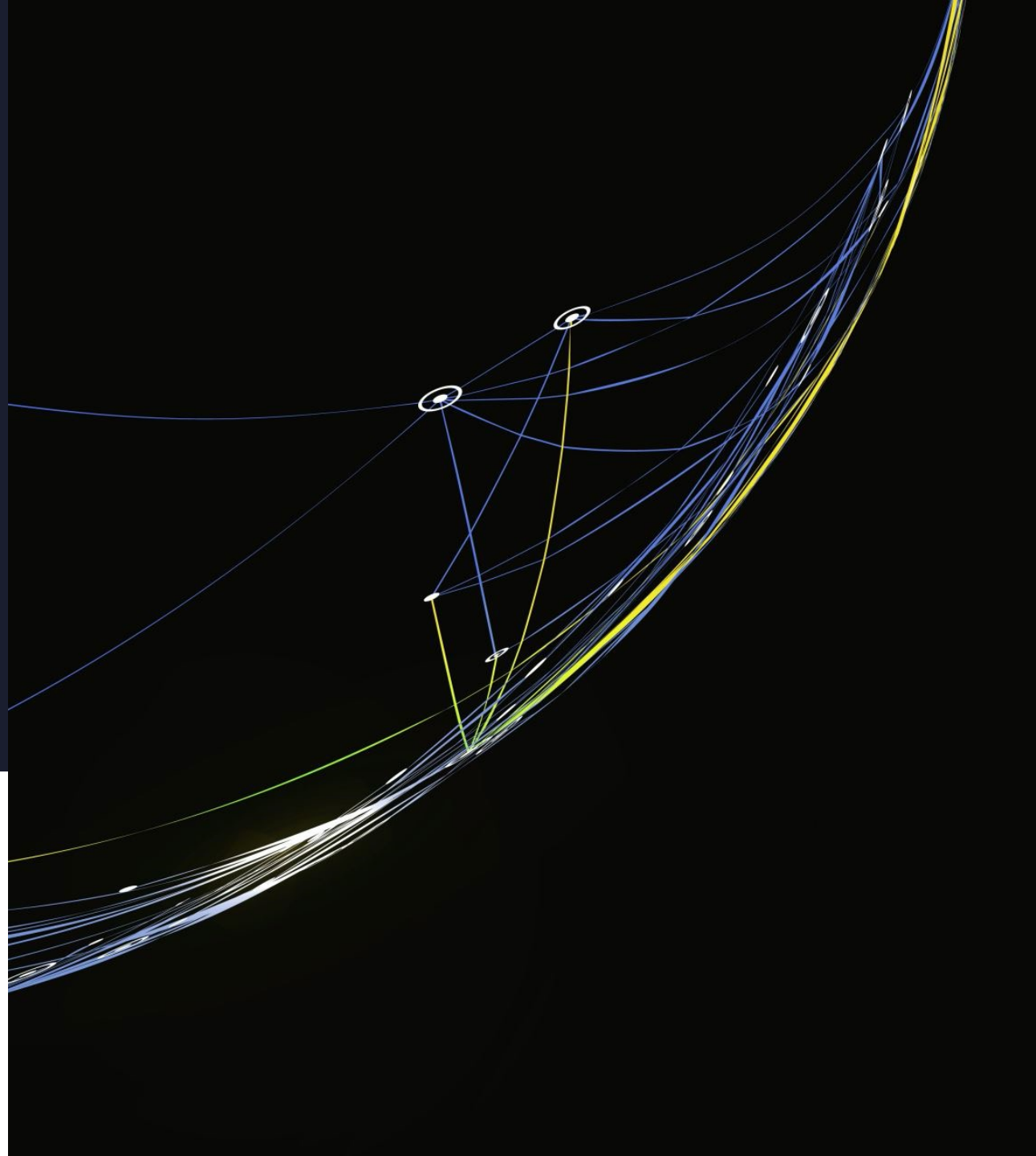
Continuous Treatment with Generalized Propensity score



By: Jasper Tsai

Why Continuous Treatment?

Causal inference traditionally focused on binary treatment variables.

Ex: **Taking a drug** versus **Not Taking a drug** .

Sometimes we are interested in the effects in continuous treatments.

Ex: consider the treatment dose to be continuous to help us decide whether taking too much of a drug may lose effectiveness and become harmful.

Propensity Scores

- Rosenbaum and Rubin (1983) defined the notion of propensity scores for binary treatments.
- Hirano and Imbens extended the idea of propensity scores to the continuous treatment cases.

The Basic Framework (Notation)

- Sample of units, indexed by $i = 1, \ldots, N$
- Each unit i assumes the existence of a set of potential outcomes denoted $Y_i(t)$.
- **T** is an interval $[t_0, t_1]$
- We are interested in the expected potential outcome, or average dose-response function defined as $\mu(t) = E[Y_i(t)]$.

The Basic Framework (Notation)

- Each unit i has a vector of covariates X_i and the treatment received T_i.
- The potential outcome corresponding to the level of treatment received is denoted by $Y_i = Y_i(T_i)$.
  - Drop the subscript i to simplify notation
- We assume that there is a collection of all potential outcomes for all individuals in the sample across all treatment levels: $\{Y(t)\}_{t \in \boldsymbol{T}}$
- T and X are defined in a common probability space.

Weak Unconfoundedness

- $Y(t) \perp T|X$ for all $t \in \boldsymbol{T}$
- By adjusting for the difference in a set of covariates the outcome will be independent to the treatment
- Assumption does not require joint independence of all potential outcomes but require conditional independence for each value of the treatment.

Generalized Propensity Score

- Let $r(t, x)$ be the conditional density of the treatment given the covariates:
  - Definition: $r(t, x) = f_{T|x}(t|x)$
  - The generalized propensity score is $R = r(T, X)$
- GPS is a function that predicts the probability of receiving a particular treatment value t given a set of pretreatment variables X (covariates).
- Like propensity score, GPS also has the balancing property: $X \perp 1\{T = t\}|r(t, X)$

Weak Unconfoundedness Given GPS

- *Theorem 1* in the paper states that if assignment to treatment is weakly unconfounded given pretreatment variables X:
- Then for every t: $f_T(t|r(t,X),Y(t)) = f_T(t|r(t,X))$
- The theorem shows that by conditioning on the GPS, the treatment assignment mechanism does not depend on the potential outcomes beyond what is already captured by the pretreatment variables X.

Bias Removal Using GPS

- Two Step Approach
  1. Estimate the conditional expectation of the outcome as a function of treatment level T and the GPS R.
     - We represent this by $\beta(t,r) = E[Y|T = t, R = r]$
  2. We estimate the dose response function at a particular level of treatment
     - We average the conditional expectation over GPS at that level of the treatment.
     - $\mu(t) = E[\beta\big(t, r(t,X)\big)]$
     - We do not average over the entire GPS $R = r(T,X)$ instead we only average GPS scores that are relevant to the specific treatment level we are interested in analyzing, $r(t,X)$.

Bias Removal Using GPS

- *Theorem 2* states that if assignment to treatment is weakly unconfounded given pretreatment variables X. Then the following is true:

$$\beta(\mathrm{t},\mathrm{r}) = \mathrm{E}[\mathrm{Y}(\mathrm{t})|\mathrm{r}(\mathrm{t},\mathrm{X}) = \mathrm{r}] = E[Y|T = t, R = r]$$
$$\mu(t) = E[\beta(t, r(t, X)]$$

Bias Removal Using GPS

Proof:

Let $f_{Y(t)|T,r(t,X)}(.|t,r)$ denote the conditional density (w.r.t some measure) of $Y(t)$ given $T = t$ and $r(t,X) = r$.

Using Bayes rule and *Theorem 1* (Weak Unconfoundedness Given GPS)

$$f_T(t|r(t,X),Y(t)) = f_T(t|r(t,X))$$

$$f_{Y(t)|T,r(t,X)}(y|t,r) = \frac{f_T(t|Y(t) = y, r(t,X) = r) f_{Y(t)|r(t,X)}(y|r)}{f_T(t|r(t,X) = r)}$$
$$= f_{Y(t)|r(t,X)}(y|r)$$

Take Expectation:

$$\mathrm{E}[Y(t)|T = t, r(t,X) = r] = \mathrm{E}[Y(t)|r(t,X) = r].$$
$$= \beta(t,r)$$

Bias Removal Using GPS

Lastly, use $\mu(t) = E[\beta\big(t, r(t, X)\big)]$ and apply iterative expectation to see that:

$$\mu(t) = E\big[\beta\big(t, r(t, X)\big)\big] = E\big[E[Y(t)|r(t, X)]\big] = E[Y(t)]$$

The above gives us the expected potential outcome, also known as the average dose response function.

Note: $\beta(t, r)$ does not have causal interpretation.

Weighting and Other GPS Methods

- Weighting observations appropriate to account for difference in covariates between treatment groups is important.
  - In continuous treatment, the weights must be “stabilized” by the marginal probabilities of treatment.
- Hirano and Imbens employ parametric assumptions on $\beta(t, r)$ instead of $\mu(t)$ to avoid the need to reweight observations.
- However, there are other method such as Marginal Structural Model (MSM) that uses a weighted scheme based on the GPS.

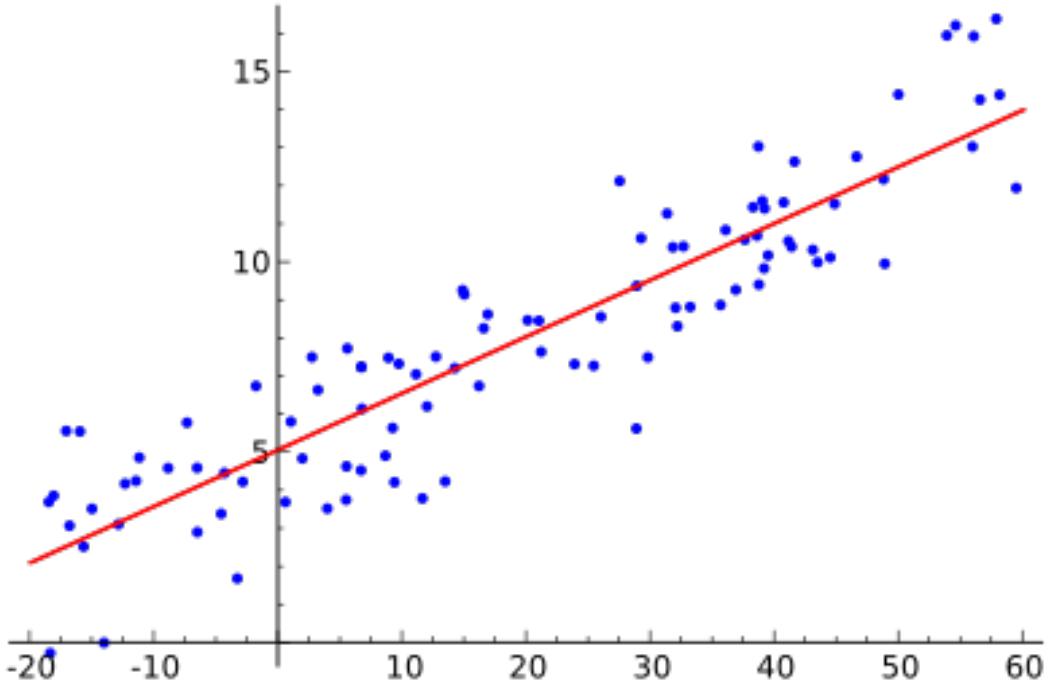
Estimation and Inference (Implementation)

1. Model	Model the continuous treatment as a function of the covariates
2. Obtain	Obtain the Generalized Propensity Score
3. Model	Model the outcomes as a function of the treatment and Generalized Propensity Score
4. Estimate	Estimate the average potential outcome at each treatment dose of interest and plot the dose-response function.

Implementation STEP 1

- Model the continuous treatment variable as a function of covariates.
- Ex: First fit the linear regression model to represent normal distribution for the treatment given covariates

$$T_i|X \sim N(\beta_0 + \beta X_i, \sigma^2)$$



Implementation STEP 2

- We obtain the Generalized Propensity Score by estimating β_0, β, σ^2 using maximum likelihood.

$$\hat{R}_i = r_i(T, X) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{2\sigma^2}(T_i - \hat{\beta}_0 - \hat{\beta}X_i)^2\right)$$

Implementation STEP 3

- Model the conditional expectation of Y_i given T_i and R_i as a flexible function of its two arguments
- We use quadratic approximation:

$$E[Y_i|T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i$$

- Estimate the above parameters by ordinary least squares using the estimated GPS $\hat{R}_i$
- There is still not causal interpretation at this step!

Implementation STEP 4

- We estimate the average potential outcome at each treatment dose of interest
- In our example, we estimate the average potential outcome at treatment level t as:

$$\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t\hat{r}(t, X_i))$$

- Repeat the calculation for each level of the treatment we are interested in to obtain an estimate of the entire dose response function

Sample Data Analysis

Background

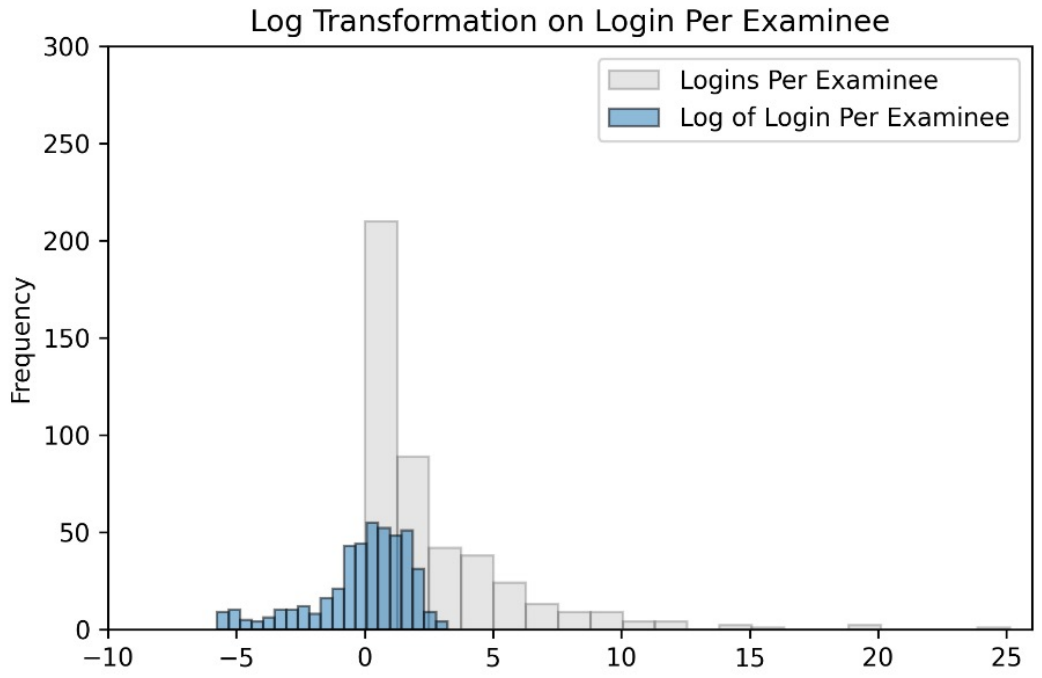
- Math Nation – aimed to improve Florida's high school students' Algebra I End-of-Course (EOC) standardize assessment scores.
- Population consisted of high school, middle/high schools, and senior high schools in all school districts in Florida.
- Data was collected from February to April 2014 and contains observations for 448 schools
- Goal: Engagement with Math Nation vs. school level average student scores for Algebra I EOC assessment.

Data

- The treatment dose is the ratio of number of logins and number of examinees in each school.
  - Total number of logins from both teacher and students
  - Count of students from the school who participated in EOC assessment during Spring 2014
- 13 relevant covariates that are related to treatment dosage and outcome
  - Ex: school mean scores on EOC exam in 2012,, student enrollment numbers, school location and binary indicators for charter, magnet, or title I institution, among others.

Procedures

1. We first regressed our treatment variable with the covariates.
   - (Log transform to adjust for skewness to satisfy normality assumption)
2. Next, we calculated the GPS using the conditional Gaussian density we described previously



$$\hat{R}_i = r_i(T, X) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} exp\left(-\frac{1}{2\sigma^2}\left(T_i - \hat{\beta}_0 - \hat{\beta}X_i\right)^2\right)$$

Procedures

- We quickly assessed covariate balance by stratifying based on GPS and fit one regression for each covariate with the treatment dose as the outcome and with GPS strata and the covariate as predictors.
- Covariate balancing has improved compared to baseline but not all achieved the desired level of balance (<0.1)

Variables	Baseline	GPS Strata
Charter	0.322	0.346
Magnet	0.46	0.319
Title_I_School	0.105	0.024
locationRural	0.322	0.034
locationSize	0.398	0.215
Students	0	0
SeniorHigh	0.064	0
numOfStud2014	0.106	0.038
meanScale2012	0.117	0.077
lev1Perc2012	0.108	0.064
lev5Perc2012	0.084	0.093
perc_free_lunch	0.037	0.015
perc_reduced_lunch	0.056	0.067

Procedures

3. With the assistance of *survey* package in R, we estimated the outcome model:

$$E[Y_i|T_i, R_i] = \alpha_0 + \alpha_1 T_i + \alpha_2 T_i^2 + \alpha_3 R_i + \alpha_4 R_i^2 + \alpha_5 T_i R_i$$

4. Lastly, we leveraged the same package to acquire the individual treatment effects:

$$\widehat{E[Y(t)]} = \frac{1}{N}\sum_{i=1}^{N}(\hat{\alpha}_0 + \hat{\alpha}_1 t + \hat{\alpha}_2 t^2 + \hat{\alpha}_3 \hat{r}(t, X_i) + \hat{\alpha}_4 \hat{r}(t, X_i)^2 + \hat{\alpha}_5 t\hat{r}(t, X_i))$$

- Commonly referred to as the average potential outcomes at treatment dose T based on the coefficients of the outcome model.

Results

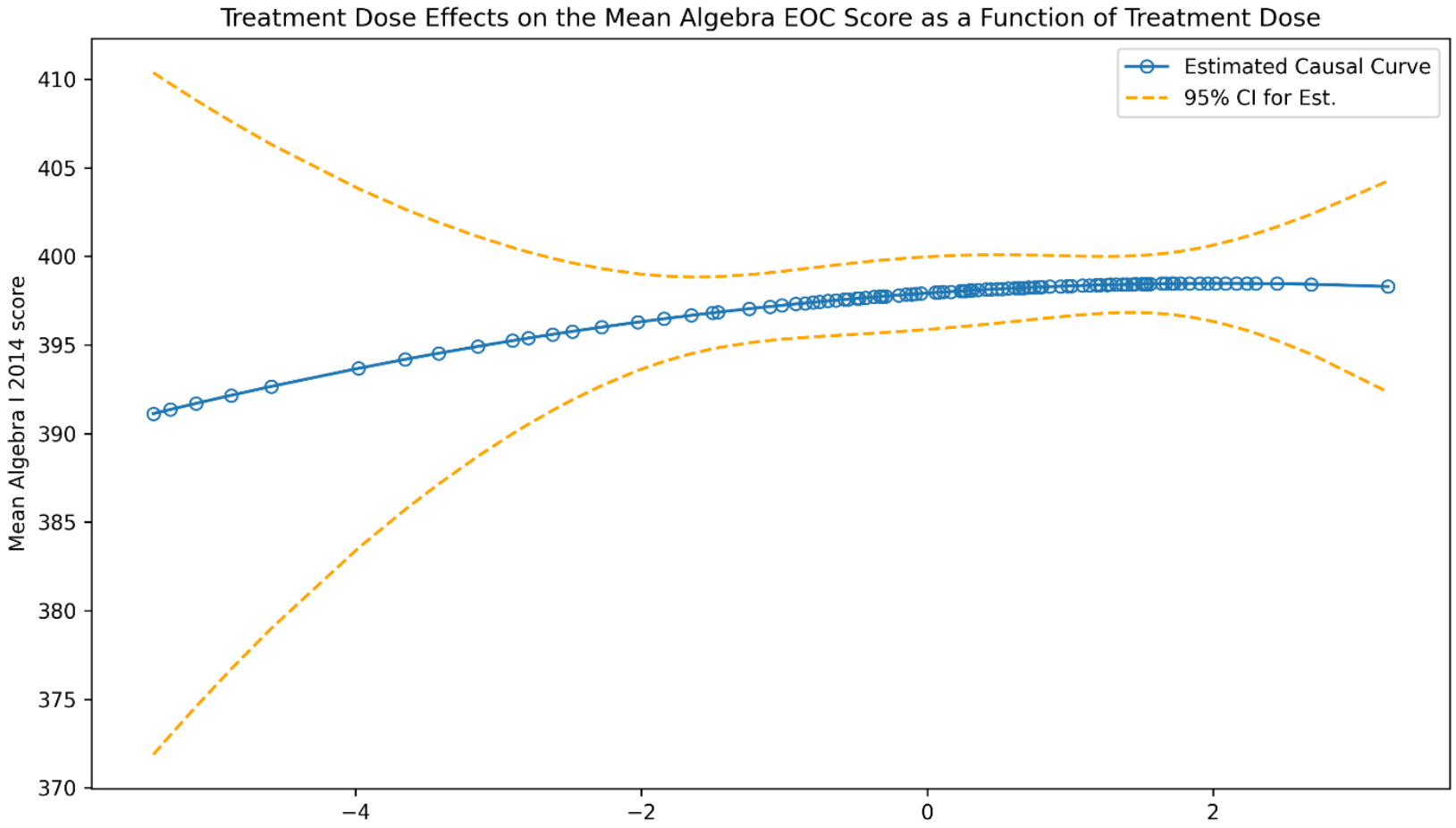
Percentile	logLoginsPerExaminee	meanScale2014	SE	lowerCL	upperCL
10%	-2.9014	395.2388	2.6813	389.9834	400.4942
20%	-1.2459	397.0411	0.9810	395.1184	398.9638
30%	-0.5532	397.5806	1.0235	395.5746	399.5866
40%	-0.1112	397.8587	1.0496	395.8015	399.9159
50%	0.2966	398.0695	1.0190	396.0722	400.0667
60%	0.6664	398.2227	0.9449	396.3707	400.0746
70%	1.0840	398.3523	0.8430	396.7001	400.0046
80%	1.4577	398.4293	0.8178	396.8265	400.0321
90%	1.8349	398.4697	0.9670	396.5744	400.3650
100%	3.2243	398.2944	3.0418	392.3324	404.2563

- We specifically calculated estimates for the average treatment effects across the range of treatment doses from the 1st percentile to the 100th percentile.
- The treatment dose effects can be interpreted as the expected values of the outcome if all the participants received each specific dose of treatment.
- It appears that all confidence intervals overlap, suggesting that there are no significant distinctions between the doses.

Results

Dose Response function

- More accurate in the middle
- Mean score increases as treatment dose increase.



Conclusion and Discussion

- The idea of Generalized Propensity Score extends the nice properties of propensity scores from the binary treatment case to continuous treatment situations.
- This paper is mostly related to formally defining Generalized Propensity Score and provided a simple parametric approach.
- There are still other methods that utilize GPS such as Marginal Structural Model (MSM).
- Hope this presentation has shed light on the significance of the Generalized Propensity Score and its potential for improving causal effect estimation in various research contexts.

References

- Hirano, K., & Imbens, G. W. (2004). The Propensity Score with Continuous Treatments. Wiley Online Library. https://scholar.harvard.edu/imbens/files/hir_07feb04.pdf
- Leite, W. (2019, December 20). Practical propensity score methods using R. Sage Research Methods. https://methods.sagepub.com/book/practical-propensity-score-methods-using-r/i852.xml