

Statistic Model to Analyze Student's Performance
Group No. 8

Question 1 (Name of all group members):

1. Xuchuan Zheng
2. Sungki Park
3. Prashant Sharma

Question 2 (Introduction):

If you have research topics in mind for this project, please briefly describe the topics and any background info I may need to understand the topic. What problems are you going to address? Why these topics are important to you? What are your goals?

We have decided to work on Student Performance dataset, to understand the impact of demographic and social factors on the performance of student in school. This dataset contains data from 2 Portuguese schools for 2 distinct subjects: Math and Portuguese, along with various other parameters like parent's education level, living situation of parents, family size, and many other.

Academic performance, though it is not the only factor but is very crucial in shaping a student's future, it not only a measure of student's current achievement but also plays an important role in their higher studies and economic progress. Getting into a good school is an oblivious choice for any parent, but that's not the only factor that determines performance of a student, otherwise every student in a good school will have great grades. Sometimes good students also get bad grades, so there must be something more than just good school and good teachers, that is impacting the performance of a student.

By analyzing the data, we are hoping to gain some insight about factors that goes beyond school hours. How the performance of a student's gets affected by their social and demographic situation.

Question 3 (Methodology):

Please briefly describe the data you have (or plan to acquire) to help answer the research topics above. Include: what type of variable or variables are included (quantitative, qualitative, etc.), how the variable or variables are measured (the measurement scale), and any other general info you may have on the variable(s)

We got the dataset from (UC Irvine Machine Learning Repository, n.d.) , it contains data for 2 subject Math and Portuguese from 2 public schools from Alentejo region of Portugal, data was collected via survey. The dataset has 33 variables and data is in 2 separate csv files. Below table provide the details about the variables:

Type of Variable	Variables Name And Description
Quantitative	<ul style="list-style-type: none"> age - student's age (numeric: from 15 to 22) Medu: mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) Fedu: father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 - 5th to 9th grade, 3 - secondary education or 4 - higher education) famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)Home to school travel time (numeric: 1 – < 15 min., 2 – 15 to 30 min., 3 – 30 min. to 1 hour or 4 – > 1 hour) studytime - weekly study time (numeric: 1 - <2 hours, 2 - 2 to 5 hours, 3 - 5 to 10 hours, or 4 - >10 hours)Number of past class failures (numeric: n if $0 \leq n < 3$, else 4) freetime - free time after school (numeric: from 1 - very low to 5 - very high) goout - going out with friends (numeric: from 1 - very low to 5 - very high) Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high) Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high) health - current health status (numeric: from 1 - very bad to 5 - very good) absences - number of school absences (numeric: from 0 to 93) G1 - first period grade (numeric: from 0 to 20) G2 - second period grade (numeric: from 0 to 20) G3 - final grade (numeric: from 0 to 20, output target)
Qualitative	<ul style="list-style-type: none"> sex - student's sex (binary: "F" - female or "M" - male)School (binary: Gabriel Pereira or Mousinho da Silveira) address - student's home address type (binary: "U" - urban or "R" - rural) Pstatus - parent's cohabitation status (binary: "T" - living together or "A" - apart) Mjob - mother's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other") Fjob - father's job (nominal: "teacher", "health" care related, civil "services" (e.g. administrative or police), "at_home" or "other") guardian - student's guardian (nominal: "mother", "father" or "other") famsize - family size (binary: "LE3" - less or equal to 3 or "GT3" - greater than 3) reason - reason to choose this school (nominal: close to "home", school "reputation", "course" preference or "other") schoolsup - extra educational support (binary: yes or no)

	<ul style="list-style-type: none"> • famsup - family educational support (binary: yes or no) • activities - extra-curricular activities (binary: yes or no) • paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no) • internet - Internet access at home (binary: yes or no) • nursery - attended nursery school (binary: yes or no) • higher - wants to take higher education (binary: yes or no) • romantic - with a romantic relationship (binary: yes or no)
--	--

Question 4 (Methodology cont.):

Is this your own data set (or the data of someone in the group) or is it "open" or "shared" data?

The dataset is from (UC Irvine Machine Learning Repository, n.d.) , it is an open dataset under the name [Student Performance](#) (Cortez, 2014) .

Question 5 (Methodology cont.):

Have you distributed the workload among your teammates? If yes, please describe the group members' workload distribution and responsibilities.

Yes, we have distributed the workload among our teammates.

Tasks	Sub tasks	Name
Introduction	Motivation	Sungki Park
	Objectives	Xuchuan Zhen
Methodology	Data	Prashant Sharma
	Approach	Sungki Park
	Workflow	Xuchuan Zhen
	Workload Distribution	Prashant Sharma
Main results of the analysis	Variable Selection Procedures	Sungki Park
	Hypothesis Statement for Individual T-tests	Xuchuan Zhen
	Higher Order Individual T-tests	Prashant Sharma
	Hypothesis Statement for Individual T-tests (Interaction Terms)	Sungki Park
	Interaction Term T-tests	Xuchuan Zhen
	Hypothesis Statement for ANOVA Test	Prashant Sharma
	<ul style="list-style-type: none"> • Multiple Regression Assumptions <ul style="list-style-type: none"> ○ Linearity Assumption ○ Independence Assumption ○ Normality Assumption ○ Equal Variance Assumption ○ Multicollinearity Tests ○ Influential Points and Outliers ○ Interpreting Coefficients 	Sungki Park, Xuchuan Zhen, Prashant Sharma

	○ Prediction	
Conclusion and Discussion		Sungki Park
Reference		Xuchuan Zhen
Appendix		Prashant Sharma

End of Project Checkpoint