Examples of CDF Gaps

David Rogosa

December 2003

Data. Stanford 9 scores Reading 4th grade 2000 and 2002. Group division into SD (socioeconomic disadvantaged) and non-SD. SD designation: student in NSLP or neither parent graduated from high school. Student score taken to be the reported national percentile rank (PR) from the reading test.

Analyses referenced to formulation and examples in "Two Measures Of Change In The Gaps Between The CDFs Of Test-Score Distributions" Paul W. Holland 1/11/02.

Unenlightened descriptive statistics. Someone unaware of the CDF gap approach would make (limited) "gapping" comparisons based on either the reported quantities PAC25, PAC50, PAC75 in the vertical direction or descriptive statistics for the percentile ranks in the horizontal direction.

Proportion at or above cut-score (PAC)

	2000				2002		
	PAC25	PAC50	PAC75		PAC25	PAC50	PAC75
SD	0.542	0.256	0.0884	SD	0.610	0.309	0.107
noSD	0.882	0.679	0.414	noSD	0.899	0.716	0.456
diff	0.339	0.423	0.325	diff	0.288	0.407	0.349

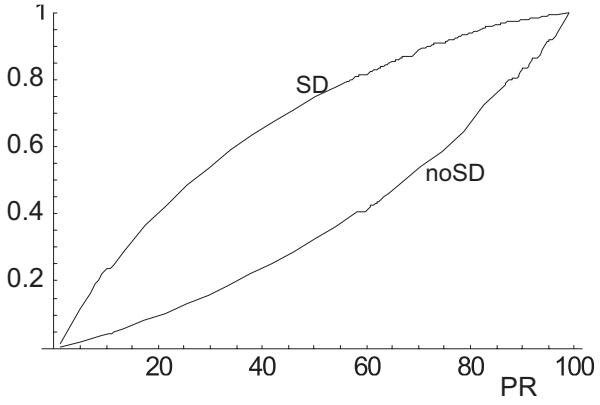
Improvement

	PAC25	PAC50	PAC75
SD2002 - SD2000			
	0.068	0.053	0.019
noSD2002 - noSD2000			
	0.017	0.037	0.043
diff2002 - diff2000			
	-0.051	-0.016	0.024

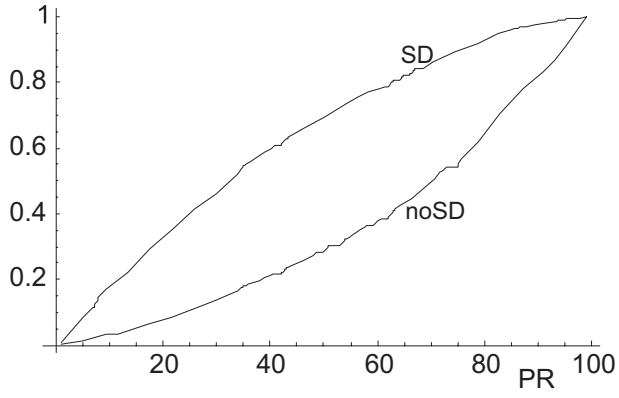
Percentiles of PR Distribution

	2000					2002				
	10th	25th	50th	75th	90th	10th	25th	50th	75th	90th
SD	4.18	11.5	27.1	50.2	72.1	5.94	15.1	31.9	56.0	75.6
noSD	20.8	40.9	67.1	84.4	94.1	23.7	45.0	70.5	85.5	94.6

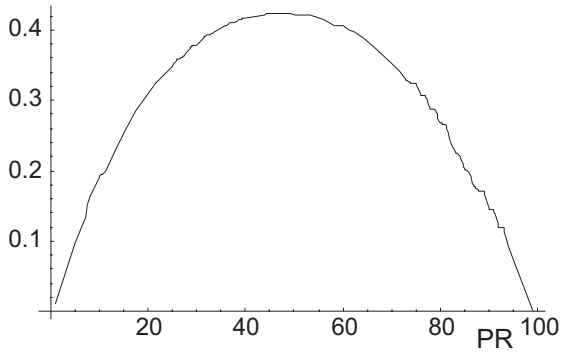
CDF 2000



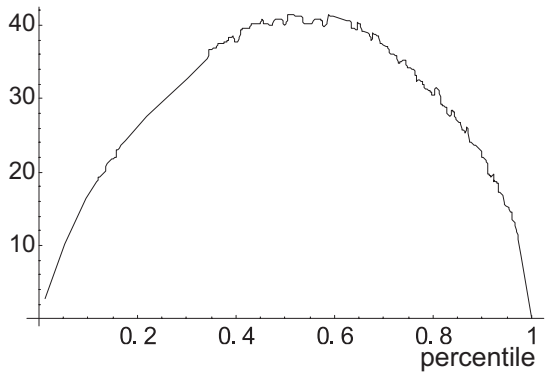
CDF 2002



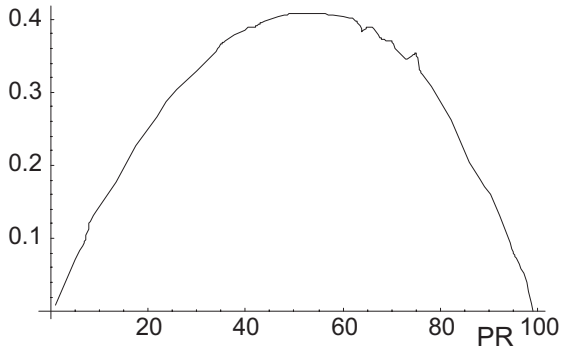
D-plot 2000



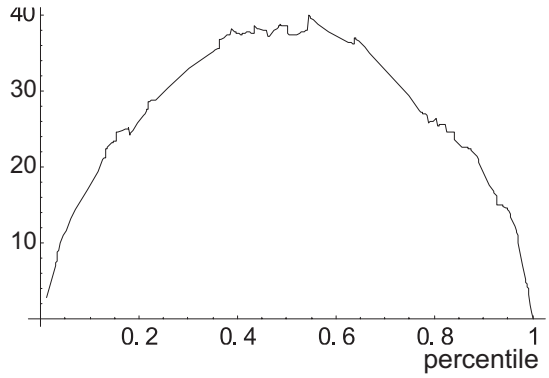
D*-plot 2000



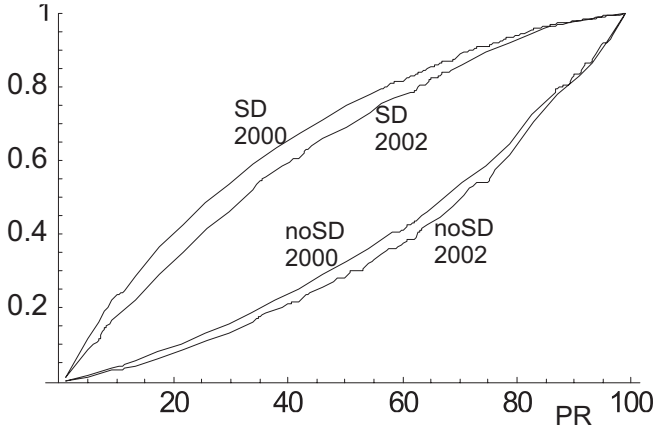
D-plot 2002



D*-plot 2002

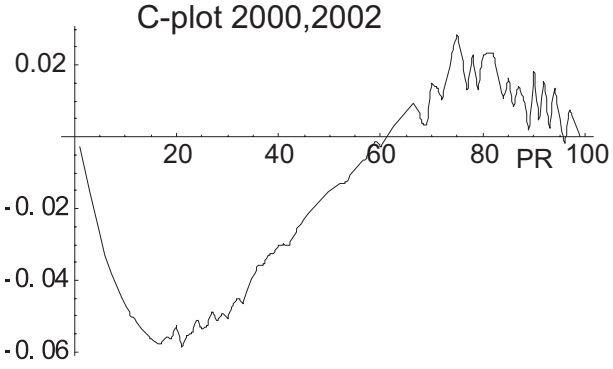


CDF Overlay 2000, 2002

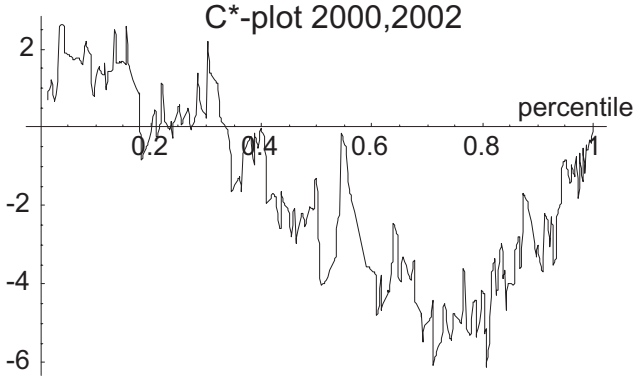


[note: plot is analog to Holland Fig 5]

C-plot 2000,2002



C*-plot 2000,2002



Notes/Comments.

The percentile rank metric (PR) for student scores removes the sigmoidal shape from the CDFs (for better or worse), perhaps allowing clearer graphical comparisons.

In terms of familiar quantities, D (vertical) represents differences between PACs, proportion of students above a stated national percentile rank, at all PR values.
D* (horizontal) provides the differences between descriptive stats, such as quartiles and median, on the national percentile ranks.

Arithmetic check of Holland Thm 1 using D* from the 2002 data.
Mean(noSD) - Mean(SD) = 27.69.
Ave(D*) = 27.85 [taken over uniform dist of percentiles].
Close enough (given the numerical details at the endpoints of inverse cdfs.)

With the PR metric D and D* plots in this example have similar humped shapes, but C and C* plots in some sense provide contradictory indications on gap closing. Does this pattern alter the recommendation favoring D* and esp C*?