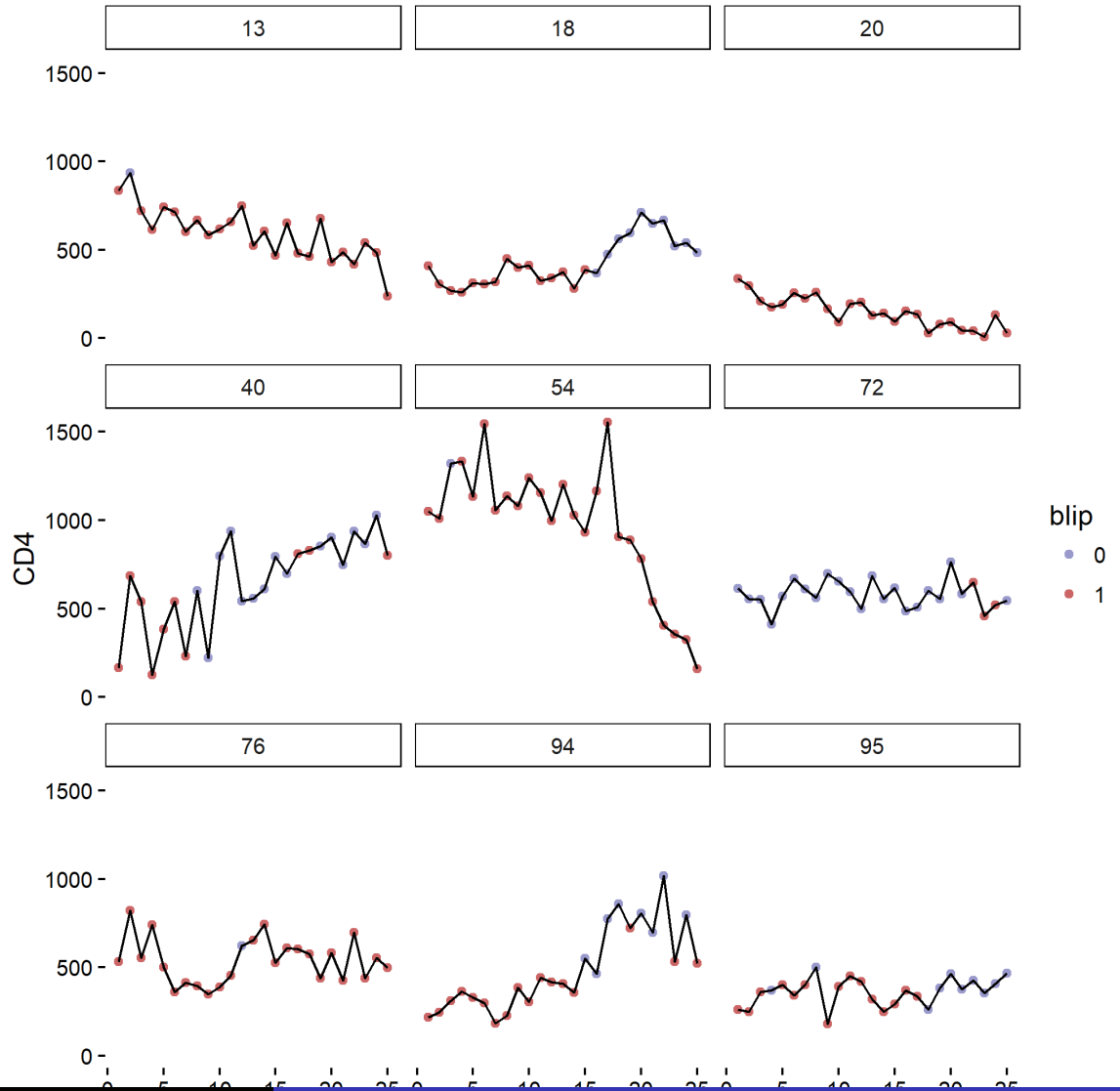
The Personalized and Population AUCs

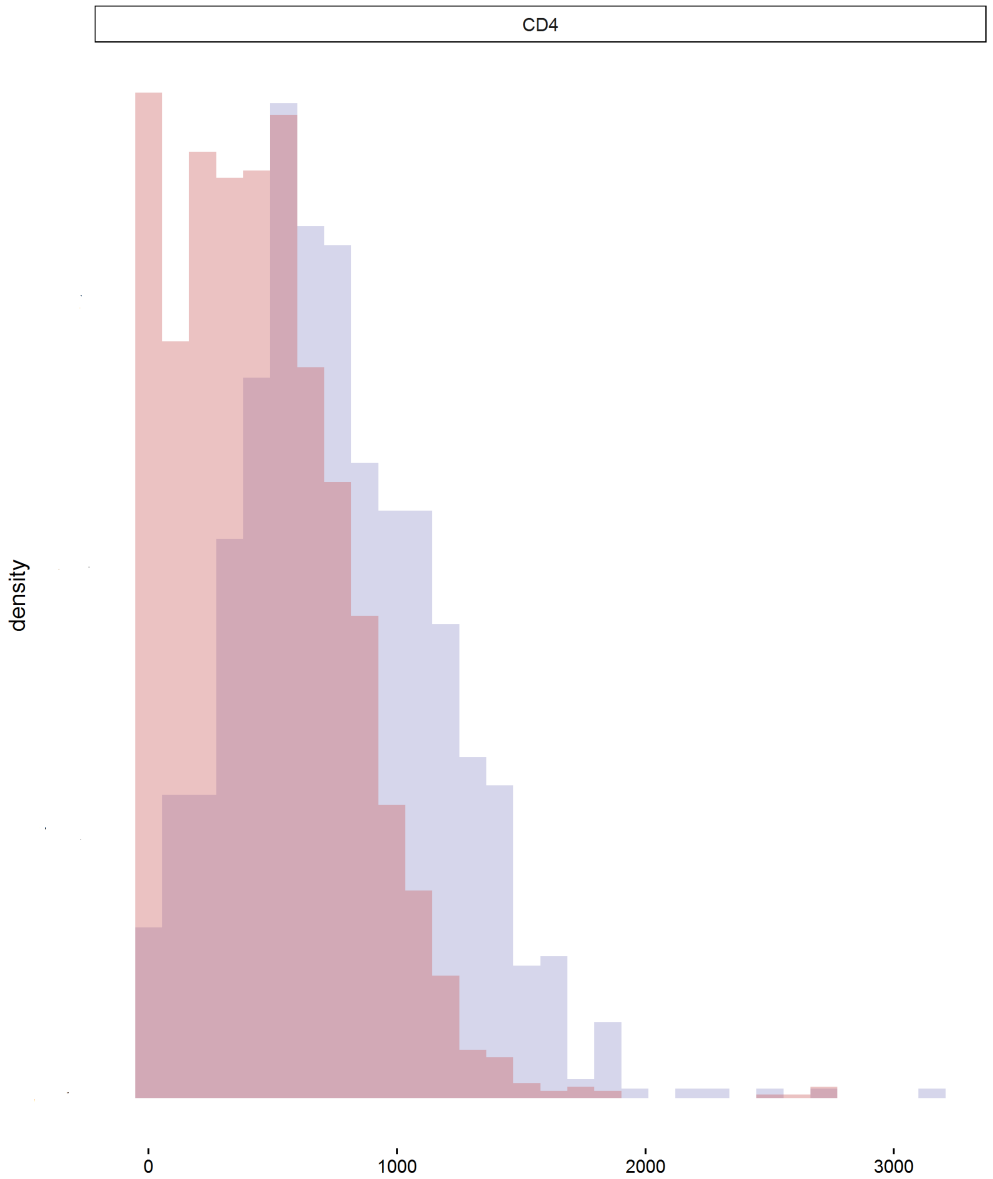
Outline

A motivating example

Data: The Yale Prospective Longitudinal HIV Cohort

Problem: Evaluate CD4 as a predictor of blip status




CD4
density
0
1000
2000
3000

	control	case
obs. # 1	X_1	
$\vdots$	$\vdots$	
obs. # k	X_k	
obs. # k+1		Y_{k+1}
$\vdots$		$\vdots$
obs. # l		Y_l

The AUC is the probability that an observation drawn from a negative/control/non-diseased subject is less than an independent observation from a positive/case/diseased subject.

$$\theta(P) = P(X < Y) = \mathrm{E}(F_X(Y))$$

$$\hat{\theta}(\vec{X}, \vec{Y}) = \frac{1}{k(l-k)} \sum_{i,j} \{X_i < Y_j\}$$

We wish to extend the AUC to clusters

- markers: longitudinal measurements of tumour antigens (CEA, CA15-3, TPS), response: progression/non-progression of breast cancer (Emir 2000)
- markers: how long an officer detains a suspect (clustered by officer), response: non-Black ("control") or Black ("case") suspect status (Ridgeway 2006)

	control	case
obs. # 1	$(X_{11}, \ldots, X_{1m_1})$	$(Y_{11}, \ldots, Y_{1n_1})$
$\vdots$	$\vdots$	$\vdots$
obs. # I	$(X_{I1}, \ldots, X_{Im_I})$	$(Y_{I1}, \ldots, Y_{In_I})$

- Let (X, Y, M, N) be a random vector with joint distribution denoted P
- X and Y are vectors of control and case observations of lengths M and N

With (X_1, Y_1, M_1, N_1) and (X_2, Y_2, M_2, N_2), being two independent draws from P, we define the population AUC as

$$\theta_{12}(P) = \frac{\mathrm{E}\,\hat{\theta}(X_1, Y_2)}{\mathrm{E}(M_1)\,\mathrm{E}(N_2)} = \frac{1}{\mathrm{E}(M_1)\,\mathrm{E}(N_2)}\,\mathrm{E}\left(\sum_{i=1}^{M_1}\sum_{j=1}^{N_2}\{X_{1i} < Y_{2j}\}\right)$$

$$(X_1, Y_1, M_1, N_1), (X_2, Y_2, M_2, N_2) \overset{\text{IID}}{\sim} P.$$

- the medical field has lately focused on personalizing treatment
- For example, in 2018 the National Academy of Medicine concluded: "The individuality of the patient should be at the core of every treatment decision. One-size-fits-all approaches to treating medical conditions are inadequate; instead, treatments should be tailored to individuals . . ."

Besides the population AUC

$$\theta_{12}(P) = \frac{\mathrm{E}\,\hat{\theta}(X_1, Y_2)}{\mathrm{E}(M_1)\,\mathrm{E}(N_2)}$$

we define the personalized AUC as:

$$\theta_{11}(P) = \mathrm{E}\left(\frac{\hat{\theta}(X_1, Y_1)}{M_1 N_1}\right) = \mathrm{E}\left(\frac{\sum_{i=1}^{M_1}\sum_{j=1}^{N_1}\{X_{1i} < Y_{1j}\}}{M_1 N_1}\right).$$

(Definition requires $M > 0, N > 0$.)

- Both the population and personalized AUC, like the usual AUC, are bounded between 0 and 1, $1/2$ represents poor discrimination, and distance from $1/2$ represents increasing discrimination.
- However, they describe distinct aspects of discrimination.
- Whereas the personalized AUC is the AUC of a typical cluster, the population AUC is the probability that a typical control observation in the population is less than a typical case observation.

Proposition

Let $(X_1, Y_1, M_1, N_1), \ldots, (X_I, Y_I, M_I, N_I)$, *be a random sample of size* I *IID according to* P. *Let* P_I *be the joint distribution of independent random selections from among the elements of* $X_1, \ldots, X_I$, *and* $Y_1, \ldots, Y_I$, *and let* $(\xi_I, \eta_I) \sim P_I$. *Then* $\theta(P_I) = Pr(\xi_I < \eta_I) + \frac{1}{2} Pr(\xi_I = \eta_I) \to \theta_{12}(P)$ *as* $I \to \infty$.

Outline

- random effects model

$$X \mid M, N \sim Z(M, N) + \xi_i^x, i = 1, \ldots, M$$
$$Y \mid M, N \sim Z(M, N) + \xi_j^y + \Delta, j = 1, \ldots, N$$

- $\Delta > 0$ is a non-random location shift between the control and case values
- Z is a random, cluster-level effect, inducing within-cluster dependence
- $\xi_i^x, \xi_j^y, i = 1, \ldots, M, j = 1, \ldots, N$, are IID individual effects

The personalized AUC is

$$\theta_{11} = P(\xi_1 - \xi_2 < \Delta)$$

The population AUC is

$$\theta_{12} = \mathrm{E}\left(\frac{M_1 N_2}{\mathrm{E}(M)\,\mathrm{E}(N)}\{Z_1 - Z_2 + (\xi^x - \xi^y) < \Delta\}\right)$$

A covariance-like term lying between 0 and 1

rug plots of fifteen clusters of data, each cluster sampled IID according to a binormal model, with the unclustered data combined at the bottom. Case observations are represented with "—" and control observations with "|"

- binary response model
- Fix the combined cluster size $M + N$

$$\text{individual effects } \vec{\xi} = (\xi_1, \ldots, \xi_{M+N}) \text{ IID}$$
$$\text{cluster effects } Z \perp\!\!\!\perp (\xi_1, \ldots, \xi_{M+N})$$
$$\text{markers } B_i = Z + \xi_i, i = 1, \ldots, M + N$$

case status indicators

$$D_i \mid \vec{Z}, \vec{\xi} \sim \text{bernoulli with parameter } \sigma(\beta_0 Z + \beta_1 \xi_i)$$

$$M = \sum_{i=1}^{M+N} (1 - D_i), \qquad N = \sum_{i=1}^{M+N} D_i.$$

- The control and case observations in a cluster, X_i and Y_i, are B_i such that $D_i = 0$ and $D_i = 1$

Suppose first that $\beta_0 > 0$ and $\beta_1 = 0$, so

$$P(D_i = 1 \mid Z, \bar{\xi}) = \sigma(\beta_0 Z)$$

- The population AUC is

$$\theta_{12} = P(Z_{11} - Z_{21} < \xi_{21} - \xi_{11} \mid D_{11} = 0, D_{21} = 1)$$

  $Z_{11} \mid D_{11} = 0$ is stochastically less than $Z_{21} \mid D_{21} = 1$,the last line is $> \frac{1}{2}$, with the difference increasing in β_0.
- The individual AUC is

$$\theta_{11} = P(\xi_{11} < \xi_{12}) = 1/2.$$

Two possible instances of the model:

1. The cluster effect Z represents a genuine signal of disease status D, such as viral load wrt HIV status, and ξ represents non-systematic measurement error on instruments measuring Z. In this case, the population AUC better matches expectations of an AUC measurement than the personalized AUC. The biomarker B isn't completely uninformative, as θ_{11} suggests.
2. The cluster effect Z is a subject's dose of a possibly ineffective drug, and larger doses are administered to sicker patients. The subject-specific measurements ξ represent non-systematic measurement error again. Here the association between the marker and disease status implied by the population AUC is spurious, and may or may not be of value to the analyst. It is possible that the personalized AUC, which does not convey any association, is preferable.

Reversing the roles of the cluster-level effect Z and within-cluster effects ξ, suppose $\beta_0 = 0$ and $\beta_1 > 0$, so that $\theta_{12} \approx 1/2$ and $\theta_{11} > 1/2$. Two instances of this model:

1. The markers B are measurements on a patient, D indicates the presence of a disease that depends little or not at all on a baseline measure Z but is indicated by the deviations ξ from the baseline. As a second example, the markers B are post-test measurements on a population that has been stratified by pre-test measurement Z. The subject effects $\xi_i = B_i - Z$ represent the difference between post-test and pre-test measurements, and the status indicators D represent an effective or ineffective intervention. Here the personalized AUC probably carries the correct interpretation.

2. A population clustered along any given dimension Z, and, analogous to 2, uptake of a possibly ineffective drug is confounded by indication. That is, sicker individuals, those for which D_i is more likely to be 1, take higher doses ξ_i of the drug. Here again a causal analysis would suggest the population AUC as less misleading than the personalized AUC, though a non-causal analysis, e.g., an intention-to-treat analysis, may point to the personalized AUC.

Outline

Proposition

Given $(X, Y, M, N) \sim P$*, suppose that* $P(X_{1k} < Y_{1l} \mid M, N)$ *and* $P(X_{1k} < Y_{2l} \mid M, N)$ *do not depend on* k, l*. Then* $\theta_{11}(P) = P(X_{11} < Y_{11})$ *and* $\theta_{12}(P) = P(X_{11} < Y_{21})$.

- In order for $\hat{\theta}_{12} \to 1$ while $\hat{\theta}_{11} \not\to 1$ in the random effects model, it was necessary that $(X, Y) \not\perp\!\!\!\perp (M, N)$.
- Consider a simpler result: For constant M and N, does $\theta_{12} < \theta_{11}$ hold?
- First consider a simple case where $M = N = 1$.

When $M = N = 1$,

$$1 - \sqrt{2(1 - \theta_{12})} \leq \theta_{11} \leq \sqrt{2\theta_{12}}$$

equivalently,

$$\frac{1}{2}\theta_{11}^2 \leq \theta_{12} \leq 1 - \frac{1}{2}(1 - \theta_{11})^2,$$

- When the personalized AUC is completely uninformative, $\theta_{11} = 1/2$, the informativity of the population AUC is limited, $1/8 \leq \theta_{12} \leq 7/8$.
- However, when the population AUC is completely uninformative, $\theta_{12} = 1/2$, the above bounds on the personalized AUC, which are tight, are vacuous, $0 \leq \theta_{11} \leq 1$.

Theorem

Let $(X, Y, M, N) \sim P$ *with* $M = m$ *and* $N = n$ *constant. Then*

$$\frac{1}{2}\left(\theta_{11} + \frac{\sum_{k,l} P(X_{1k} = Y_{1l})}{2mn}\right)^2 \leq \theta_{12}$$
$$\leq 1 - \frac{1}{2}\left(1 - \theta_{11} + \frac{\sum_{k,l} P(X_{1k} = Y_{1l})}{2mn}\right)^2$$

- Situations where the population AUC $\to 1$ while the personalized AUC $\to 1/2$, may require some dependence between M, N and X, Y

Outline

Let $\psi : V \times V \to \mathbb{R}$, $W = (X, Y, M, N) \sim P$ with $(X, Y) \in V \times V$, $\psi \in L^2(P)$, M and N counting numbers > 0 with finite means. Then:

$$\sqrt{I}(\hat{\theta}_{12} - \theta_{12}, \hat{\theta}_{11} - \theta_{11}) \rightsquigarrow \mathcal{N}(0, \Sigma)$$

with

$$\Sigma_{11} = \lim_{I \to \infty} I \operatorname{Var}(\hat{\theta_{12}}) = \mathrm{E}\left(\frac{\mathrm{E}(\psi_{12} \mid W_1) + \mathrm{E}(\psi_{21} \mid W_1)}{\mathrm{E}\, M \, \mathrm{E}\, N} - \theta_{12} \left(\frac{M_1}{\mathrm{E}\, M} + \frac{N_1}{\mathrm{E}\, N} \right) \right)^2$$

$$\Sigma_{22} = \lim_{I \to \infty} I \operatorname{Var}(\hat{\theta_{11}}) = \operatorname{Var}(\psi_{11}/(M_1 N_1))$$

$$\Sigma_{12} = \lim_{I \to \infty} I \operatorname{Cov}(\hat{\theta_{12}}, \hat{\theta_{11}}) = \theta_{12} \, \mathrm{E}\left(\frac{\psi_{11}}{M_1 N_1} \left(\frac{\psi_{12} + \psi_{21}}{\mathrm{E}\, \psi_{12}} - \frac{M_1}{\mathrm{E}\, M} - \frac{N_1}{\mathrm{E}\, N} \right) \right)$$

Simulation

- M and N are sampled as the negative and positive values in a correlated normal sample.
- The greater the correlation, the greater the imbalance between case and control observations within the clusters

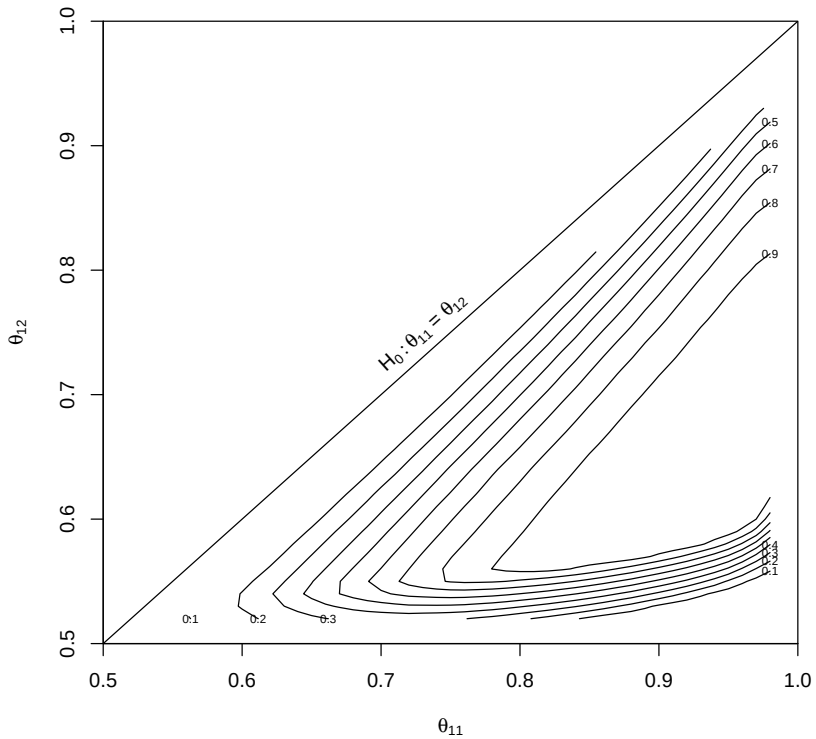
Two models for the observation vectors (X, Y)

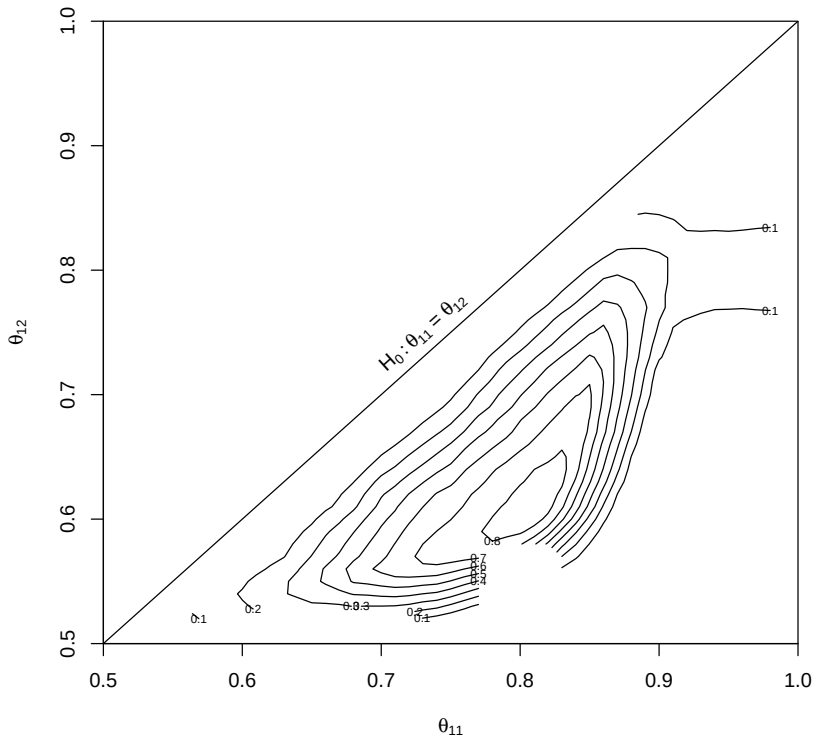
- binormal model:

$$(X, Y) \mid (M, N) \sim \mathcal{N}_{M+N}\left(\begin{pmatrix} 0 \cdot \mathbb{1}_M \\ \Delta \cdot \mathbb{1}_N \end{pmatrix}, \begin{pmatrix} 1 & \rho & \cdots & \rho \\ & \ddots & \ddots & \\ \rho & \cdots & \rho & 1 \end{pmatrix}\right)$$

- $\theta_{12}(P) = \Phi\left(\frac{\Delta}{\sqrt{2}}\right), \theta_{11}(P) = \Phi\left(\frac{\Delta}{\sqrt{2(1-\rho)}}\right)$
- Censored binormal model: Sample $(\overline{X}, \overline{Y})$, then clip the observations to $\pm a$

- Empirical power function of the test of $H_0 : \theta_{12} = \theta_{11}$ versus $\theta_{12} < \theta_{11}$ using the asymptotic estimator
- same as $H_0 : |\theta_{12} - 1/2| = |\theta_{11} - 1/2|$, equal informativity
- normal data with and without clipping





Outline

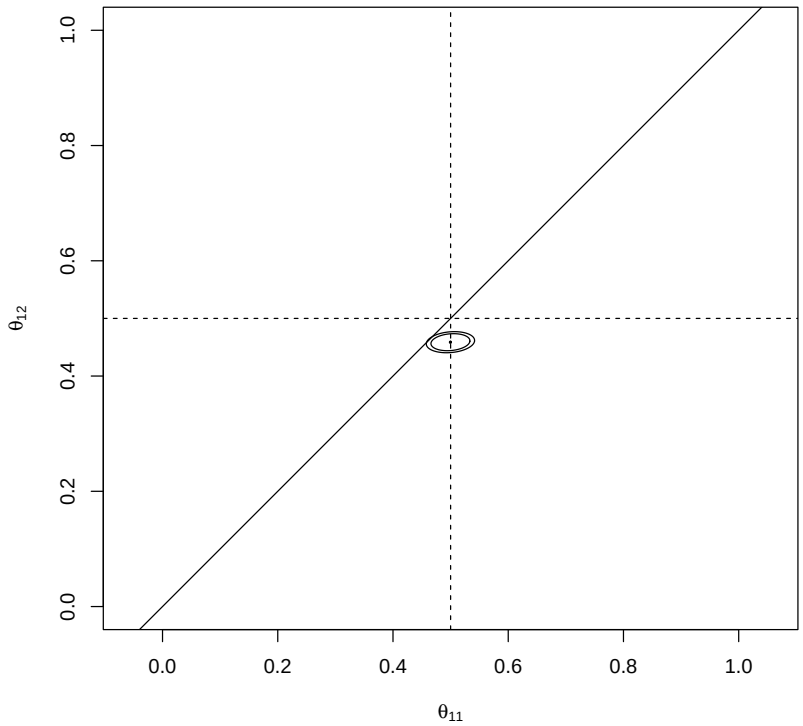
- The data consists of Terry stops in New York City and Boston.
- The analysis here focuses on the relationship between the duration of the stop and race of the suspect.
- We cluster the stops according to precinct, in the case of NYC, and according to the officer conducting the stop, in the case of Boston.

	NYC			Boston		
group	mean duration (SD)	count	freq.	mean duration (SD)	count	freq.
Asian	14.24 (21.16)	1139	0.02	25.00 (24.22)	53	0.01
Black Hispanic	11.01 (17.12)	4675	0.09	15.28 (18.73)	391	0.06
Black non-Hispanic	10.99 (16.78)	31588	0.58	19.06 (28.93)	3448	0.55
White Hispanic	11.21 (15.15)	11486	0.21	15.63 (15.96)	578	0.09
White non-Hispanic	12.85 (16.18)	4854	0.09	21.74 (33.01)	1760	0.28
other	11.84 (17.70)	261	0.00	20.89 (23.90)	93	0.01

Table: Summary estimates on the duration of Terry stops by racial group.

$\theta_{12} < \theta_{11}$

- With Black race as the binary classification, the AUC analysis looks for a difference in location between the distribution of stop durations of non-Black ("control") and Black ("case") suspects.
- For the NYC data, the population AUC estimate is $\hat{\theta}_{12} = 0.46$ with 95% CI 0.45—0.47, significantly different from the null value of $1/2$. The personalized AUC estimate is $\hat{\theta}_{11} = 0.50$ with a 95% CI 0.47—0.53.
- A test of equality $H_0 : \theta_{12} = \theta_{11}$ against $\theta_{12} < \theta_{11}$ returns a p-value of .05%. The Boston data is similar.


1.0
0.8
0.6
0.4
0.2
0.0
θ_{12}
0.0
0.2
0.4
0.6
0.8
1.0
θ_{11}

$\theta_{11} < \theta_{12}$

- For the Boston data, the personalized AUC, 0.46 [0.40, 0.53], is more informative than the population AUC, 0.52 [0.48, 0.55],
- the test of equality versus $\theta_{11} < \theta_{12}$ returning a p-value of 2.5%.

No significant difference between θ_{12} and θ_{11}.

- duration of the stop between non-Hispanic ("control") and Hispanic ("case") suspects: For both the NYC and Boston data, neither the population AUC nor personalized AUC is significantly different from the null value $1/2$, and the test of equality of the two AUCs fails to reject.
- For the Boston data, whether one takes the case status to be non-Hispanic Black or non-Hispanic White, the two AUCs are statistically indistinguishable from each other and each is indistinguishable from the null value $1/2$.

Haben Michael, Lu Tian. The Population and Personalized AUCs. Forthcoming. Available at:
`https://haben-michael.github.io/`