W200 Python Fundamentals – Section 2

Project 2 Proposal

Team 3 – Siobhan Harrington, Andres Kodaka, Jeff Braun

The objective of the PISA project is to rate educational performance of countries in the areas of reading, science and math, as well as multiple sub-topic areas. Fifteen year old students, their parents and the administrators of their schools are surveyed along multiple dimensions to build a picture of the school and home environments, economic conditions, and cultural norms that may contribute to differences in academic performance.

Our hypothesis is that certain of the environmental, economic and cultural factors surveyed will show a strong correlation to the differences in student performance by country. For example, differences in the length of a school day, or access to the Internet in school and at home, may show a strong correlation to differences in student performance by country in reading, science and math. Some of the correlations may differ based on gender, so we will apply a gender view to the analysis we perform.

Data To Be Used in Analysis

The PISA data we plan to analyze are as follows:

Dependent variables:

- Plausible* Value in Reading
- Plausible* Value in Science
- Plausible* Value in Math

**Note that students do not answer all test questions. Each student answers a subset. A Rasch statistical model is then used by PISA to predict the student's answers to all other questions. For the purpose of this project we are assuming the mean score of the 10 plausible subscores per subject area is within the margin of error as documented by PISA.*

Independent variables:

Demographic: How do demographics impact student scores?

- Student (Standardized) Birth - Month
- Student (Standardized) Gender
- Age

Socioeconomic:: The survey asked socioeconomic questions in a variety of ways. **How does income/socioeconomic status impact student scores as measured in these survey question?** These questions will require extra data validation. For example, a household income field is provided. If this field is particularly high or low and conflicts with some of the other information (ie: amount paid to tutors) then this data might need further validation.

- What is the <highest level of schooling> completed by your mother?
- What is the <highest level of schooling> completed by your father?
- In your home: A desk to study at
- In your home: A room of your own

- In your home: A quiet place to study
- In your home: A computer you can use for school work
- In your home: A link to the Internet
- How many in your home: Musical instruments (e.g. guitar, piano)
- How many books are there in your home?
- After leaving school did you: Work in the household or take care of other family members
- After leaving school did you: Work for pay
- Digital devices available at school
- What is your annual household income?
- Eat <the main meal> with my child around a table.
- In the last twelve months, about how much would you have paid to educational providers for services?

Culture/Motivation/Study Habits/Family Dynamic: The survey asked for rich information about students' motivation, learning environment, and family dynamics. We will study this section of questions to understand how home/school life impact scores. This section is vast but we suspect some variables will correlate more than others as we become more familiar with the data.

- I want top grades in most or all of my courses.
- I want to be able to select from among the best opportunities available when I graduate.
- Number of <class periods> required per week in <test language>
- Number of <class periods> required per week in mathematics
- Number of <class periods> required per week in <science>
- In a normal, full week at school, how many <class periods> are you required to attend in total?
- How many minutes, on average, are there in a <class period>?
- In the last two full weeks of school, how often: I <skipped> a whole school day
- In the last two full weeks of school, how often: I <skipped> some classes
- In the last two full weeks of school, how often: I arrived late for school
- This school year, approximately how many hours per week do you spend learning in addition? <School Science>
- On avg, how many days do you attend physical education classes each week?
- How often does this happen in your <school science> lessons? Students don't listen to what the teacher says.
- How often does this happen in your <school science> lessons? There is noise and disorder.
- How often does this happen in your <school science> lessons? The teacher waits long for students to quiet down.
- How often does this happen in your <school science> lessons? Students cannot work well.
- How often does this happen in your <school science> lessons? Students don't start working for a long time after.
- Many things I learn in my <school science> subject(s) will help me to get a job.
- Approx how many hrs\week attend add. instruct in the follow. domains? <School science> or <broad science>
- Approx how many hrs\week attend add. instruct in the follow. domains? Mathematics
- Approx how many hrs\week attend add. instruct in the follow. domains? <Test language>

- In your family, who helps you regularly with your homework or private study? Nobody (verification with other “help” questions”)
- Discuss how well my child is doing at school.
- Spend time just talking to my child.
- Ask how my child is performing in science class.

Analysis Method

Analyses we will perform include:

1. Start off with extensive cleaning of data (dropping rows where data was incomplete/unavailable), and data transformation (consolidating repetitive questions to make it more useful to our analysis).
2. Examine dependent and independent variables to count missing values. Drop observations or variables that have too many missing values to allow for meaningful analysis. For example, we will drop the observations for an entire country if a majority of its Plausible Value dependent variable values are missing. We will also drop an independent variable if there are not enough observations for that variable within each country to allow us to reach conclusions about that independent variable.
3. Generate histograms and box plots for the independent variables that remain after step 1 above in order to assess their distribution of values. We may drop an independent variable if the distribution of values leads us to suspect we do not have an unbiased set of sample data for that variable.
4. Calculate correlations for all dependent and independent variables after step 2 is completed. Group data by country for this and remaining steps. Do a separate set of correlations for male and female students to assess potential differences in correlation strength based upon gender.
5. For independent variables showing strong correlation to student performance, create line plots to identify potential linear trends between the independent variables and the performance variables

For example, the PISA test offered new evidence for the stereotype that girls are better at reading - the top deciles contain significantly more females than males. On the other hand, males did better at math/science. However, we want to dig into this data further to understand if this is true for every country and how other dependent variables impact/correlate the final scores. The rich survey data about home and school life for students around the world gives us a lot of opportunity to find common factors across countries that lead to high scores, assuming this self reported data is complete and accurate.

In the final report, we will describe the interim conclusions we drew from each step of the analysis. We will also describe adjustments we made to our planned analysis based on our interim conclusions, including any adjustments to the data being used in the analysis. Finally, we will state whether our analysis tends to support or refute our hypothesis.

Appendix

Other interesting variables available in the PISA dataset (used if time allows).

- Before going to school did you: Eat breakfast
- Before going to school did you: Study for school or homework
- Before going to school did you: Watch TV\<DVD>\Video
- Before going to school did you: Read a book\newspaper\magazine
- Before going to school did you: Internet\Chat\Social networks (e.g. <Facebook>, <country-specific social network>)
- Before going to school did you: Play video-games
- Before going to school did you: Meet friends or talk to friends on the phone
- Before going to school did you: Talk to your parents
- Before going to school did you: Work in the household or take care
- Before going to school did you: Work for pay
- Before going to school did you: Exercise or practice a sport
- After leaving school did you: Eat dinner
- After leaving school did you: Study\school\hmk
- After leaving school did you: Watch TV\<DVD>\Video
- After leaving school did you: Read a book\newspaper\magazine
- After leaving school did you: Internet\Chat\Social net (e.g. <Facebook>)
- After leaving school did you: Play video-games
- After leaving school did you: Meet friends or talk to friends on the phone
- After leaving school did you: Talk to your parents
- Available for you to use at home: Desktop computer
- Available for you to use at home: Portable laptop, or notebook
- Available for you to use at home: <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook>)
- Available for you to use at home: Internet connection
- Available for you to use at home: <Video games console>, e.g. <Sony® PlayStation®>
- Available for you to use at home: <Cell phone> (without Internet access)
- Available for you to use at home: <Cell phone> (with Internet access)
- Available for you to use at home: Portable music player (Mp3\Mp4 player, iPod® or similar)
- Available for you to use at home: Printer
- Available for you to use at home: USB (memory) stick
- Available for you to use at home: <ebook reader>, e.g. <Amazon® Kindle>
- *Digitial devices available at school: Desktop computer*
- *Digitial devices available at school: Portable laptop or notebook*
- *Digitial devices available at school: <Tablet computer> (e.g. <iPad®>, <BlackBerry® PlayBook>)*
- *Digitial devices available at school: Internet connected school computers*
- *Digitial devices available at school: Internet connection via wireless network*
- *Digitial devices available at school: Storage space for school-related data, e.g. a folder for own documents*
- *Digitial devices available at school: USB (memory) stick*
- *Digitial devices available at school: <ebook reader>, e.g. <Amazon® Kindle>*
- *Digitial devices available at school: Data projector, e.g. for slide presentations*
- *Digitial devices available at school: Interactive Whiteboard, e.g. <Smartboard®>*

- In your family, who helps you regularly with your homework or private study? Father or other male guardian
- In your family, who helps you regularly with your homework or private study? Sister(s)\brother(s)
- Help my child with his\her science homework.
- In your family, who helps you regularly with your homework or private study? Grandparents
- In your family, who helps you regularly with your homework or private study? Other relatives
- In your family, who helps you regularly with your homework or private study? Other person
- In your family, who helps you regularly with your homework or private study? Mother or other female guardian
- During a typical weekday, for how long do you use the Internet at school?
- During a typical weekday, for how long do you use the Internet outside of school?
- On a typical weekend day, for how long do you use the Internet outside of school?
- How old were you when you first used a computer?
- How old were you when you first accessed the Internet?
- The school is at a short distance to home.
- Moderate physical activities for a total of at least 60 minutes per day
- Vigorous physical activities for at least 20 minutes per day that made you sweat and breathe hard
- After leaving school did you: Exercise or practice a sport
- Approx how many hrs\week attend add. instruct in the follow. domains? <Foreign languages>
- Approx how many hrs\week attend add. instruct in the follow. domains? Social sciences
- Approx how many hrs\week attend add. instruct in the follow. domains? Music
- Approx how many hrs\week attend add. instruct in the follow. domains? Sports
- Approx how many hrs\week attend add. instruct in the follow. domains? Performing arts
- Approx how many hrs\week attend add. instruct in the follow. domains? Visual arts
- Approx how many hrs\week attend add. instruct in the follow. domains? Other