Benchmarking Justice Performance

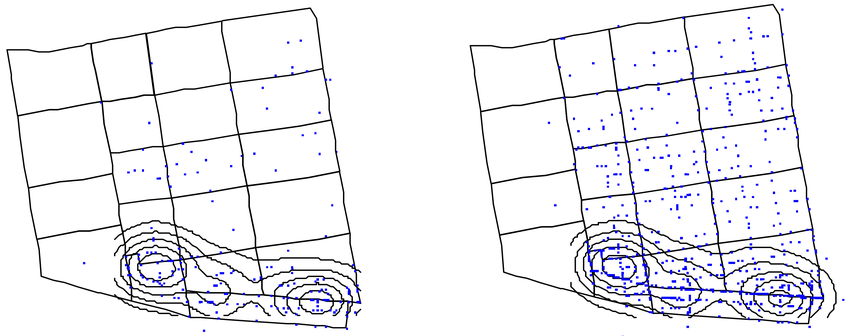
Greg Ridgeway

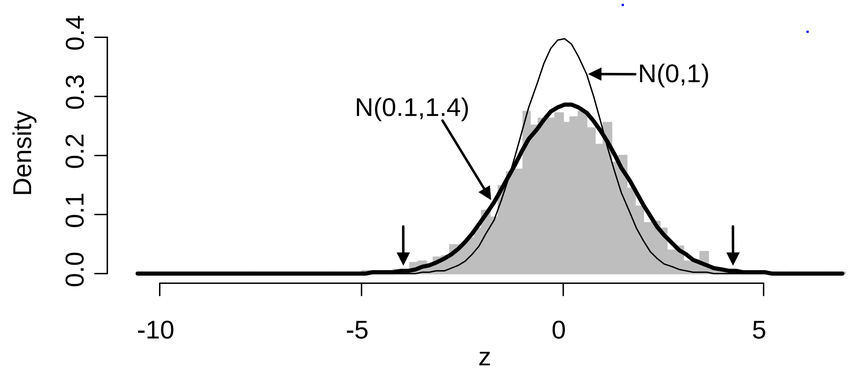
Associate Professor of Criminology and Statistics

gridge@sas.upenn.edu

Benchmarking Cops

Stop features		Officer 314 (%)	Benchmark (%)
		n = 392	ESS = 3,676
Month	January	3	3
	February	4	4
Day of the week	Monday	13	13
	Tuesday	11	10
Time of day	[12-2a.m.]	11	11
	(2-4a.m.]	5	5
Assignment	Patrol	100	100
In uniform	Yes	99	97
Radio run	Yes	1	3





Officer		Benchmark		
Black (%)	Stops (n)	Black (%)	Stops (n)	fdr
86	151	55	773	0.03
85	218	67	473	0.38
77	237	56	1,081	0.14
75	178	51	483	0.22
64	59	20	695	0.02

Methodology

- Aim: Construct benchmarks for assessing the performance of criminal justice system entities while accounting for confounding differences
- Step 1: Use propensity score weighting to construct a statistical benchmark from a comparison group so the features of an entity and the benchmark match

Find residents in other Cincinnati neighborhoods similar to Over-the-Rhine residents

Find stops made by other NYPD officers in the same time, place, and context

  - Weight benchmark elements as

$$w(\mathbf{x}) = P(t = 1|\mathbf{x})/P(t = 0|\mathbf{x})$$

- Step 2: Compare the entity and its benchmark on outcomes of interest

Compare the neighborhood with its benchmark on satisfaction with police service

Compare the officer with the benchmark for the percentage of black pedestrians stopped

  - Computed as the doubly robust estimator

$$\theta = \frac{\sum_{i=1}^{n} t_i\left(y_i - \hat{\beta}'\mathbf{x}_i\right)}{\sum_{i=1}^{n} t_i}$$

$\hat{\beta}$ maximizes $\sum w_i \ell(y_i, \beta'\mathbf{x}_i + \delta t_i)$

- Step 3: Repeat Step 1 and Step 2 for all entities. Collect and standardize all the benchmark comparisons. Flag outliers

Use Hochberg step-down to identify differences that are unusually large

Or calculate the false discovery rate, the probability of incorrectly flagging as an outlier

- Use with judges, prosecutors, defense counsel, courthouses, correctional officers to flag unusual sentencing practices, anomalous practice, excessive inmates injuries

- G. Ridgeway and J.M. MacDonald (2014). "A Method for Internal Benchmarking of Criminal Justice System Performance," *Crime & Delinquency* 60(1):145-162.
- G. Ridgeway and J.M. MacDonald (2009). "Doubly Robust Internal Benchmarking and False Discovery Rates for Detecting Racial Bias in Police Stops," *Journal of the American Statistical Association* 104(486):661–668.

Benchmarking Neighborhoods

Respondent features	Over-the-Rhine residents (n=146)	Benchmark residents in other neighborhoods (ESS=422)
High school or less (%)	21	21
Black (%)	66	65
$20K or less (%)	47	45
Employed (%)	60	58
Married (%)	15	16
Male (%)	43	42
Age 18-21 (%)	3	3
Homeowner (%)	20	21
Children at home (%)	40	38
Disorder (scale)	2.83	2.84
Fear of crime (scale)	2.64	2.69
Neighbors meet (scale)	2.38	2.38
Know an officer (%)	46	45