EMPIRICAL PREDICTION OF STUDENT ATTENDANCE AT DENNIS GABOR COLLEGE

Elemér Károly Nagy

1. ABSTRACT

Reliably predicting student attendance is a powerful tool both for scheduling courses and for improving the quality of lectures through scheduling mid-term assignments and laboratory tasks. This article analyses student attendance data collected by the author for six semesters and data collected by other lecturers, for a variety of courses.

The aim of this analysis is to provide reliable real-world answers for a single question: What percentage of students will attend the N^{th} lecture / laboratory / workshop?

2. THE PROBLEM

This question is quite easy to answer – even a quick Google search reveals that the answer in Australia, for example, is “32% to 95%”, according to [1].

However, pragmatically, we are interested only in the student attendance at Dennis Gabor College, especially laboratory attendance, as laboratory workspaces are more expensive to provide. The answer also has to has a low enough variance and/or error to be useful for classroom and test scheduling.

2.1 Variance in the answers – variables in the problem

The high variance in the known answers is caused by the high number of variables in the problem. Some of these variables are constants for years, like education type (regular or distance learning, BSc or expert certification) or course type (laboratory or theory, introduction or expert course). Some variables are practically constants for a given semester, like academic policies regarding whether attendance is mandatory or not. Some variables are changing slowly (like the amount and quality of distance learning materials, the teacher’s personality) and some are changing rapidly from lecture to lecture (student’s timetable, weather, mid-term assignment and the students’ other courses).

2.2 The goal – our motivation

There are questions that resurface from semester to semester that are the derivatives of the problem stated above. Answering the original question will yield answers to these questions and provide an opportunity to improve the efficiency and/or the quality of the courses. These questions are:

1. Which classroom should be scheduled for the course? (so that every student gets a lab workspace)
2. How many laboratory groups should be planned for X students if the laboratory has Y workspaces? (so that every student has >90% to get a workspace, one workspace can serve Z students)
3. How many lectures should be sacrificed for mid-semester tests? (if the test needs lab workspace)
4. When should mid-term assignments start? (the later they start, the less lab workspace is needed?
5. How feasible is it to merge laboratory groups mid-term? When should it happen?

3. THE DATA

The attendance data was collected by the author for six semesters from a variety of courses, including both regular and distance learning courses, both BSc and expert certification courses, both first-year and last-year courses. After the initial analysis, attendance data was added from three other lecturers to add freshman data, 0^{th}-semester data, data from the field of business and management (as the author teaches ICT), and data for “strictly mandatory attendance” courses.

3.1 Data requirements

Only courses with a clean separation between “theory” and “laboratory” lectures were considered. In laboratory/workshop courses, only

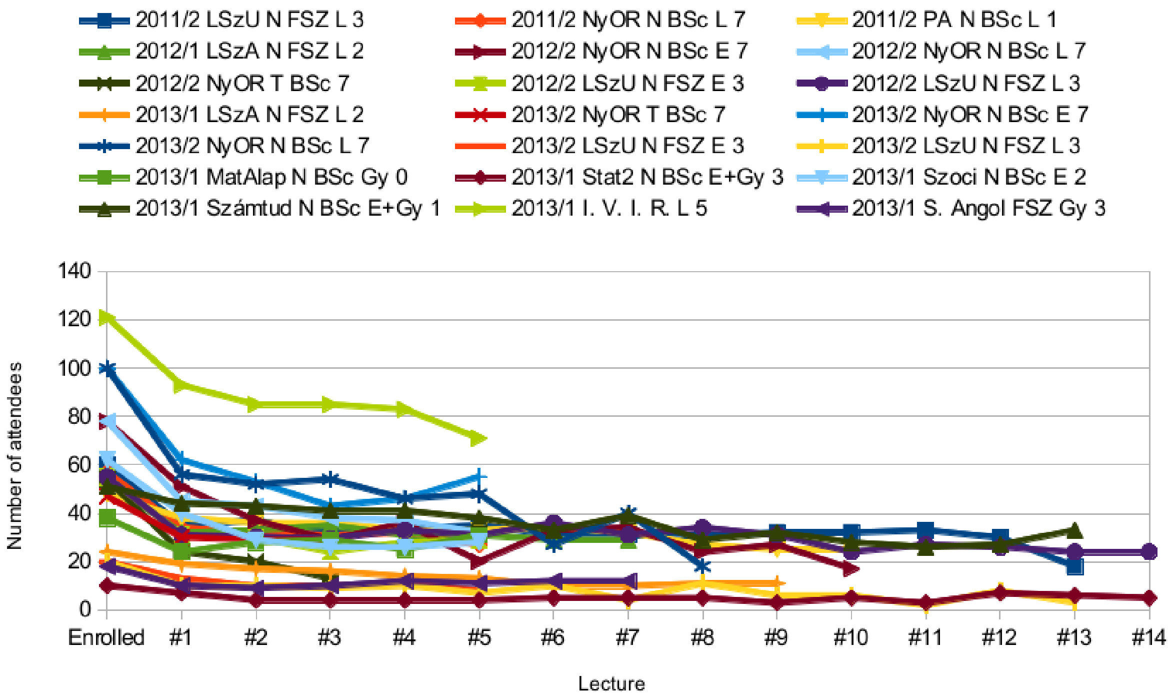

Figure 1.
Keyed in data for 21 courses from 4 lecturers from a variety of courses

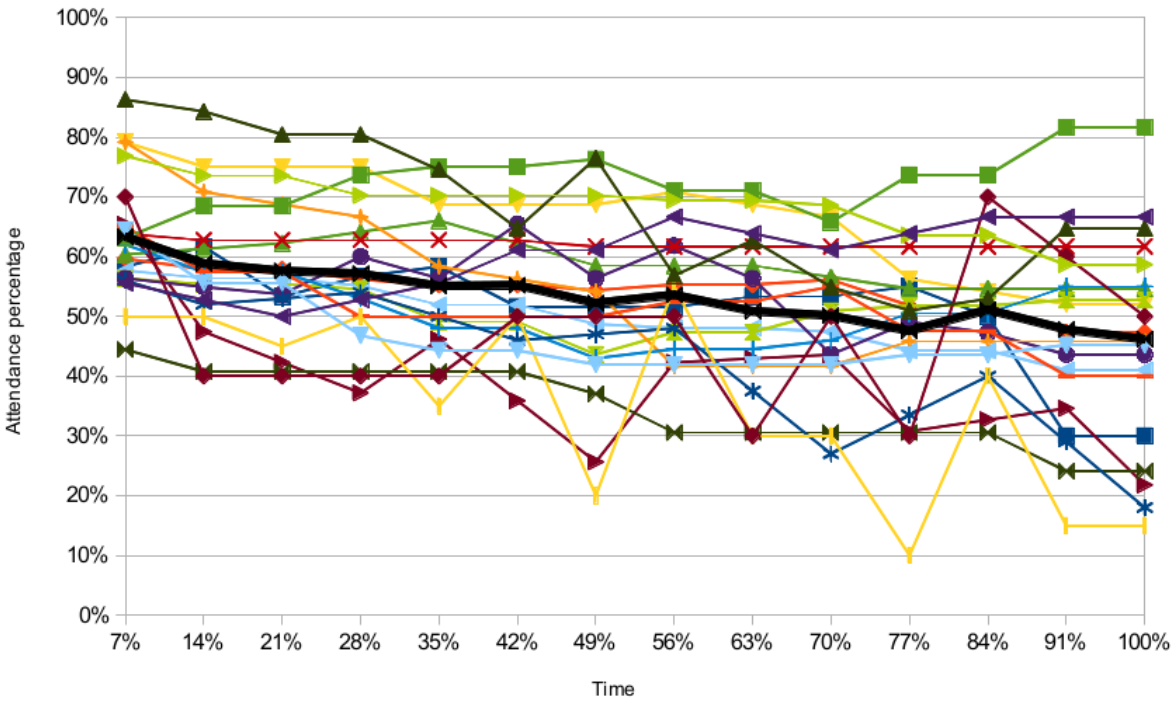

Figure 2.
Student attendance in percent, as the course progresses (for legend, see Fig. 1. – thick line is average)

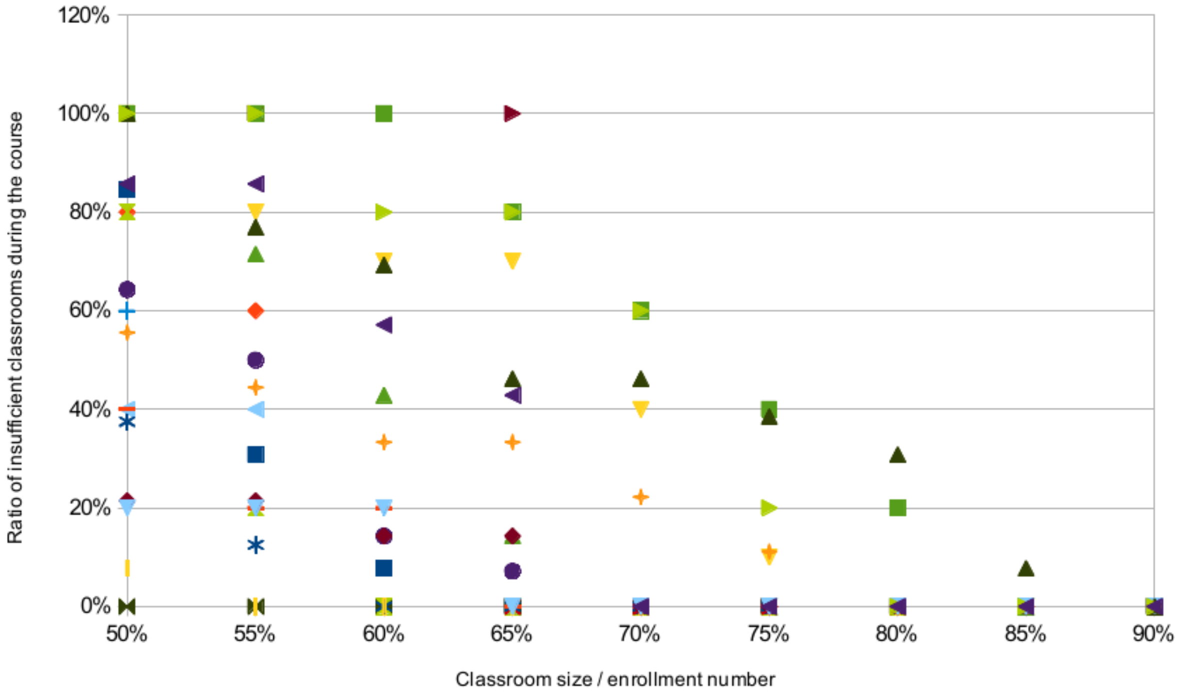

Figure 3.
In what percentage of the lectures will a smaller lab be insufficient? (for legend, see Fig. 1)

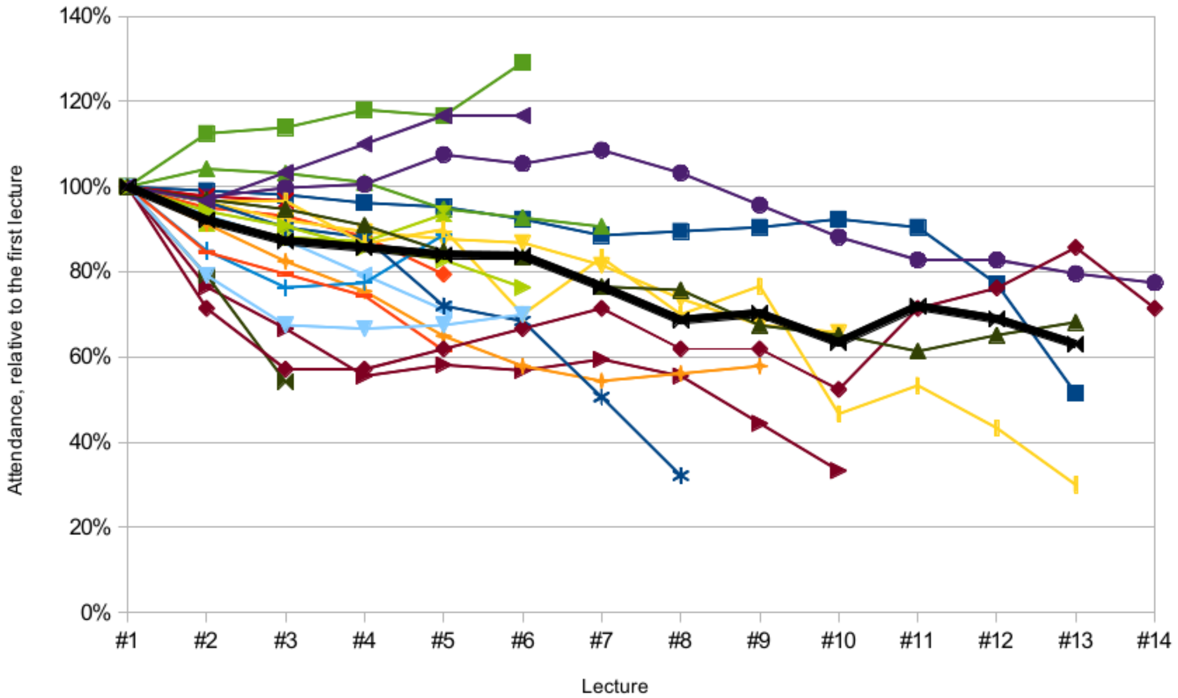

Figure 4.
Attendance during the course, relative to the first lecture (for legend, see Fig. 1. – thick line is average)

attendance data verified by a teacher was acceptable. Every course needed a snapshot of the "official enrolled students list" at the time of the first lecture. Courses with less than 10 students at the first lecture were not considered. Only courses with full attendance statistics for the first 5 lectures were considered. Courses needed to have attendance statistics for at least 90% of the lectures.

3.2 Input data

The input data was originally provided in several different formats, including printed attendance sheets with the students' signatures, lecturer's notebook with marks next to the students' names and LibreOffice spreadsheets.

3.3 Data processing

All input data was keyed into a spreadsheet.

Some data points were omitted and some data points were merged as consulting with the lecturers revealed that mid-term assignments and tests caused both positive and negative spikes in the attendance data. Positive spikes at tests are quite natural as the majority of students consider the tests more important than the lectures – yet, as this analysis aims to provide input for scheduling test, the positive spike caused by scheduled tests should be removed from the data. Negative spikes are a bit more interesting – some lecturers split the class to smaller groups for mid-semester tests so that two or more data points should be merged into a single data point to restore the real student attendance data.

4. ANALYSIS

The goal of the analysis was less quite pragmatical – to provide "rules of thumb" for scheduling classrooms and mid-semester tests. Originally, the author hoped to establish a numerical formula for attendance prediction with only a few variables – however, the variance in the data disproved each and every attempt to provide such a formula.

No common curve was found. For example, course "2013/2 LSzU N FSZ L 3" is having plenty of ripples caused by the timetable of the course – two 3*45 minute blocks separated by a lunch break resulted in many students only attending the morning lectures. Course "2012/1 LSzA N FSZ L 2", on the other hand, has a quiet flat attendance curve that increases at the beginning, in contrary of the majority of attendance curves. Course "2013/1 MatAlap N BSc Gy 0" is a 0^{th} semester course, therefore enrollment was continuous during the whole course, resulting a much higher attendance rate at the end of the course than at the beginning of the course.

5. CONCLUSIONS

Average student attendance is about 60% at the beginning of the course and is about 45% at the end of the course – with laboratory courses, freshman courses and Math courses having a higher attendance.

Most courses will never have 70% or more of enrolled students present and only Math classes will have attendance above 80%.

Theory lectures can be carried out in classrooms with a capacity of only 70% of the enrolled students, except for Math lectures, in which case the needed capacity is 85%-90%. For laboratory lectures, this ratio is 80%.

In the majority of the cases, the attendance is the highest at the first lecture – and with the exception of Math, attendance will never be more that 110% of the attendance of the first lecture.

Attendance relative to the first lecture varies wildly – 71% to 113% on the third lecture, 54% to 114% on the second lecture. On average, it decreases about 3% per lecture.

6. REFERENCES

[1] http://www.education.nt.gov.au/students/at-school/enrolment-attendance/enrolment-attendance-statistics