

Exercise cashschool data – SBDM course

The data cashschool are from all 420 K-6 and K-8 districts in California with data available for 1998 and 1999. Test scores are on the Stanford 9 standardized test administered to 5th grade students. School characteristics (averaged across the district) include enrollment, number of teachers (measured as “full-time equivalents”, number of computers per classroom, and expenditures per student. Demographic variables for the students 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).

The data frame contains 420 observations on 14 variables.

Variables name:

- District: character. District code.
- School: character. School name.
- County: factor indicating county.
- Grades: factor indicating grade span of district.
- Students: Total enrollment.
- Teachers: Number of teachers.
- Students teach ratio (str): ratio Student/teacher
- Calworks: Percent qualifying for CalWorks (income assistance).
- Lunch: Percent qualifying for reduced-price lunch.
- Computer: Number of computers.
- Expenditure: Expenditure per student.
- Income (avginc): District average income (in USD 1,000).
- English: Percent of English learners.
- Read: Average reading score.
- Testscr: total test score

What you have to do: load cashschool.csv data

By using R, compute the following operations:

1. observe the data structure.
2. focusing on the test score variable (var.name= testscr), calculate the main position indicators (the mean, quantiles, min, max, mode), range, interquartile difference, asymmetry indicators.
3. Compute the mean of the test score variable by Grades; this is factor indicating grade span of district (var.name=gr_span).
4. Compute the variance, the standard deviation and coefficient of variation of the variable test score.
5. Compute variance and standard deviation of the variable test score by gr_span.

6. Represent the variable test score graphically using a boxplot.
7. Identify the presence of outliers in the variable student teacher ratio
8. Represent graphically income variable (var.name: avginc) with a histogram.