and universities at one time or another during the 20th century. This gave rise to our standardized courses, with their standardized titles and standardized course descriptions. Our aim was to be as comprehensive as possible, so that, any course actually encountered in a catalog would match up well with a standardized course on our list without the use of *Other*.

We assembled our standardized courses and course descriptions from the following resources:
a) the standardized courses in the MAA Catalog Survey of 1960, with minor modifications,

[4] Briefer calculus, usually taught for students in the social sciences or business is often just one term, but honored with the title sequence anyhow as it sometimes stretches over two. In general we provide for more courses in a sequence than some departments will have.

b) additional standardized courses, not found in the MAA list, but found in either
1. Stanford University curricula 1900-2000
2. the CBMS 2000 course list
3. Cajori's 1890 course list

c) a small number of our own suggestions
d) courses actually found in the catalogs of the institutions in our sample but not occurring in a), b) or c).

Size of our list of courses

Our method produced a large number of standardized courses, large enough that each course found in the departments in the sample could be mapped to some standardized course on our list without excessive inaccuracy, and only a minimal number would have to be mapped (maddeningly) to *Other*. But our methods of list assembly probably produced courses not found anywhere in our sample.

Problems in Mapping

Combination courses

It is surprisingly common to find course titles composed of two commonly occurring topics, but where the combination is a bit unusual (but not bizarre). Examples:

1. Ordinary and partial differential equations
2. Vector spaces and vector analysis
3. Abstract Algebra and Geometry

To try to list all such combination titles that might someday somewhere be found, and then to enter them into our inventory of courses, seemed foolishly ambitious. We did list some that we knew would be found such as "Abstract/Modern Algebra with Linear Algebra". To handle others that might occur, we will do the classification as follows:

a) Determine from the course description (if possible) which topic takes up the larger portion of the course. Classify the course under the standardized course that best fits that dominant topic. Use the other topic in the title as a flavor (either using a flavor abbreviation if there is an abbreviation for that flavor, or simply write out a word or two describing the flavor, e.g., "fl: geometry", or using an additional comment, e.g., "ac: includes category theory".)

b) If it is not possible to find, with any confidence, which topic is dominant, use the most advanced topic to classify the course and the other as a flavor.

c) If neither method a) nor method b) gives a verdict, and no other classification approach makes sense, use the first-named topic in the catalog course title as the dominant topic.

Courses with unspecified content

This is a category that we use for: reading courses, conference courses, internships, seminars, thesis work, etc. If content is restricted, as in, *Readings in Abstract Algebra*, record such a course as *Other* in the more restricted category. So *Readings in Abstract Algebra* is best recorded in the *Higher and Abstract Algebra, Linear Algebra and Number Theory* category.

Presenting the Data in a Table and Subtables

Rows, columns, cells

We present the results of classifying the courses in our sample in what we call the Abbreviated Catalog Data. It takes the form of a table with many subtables. For each department or combination we have a subtable. The rows of the table are the labeled by the standardized courses, grouped into clusters, the clusters grouped into categories. The columns of the table are labeled by the years of our sample. In a cell at the intersection of a row and column, we record whether there is a course in the given department or combination, in the year of the column, which would map to the standardized course of the row. If there is none, we leave the cell blank. If there is one or more, we enter highly abbreviated information about the course or courses in the cell. The subtables for the various departments or combinations are large and sparse and for this reason we use Excel workbooks to record the table rather than paper.

What's in a cell?

For each course recorded in a cell, we note the following in the cell:

Catalog course number,
Course weight (credits, hours, etc.),
Prerequisites[5]
Text in use (if known)

The following additional information maybe recorded in the cell, when it is known, and at the option of the coder:

[5] Especially in the early years of the century, prerequisites are not listed even for courses which surely have them. Thus, if no prerequisite information is given, the reader should not conclude that there were none.

Course flavors;
Title (when the standardized course title labeling this row does not tell the story adequately);
Additional comments.

Note that, for brevity, we do not routinely record the catalog title of the course – just the course number. However, there are cases where course do not get numbers, and in this case we record the title.

Multiple Courses in a Cell

When there is more than one course in the catalog that maps to the same standardized course in the year in question, we put information about both courses in the cell. We call the various courses in a cell *instances* of the standardized course labeling the row of that cell.

For example, say there is the usual course in linear algebra but also one for business and social science majors, both 1-term courses. We do not have a standardized course in linear algebra for business and social science students, so we put both these linear algebra courses in the same cell and distinguish the second from the first by indicating the intended audience in a flavor[6] or additional comment. In effect, we are making fine distinctions within the equivalence relation.

Coding multi-term sequences.

Another example of multiple instances in a cell occurs when there is a "part 2" to the course. Suppose a department or combination has a 2-term course in linear algebra. We do not have standardized courses (equivalently, rows of our Excel table) for each separate term of a 2-term course in linear algebra[7]. We just have a 1-term course in linear algebra in our inventory. So we use that row of the table twice. (See elsewhere for how this is done.) In simple language: we have two 1-term courses in linear algebra listed in the same cell and the prerequisite structure will show that they are a sequence.

Instances, presences, sequences.

[6] See the document *How To Fill In a Cell* for a description and listing of flavors.

[7] We do have separate standardized courses for the terms of a basic calculus sequence because it is nearly universal that basic calculus is taught as a sequence.

The instances of a standardized course that are found in a single cell may be involved in one or more sequences.

Consider the following three courses, hypothetically found at one department:

1. Discrete Mathematics I (should not be taken if course 2 is taken)

2. Discrete Mathematics I With Lab (should not be taken if course 1 is taken)

3. Discrete Mathematics II (Prerequisite: 1 or 2.)

We only have one course in Discrete Mathematics in our inventory, so all three above would map to that single standardized course and go into the same cell. They are 3 *instances* of the same course. Courses 1 and 3 are clearly a sequence as are courses 2 and 3. There is no need to make any special note of that: the listing of prerequisites tells the sequence story. The "lab" nature of course 2 should be noted in an additional comment ("ac".) The prohibition against taking #1 and #2 could be noted, but this is a judgment call for the coder. Cajori Two is more interested in what is available to students than in what is not possible.

No matter how many instances of Discrete Mathematics a department or combination has (3 in our example), we would say that this course has a *presence* at this department or combination. Thus, if there is at least one instance, then there is a presence. Presence is a 0/1 variable – a Boolean variable.

Excel-based Data Analysis Software

Presence is a Boolean variable, but in our tables we show it with the numerical variable (1 if there is a presence, 0 otherwise.) This allows us to add the presences in all the departments or combinations. If that total comes to 14, this means there were 14 departments or combinations in which there was at least one Discrete Mathematics course in the year in question. We can, of course, also add the instances. Summation and tabulation of instances and presences is carried out by the Excel-based Data Analysis Software we have created for this project. It is also carried out by the MySQL software invoked by our website.

Documentation for the Excel-based Data Analysis Software is found elsewhere.

Decision Tree for Classifying a Course Found in a Catalog