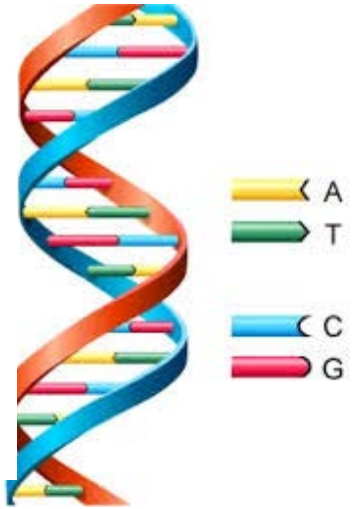
Efficient Identification of Heterogeneous Treatment Effects via Anomalous Pattern Detection

Edward McFowland III
emcfowla@umn.edu

Joint work with (Sriram Somanchi & Daniel B. Neill)

Treatment Effects

+

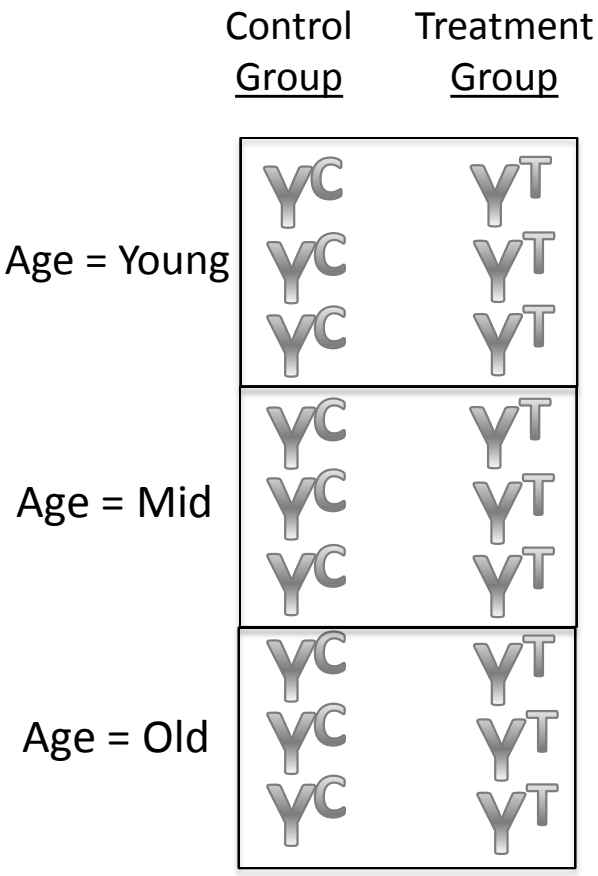


=

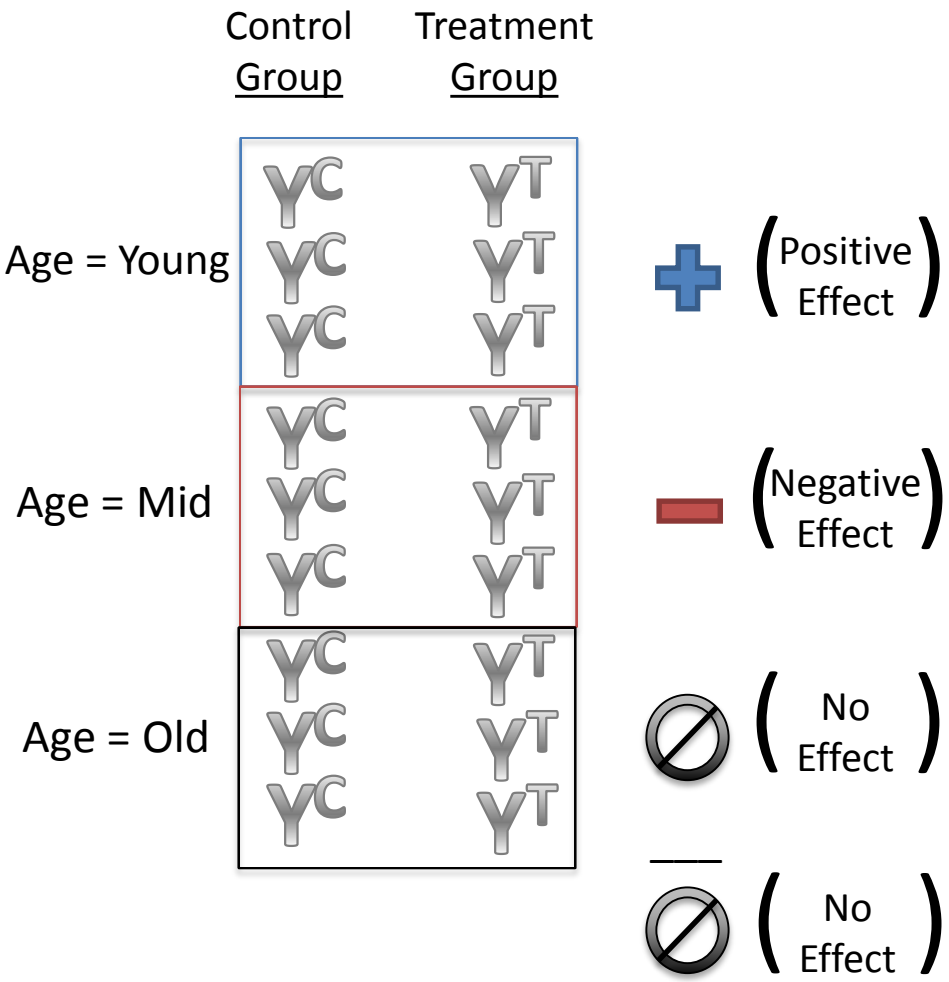
?

Treatment Effects

Treatment Effect Heterogeneity

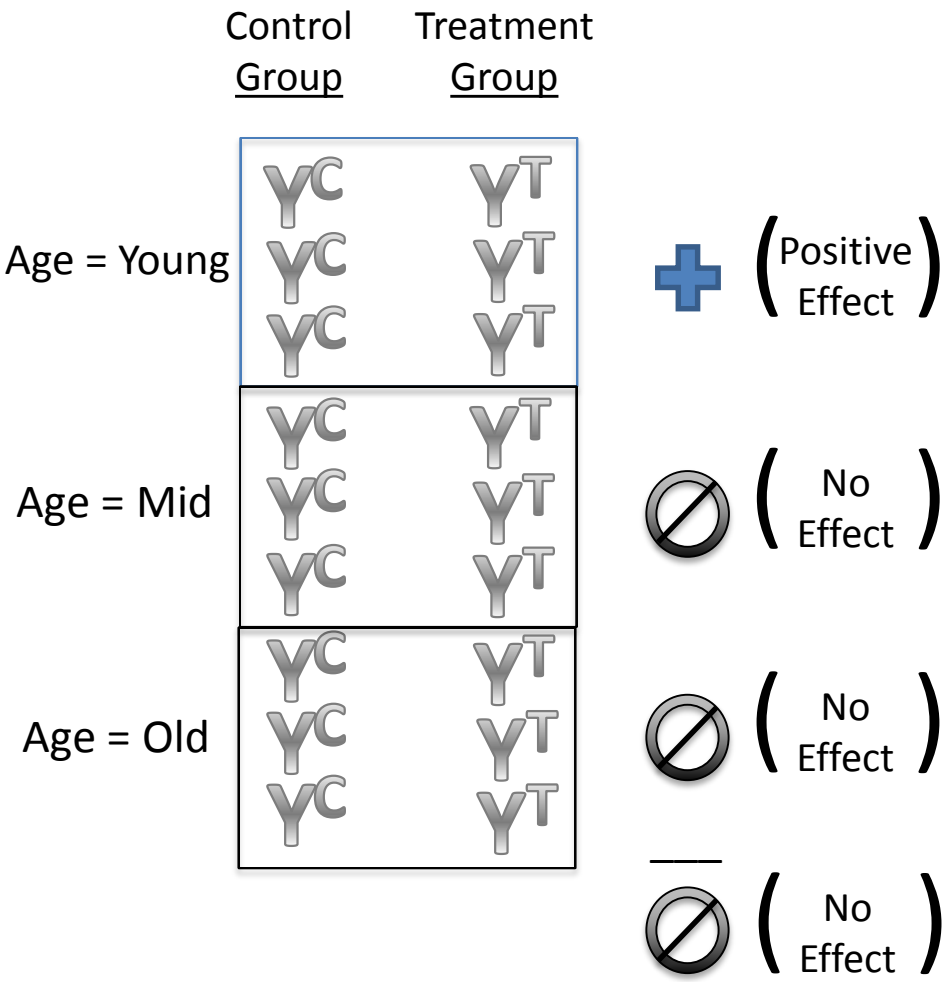


Treatment Effect Heterogeneity



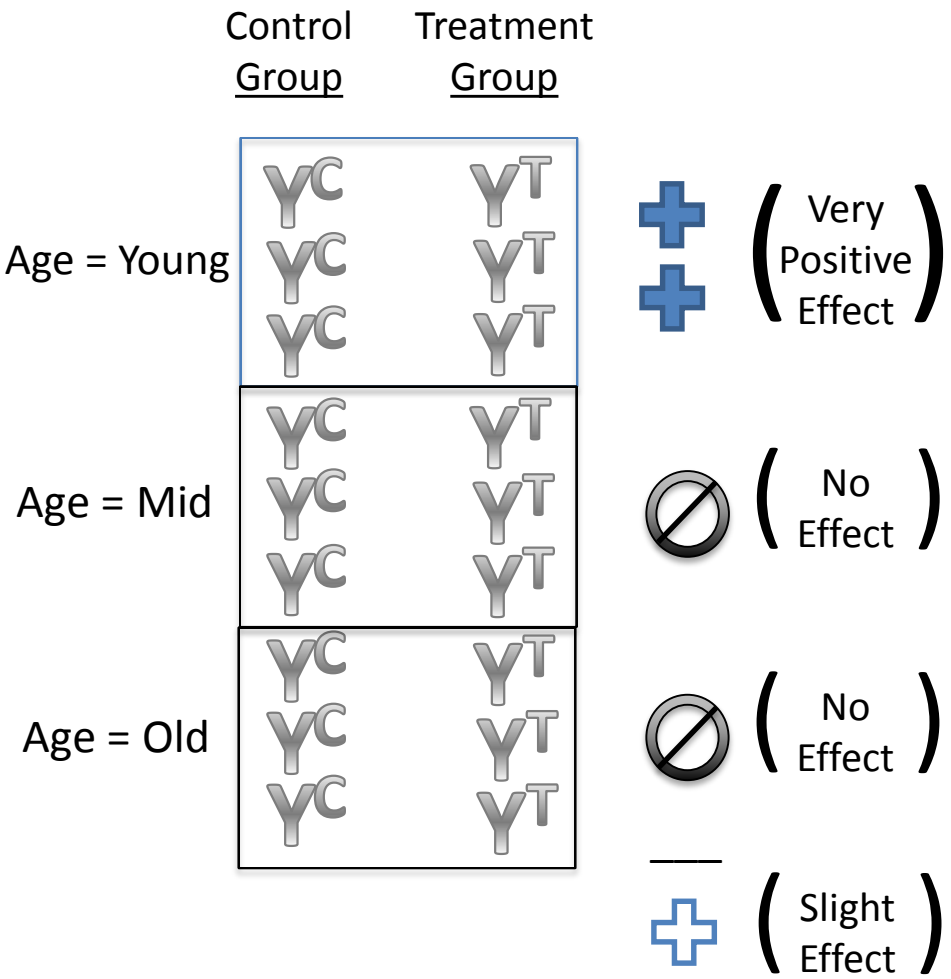
- Positive and negative effects can cancel

Treatment Effect Heterogeneity



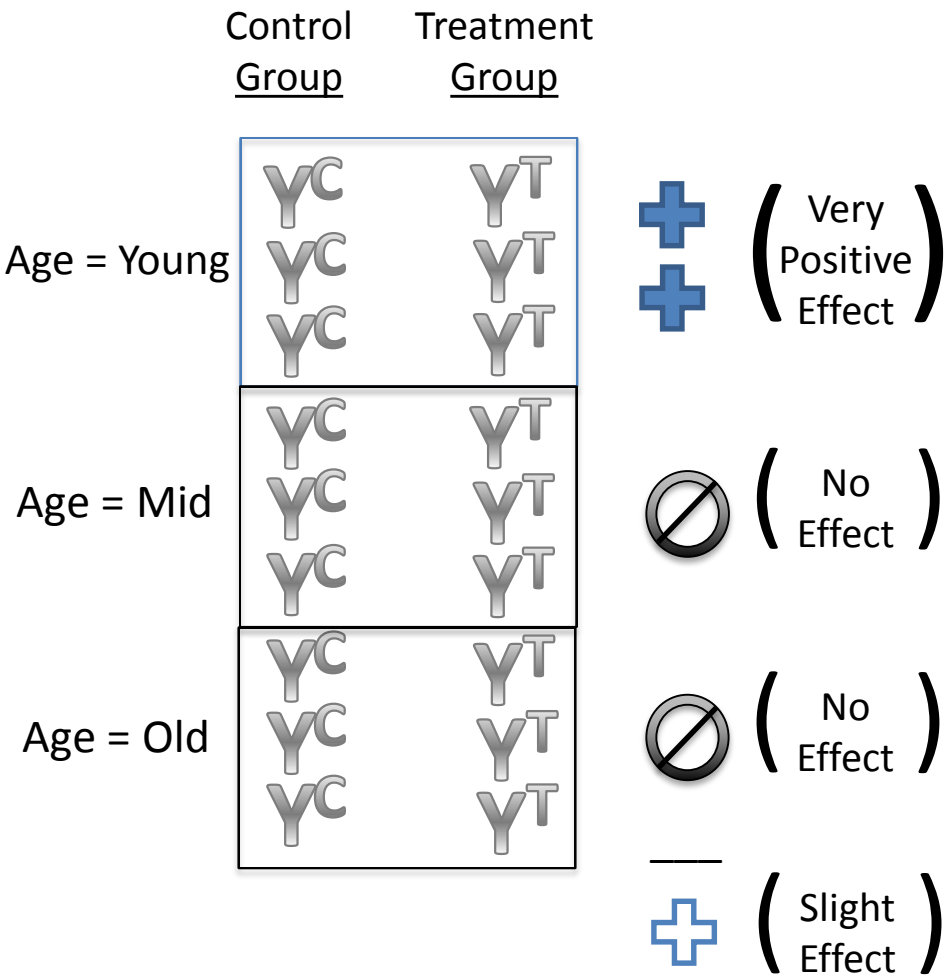
- Positive and negative effects can cancel
- True effect can be masked

Treatment Effect Heterogeneity



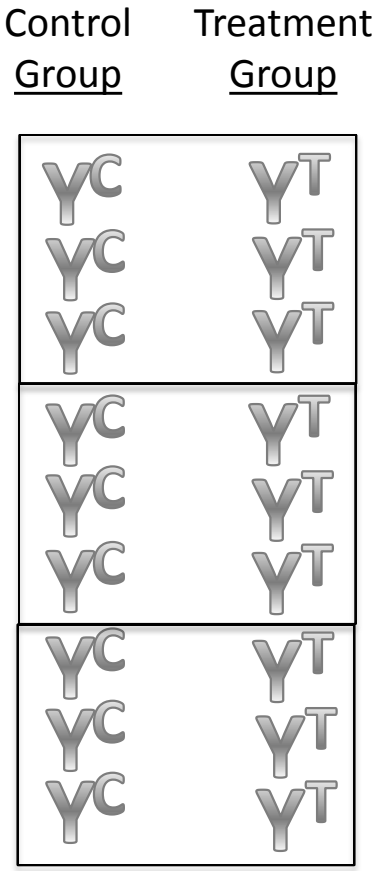
- Positive and negative effects can cancel
- True effect can be masked
- Effects could really be driven by a subpopulation

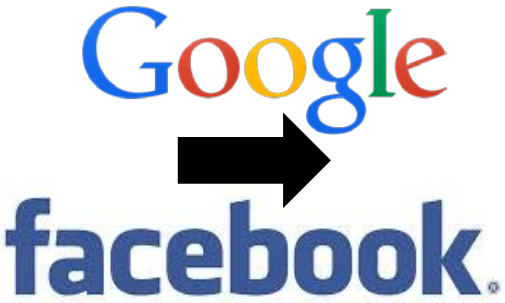
Treatment Effect Heterogeneity

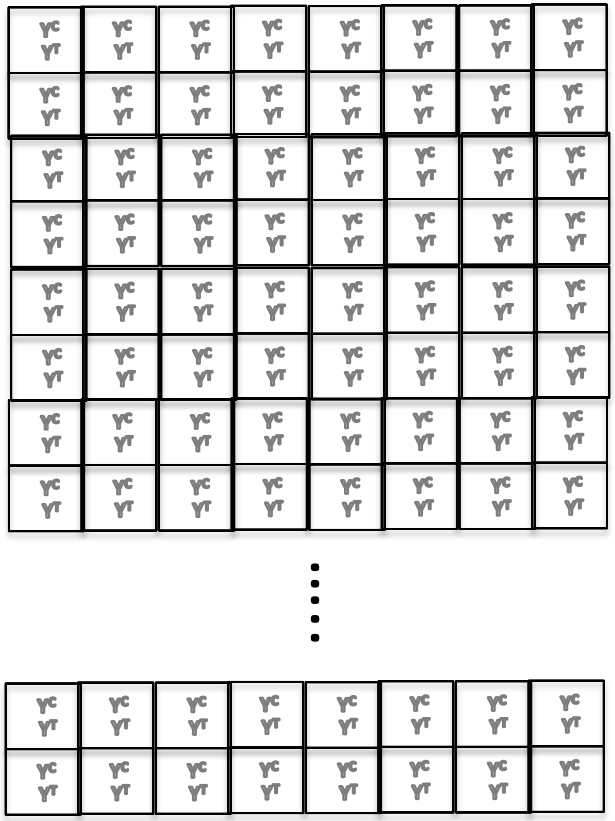


- Positive and negative effects can cancel
- True effect can be masked
  - Ex: FDA Approved BiDil Drug
- Effects could really be driven by a subpopulation
  - Ex: Perry Preschool

Treatment Effect Heterogeneity (Big Data)







Machine Learning's Contributions

- Regression Methods
  - OLS and Regularized Regression (e.g., LASSO)*
  - Imai and Ratkovic (2013)

- Single Tree Methods
  - Su et al (2009)
  - Imai and Strauss (2011)
  - Athey and Imbens (2015)*

- Ensemble Methods
  - Green and Kern (2012)
  - Wager and Athey (2012)*

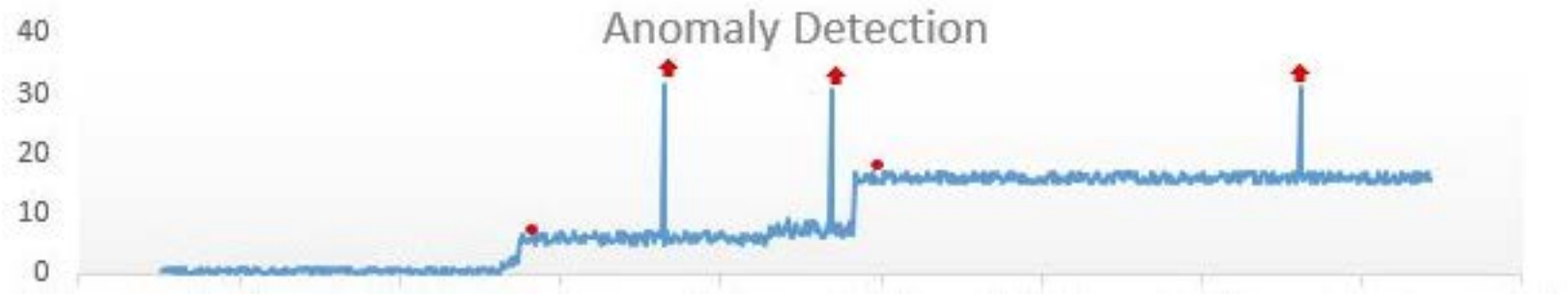
* provide asym confidence intervals, for inference of effects significance

Limitations

- Regression Methods
  - Pre-specification of the model
- Single Tree Methods
  - Greedy and unstable
- Ensemble Methods
  - Fairly uninterpretable/no natural subpopulations
- General Limitations
  - The mean and only the mean
    - Other moments can be effected
    - Simpsons Paradox
  - Risk minimization not effect maximization
    - Small number of subpopulations considered
    - No guarantee on their “interestingness”
  - No “discovery”, only model inspection

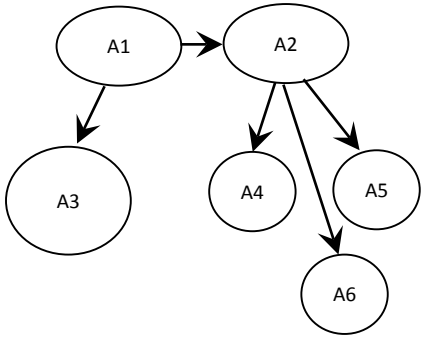
Anomaly Detection Paradigm

- Identifying when a “system” deviates away from its expected behavior.



Anomalous Pattern Detection Procedure

Normal Activity (M_0)



Test Data

PPORT	DSPORT	COUNTRY	SLINE	VESSEL	SHIPPER NAME	F NAME	COMMODITY	SIZE	MTONS	VALUE
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	AMERICAN TRANSNET EXPRESS TRANSNET		EMPTY RACK	1	5.6	2757
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	ORDER		ORDER 0 USED TIRE	2	13.43	9497
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	ORDER		ORDER 0 USED TIRE	2	13.43	9497
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	AMERICAN TRANSNET EXPRESS TRANSNET		CRUDE ICONE PARTY	1	17.38	23151
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	NEW WAVE TRANSPORT	JIT	PANELS F MODEL 38	1	38.57	67195
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	NEW WAVE TRANSPORT	JIT	PANELS F MODEL 38	1	38.57	67195
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	NEW WAVE TRANSPORT	JIT	PANELS F MODEL 38	1	38.57	67195
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	ORDER		ORDER 0 USED TIRES	2	13.43	9497
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	CHINA OCEAN SHIPPING		CHINA OCEAN EMPTY CONTAINERS	1	1	0
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	CHINA OCEAN SHIPPING		CHINA OCEAN EMPTY CONTAINERS	1	1	0

Detect Anomalous Pattern Given M_0

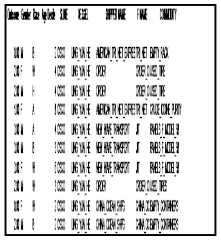
Novel Pattern

YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	ORDER		ORDER 0 USED TIRE	2	13.43	9497
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	ORDER		ORDER 0 USED TIRE	2	13.43	9497
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	AMERICAN TRANSNET EXPRESS TRANSNET		CRUDE ICONE PARTY	1	17.38	23151
YOKOHAMA	SEATTLE	JAPAN	OSCO	LING YUN HE	NEW WAVE TRANSPORT	JIT	PANELS F MODEL 38	1	38.57	67195

HTE Pattern Detection

Control Group

Treatment Group



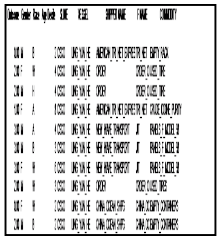
Detect Anomalous Pattern Given M_0

Novel Pattern

HTE Pattern Detection

Control Group

Treatment Group



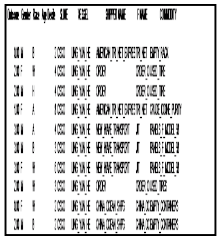
Detect Anomalous Subpopulation Given M_0

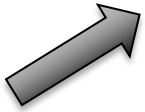
Novel Pattern

HTE Pattern Detection

Control Group

Treatment Group

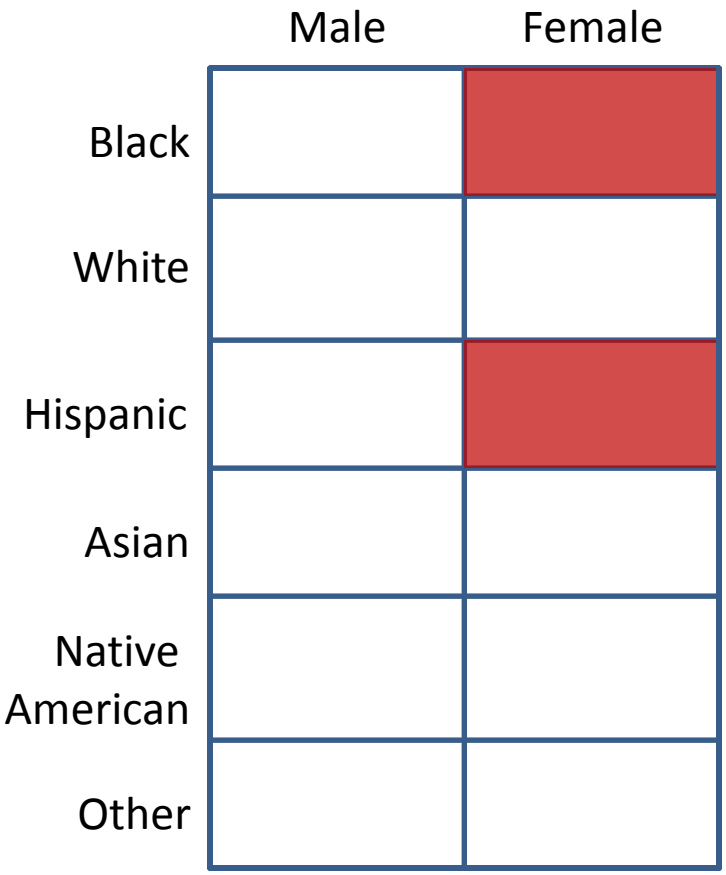




Detect Anomalous Subpopulation Given M_0

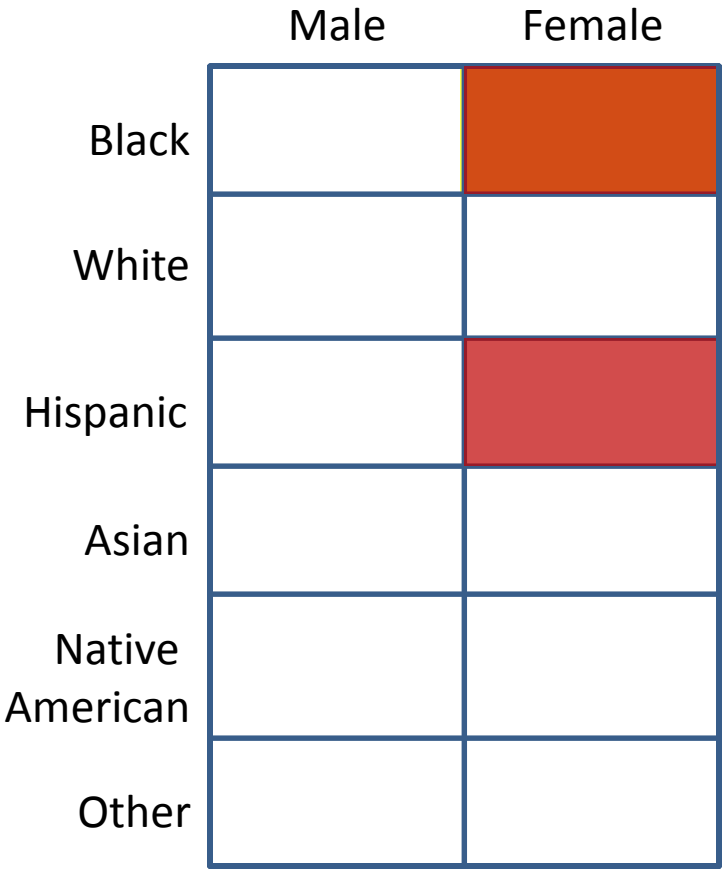
Subpopulation

The Goal



Detect a subpopulation (subsets of attribute values), which correspond to anomalous outcomes for subjects in the treatment group

The Goal

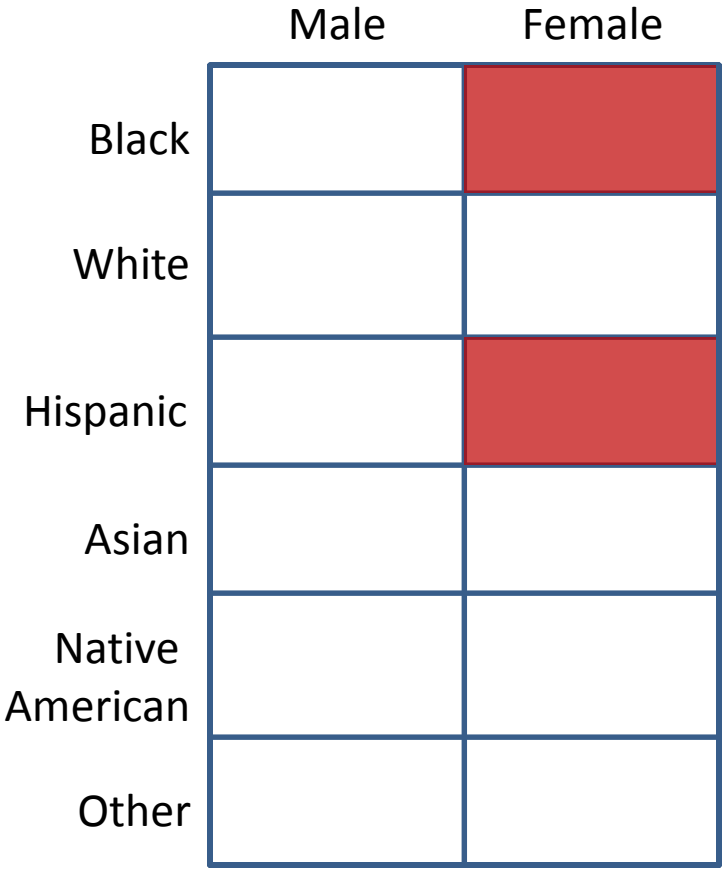


Detect a subpopulation (subsets of attribute values), which correspond to anomalous outcomes for subjects in the treatment group

The Optimization

$s_1 \subseteq \{a_1 \ldots a_t\}, \ldots, s_M \subseteq \{a_1 \ldots a_p\}$

The Goal



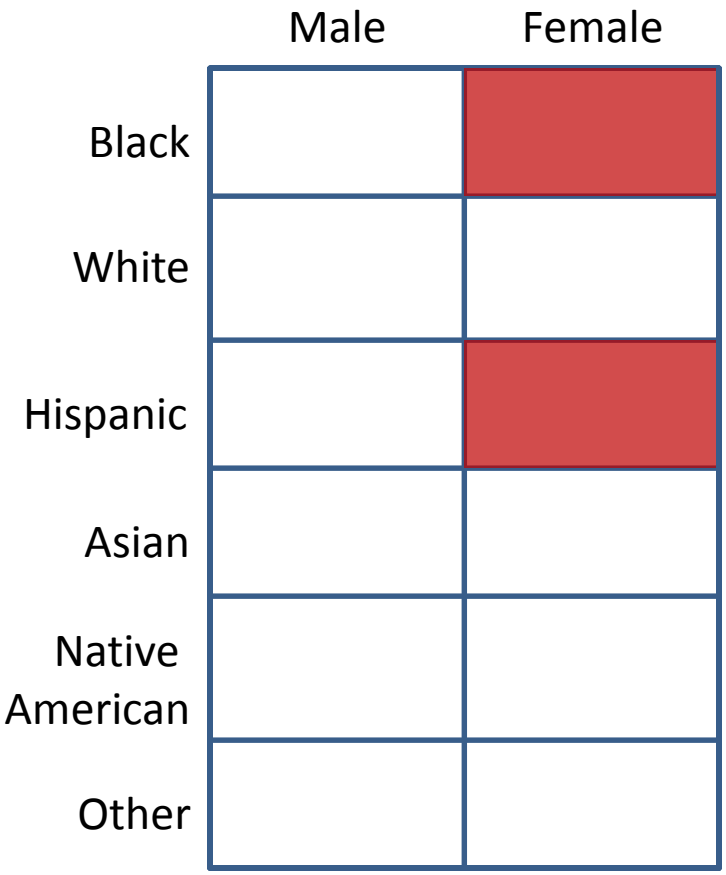
Detect a subpopulation (subsets of attribute values), which correspond to anomalous outcomes for subjects in the treatment group

The Optimization

$s_1 \subseteq \{a_1 \ldots a_t\}, \ldots, s_M \subseteq \{a_1 \ldots a_p\}$

$S = s_1 \times \ldots \times s_M$

The Goal



Detect a subpopulation (subsets of attribute values), which correspond to anomalous outcomes for subjects in the treatment group

The Optimization

$s_1 \subseteq \{a_1 \ldots a_t\}, \ldots, s_M \subseteq \{a_1 \ldots a_p\}$

$S = s_1 \times \ldots \times s_M$

$S^* = \textit{argmax}_S F(S)$

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	Y^{BM}	Y^{BF}
White	Y^{WM}	Y^{WF}
Hispanic	Y^{HM}	Y^{HF}
Asian	Y^{AM}	Y^{AF}
Native American	Y^{NM}	Y^{NF}
Other	Y^{OM}	Y^{OF}

I. Compute the statistical anomalousness of each treatment group subject

II. Detect subpopulation that is collectively the most anomalous

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

I. Compute the statistical anomalousness of each treatment group subject

-- This measurement will be a p-value

II. Detect subpopulation that is collectively the most anomalous

-- Many subjects with significant p-values

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	Y^{BM}	Y^{BF}
White	Y^{WM}	Y^{WF}
Hispanic	Y^{HM}	Y^{HF}
Asian	Y^{AM}	Y^{AF}
Native American	Y^{NM}	Y^{NF}
Other	Y^{OM}	Y^{OF}

Control Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	Y^{BM}	Y^{BF}
White	Y^{WM}	Y^{WF}
Hispanic	Y^{HM}	Y^{HF}
Asian	Y^{AM}	Y^{AF}
Native American	Y^{NM}	Y^{NF}
Other	Y^{OM}	Y^{OF}

Treatment Group

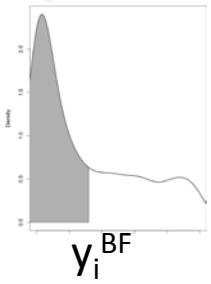
I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	Y^{BM}	Y^{BF}
White	Y^{WM}	Y^{WF}
Hispanic	Y^{HM}	Y^{HF}
Asian	Y^{AM}	Y^{AF}
Native American	Y^{NM}	Y^{NF}
Other	Y^{OM}	Y^{OF}

Treatment Group



I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values
   i. Maps each bin's distribution to the same interval

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values
   i. Maps each bin's distribution to the same interval
   ii. P^{ij} ~ Uniform[0,1] under H_0

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values
   - i. Maps each bin's distribution to the same interval
   - ii. P^{ij} ~ Uniform[0,1] under H_0
   - iii. For any N p-values, we expect N*α to be significant at level α

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values
   - i. Maps each bin's distribution to the same interval
   - ii. P^{ij} ~ Uniform[0,1] under H_0
   - iii. For any N p-values, we expect $N*\alpha$ to be significant at level α

Higher Criticism:

$$F(S) = \max_{\alpha} \frac{N_{\alpha} - N\alpha}{\sqrt{N\alpha(1-\alpha)}}$$

Treatment Effects Subset Scan (TESS)

	Male	Female
Black	P^{BM}	P^{BF}
White	P^{WM}	P^{WF}
Hispanic	P^{HM}	P^{HF}
Asian	P^{AM}	P^{AF}
Native American	P^{NM}	P^{NF}
Other	P^{OM}	P^{OF}

Treatment Group

I. Compute the statistical anomalousness of each treatment group subject

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $S_1 \times \ldots \times S_M$
   - Naïve search is infeasible $O(2^{\sum |A_i|})$

Treatment Effects Subset Scan (TESS)

Nonparametric Scan Statistic (NPSS)

Have: $S \subseteq \{A_1 \times \ldots \times A_M\}$
$= \{s_1 \times \ldots \times s_M\}$

Select: F(S)

Want: max_S F(S)

I. Compute the statistical anomalousness of each treatment group subject

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - Naïve search is infeasible $O(2^{\sum |A_i|})$

Treatment Effects Subset Scan (TESS)

Nonparametric Scan Statistic (NPSS)

Have: $S \subseteq \{A_1 \times \ldots \times A_M\}$
$= \{s_1 \times \ldots \times s_M\}$

Select: F(S)

Want: max_S F(S)

There Exist: $G(a_i)$

I. Compute the statistical anomalousness of each treatment group subject

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - Naïve search is infeasible $O(2^{\sum |A_i|})$

Treatment Effects Subset Scan (TESS)

Nonparametric Scan Statistic (NPSS)

Have: $S \subseteq \{A_1 \times \ldots \times A_M\}$
$= \{s_1 \times \ldots \times s_M\}$

Select: F(S)

Want: max_S F(S)

There Exist: $G(a_i)$

Such That: $\max_{s_j \subseteq \{a_1, \ldots, a_t\}} F(s_j | A_{-j}) = \max_{i=1\ldots t} F(\{a_{(1)} \ldots a_{(t)}\} | A_{-j})$

Only Consider: {Black}
{Black, Hispanic}
{Black, Hispanic, Asian}
⋮
{Black, Hispanic, Asian, …, White }

I. Compute the statistical anomalousness of each treatment group subject

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - Naïve search is infeasible $O(2^{\sum |A_i|})$

Treatment Effects Subset Scan (TESS)

Nonparametric Scan Statistic (NPSS)

Have: $S \in \{A_1 \times \ldots \times A_M\}$

$= \{s_1 \times \ldots \times s_M\}$

Select: F(S)

Want: max_S F(S)

There Exist: $G(a_i)$

Such That: $\max_{s_j \subseteq \{a_1, \ldots, a_t\}} F(s_j | A_{-j}) = \max_{i=1\ldots t} F(\{a_{(1)} \ldots a_{(t)}\} | A_{-j})$

Only Consider: {Black}

{Black, Hispanic}

{Black, Hispanic, Asian}

⋮

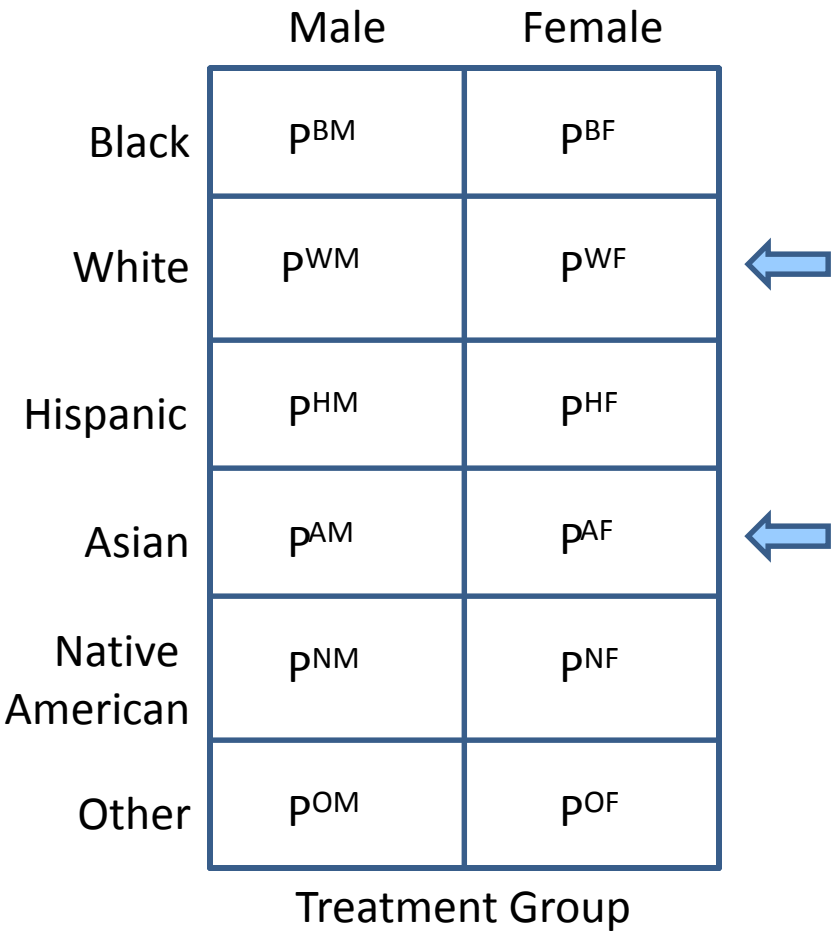
{Black, Hispanic, Asian, …, White }

I. Compute the statistical anomalousness of each treatment group subject

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - NPSS over an attribute in O(t log t)

TESS Search Procedure



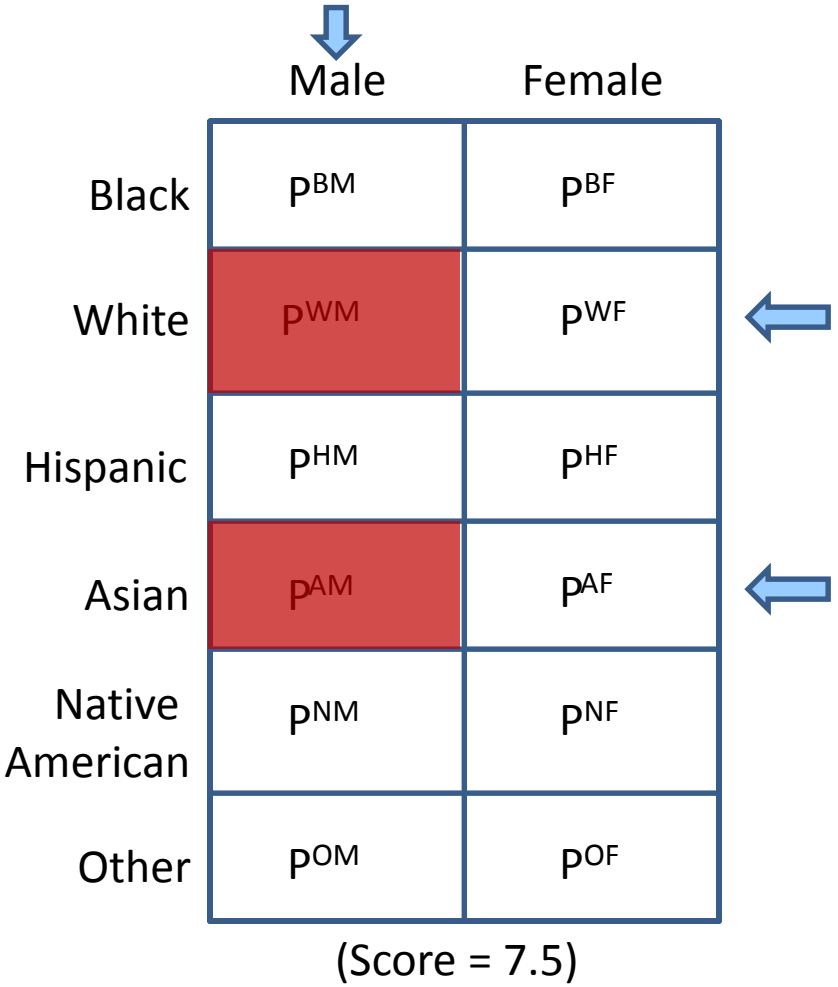
I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - NPSS over an attribute in O(t log t)

TESS Search Procedure



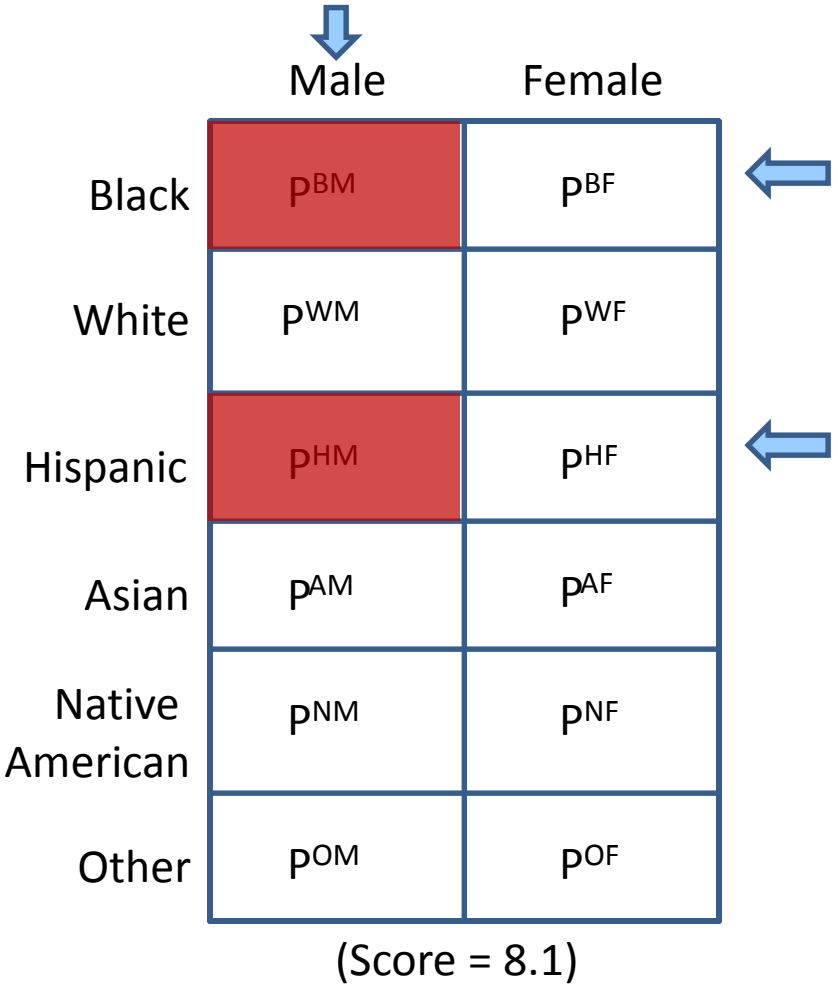
I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $S_1 \times \ldots \times S_M$
   - NPSS over an attribute in O(t log t)

TESS Search Procedure



(Score = 8.1)

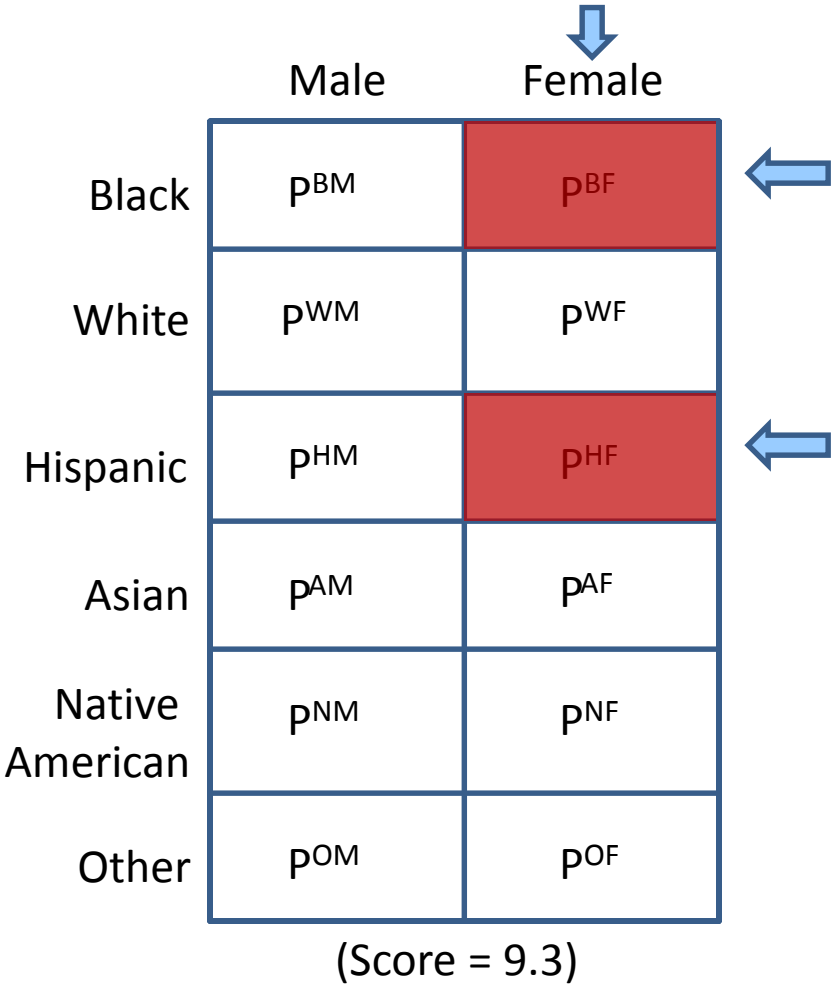
I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $S_1 \times \ldots \times S_M$
   - NPSS over an attribute in O(t log t)

TESS Search Procedure



(Score = 9.3)

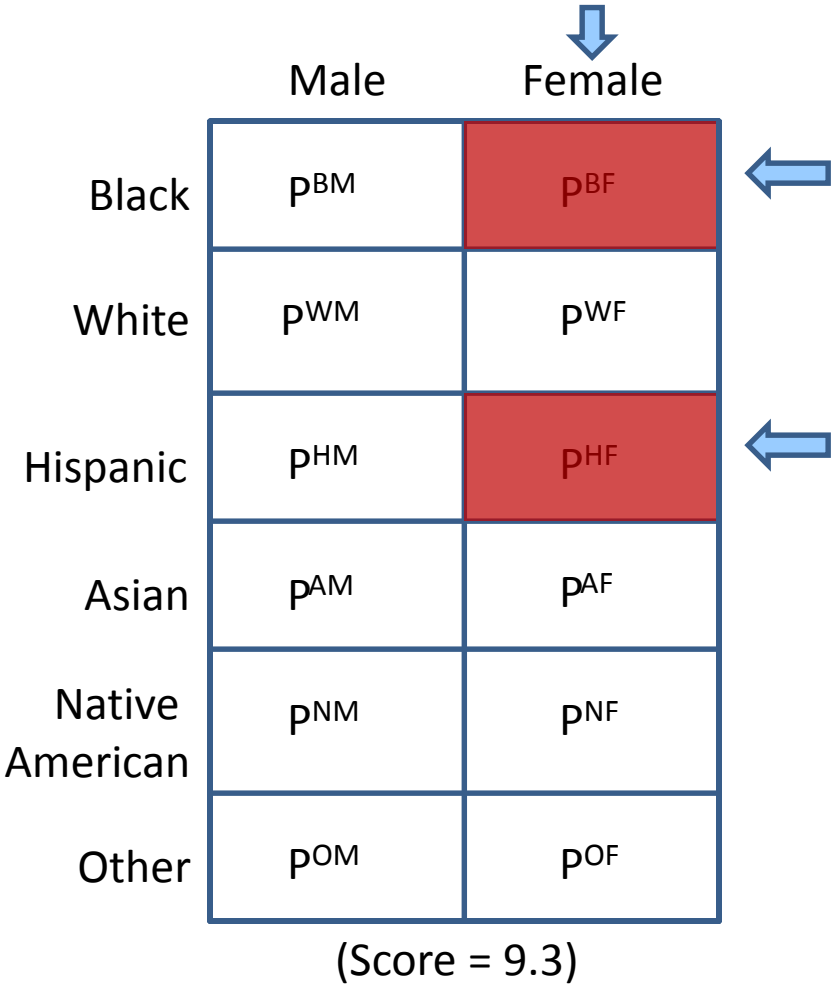
I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - NPSS over an attribute in O(t log t)

TESS Search Procedure



(Score = 9.3)

I. Compute the statistical anomalousness of each treatment group subject

1. Estimate Conditional Distribution Under H_0
2. Compute empirical p-values

II. Discover subsets of attribute values that define the most anomalous outcomes

1. Maximize F(S) over all subsets of $s_1 \times \ldots \times s_M$
   - NPSS over an attribute in O(t log t)

Significance of our subpopulation

Compare subpopulation score to maximum scores of simulated datasets under H_0

Tennessee Star Analysis (1985)

- Effect of classrooms size on achievement (test scores)
- 4 year panel (kindergarten to 3rd grade)
- 6,500 students, 330 classrooms, 80 schools
  - Total of over 11,000 records
- Treatment Conditions (randomized within school)
  - Regular Size Class (20-25 students)
  - Regular Size + Aide Class (20-25 students)
  - Small Size Class (13-17 students)

Tennessee Star Analysis (1985)

read	math	gender	ethnicity	lunch	grade	school	experience	degree	tethnicity	schoolid
439	463	male	afam	free	kindergarten	inner-city	0	bachelor	cauc	19
448	559	male	cauc	non-free	kindergarten	rural	16	bachelor	cauc	69
431	454	male	cauc	free	kindergarten	rural	8	bachelor	cauc	5
395	423	female	afam	free	kindergarten	inner-city	17	master	cauc	16
451	500	female	cauc	non-free	kindergarten	rural	3	bachelor	afam	56
430	473	male	cauc	non-free	kindergarten	rural	13	master	cauc	38
437	468	male	cauc	non-free	kindergarten	rural	6	master	cauc	69
490	528	male	cauc	non-free	kindergarten	suburban	18	bachelor	cauc	52
439	484	male	cauc	non-free	kindergarten	suburban	13	master	cauc	54
424	459	female	cauc	free	kindergarten	rural	12	bachelor	cauc	12
437	528	female	afam	free	kindergarten	suburban	1	bachelor	afam	21
424	559	male	cauc	free	kindergarten	rural	13	bachelor	cauc	79
431	454	male	cauc	non-free	kindergarten	rural	13	master	cauc	8
451	473	male	cauc	non-free	kindergarten	rural	3	bachelor	cauc	66
421	459	female	afam	free	kindergarten	inner-city	11	bachelor	cauc	31

Tennessee Star Analysis

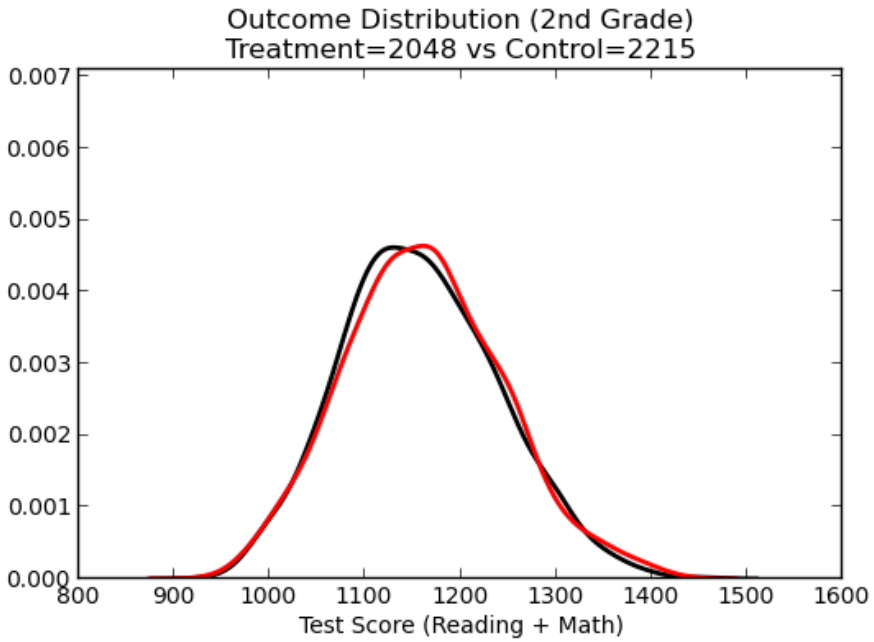
	(1)	(2)
Treatment	3.4791	-0.2909
	(2.547)	(2.277)
Sample	All 2nd Grade	All 3rd Grade
R-squared	0.000	0.000
Observations	4263	4063

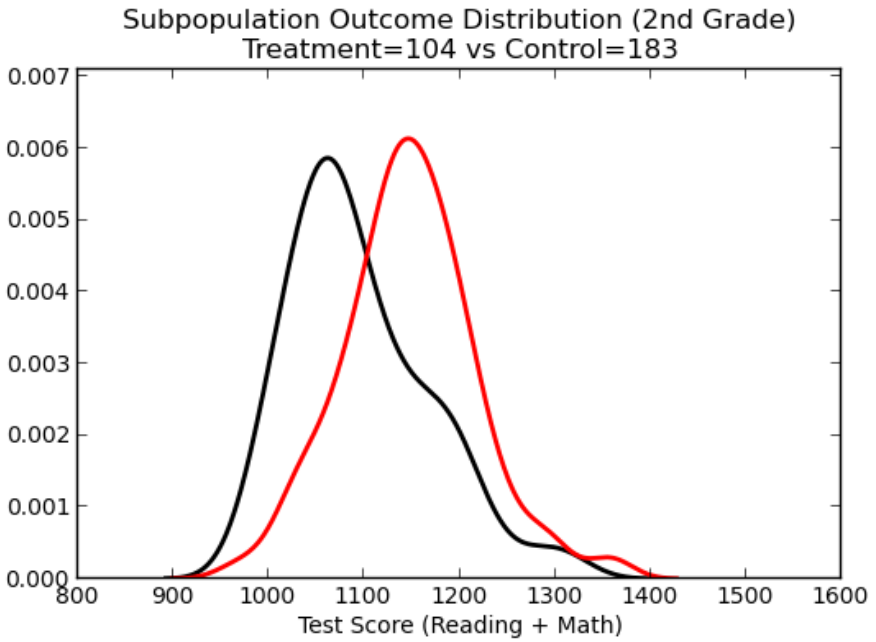
Notes: All estimates are from OLS models.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Tennessee Star Analysis (1985)

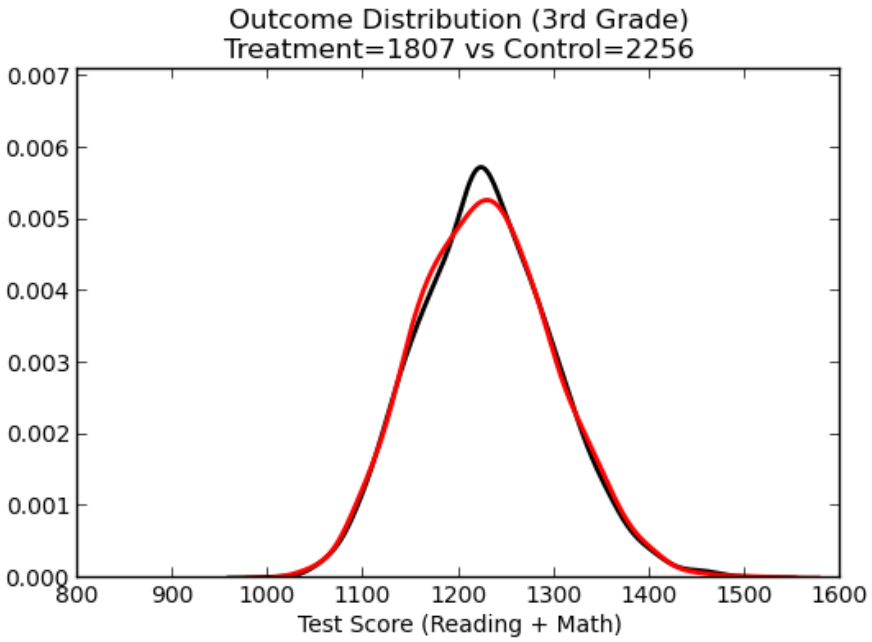
- Detected Subpopulation
  - grade:
    - 2nd or 3rd
  - school:
    - inner-city or urban
  - experience:
    - [10, infinity)
  - other features are considered irrelevant

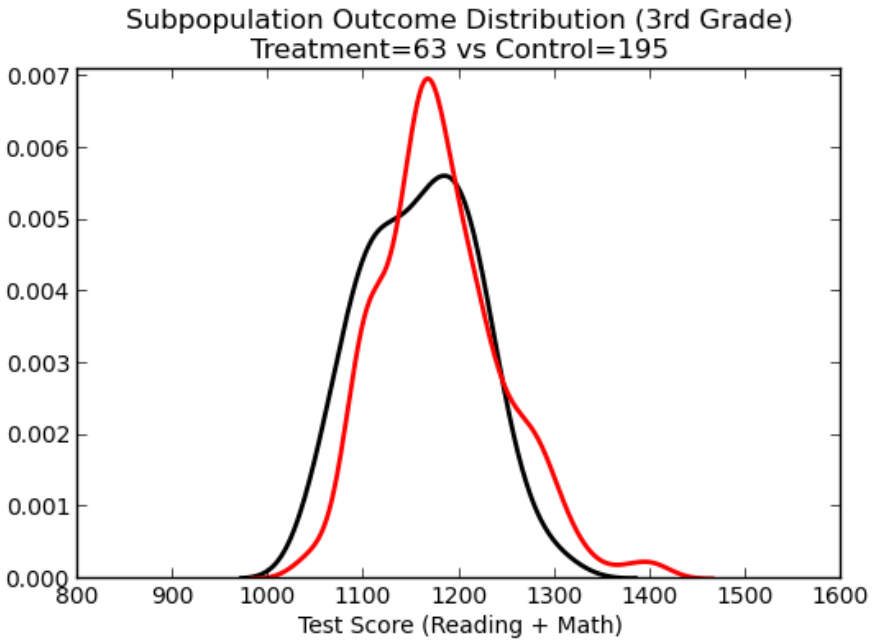
Tennessee Star Analysis





Tennessee Star Analysis





Tennessee Star Analysis

	(1)	(2)	(3)
Treatment	3.4791	51.2497***	1.4532
	(2.547)	(8.727)	(2.639)
Sample	All 2nd Grade	Detected Subpopulation	Undetected Subpopulation
R-squared	0.001<	0.108	0.001<
Observations	4263	287	3976

Notes: All estimates are from OLS models.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Tennessee Star Analysis

	(1)	(2)	(3)
Treatment	-0.2909	21.822**	1.6851
	(2.277)	(9.154)	(2.318)
Sample	All 3rd Grade	Detected Group (3rd Grade)	Undetected Group (3rd Grade)
R-squared	0.001<	0.022	0.001<
Observations	4063	258	3805

Notes: All estimates are from OLS models.
Standard errors are in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Conclusion

- Discovering subpopulations with significant treatment effects can be paramount
- Machine Learning can flexibly estimate effects but it limited when goal is to identify subpopulations with large effects
- Anomalous Pattern Detection paradigm offers overcome some abilities to overcome these limitations
  - Maintain high power to detect by searching over and combining signal across various subpopulations

References

- Athey, S., & Imbens, G. W. (2015). Machine Learning Methods for Estimating Heterogeneous Causal Effects. arXiv.org.
- Green, D. P., & Kern, H. L. (2012). Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees. Public Opinion Quarterly, 76(3), 491–511.
- Imai, K., & Ratkovic, M. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. The Annals of Applied Statistics, 7(1), 443–470.
- Imai, K., & Strauss, A. (2011). Estimation of Heterogeneous Treatment Effects from Randomized Experiments, with Application to the Optimal Planning of the Get-Out-the-Vote Campaign. Political Analysis, 19(1), 1–19.
- McFowland, E., III, Speakman, S. D., & Neill, D. B. (2013). Fast generalized subset scan for anomalous pattern detection. The Journal of Machine Learning Research, 14(1), 1533–1561.
- Mehlig, Kirsten et al. CETP TaqIB genotype modifies the association between alcohol and coronary heart disease: The INTERGENE case-control study. Alcohol, 48(7), 695 - 700.
- Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., & Li, B. (2009). Subgroup Analysis via Recursive Partitioning. The Journal of Machine Learning Research, 10, 141–158.
- Wager, S., & Athey, S. (2015). Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. arXiv.org.