



Assignments	14-15		
Title	Student Performance Data Set		
Data files	14-15_student.zip		
Sample size	649	Number of variables	33

This dataset approach student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. Two datasets are provided regarding the performance in two distinct subjects: Mathematics (mat) and Portuguese language (por). In [Cortez and Silva, 2008], the two datasets were modeled under binary/five-level classification and regression tasks. Important note: the target attribute G3 has a strong correlation with attributes G2 and G1. This occurs because G3 is the final year grade (issued at the 3rd period), while G1 and G2 correspond to the 1st and 2nd period grades. It is more difficult to predict G3 without G2 and G1, but such prediction is much more useful (see paper source for more details)

Attribute Information:

Attributes for both student-mat.csv (Math course) and student-por.csv (Portuguese language course) datasets:

- 1 school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
- 2 sex - student's sex (binary: 'F' - female or 'M' - male)
- 3 age - student's age (numeric: from 15 to 22)
- 4 address - student's home address type (binary: 'U' - urban or 'R' - rural)
- 5 famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
- 6 Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
- 7 Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- 8 Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education)
- 9 Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
- 10 Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at_home' or 'other')
- 11 reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
- 12 guardian - student's guardian (nominal: 'mother', 'father' or 'other')
- 13 traveltime - home to school travel time (numeric: 1 - <15 min., 2 - 15 to 30 min., 3 - 30 min. to 1 hour, or 4 - >1 hour)
- 14 studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)
- 15 failures - number of past class failures (numeric: n if $1 \leq n < 3$, else 4)

16 schoolsup - extra educational support (binary: yes or no)
 17 famsup - family educational support (binary: yes or no)
 18 paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
 19 activities - extra-curricular activities (binary: yes or no)
 20 nursery - attended nursery school (binary: yes or no)
 21 higher - wants to take higher education (binary: yes or no)
 22 internet - Internet access at home (binary: yes or no)
 23 romantic - with a romantic relationship (binary: yes or no)
 24 famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
 25 freetime - free time after school (numeric: from 1 - very low to 5 - very high)
 26 goout - going out with friends (numeric: from 1 - very low to 5 - very high)
 27 Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
 28 Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
 29 health - current health status (numeric: from 1 - very bad to 5 - very good)
 30 absences - number of school absences (numeric: from 0 to 93)

these grades are related with the course subject, Math or Portuguese:

31 G1 - first period grade (numeric: from 0 to 20)
 31 G2 - second period grade (numeric: from 0 to 20)
 32 G3 - final grade (numeric: from 0 to 20, output target)

References

- P. Cortez and A. Silva (2008). Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., *Proceedings of 5th FUTURE BUSINESS TECHNOLOGY CONFERENCE (FUBUTEC 2008)* pp. 5-12, Porto, Portugal, EUROESIS, ISBN 978-9077381-39-7.
Available at: <http://www3.dsi.uminho.pt/pcortez/student.pdf>.
- Lichman, M. (2013). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml>]. Irvine, CA: University of California, School of Information and Computer Science..

RESEARCH TASK FOR ASSIGNMENT 14

Analyze the above data and setup a model to predict the final and the intermediate grades for Maths.

RESEARCH TASK FOR ASSIGNMENT 15

Analyze the above data and setup a model to predict the final and the intermediate grades for Portuguese.

RESEARCH TASKS FOR BOTH ASSIGNMENTS 14 & 15

- Perform an exploratory analysis for your dataset.
- Use visual methods to describe the most important findings.
- Study the pairwise associations between the variables
- Construct regression models to assess your responses.
- Both Assignments (with collaboration):** Compare the two models. What are the differences between the two subjects?