

Case Study 2: Students' Grades

Ling-Chieh Kung*

Note. For this case study, you will be given *real* data provided by two European high school. This data set is organized by the instructor specifically for this course. Therefore, they may be noisy and not “perfect” for a class practice. Nevertheless, they give us a chance to taste how data analysis may really help business decision makers in practice.

1 Introduction

We all care about education, and we all (somewhat) care about grades. Typically we rely on grades to evaluate the effectiveness of an education system, ability/effort of an instructor, and performance of students. In many cases, educators want to understand what makes students perform differently, i.e., to find factors to *explain* the variability of students' grades. Regression is typically the first (and in many cases the only) method to use.

In this case study, you will be given a real data set from two public senior high schools in a country at Europe. Students voluntarily participated in the study are given questionnaires for them to provide some information about themselves. These information, as well as their final grades in mathematics and literature in a semester, are given to you (with students' identities removed). Your task is, for each subject, to explain why some students did well while some did not. In other words, you should identify factors that significantly affect students' grades in each subject. It is also interesting to see whether the factors are different for different subjects. Finally, given your findings, please make some suggestions to the educators in the two schools.

*Department of Information Management, National Taiwan University; lckung@ntu.edu.tw.

2 Data

The data are given in the MS Excel file “SDA106-1_case2_data.xlsx,” where the sheets “Math” and “Literature” contain the data for mathematics and literature, respectively. For mathematics, we have 395 students participated in the study. For literature, we have 649. Please note that because some students participated in both subjects, you may see two rows in the two sheets containing identical background information. While the two rows may come from one single student, you have no proof for that.

1. *school*: student’s school (“GP” for school 1 or “MS” for school 2).
2. *gender*: student’s gender (“F” for female or “M” for male).
3. *Pstatus*: parents’ cohabitation status (“T” for living together or “A” for apart).
4. *Medu*: mother’s education (0 for none, 1 for elementary school, 2 for junior school, 3 for senior high school, or 4 for higher education).
5. *Fedu*: father’s education (0 for none, 1 for elementary school, 2 for junior school, 3 for senior high school, or 4 for higher education).
6. *traveltime*: home to school travel time (1 for less than 15 minutes, 2 for 15 to 30 minutes, 3 for 30 minutes to 1 hour, or 4 for above 1 hour).
7. *studytime*: weekly study time (1 for less than 2 hours, 2 for 2 to 5 hours, 3 for 5 to 10 hours, or 4 for above 10 hours).
8. *paid*: extra paid classes within the course subject (“yes” or “no”).
9. *higher*: wants to take higher education (“yes” or “no”).
10. *internet*: Internet access at home (“yes” or “no”).
11. *romantic*: with a romantic relationship (“yes” or “no”).
12. *famrel*: quality of family relationships (from 1 for very bad to 5 for excellent).
13. *freetime*: free time after school (from 1 for very low to 5 for very high).
14. *Dalc*: weekday alcohol consumption (from 1 for very low to 5 for very high).

15. *Walc*: weekend alcohol consumption (from 1 for very low to 5 for very high).
16. *health*: current health status (from 1 for very bad to 5 for very good).
17. *absences*: number of missed class hours (for each subject, there are 93 class hours in total).
18. *grades*: final grades (from 0 to 20).

3 Tasks

The educators in the two schools are waiting for your report and/or presentation, which should address the following:

1. Do regression analysis to find factors that affect students' grades in each subject. Use those factors to explain the performance variability among students. For each factors, gives qualitative or quantitative interpretations as much as possible.
2. Given some practical suggestions to the educators about improving students' grades.

Please do not forget that while the key part of your analysis is regression, you always start from descriptive statistics. Include some figures and numbers that you believe important in your report/presentation.

4 Submission rules

Each team should have at most six students unless a special approval is obtained from the instructor. Each team must submit a written report to address the above tasks. You do not need to include the details of your solution process, but you may want to summarize your procedure. Focus more on the presenting your results.

The report, including everything (cover page, appendix, etc.), cannot be longer than **12 pages** (i.e., six double-sided pieces of paper). A report is considered good if it addresses the tasks correctly, precisely, and concisely. Including managerial implications found by analyzing the data is always a plus. Obviously, a well-formatted report is expected.

A hard copy of the written report must be submitted in class by **6:45 pm, December 13**. Electronic copies of the report (for all teams) and slides (for the presenting teams) should be uploaded to CEIBA by the same due time. For each team, only one member should upload the files. Submissions between 6:45 pm and 7:45 pm on the due date will get 20% off as a penalty. Submissions later than 7:45 pm are not accepted.

The report (for all teams) counts for 10% of the semester grades. The oral presentation (for presenting teams) counts for another 10%.