Fatemeh Mardi 12/31/2021

Final Project  
Checkpoint 1: Selecting your business issue and dataset

Put your name in the right-hand corner and type up your business issue and provide the link to your chosen dataset. Submit your document on the Canvas submission page.

Business issue: There is more to math than just a score. While educators and administrators know this, the impact of influential factors on math achievement scores is not clear enough for action to take place.

I want to explore public data collected from students across the nation to explore the constructs below and how they influence the math proficiency of students (across gender, race, and socio-economic status).

A picture containing text

Description automatically generated

Data: The dataset can be found here: <https://nces.ed.gov/onlinecodebook>  
Because of the large size (880MB) of the file, I have selected specific variables and created a smaller data subset which can be [viewed here](https://drive.google.com/file/d/1RaJlXvm-sHTD09ye1qqcQcJGRtOg75f_/view?usp=sharing).

These are the columns in the selected data set:

* Student ID
* Sex
* Race
* SES quintile
* Math score quintile X2TXMQUINT
* Math identity (X2MTHID)
* Math utility
* Math efficacy
* Math interest X2MTHINT

Variable name: X2SESQ5 Record #1, Position:963 - 964, Format: N2. Variable label: X2 Quintile coding of X2SES composite Variable description: - This variable is the quintile of X2SES, weighted using the student weight (W2STUDENT). For more information on this variable, please refer to section 7.3.2.2 of the HSLS:09 F1 Data File Documentation (NCES 2014-361).

Table

Description automatically generated

X2SEX is the sample member's sex. The composite is based on data from the BY student questionnaire, parent questionnaire, and/or school-provided sampling roster, and then updated when missing with data from the F1 student questionnaire.

Table

Description automatically generated with medium confidence

<https://nces.ed.gov/pubs2015/2015036_student.pdf>

X2RACE characterizes the sample member’s race/ethnicity by summarizing the following six dichotomous race/ethnicity composites: X2HISPANIC, X2WHITE, X2BLACK, X2ASIAN, X2PACISLE, and X2AMINDIAN. Each of these dichotomous composites is pulled from the base year composite when not imputed, otherwise it is based on data from the F1 student questionnaire; if still missing, they are based on the presence of the race/ethnicity from the F1 parent questionnaire (if parent questionnaire data includes race/ethnicity information for biological parents). X2RACE is derived from the six dichotomous race/ethnicity variables listed above (though the imputed values of X2WHITE, X2BLACK, X2ASIAN, X2PACISLE, and X2AMINDIAN are not stored on the data file).

Table

Description automatically generated

Table

Description automatically generated

Variable name: X2TXMQUINT Record #1, Position:803 - 804, Format: N2. Variable label: X2 Mathematics quintile score Variable description: - The math quintile score is a norm-referenced measure of achievement. The quintile score divides the weighted (population estimate) achievement distributions into five equal groups, based on math score (X2TXMTSCOR). Quintile 1 corresponds to the lowest-achieving one-fifth of the population, quintile 5 the highest. To determine the quintile cut-points, the weighted distribution of the standardized scores was divided at the 20th, 40th, 60th, and 80th percentiles. See Chapter 2 of the HSLS:09 F1 Data File Documentation (NCES 2014-361) for more information on the derivation of the math quintile score

Table

Description automatically generated

1. Your business issue and how you chose your dataset
2. Anything of note you learned in the EDA process
3. Anything of note you learned in the cleaning process
4. Your Tableau story

Variable name: X2MTHID Record #1, Position:998 - 1002, Format: N5.2 Variable label: X2 Scale of student's mathematics identity Variable description: - This variable is a scale of the sample member's math identity. Sample members who tend to agree with the statements "You see yourself as a math person" and/or "Others see me as a math person" will have higher values for X2MTHID. This variable was created through principal components factor analysis (weighted by W2STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were S2MPERSON1 and S2MPERSON2. Only respondents who provided a full set of responses were assigned a scale value. The coefficient of reliability (alpha) for the scale is .65. For more information on this scale score, please see chapter 5 of the F1 Data File Documentation (NCES 2014-361).

Table

Description automatically generated

Variable name: X2MTHUTI Record #1, Position:1003 - 1007, Format: N5.2 Variable label: X2 Scale of student's mathematics utility Variable description: - This variable is a scale of the sample member's perception of the utility of mathematics; higher values represent perceptions of greater mathematics utility. Variable was created through principal components factor analysis (weighted by W2STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were S2MUSELIFE, S2MUSECLG, and S2MUSEJOB. Only respondents who provided a full set of responses were assigned a scale value. The coefficient of reliability (alpha) for the scale is .65. For more information on this scale score, please see chapter 5 of the F1 Data File Documentation (NCES 2014-361).

Table

Description automatically generated

Variable name: X2MTHEFF Record #1, Position:1008 - 1012, Format: N5.2 Variable label: X2 Scale of student's mathematics self-efficacy Variable description: - This variable is a scale of the sample member's math self-efficacy; higher X2MTHEFF values represent higher math self-efficacy. Variable was created through principal components factor analysis (weighted by W2STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were S2MTESTS, S2MTEXTBOOK, S2MSKILLS, and S2MASSEXCL. Only respondents who provided a full set of responses were assigned a scale value. The coefficient of reliability (alpha) for the scale is .65. For more information on this scale score, please see chapter 5 of the F1 Data File Documentation (NCES 2014-361).

Table

Description automatically generated

Variable name: X2MTHINT Record #1, Position:1013 - 1017, Format: N5.2 Variable label: X2 Scale of student's interest in fall 2009 math course Variable description: - This variable is a scale of the sample member's interest in their base-year math course; higher values represent greater interest in their base-year math course. Variable was created through principal components factor analysis (weighted by W2STUDENT) and standardized to a mean of 0 and standard deviation of 1. The inputs to this scale were S2MENJOYING, S2MWASTE, S2MBORING, S2FAVSUBJ, and S2MENJOYS. Only respondents who provided a full set of responses were assigned a scale value. The coefficient of reliability (alpha) for the scale is .65. For more information on this scale score, please see chapter 5 of the F1 Data File Documentation (NCES 2014-361).

Table

Description automatically generated