Analysis Write-up

Gleb Furman[1]

[1] Who Kneads a PH.D. Bakery

Analysis Write-up

Cluster Analysis

Population Frame

The population frame is composed of data from three sources: (1) the Common Core of Data (CCD), (2) publically available accountability data, and (3) the U.S. Census. The CCD is a comprehensive database housing anually collected national statistics of all public schools and districts. Accountability data was used to calculate the proportion of students within each school performing at or above proficiency in Math and ELA. Finally, local median income was obtained from the U.S. Census and was matched to each school by zipcode. School level data was aggregated to get district level variables. These are reported in Table @ref(tab:tbl_desc)

SUBS. Stratification using balanced sampling (SUBS) was performed prior to simulation because the group of schools in each strata would be static across condtions except where the balancing model is manipulated. The set of covariates in both the full model (SUBS-F) and the omitted variable model (SUBS-OV) include binary indicator variables ()

Table 1

Descriptives of variables

Variables	School		District Weighted		District Unweighted	
	Mean	SD	Mean	SD	Mean	SD
Number of Schools	NA	NA	9,875.00	0.00	4.84	12.72
School Size	579.07	203.19	534.77	225.34	534.77	225.34
Median Income	60,084.98	25,007.61	56,710.63	20,804.82	56,648.44	20,750.06
Average Proportions						
ELA Proficiency	0.59	0.23	0.60	0.20	0.60	0.20
Math Proficiency	0.53	0.29	0.54	0.26	0.54	0.26
Economically Disadvantaged	0.59	0.29	0.54	0.24	0.54	0.24
English Language Learners	0.23	0.21	0.14	0.17	0.14	0.17
Minority Status	0.65	0.30	0.46	0.32	0.46	0.32
Total/Free/Reduced Lunch	0.59	0.29	0.53	0.24	0.53	0.23
Indicators						
Urban	0.40	0.49	0.15	0.33	0.15	0.33
Suburban	0.41	0.49	0.33	0.44	0.33	0.44
Town or Rural	0.19	0.39	0.51	0.48	0.51	0.48

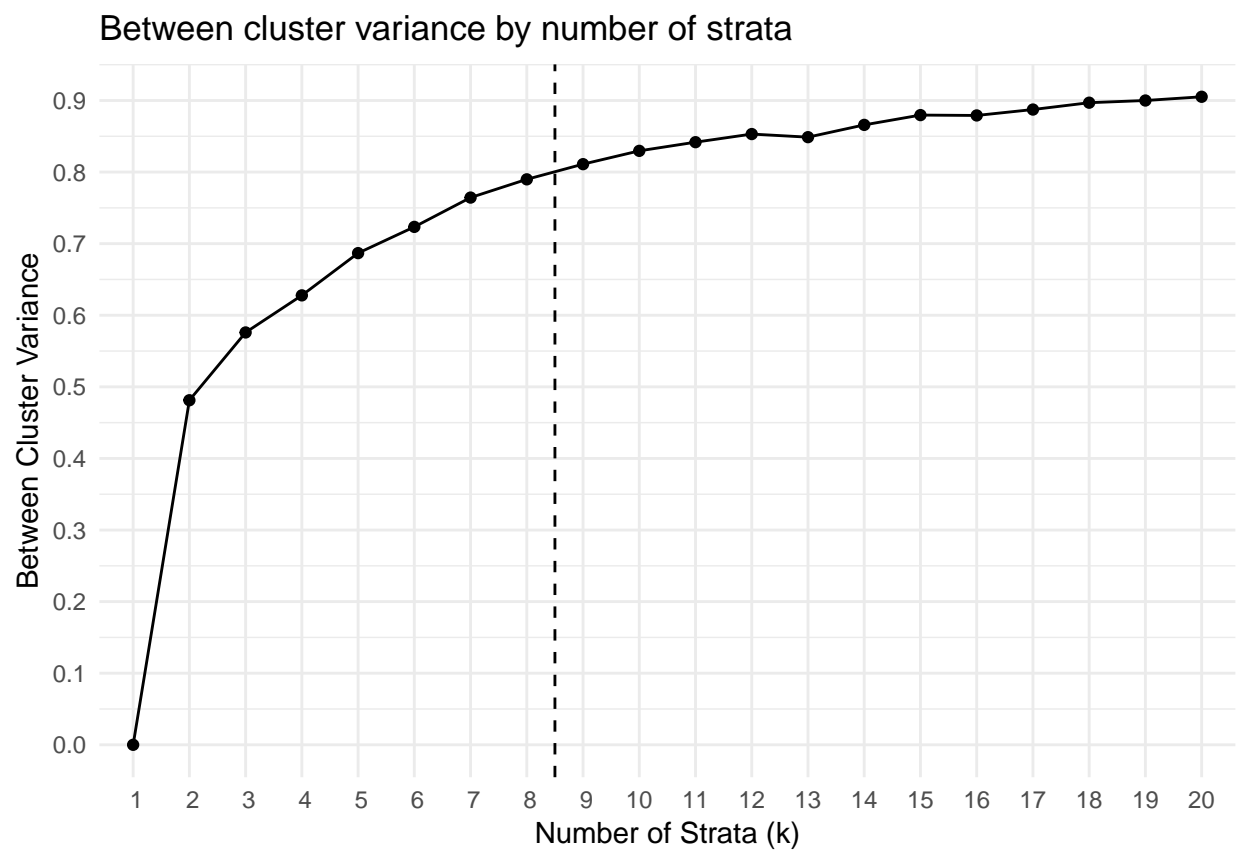*Note.* District variables are derived as aggregate means of school variables

Figure 1

References