Empirical exercise – Data definition

The California Standardized Testing and Reporting (STAR) dataset contains data on test performance, school characteristics and student demographic backgrounds. The data used here are from all 420 K-6 and K-8 districts in California with the data collected in 1999. All of these data were obtained from the California Department of Education (www.cde.ca.gov).

Test scores are the average of the reading and math scores on the Stanford 9 standardised test administered to 5th grade students.

School characteristics (averaged across the district) include enrolment, number of teachers (measured as full time-equivalents), number of computers per classroom, and expenditures per student.

The student-teacher ratio used here is the number of full-time equivalent teachers in the district, divided by the number of students.

Demographic variables for the students also are averaged across the district. The demographic variables include the percentage of students in the public assistance program CalWorks (formerly AFDC), the percentage of students that qualify for a reduced price lunch, and the percentage of students that are English Learners (that is, students for whom English is a second language).

Series in Data Set:

county: County
district: District
dist_code: District code
avginc: District average income (in $1000's)
read_scr: Average reading score
math_scr: Average math score
testscr: Average test score (=(read_scr+math_scr)/2)
enrl_tot: Total enrollment
teachers: Number of teachers
str: Student teacher ratio (enrl_tot/teachers)
computer: Number of computers
comp_stu: Computers per student (=computer/enrl_tot)
gr_span: Grade span of district
expn_stu: Expenditures per student ($'s)
el_pct: Percent of English learners
meal_pct: Percent qualifying for reduced-price lunch
calw_pct: Percent qualifying for CalWorks.