COURSE PROJECT – Phase 1 Data Selection by Tabitha Hagen

Phase 1 Instructions:

In Phase 2 you will conduct a descriptive analysis and create a dashboard. In this phase you will select an appropriate dataset. Your dataset must be approved by Dr. Hill.

Dataset requirements:

- You must be able to **develop three questions** for your dashboard (see Phase 2 description) from your dataset
- At least 100 rows **220 observations**
- At least 5 columns **16 variables**
- A mixture of quantitative and categorical data (at least one column of each type) **Municipality is character and the rest are numeric**
- If you use a dataset from your work, please make sure that the dataset does NOT contain any proprietary information **N/A**

Write a short description of your dataset in a Word document (including three possible questions that might be asked of the data and answered with a dashboard).

Questions:

1. **How much money is being spent per student in different demographics?**
2. **How does the Student-to-teacher-ratio effect scores in 4th grade?**
3. **Which municipalities show the highest scores in 8th grade?**

Description of Dataset:

The dataset is self-described as containing "data on test performance, school characteristics, and student demographic backgrounds for school districts in Massachusetts from 1998. It was obtained from Stock and Watson to accompany their *Introduction to Econometrics*, 2nd edition. There are multiple demographic stats including English-learners, Special Education Learners, as well as Low-income Free or Reduced Lunches. They have included two gateway testing scores from the Massachusetts Comprehensive Assessment System (MCAS) test which was mandatory for all MA public schools. The Data on the average district income came from the 1990 US Census.

This data was found using a Google Search which led me to a site at Githib.io located at https://vincentarelbundock.github.io/Rdatasets/doc/AER/MASchools.html .

COURSE PROJECT – Phase 1 Data Selection by Tabitha Hagen

MASchools R Documentation

Massachusetts Test Score Data

Description

https://vincentarelbundock.github.io/Rdatasets/doc/AER/MASchools.html

The dataset contains data on test performance, school characteristics and student demographic backgrounds for school districts in Massachusetts.

Usage

```
data("MASchools")
```

Format

A data frame containing 220 observations on 16 variables.

district

character. District code.

municipality

character. Municipality name.

expreg

Expenditures per pupil, regular.

expspecial

Expenditures per pupil, special needs.

expbil

Expenditures per pupil, bilingual.

expocc

Expenditures per pupil, occupational.

exptot

Expenditures per pupil, total.

scratio

Students per computer.

special

Special education students (per cent).

lunch

Percent qualifying for reduced-price lunch.

stratio

Student-teacher ratio.

income

Per capita income.

score4

4th grade score (math + English + science).

score8

8th grade score (math + English + science).

salary

Average teacher salary.

english

Percent of English learners.

Details

The Massachusetts data are district-wide averages for public elementary school districts in 1998. The test score is taken from the Massachusetts Comprehensive Assessment System (MCAS) test, administered to all fourth graders in Massachusetts public schools in the spring of 1998. The test is sponsored by the Massachusetts Department of Education and is mandatory for all public schools. The data analyzed here are the overall total score, which is the sum of the scores on the English, Math, and Science portions of the test. Data on the student-teacher ratio, the percent of students receiving a subsidized lunch and on the percent of students still learning english are averages for each elementary school district for the 1997–1998 school year and were obtained from the Massachusetts department of education. Data on average district income are from the 1990 US Census.

Source

Online complements to Stock and Watson (2007).

References

Stock, J. H. and Watson, M. W. (2007). *Introduction to Econometrics*, 2nd ed. Boston: Addison Wesley.

See Also

https://vincentarelbundock.github.io/Rdatasets/doc/AER/MASchools.html

`StockWatson2007`, `CASchools`

Examples

```
## Massachusetts
data("MASchools")
```

COURSE PROJECT – Phase 1 Data Selection by Tabitha Hagen

```
## compare with California
data("CASchools")
CASchools$stratio <- with(CASchools, students/teachers)
CASchools$score4 <- with(CASchools, (math + read)/2)

## Stock and Watson, parts of Table 9.1, p. 330
vars <- c("score4", "stratio", "english", "lunch", "income")
cbind(
  CA_mean = sapply(CASchools[, vars], mean),
  CA_sd   = sapply(CASchools[, vars], sd),
  MA_mean = sapply(MASchools[, vars], mean),
  MA_sd   = sapply(MASchools[, vars], sd))

## Stock and Watson, Table 9.2, p. 332, col. (1)
fm1 <- lm(score4 ~ stratio, data = MASchools)
coeftest(fm1, vcov = vcovHC(fm1, type = "HC1"))

## More examples, notably the entire Table 9.2, can be found in:
## help("StockWatson2007")
```

https://vincentarelbundock.github.io/Rdatasets/doc/AER/MASchools.html