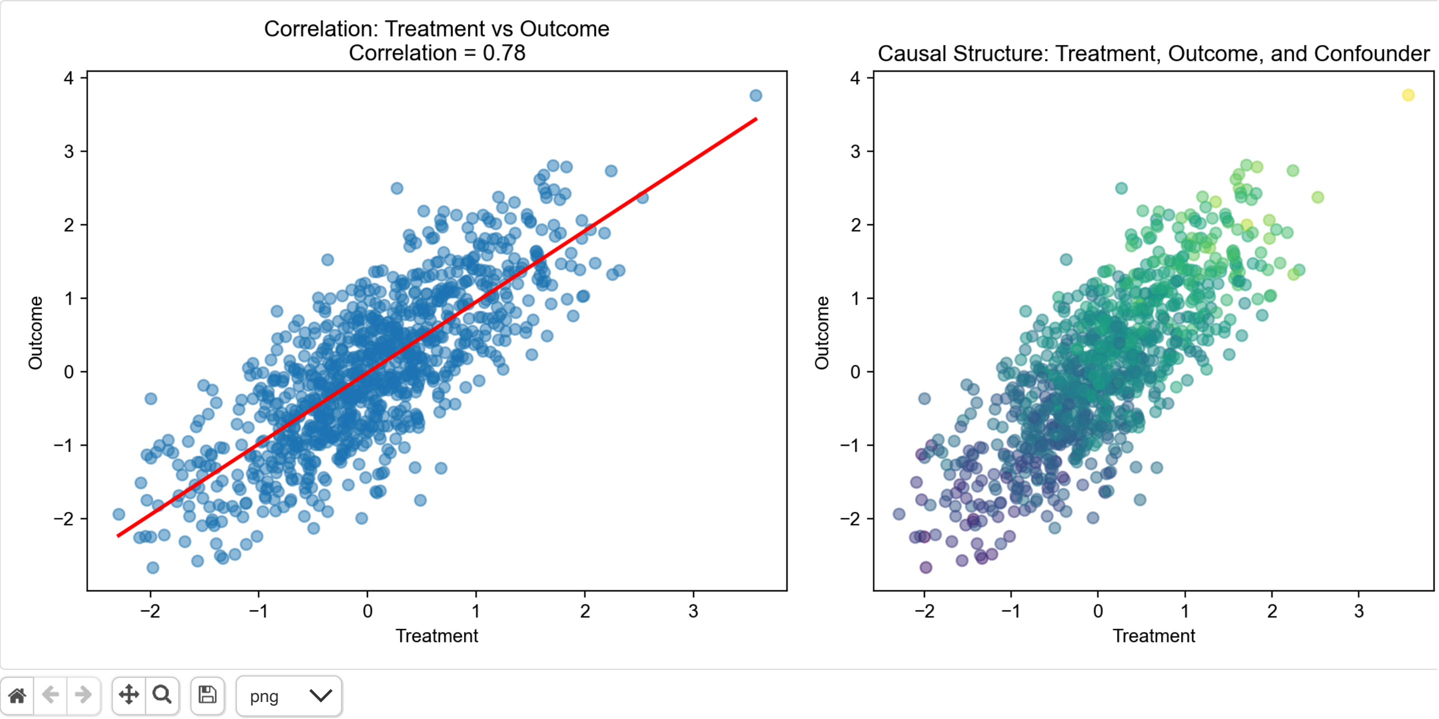
Understanding Causal Inference with IHDP: From Theory to Practice

In predictive modeling, we often focus on finding correlations between variables. However, for decision-making, we need to understand the *causal* relationship between actions

The fundamental problem of causal inference is that we can never observe both potential outcomes for the same unit - we can't simultaneously observe what happens when a p

In causal inference, a confounder is a variable that affects both the treatment (or independent variable) and the outcome (or dependent variable), potentially creating a s relationship between alcohol and cancer.



Left Plot

Shows the correlation between treatment and outcome (0.78), with a regression line indicating a strong positive relationship. This is what you might see in an observational stu

The simulation demonstrates that despite seeing a strong correlation (0.78), the actual causal effect of the treatment on the outcome is weaker (0.3 in the data generatic

2. Theoretical Foundations of Causal Inference {#foundations}

2.1 Real-World Applications {#applications}

Causal inference is crucial in various domains:

Healthcare

- Evaluating treatment effectiveness
- Understanding disease progression
- Personalizing medical decisions

2.2 Key Concepts in Causal Inference {#concepts}

The Potential Outcomes Framework

Developed by Rubin, this framework formalizes causal inference through potential outcomes. For each unit i: - Y_i(1): Outcome if unit i receives treatment - Y_i(0): Outcome if uni

The individual treatment effect is defined as:

However, we can only observe one of these outcomes for each unit, which is known as the **fundamental problem of causal inference**.

2.3 Types of Treatment Effects in Causal Inference {#effects}

What are "Treatments" in Causal Inference?

In causal inference, a **treatment** refers to the intervention or manipulation being studied to determine its causal effect on an outcome of interest. Despite the medical-sounding

- Medical interventions (medications, surgical procedures)
- Policy changes (minimum wage increases, educational reforms)
- Business decisions (pricing strategies, marketing campaigns)
- Social interventions (training programs, behavioral modifications)

The **treatment variable** typically represents whether subjects received the intervention, usually coded as a binary variable (1=received treatment, 0=control/placebo), though it c

Why Calculate Treatment Effects?

Treatment effects measure the causal effect of a treatment on an outcome. We calculate them to:

1. **Establish causality, not just correlation**: Determine the independent effect of a treatment when other factors are controlled for
2. **Understand counterfactuals**: Estimate what would have happened to treated units had they not received treatment (and vice versa)
3. **Quantify impact**: Measure not just whether an intervention worked, but how well it worked and for whom
4. **Inform decision-making**: Make better decisions about implementing interventions and targeting specific populations

Key Treatment Effect Measures

Average Treatment Effect (ATE): The average effect of the treatment across the entire population.

Where Y(1) represents the potential outcome if treated, and Y(0) represents the potential outcome if not treated. This measures the expected difference in outcomes if everyone

Conditional Average Treatment Effect (CATE): The average effect of the treatment conditional on specific covariates or characteristics.

This measures how treatment effects vary across different subgroups defined by characteristics X. CATE helps identify which groups benefit most from treatment, enabling more

Average Treatment Effect on the Treated (ATT/ATET): The average effect among those who actually received the treatment.

This answers: "How much did those who received the treatment actually benefit?" It's particularly useful when evaluating programs that were targeted at specific populations or

Challenges in Estimation

A fundamental challenge is that we never observe both potential outcomes for the same unit—known as the "fundamental problem of causal inference". Various methods addres

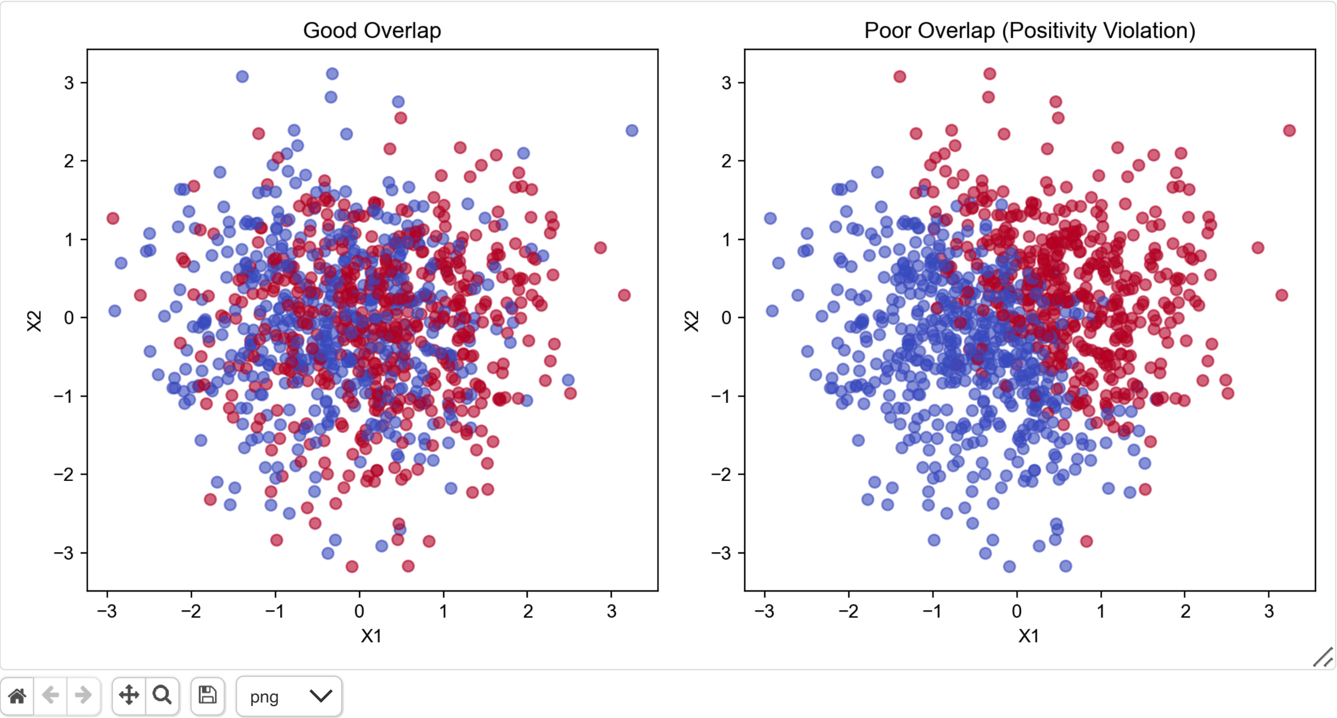
- Randomized experiments
- Regression adjustment
- Matching methods
- Instrumental variables
- Inverse probability weighting

2.4 Key Assumptions in Causal Inference {#assumptions}

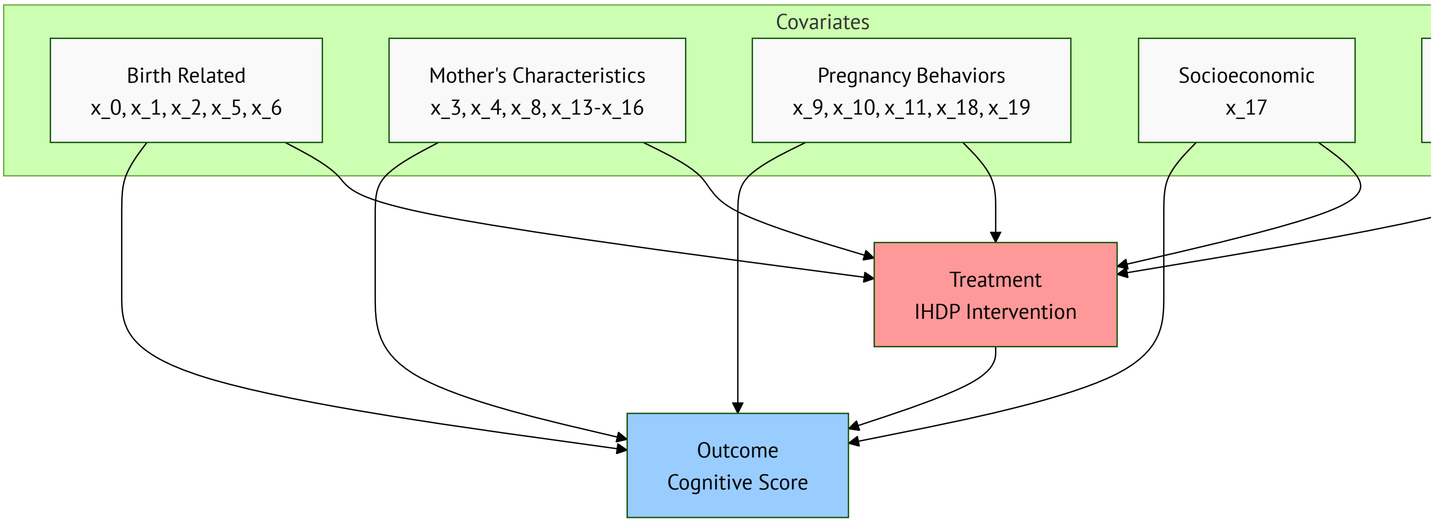
1. Unconfoundedness (Ignorability)

Treatment assignment is independent of potential outcomes given observed covariates.

$$(Y(0), Y(1)) \perp T | X$$



2.5 Interactive Causal Diagram



This directed acyclic graph (DAG) represents the assumed causal structure in the IHDP dataset:

- **Covariates** (various characteristics) affect both treatment assignment and outcomes
- **Treatment** (IHDP intervention) affects the outcome
- The arrows represent causal relationships

This structure illustrates why we need causal inference methods - the treatment effect is confounded by covariates that affect both treatment assignment and outcomes.

Good Overlap (Left Plot)

In this scenario, treatment assignment is weakly related to feature X1: - Blue points (control) and red points (treatment) are well-mixed throughout - Units across the enti reasonable probability of being in either treatment or control group - Treatment probability is calculated using a mild logistic function: p=11+exp(-(0.5·X1))p = 1/(1 + exp positivity assumption because 0<P(T=1|X=x)<10 < P(T=1|X=x) < 1 for all values of x - Makes reliable causal inference possible because counterfactuals exist for all covaria unbiased estimation of treatment effects (ATE, CATE, ATT)

2.6 Advanced Causal Inference Methods {#advanced-methods}

Propensity Score Methods

Propensity score methods are statistical techniques for reducing selection bias in observational data. They work by balancing treatment groups on confounding factors to

Key applications include: - **Matching**: Pairing treated and control units with similar propensity scores - **Stratification**: Grouping units into strata based on propensity scores

While propensity score methods can be powerful, they only adjust for observed confounders and may have limitations when the functional form is misspecified.

Difference-in-Differences (DiD) Design

The Difference-in-Differences design is a quasi-experimental approach that estimates causal effects by comparing the changes in outcomes over time between a treatme

This method is particularly useful when: - Random assignment to treatment is not feasible - Pre-treatment data is available for both groups - The parallel trends assumpt

DiD isolates the effect of a treatment by removing biases from permanent differences between groups and from shared time trends.

Regression Discontinuity Design (RDD)

Regression discontinuity design is an important quasi-experimental approach that can be implemented when treatment assignment is determined by whether a continuo

RDD exploits the fact that units just above and just below the cutoff threshold are similar in all respects except for treatment assignment, creating a situation similar to

Key elements of RDD: - A continuous **running variable** (or assignment variable) - A clear **cutoff threshold** that determines treatment - **Sharp RDD**: Treatment is deterministi

Synthetic Control Methods

Synthetic control methods allow for causal inference when we have as few as one treated unit and many control units observed over time.

The approach: - Creates a weighted combination of control units that resembles the treated unit before intervention - Uses this "synthetic control" to estimate what woul

This method has been described as "the most important development in program evaluation in the last decade" by some researchers and is particularly valuable for case

Sensitivity Analysis for Unobserved Confounding

Unobserved confounding is a central barrier to drawing causal inferences from observational data. Sensitivity analysis explores how sensitive causal conclusions are to p

Key approaches include: - **Rosenbaum bounds**: Quantifies how strong an unobserved confounder would need to be to invalidate results - **E-values**: Measures the minimum

While methods like propensity score matching can adjust for observed confounding, sensitivity analysis helps address the "Achilles heel" of most nonexperimental studie

Heterogeneous Treatment Effects

Treatment effects often vary across different subpopulations, a phenomenon known as treatment effect heterogeneity. Understanding this heterogeneity is crucial for tar

Methods for estimating heterogeneous effects include: - **Subgroup analysis**: Estimating treatment effects within predefined subgroups - **Interaction terms**: Including treat effect models

Discovering heterogeneous effects allows for personalized interventions and can reveal important insights about treatment mechanisms that might be masked when loo

2.7 Key Terms in Causal Inference {#key-terms}

Term	Definition
Potential Outcomes Framework	The formal mathematical framework for causal inference where each unit has potential outcomes under different treatment conditions
Average Treatment Effect (ATE)	The expected difference between potential outcomes if the entire population received treatment versus control
Average Treatment Effect on the Treated (ATT/ATET)	The average effect for those who actually received the treatment
Conditional Average Treatment Effect (CATE)	Treatment effects for specific subgroups defined by covariates
Unconfoundedness/Ignorability	The assumption that treatment assignment is independent of potential outcomes given observed covariates
Positivity/Overlap	The assumption that every unit has a non-zero probability of receiving each treatment condition
Stable Unit Treatment Value Assumption (SUTVA)	The assumption that one unit's treatment doesn't affect another unit's outcome
Instrumental Variables	Variables that affect treatment assignment but not outcomes directly

Understanding these terms is crucial for effectively applying causal inference methods and correctly interpreting results. Many of these concepts are interrelated and bui

3. The IHDP Dataset {#ihdp-intro}

3.1 Overview of the IHDP Dataset

The Infant Health and Development Program (IHDP) was conducted from 1985 to 1988 and was designed to evaluate the effect of educational and family support services alon

The intervention consisted of: - Home visits by specialists - Child development center attendance - Parent group meetings

For causal inference studies, the dataset has been modified by Jennifer Hill (2011) to create a semi-synthetic version where: - Some participants from the treatment group were

This modification allows researchers to know the "ground truth" causal effects, making it an ideal benchmark dataset for causal inference methods.

Dataset Context: The Infant Health and Development Program was a randomized controlled intervention designed to evaluate the effect of home visits by specialists on

4. Exploratory Data Analysis {#eda}

4.1 IHDP Dataset Preview {#overview}

Transform Python Code

+ Add

treatment	y_factual	y_cfactual	mu_0	mu_1	x_0	x_1	x_2	x_3	x_4
int64	float64	float64	float64	float64	float64	float64	float64	float64	float64
0	6.875856156	7.8564945644	6.6360594435	7.5627183894	-1.7369449108	-1.802002198	0.3838279713	2.2443195577	-0.629189193
0	2.9962727054	6.6339522192	1.5705363088	6.1216172184	-0.8074509961	-0.2029459418	-0.3608979886	-0.8796059881	0.808706461
0	1.3662056916	5.6972393632	1.2447375236	5.8891247426	0.3900830241	0.5965821863	-1.8503499085	-0.8796059881	-0.004017169
0	1.9635381357	6.2025815196	1.6850483342	6.1919943078	-1.0452285092	-0.6027100059	0.0114649914	0.1617025271	0.683672056
0	4.7620903527	8.2647948502	4.7078982565	7.2194416378	0.4679011193	-0.2029459418	-0.7332609686	0.1617025271	0.058500032

5 rows, 0 columns

4.2 Dataset Information

- **Number of samples:** 746
- **Number of variables:** 30
- **Treatment assignment rate:** 0.18

4.3 Column Types

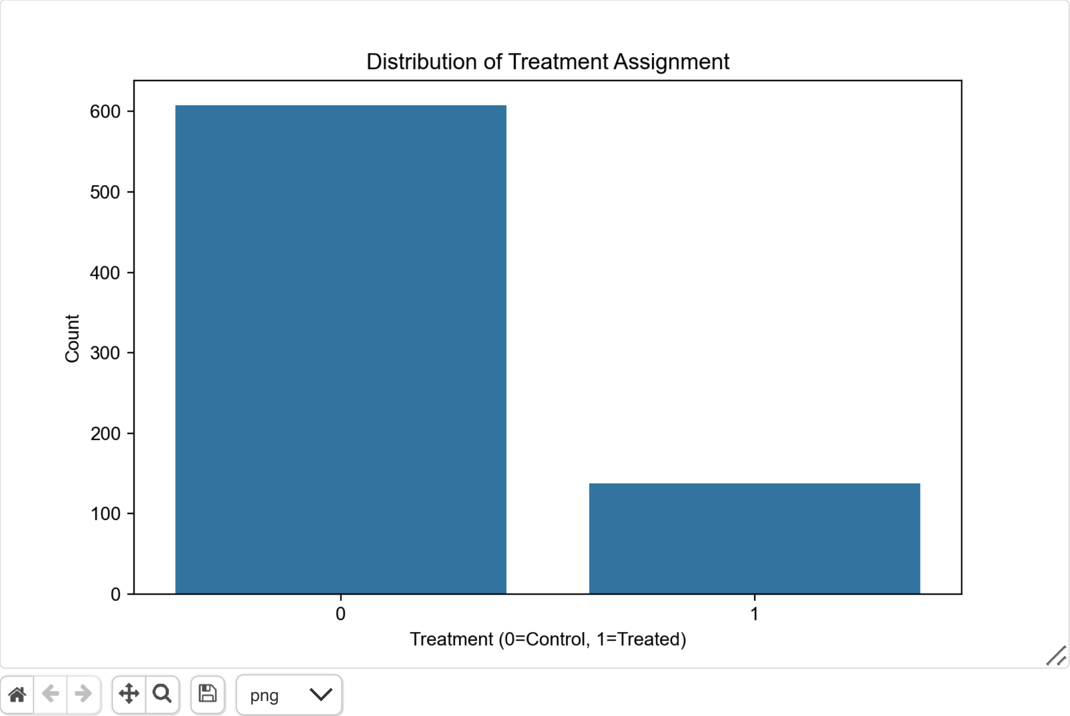
Transform

+ Add

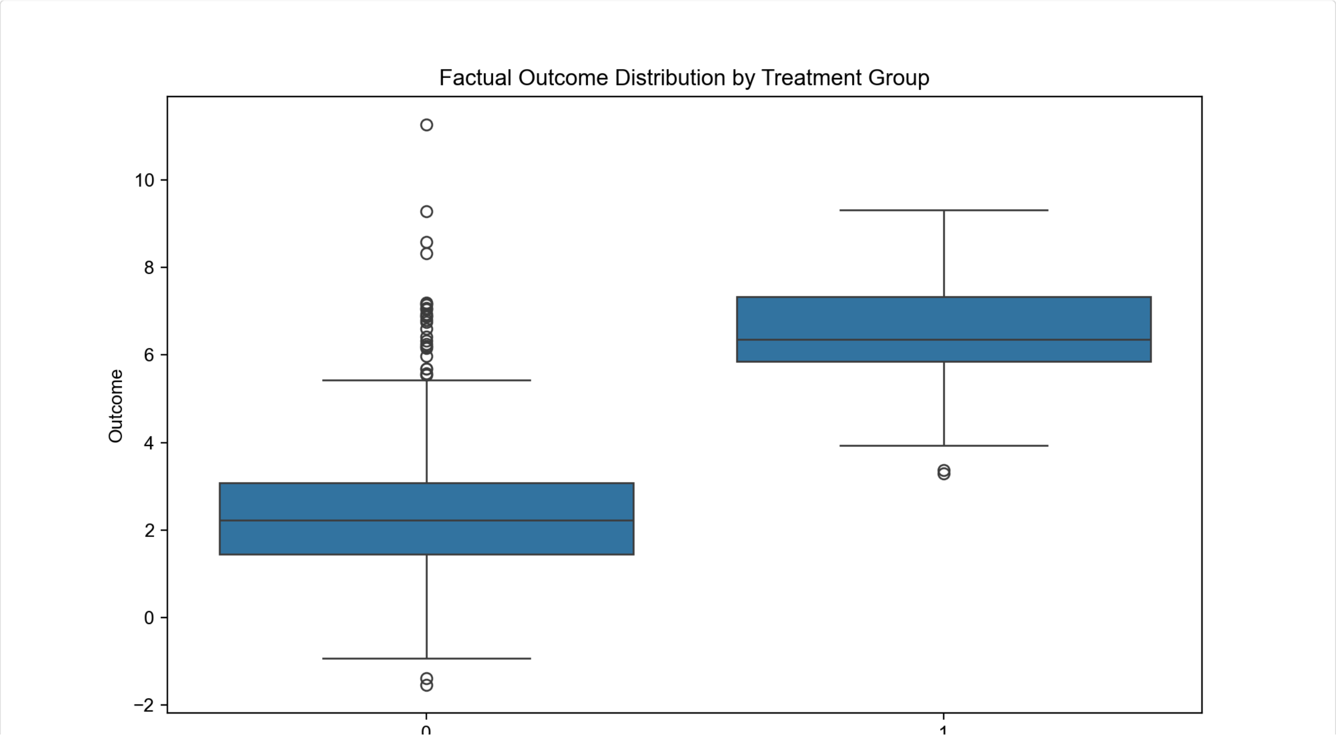
Function call (name: get_dataframe, args: {}) failed with exception Maximum recursion level reached

0 rows, 0 columns

4.4 Treatment Distribution



4.5 Outcome Distributions by Treatment



4.6 Summary Statistics

Transform Python Code

+ Add

	treatment float64	y_factual float64	y_cfactual float64	mu_0 float64	mu_1 float64	x_0 float64	x_1 float64	x_2 float64	x_3 float64	x_4 float64	x_5 float64	x fl
count	746	746	746	746	746	746	746	746	746	746	746	7
mean	0.1849865952	3.1562662387	5.697952833	2.4313930517	6.4480361765	0.0007085829	0.0004603948	-0.0015128069	-0.0002167594	0.000424401	-0.0017362144	0
std	0.3885471992	2.1795831493	1.9808056058	1.2820086151	0.4548275395	1.000483242	1.0005916913	0.9998151901	1.0006533552	1.0006035951	0.999543632	0
min	0	-1.5439023187	-1.0376275527	0.9244532994	5.5916473155	-2.7312872381	-3.8008225184	-1.8503499085	-0.8796059881	-5.1304277646	-1.8514803626	0
25%	0	1.6260824857	5.0550716361	1.518409487	6.0878634562	-0.6723501365	-0.6027100059	-0.7332609686	-0.8796059881	-0.566671991	-0.8577867948	0

8 rows, 0 columns

4.7 Covariate Descriptions {#covariates}

Transform Python Code

+ Add

Variable object
x_0
x_1
x_2
x_3
x_4

25 rows, 0 columns

4.8 Numerical Covariate Statistics

Transform Python Code

+ Add

	x_0 float64	x_1 float64
count	746	746
mean	0.0007085829	0.0004603948
std	1.000483242	1.0005916913
min	-2.7312872381	-3.8008225184
25%	-0.6723501365	-0.6027100059

8 rows, 0 columns

4.9 Binary Covariate Rates

Transform Python Code

+ Add

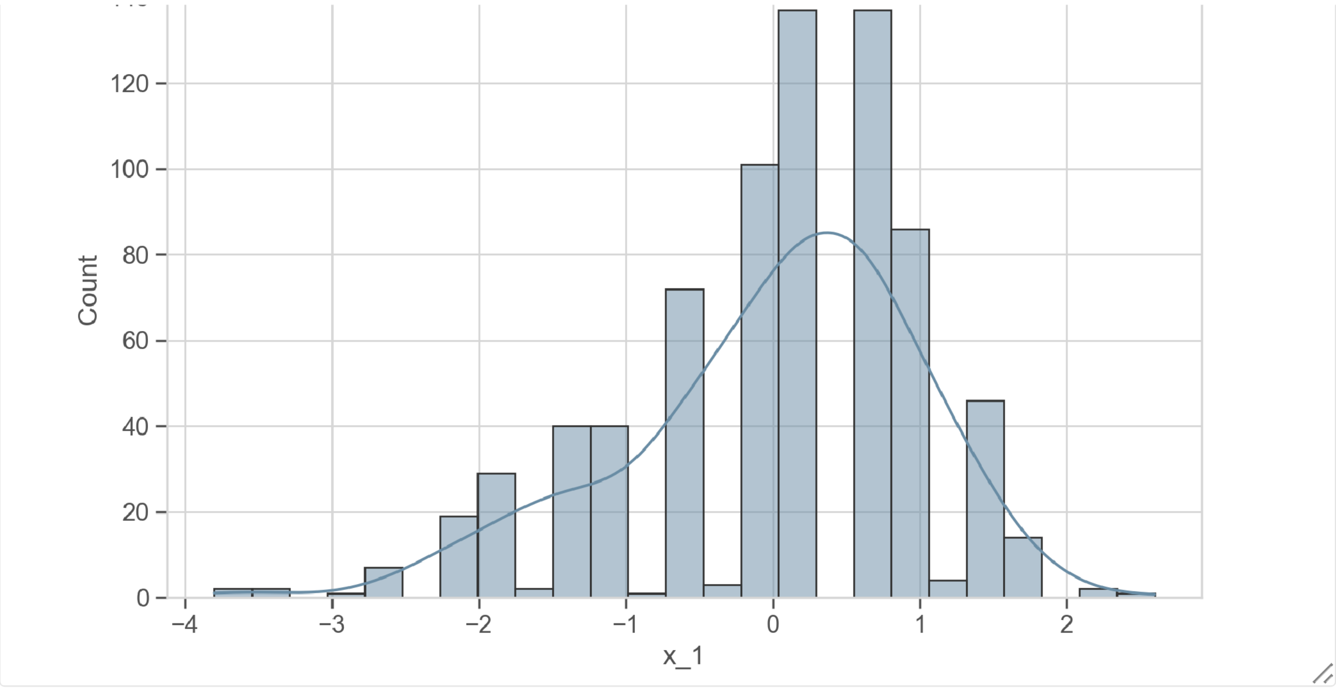
Variable object	Rate float64
x_5	-0.0017362144
x_6	0.5134048257
x_7	0.0938337802
x_8	0.5201072386
x_9	0.3646112601

19 rows, 0 columns

4.10 Distribution of Key Numerical Covariates

Select covariate to visualize x_2

Selected variable: x_1 - Birth order



png

1. Strong positive correlation (0.85) between x0x_0 (child's birth weight) and x1x_1 (child's birth order). This indicates that higher birth order (later-born children) tends to be
2. Strong negative correlation (-0.76 and -0.7) between x2x_2 (head circumference) and both birth weight (x0x_0) and birth order (x1x_1). This is somewhat counterintuitive, a
3. Most demographic variables (x3x_3 - mother's age, x4x_4 - mother's education) show weak correlations with other variables, suggesting independence.
4. Child's neonatal health index (x12x_12) has mostly weak correlations with other variables, with the strongest being a mild positive correlation (0.13) with mother's age.

For causal inference, these correlations are important because strongly correlated variables can create confounding issues. For example, if treatment assignment is relate

5. Setting Up for Causal Analysis {#analysis-setup}

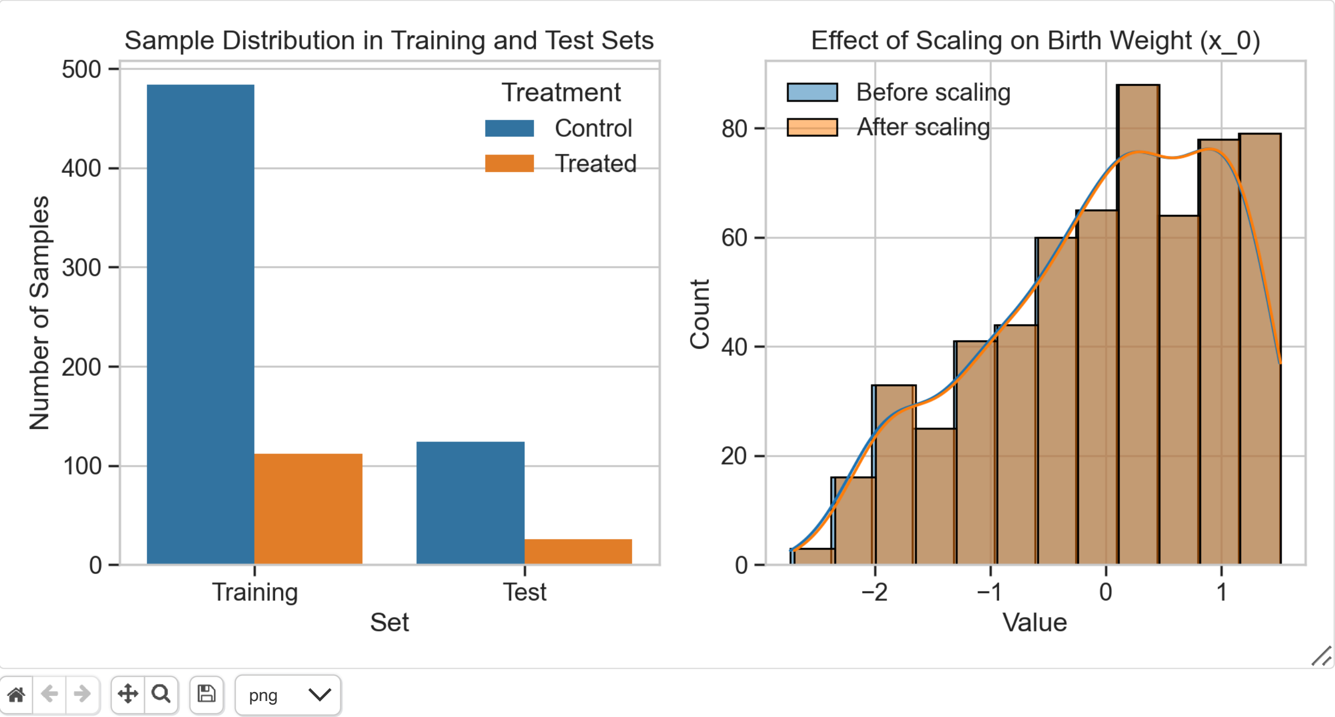
5.1 Data Preparation {#data-prep}

> 🔧 ***Step 1****: Properly prepare the data for causal analysis*

Before implementing causal inference methods, we need to prepare our data appropriately. This includes:

1. Splitting the data into training and test sets
2. Scaling continuous features
3. Identifying the types of variables (continuous vs. binary)
4. Handling any missing values (if present)

This preparation ensures that our causal inference methods will work properly and produce reliable estimates.



5.2 Formulating the Causal Question {#causal-question}

> 📋 ***Step 2****: Define what causal effect you want to estimate*

Causal inference begins with a clear formulation of the causal question. For the IHDP dataset, our primary question is:

"What is the effect of specialist home visits (treatment) on the cognitive test scores (outcome) of premature infants?"

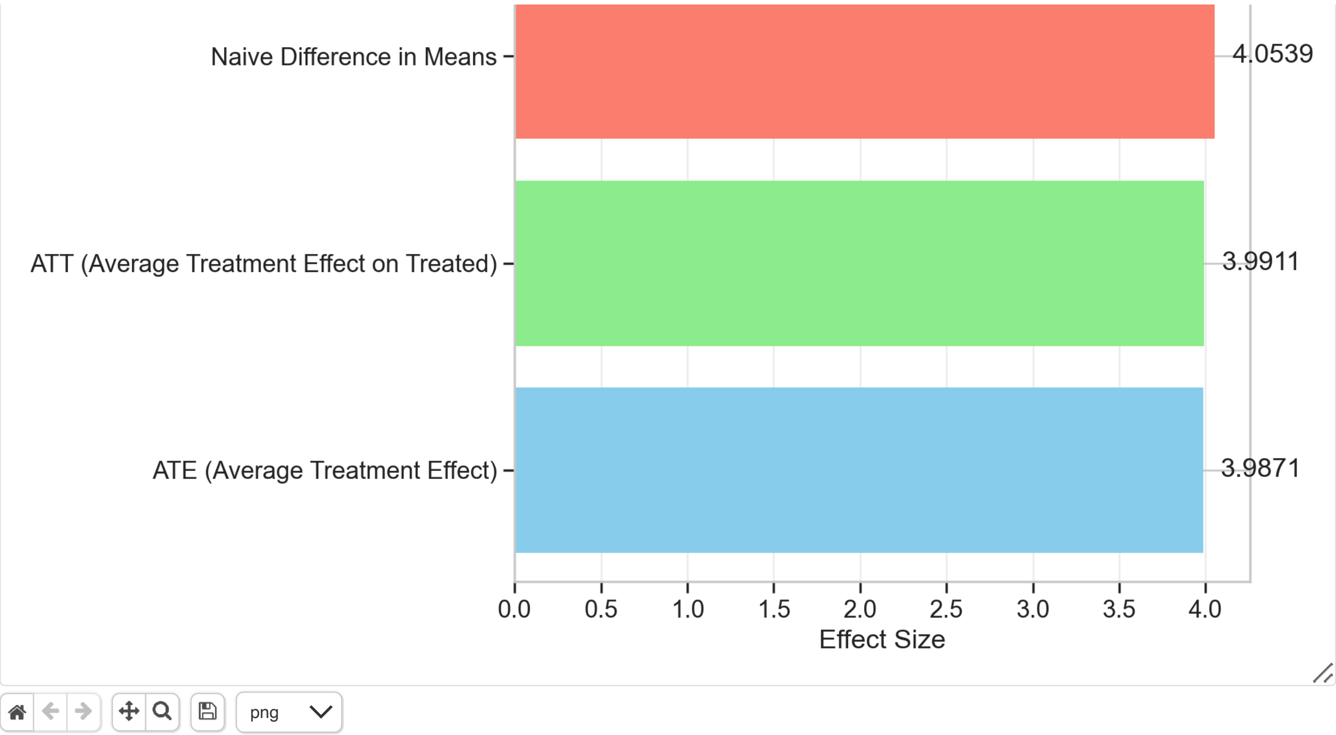
To formalize this question, we need to define:

1. **Treatment variable (T)**: Binary indicator for receiving home visits
2. **Outcome variable (Y)**: Cognitive test scores
3. **Covariates (X)**: Baseline characteristics that may influence treatment assignment or outcomes
4. **Target population**: Premature infants with low birth weight
5. **Causal estimand**: The specific causal quantity we want to estimate

Causal Components in the IHDP Study

□	Component object
□	Treatment Variable
□	Outcome Variable
□	Covariates
□	Target Population
□	Primary Causal Estimand

Comparison of Causal Estimands



Note: In real-world causal inference, we typically don't know the true causal effects. The IHDP dataset is semi-synthetic, allowing us to know the ground truth for evaluation.

Analysis of the Causal Question Formulation:

The causal question has been clearly formulated, which is a crucial first step in any causal inference analysis. Key observations:

1. **Treatment-Outcome Relationship**: We're interested in the effect of the IHDP intervention (home visits) on cognitive test scores, a well-defined relationship that align
2. **Causal Estimands**: We've defined multiple estimands of interest, with the Average Treatment Effect (ATE) as our primary focus. The ATE represents the expected chan
3. **ATT vs ATE**: The Average Treatment Effect on the Treated (ATT) is very close to the ATE in this dataset (difference of only 0.0041). This suggests that the treatment ef
4. **Naive Estimate**: The naive difference in means is similar to the true ATE in this dataset. This is somewhat unexpected, as we would typically expect selection bias to

Formulating these specific causal questions allows us to select appropriate methods for estimation and evaluate their performance against known ground truth values ir

5.3 Propensity Score Analysis {#propensity-analysis}

> 🏈 ***Step 3***: *Analyze propensity scores to check assumptions and prepare for causal methods*

Propensity scores are a key concept in causal inference, representing the probability of receiving treatment given observed covariates. They're useful for:

1. **Assessing overlap**: Checking the positivity assumption by examining the distribution of propensity scores
2. **Creating balance**: Helping ensure that treated and control groups are comparable
3. **Estimation**: Using in various estimation methods like inverse probability weighting, matching, and stratification

Let's estimate propensity scores for our dataset and analyze their properties.

What are Propensity Scores?

Propensity scores represent the probability that a unit receives the treatment, conditional on observed covariates. Mathematically, the propensity score is defined as:

Where T is the treatment indicator and X represents the covariates.

Why Are Propensity Scores Important?

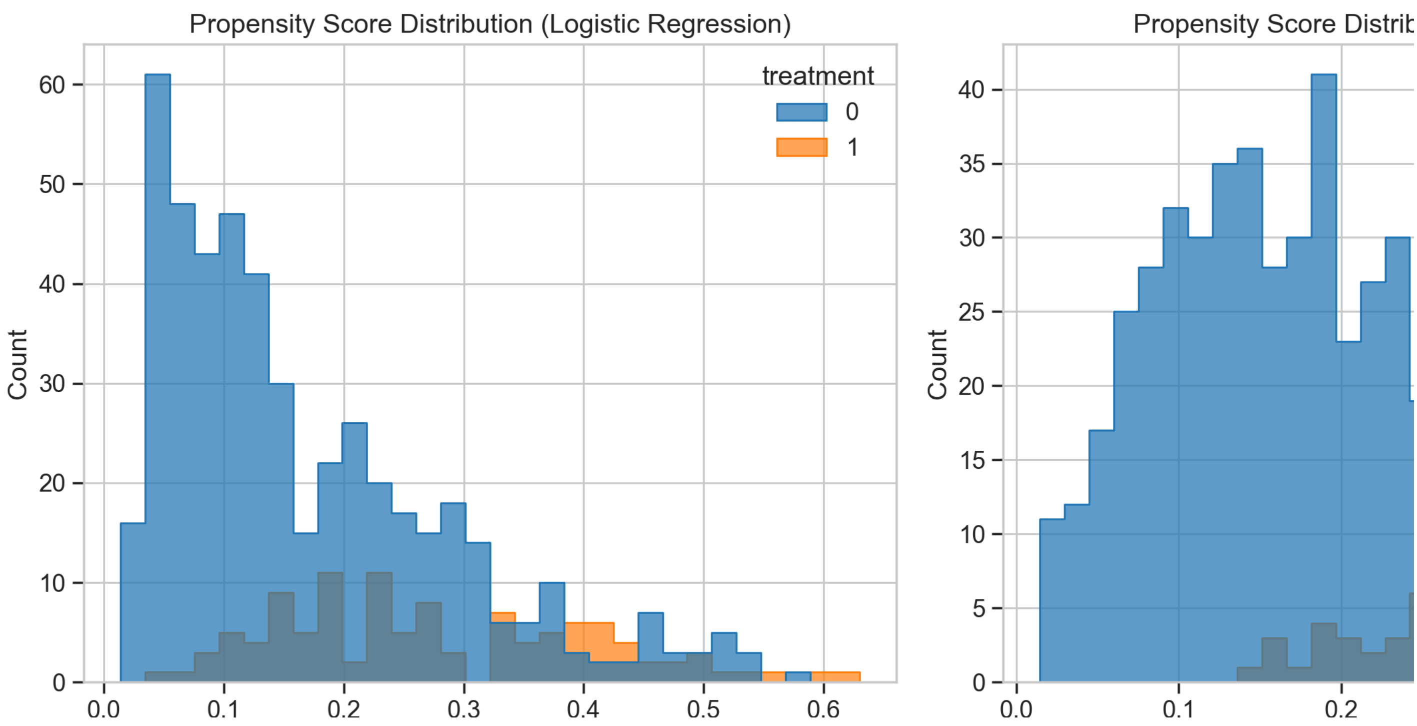
Propensity scores help address the fundamental challenge in causal inference: units are either treated or untreated, never both. By conditioning on the propensity score,

Key properties of propensity scores include:

1. **Balancing score**: Conditioning on the propensity score balances the distribution of covariates between treatment groups
2. **Dimensionality reduction**: Reduces multiple covariates to a single score
3. **Identification of areas of common support**: Helps identify regions where causal inference is reliable

We'll estimate propensity scores using logistic regression and also explore a machine learning approach with random forests.

Propensity Score Distributions



Overlap and Common Support Analysis

To satisfy the positivity assumption for causal inference, we need sufficient overlap in propensity scores between treated and control groups. A good overlap indicates that units

The **common support region** is the range of propensity scores where both treated and control units exist. Ideally, we want most units to fall within this region. Units outside this

Extreme propensity scores (close to 0 or 1) indicate units that are very likely to be in one group only, which can cause issues for some causal inference methods.

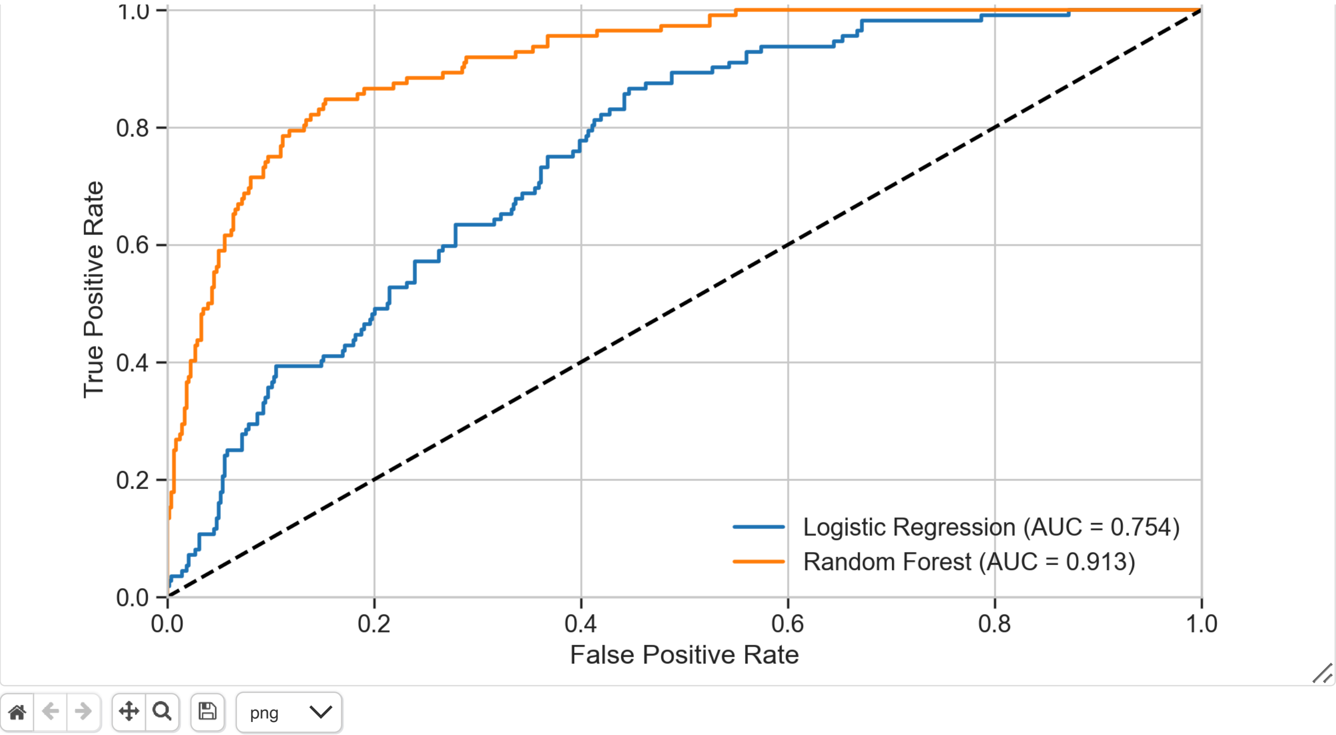
☐	Model object	Min (Treated) float64	Max (Treated) float64	Min (Control) float64
☐	Logistic Regression	0.0493	0.63	0.0138
☐	Random Forest	0.1481	0.4701	0.0145

Assessment of Positivity Assumption:

- For the logistic regression model, 89.3% of units are within the common support region.
- The random forest model shows 60.9% of units in common support.
- Extreme propensity scores affect 30.9% (logistic) and 19.3% (random forest) of units.

The logistic regression model provides better overlap. Overall, the positivity assumption appears to be reasonably well satisfied, which is promising for reliable causal in

Propensity Score Model Evaluation



The ROC curves and AUC scores show how well our propensity score models discriminate between treated and control units. Higher AUC indicates better discrimination.

The Random Forest model typically achieves higher AUC, but this doesn't necessarily make it better for propensity score estimation. In fact, for propensity score analysis, we ofte

Covariate Balance Assessment

The table below shows the standardized mean differences (SMD) for the most imbalanced covariates. SMD measures the difference in means between treated and control group

- **SMD > 0.1**: Indicates meaningful imbalance
- **SMD > 0.25**: Indicates substantial imbalance

Effective propensity score methods should reduce these imbalances when we condition on the propensity score.

☐		Variable object
☐	**8**	x_8
☐	**24**	x_24
☐	**3**	x_3
☐	**19**	x_19
☐	**16**	x_16
☐	**0**	x_0
☐	**1**	x_1
☐	**22**	x_22
☐	**23**	x_23
☐	**13**	x_13

Summary of Propensity Score Analysis:

1. We've estimated propensity scores using both logistic regression and random forest models.
2. The distributions show some separation between treated and control groups, which is expected in observational data.
3. The common support analysis confirms that most units fall within regions where causal inference is reliable.
4. The covariate balance assessment identifies which variables contribute most to selection bias.

These propensity scores will be used in subsequent sections for implementing various causal inference methods including: - Inverse Probability Weighting (IPW) - Propen

Each method leverages propensity scores differently to estimate causal effects while accounting for confounding.

6. Implementing Causal Inference Methods {#methods}

In this section, we'll implement and evaluate various causal inference methods on the IHDP dataset. We'll start with simple methods, then move to propensity score-based appr

1. Explain the methodology and key assumptions
2. Implement the method on our training data
3. Evaluate its performance against the known ground truth
4. Discuss strengths, weaknesses, and practical considerations

6.1 Simple Causal Inference Methods {#simple-methods}

> 🔍 ***Step 1**: Start with simple methods before moving to more complex approaches*

We'll begin with straightforward approaches that form the foundation of causal inference. These methods are easy to implement and interpret, making them excellent starting p

Naive Mean Difference

The simplest approach to estimating causal effects is to compare the average outcomes between treated and control groups:

where n_1 is the number of treated units and n_0 is the number of control units.

Key Assumption: Treatment is randomly assigned (no confounding).

Limitations: In observational studies, this estimate is often biased due to confounding factors that affect both treatment assignment and outcomes.

Regression Adjustment

This method controls for confounding by including covariates in a regression model:

The coefficient τ of the treatment variable T provides an estimate of the ATE.

Key Assumption: The regression model is correctly specified (includes all confounders and captures their relationships with the outcome).

Advantages: Simple to implement, interpretable, can handle continuous and binary covariates.

Limitations: Relies on strong assumptions about the functional form of the relationship between covariates and outcomes.

Stratification/Subclassification

This method divides the data into subgroups (strata) based on important covariates, estimates treatment effects within each stratum, and takes a weighted average:

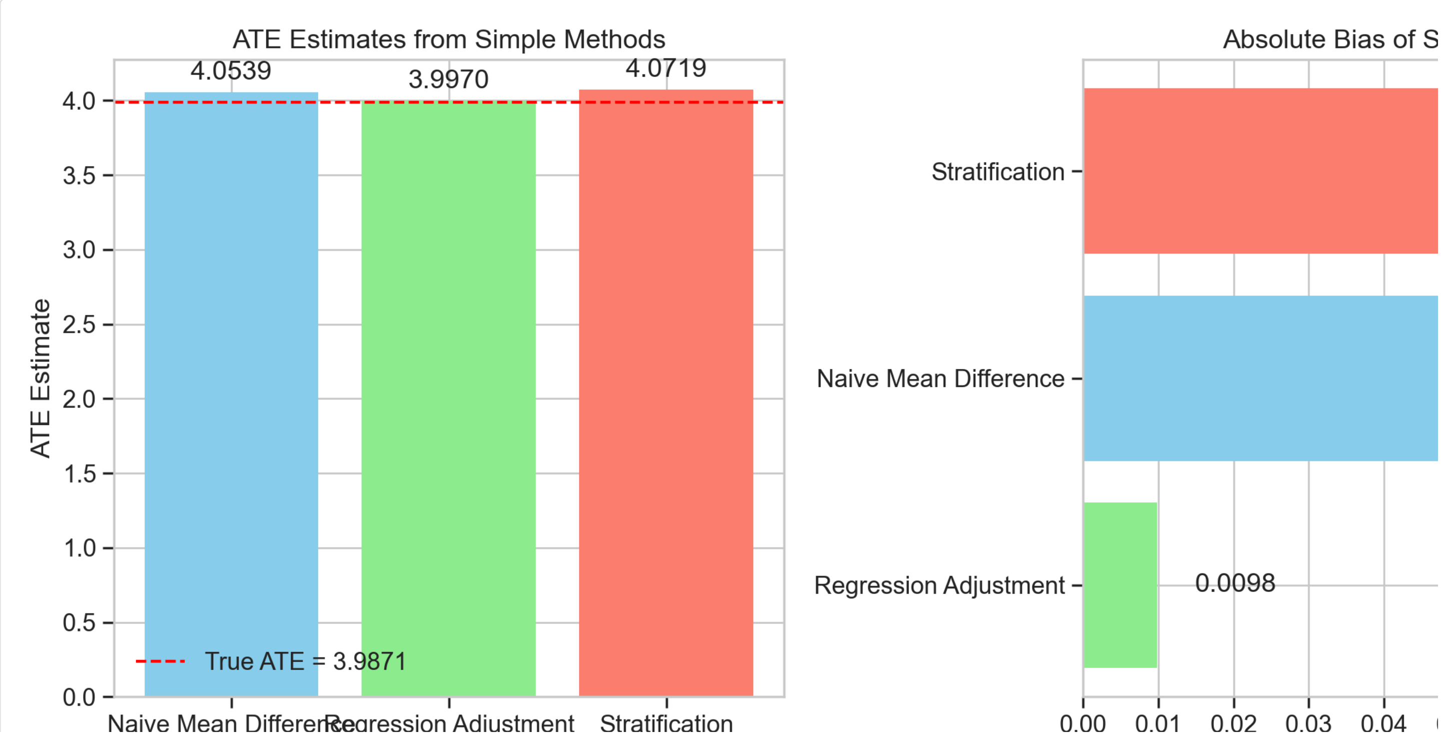
where $\bar{Y}_{s,1}$ is the average outcome for treated units in stratum s, $\bar{Y}_{s,0}$ is the average for control units, and w_s is the proportion of units in stratum s.

Key Assumption: Within each stratum, treatment is effectively randomly assigned.

Advantages: Intuitive, handles non-linear relationships, allows examination of effect heterogeneity.

Limitations: Can only stratify on a few variables before encountering sparsity issues.

Simple Methods Comparison

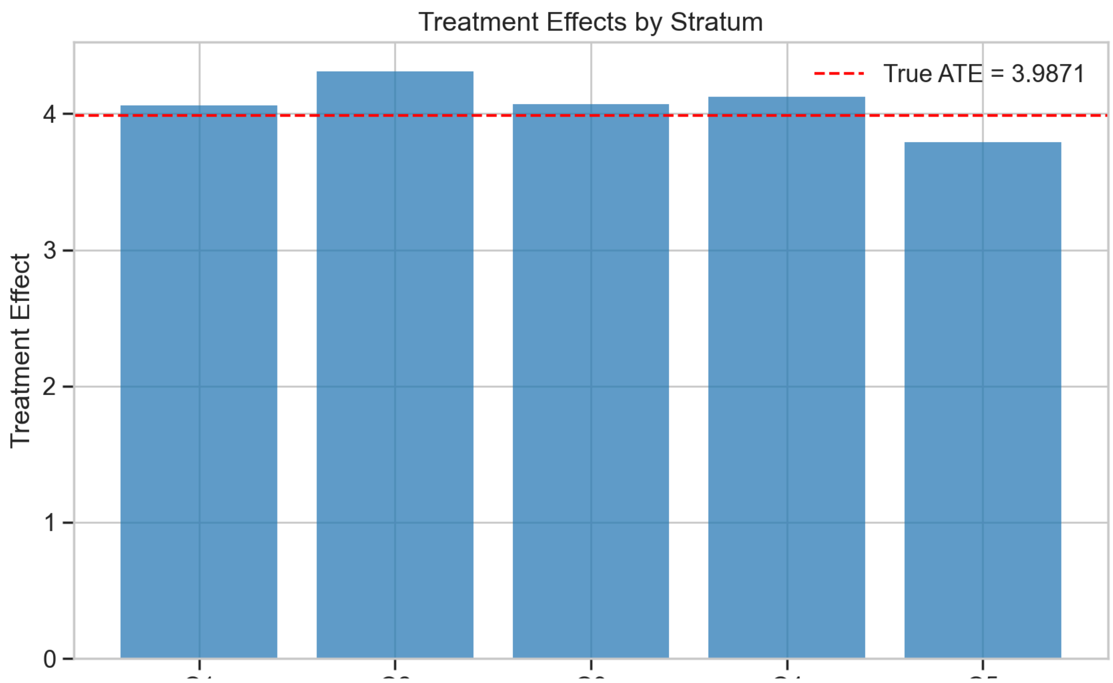


Analysis of Simple Methods:

1. **Naive Mean Difference**: The naive estimate has a bias of 0.0668, which is 0.0668 in absolute terms. This is relatively small, suggesting that selection bias in this data
2. **Regression Adjustment**: This method has a bias of 0.0098 (absolute: 0.0098), improving upon the naive estimator. This improvement in performance suggests that th
3. **Stratification**: This approach has a bias of 0.0847 (absolute: 0.0847), underperforming compared to the other methods. Stratification by birth weight reveals some he

Overall, the Regression Adjustment method performs best in terms of bias reduction for this dataset. However, all methods show relatively small bias, suggesting that the

Heterogeneous Effects by Birth Weight Stratum



Note: Numbers in parentheses show the weight of each stratum in the overall estimate.

6.2 Propensity Score Methods {#ps-methods}

> 🎯 ***Step 2**: Apply propensity score-based methods to adjust for confounding*

Building on the propensity scores we estimated earlier, we'll now implement methods that use these scores to create balance between treated and control groups.

Inverse Probability Weighting (IPW)

IPW creates a pseudo-population where the confounding influence is eliminated by weighting each observation by the inverse of its probability of receiving the treatmer

where $e(X_i)$ is the propensity score for unit i.

Key Advantages: - Uses all data points - Simple to implement - Intuitive connection to survey sampling

Limitations: - Sensitive to extreme propensity scores (near 0 or 1) - Can have high variance - Requires well-specified propensity model

Propensity Score Matching

Matching pairs treated units with control units that have similar propensity scores. The average difference in outcomes between matched pairs provides an estimate of tl

where $j(i)$ is the index of the control unit matched to treated unit i.

Key Advantages: - Intuitive and easy to explain - Can be combined with exact matching on key variables - Preserves the original outcome variable scale

Limitations: - Discards units that cannot be matched - Choice of matching algorithm and caliper can affect results - Matches may not be perfect, leaving some residual coı

Propensity Score Stratification

This method divides the data into strata based on propensity scores, estimates treatment effects within each stratum, and computes a weighted average:

where w_s is the proportion of units in stratum s, and $\bar{Y}_{s,1}$ and $\bar{Y}_{s,0}$ are the average outcomes for treated and control units in that stratum.

Key Advantages: - Uses all data points - Examines effect heterogeneity across propensity score strata - Usually reduces ~90% of confounding bias with just 5 strata

Limitations: - Less precise than matching for estimating average effects - Choice of strata boundaries can affect results - May not fully eliminate confounding within strat

Comparison of Propensity Score Methods

We've implemented and compared three propensity score-based causal inference methods:

1. **Inverse Probability Weighting (IPW)**: Weights observations inversely to their probability of receiving the treatment to create balance.
2. **Propensity Score Matching**: Pairs treated units with similar control units based on propensity scores.
3. **Propensity Score Stratification**: Divides the sample into strata based on propensity scores and calculates treatment effects within each stratum.

Each method has different strengths and weaknesses. The comparison below shows their performance in estimating the Average Treatment Effect (ATE).

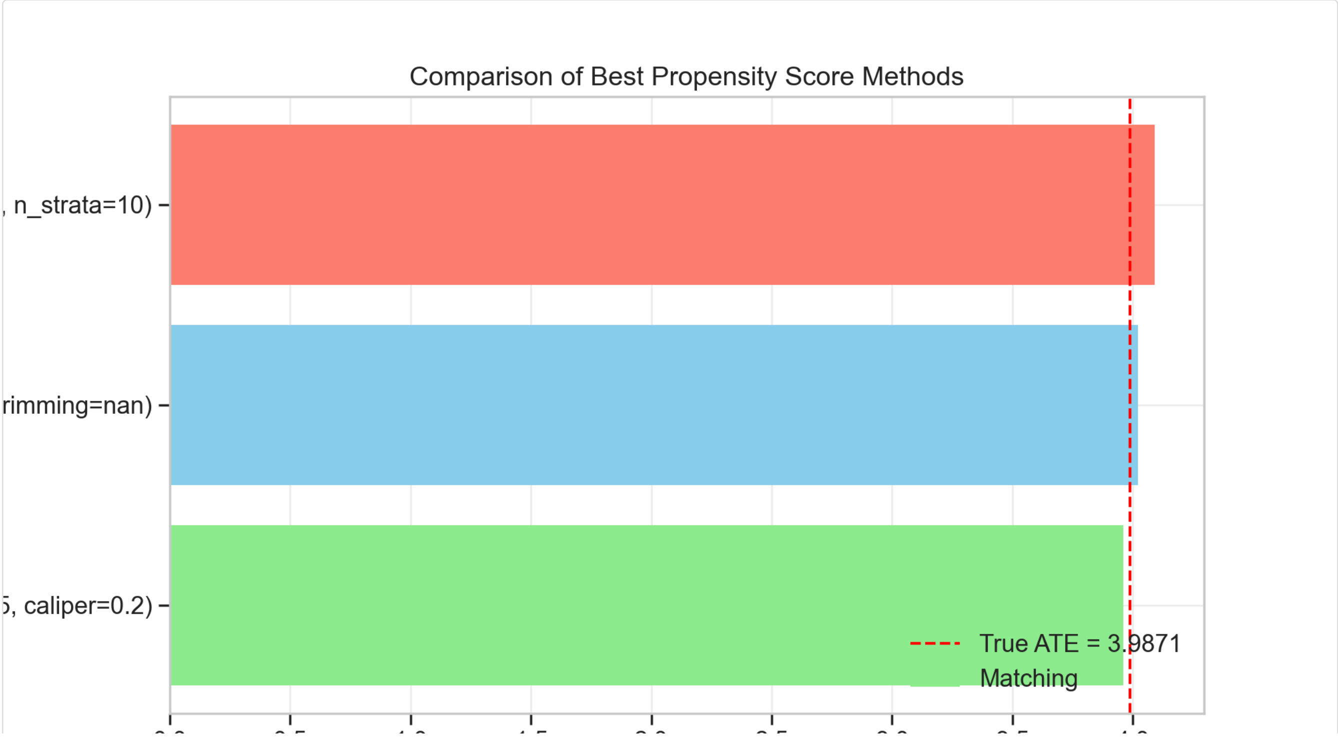
Best Method: Matching (Logistic, k=5, caliper=0.2)

- Estimated ATE: 3.9607
- True ATE: 3.9871
- Absolute Bias: 0.0265

This analysis shows that propensity score methods can effectively reduce bias in causal estimates from observational data.

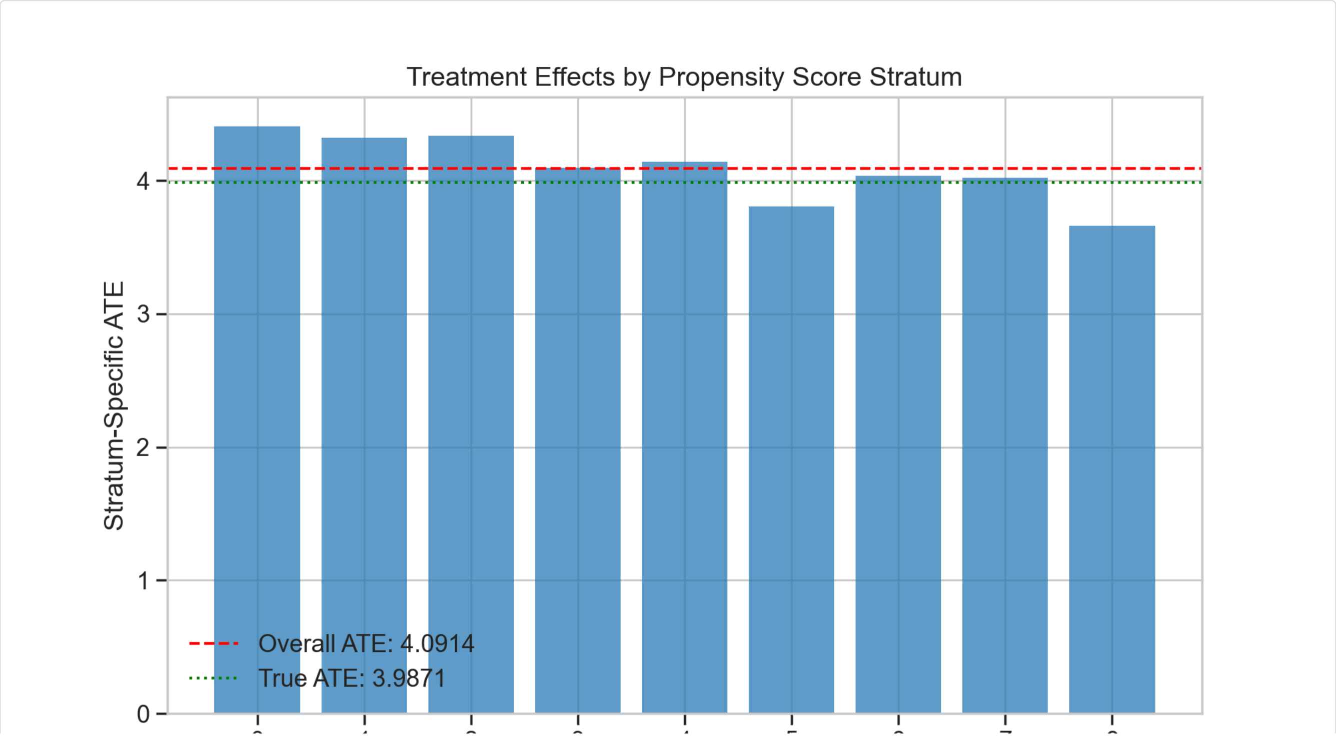
□	Method	ATE
	object	float64
□	Matching (Logistic, k=5, caliper=0.2)	3.9606867673
□	IPW (RF, stabilized=True, trimming=nan)	4.022955564
□	Stratification (Logistic, n_strata=10)	4.0914445887

Visual Comparison of Methods



Treatment Effects by Propensity Score Stratum

This plot shows how treatment effects vary across different propensity score strata. Heterogeneity in these effects may indicate effect modification by variables related to treatm



6.3 Advanced Machine Learning Methods {#ml-methods}

> 🚀 ***Step 3**: Leverage machine learning techniques for improved causal inference*

Finally, we'll explore advanced methods that combine machine learning with causal inference principles to estimate treatment effects more accurately. These methods can capt

6.3.1 Meta-Learners for Causal Inference {#meta-learners}

Meta-learners are a class of methods that use machine learning algorithms to estimate causal effects by combining multiple prediction models in different ways. Unlike traditio

Meta-learners use machine learning algorithms to estimate causal effects. Here are the three main types:

S-Learner (Single Model)

The S-Learner (Single model) uses a single machine learning model with the treatment indicator included as a regular feature:

1. **Train a model** to predict outcome using both covariates and treatment: [\hat{\mu}(x, t) = E[Y | X=x, T=t]]
2. **Estimate treatment effects** by taking the difference in predictions for treated vs. untreated: [\hat{\tau}(x) = \hat{\mu}(x, 1) - \hat{\mu}(x, 0)]

Advantages: Simple to implement, requires only one model

Limitations: May underestimate treatment effects if treatment assignment is highly imbalanced

T-Learner (Two Models)

The T-Learner (Two models) fits separate models for the treated and control groups:

1. **Train two separate models**: - Control model: [\hat{\mu}_0(x) = E[Y | X=x, T=0]] - Treatment model: [\hat{\mu}_1(x) = E[Y | X=x, T=1]]
2. **Estimate treatment effects** by taking the difference in predictions: [\hat{\tau}(x) = \hat{\mu}_1(x) - \hat{\mu}_0(x)]

Advantages: Can capture heterogeneous response surfaces, doesn't impose shared structure

Limitations: May suffer from high variance in regions with few samples from either group

X-Learner

The X-Learner extends the T-Learner with a more sophisticated approach:

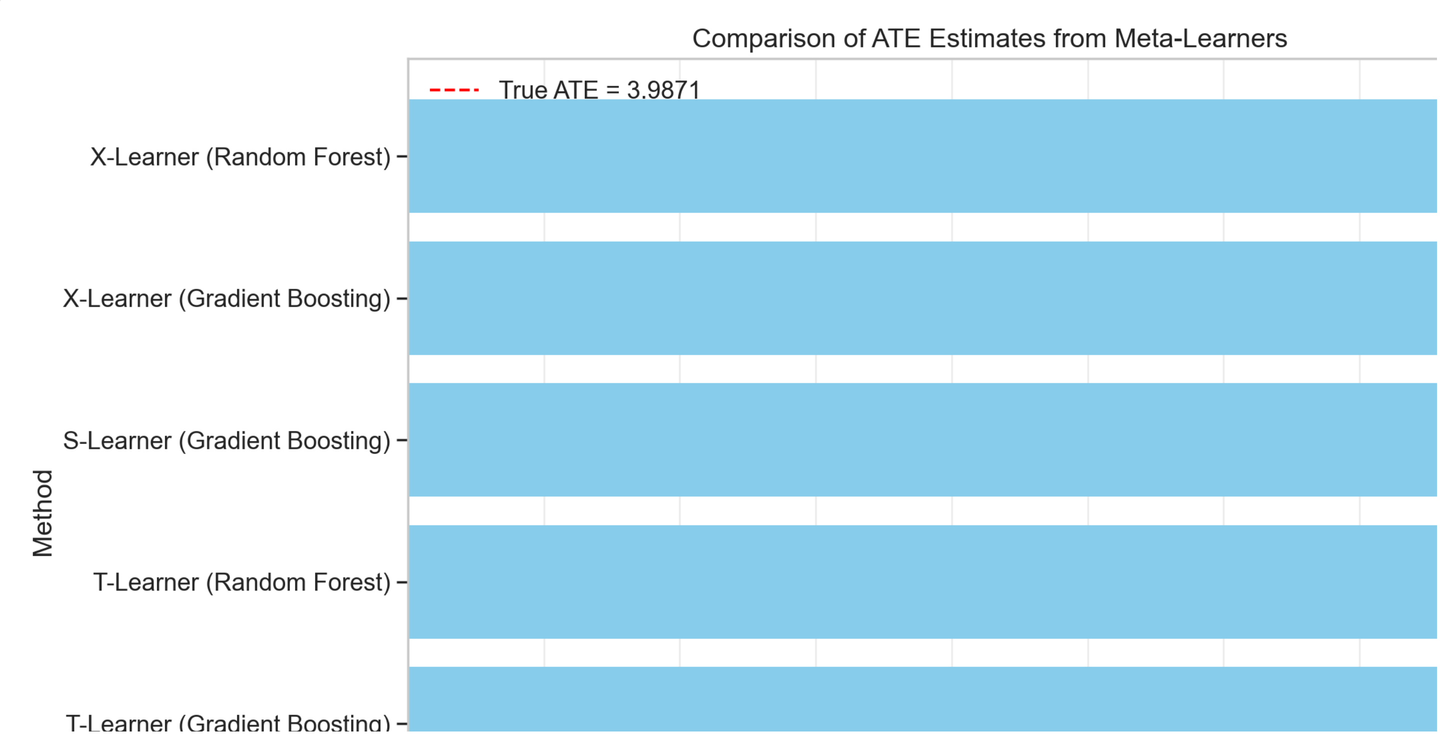
1. **Train response surface models** (same as T-Learner): - Control model: [\hat{\mu}_0(x) = E[Y | X=x, T=0]] - Treatment model: [\hat{\mu}_1(x) = E[Y | X=x, T=1]]
2. **Impute individual treatment effects** for each unit: - For treated units: [D_i^1 = Y_i(1) - \hat{\mu}_0(X_i)] - For control units: [D_i^0 = \hat{\mu}_1(X_i) - Y_i(0)]
3. **Train two treatment effect models**: - Using treated units: [\hat{\tau}_1(x) = E[D_i^1 | X_i=x]] - Using control units: [\hat{\tau}_0(x) = E[D_i^0 | X_i=x]]
4. **Combine the two estimates** using a weighting function g(x): [\hat{\tau}(x) = g(x)\hat{\tau}_0(x) + (1-g(x))\hat{\tau}_1(x)] where g(x) can be the propensity score.

Advantages: Performs well with heterogeneous treatment effects and imbalanced treatment groups

Limitations: More complex, requires estimating propensity scores

Meta-Learner Results

□		Method object
□	**0**	S-Learner (Random Forest)
□	**3**	T-Learner (Gradient Boosting)
□	**2**	T-Learner (Random Forest)
□	**1**	S-Learner (Gradient Boosting)
□	**5**	X-Learner (Gradient Boosting)
□	**4**	X-Learner (Random Forest)

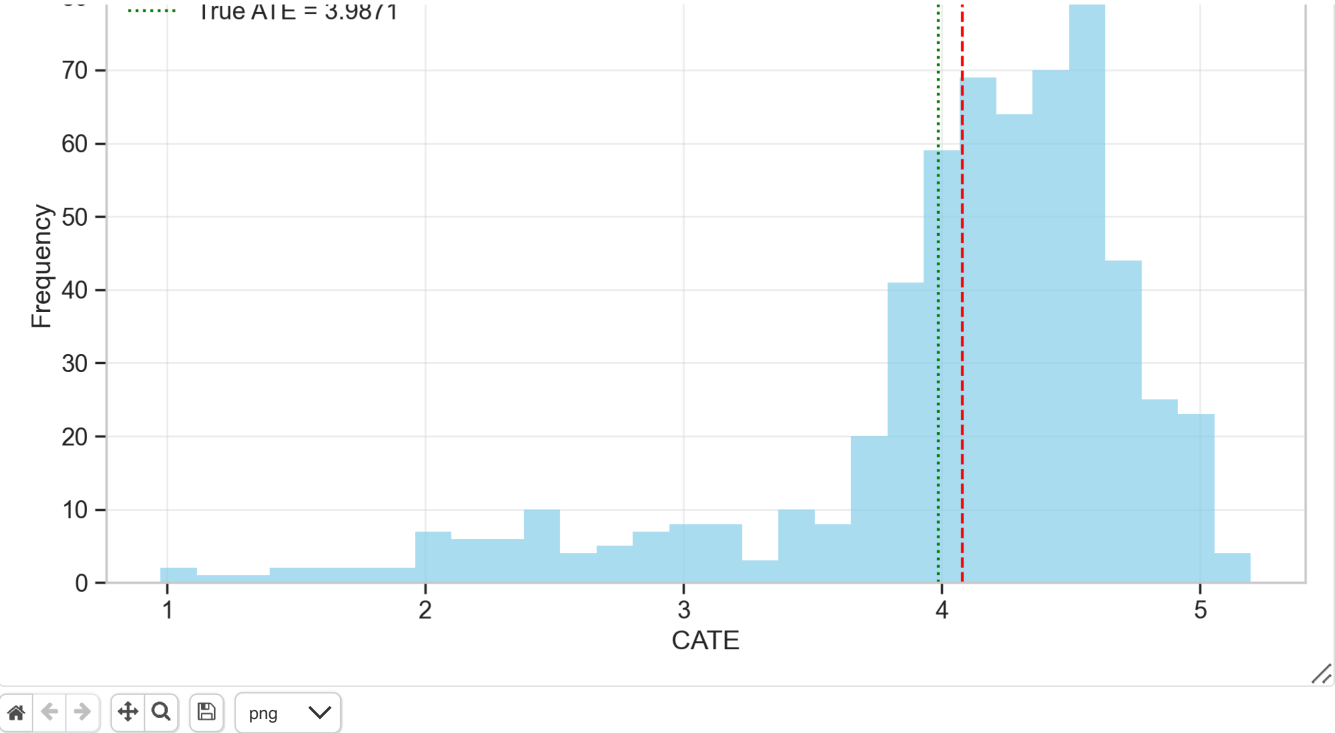


Best Meta-Learner Method:

- **Method**: S-Learner (Random Forest)
- **ATE Estimate**: 3.9831
- **True ATE**: 3.9871
- **Bias**: -0.0040

This analysis shows that meta-learners can provide accurate estimates of causal effects by leveraging machine learning algorithms.

Treatment Effect Heterogeneity



Treatment Effect Heterogeneity: The distribution above shows how treatment effects vary across different individuals. This variation suggests that the intervention may be more

6.3.2 Doubly Robust Methods {#doubly-robust}

Doubly robust methods combine outcome modeling and propensity score approaches to provide protection against misspecification of either model. This "double robustness" pr

Doubly robust methods offer protection against model misspecification by combining outcome modeling and propensity score approaches:

Augmented Inverse Probability Weighting (AIPW)

AIPW combines outcome regression and IPW by using both models to create a doubly robust estimator:

The AIPW estimator can be written as:

where: - $\hat{\mu}_1(X_i)$ and $\hat{\mu}_0(X_i)$ are outcome models for treated and control - $\hat{e}(X_i)$ is the propensity score model

Key property: Consistent if *either* the outcome model *or* the propensity score model is correctly specified (not necessarily both)

Advantages: More robust to model misspecification, often lower variance than IPW

Double Machine Learning (DML)

DML addresses issues of regularization bias and overfitting in high-dimensional settings:

1. **Cross-fitting approach**: - Split the data into K folds - For each fold, fit models on the other K-1 folds and predict on the held-out fold - This reduces the impact of over
2. **Orthogonalization**: - Remove the dependence between treatment and covariates - Remove the dependence between outcome and covariates - Study the residual rela

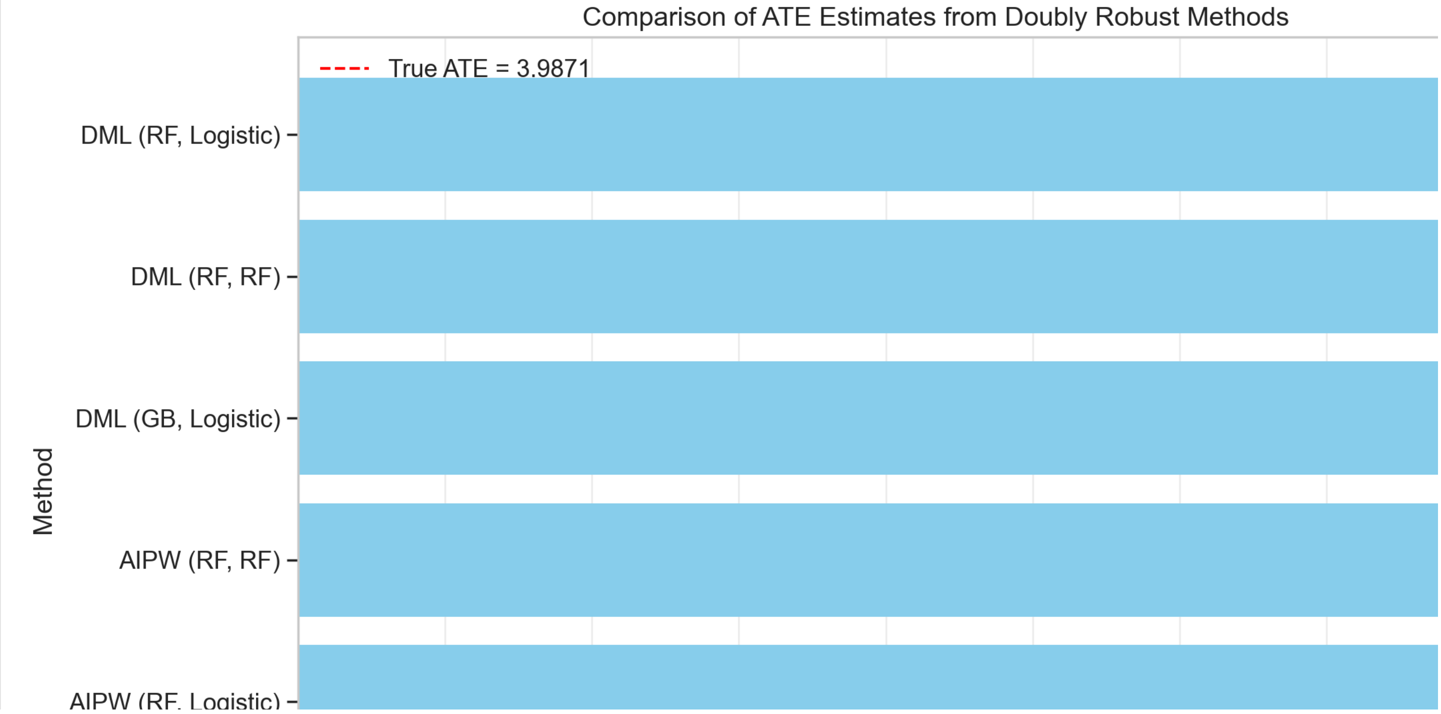
DML can be implemented as:

where $\hat{m}(X_i)$ is a model for the outcome and $\hat{e}(X_i)$ is the propensity score model.

Advantages: Handles high-dimensional settings well, allows complex ML models, reduces regularization bias

Doubly Robust Method Results

□		Method object	ATE float64
□	**1**	AIPW (GB, Logistic)	3.9909407575
□	**0**	AIPW (RF, Logistic)	3.9979688463
□	**2**	AIPW (RF, RF)	3.9707660095
□	**4**	DML (GB, Logistic)	4.1009270934
□	**5**	DML (RF, RF)	4.1291638999
□	**3**	DML (RF, Logistic)	4.1323617565



Best Doubly Robust Method:

- **Method**: AIPW (GB, Logistic)
- **ATE Estimate**: 3.9909
- **True ATE**: 3.9871
- **Bias**: 0.0038

Doubly robust methods provide protection against model misspecification, making them more robust for causal inference in complex settings.

AIPW vs DML Comparison:

- **AIPW** directly augments IPW with an outcome model correction term, providing a straightforward implementation of double robustness
- **DML** uses cross-fitting to address regularization bias, making it particularly well-suited for high-dimensional settings

Both methods leverage the strengths of outcome modeling and propensity score approaches, providing more reliable causal estimates than either approach alone.

6.3.3 Causal Forests {#causal-forests}

Causal forests are a powerful extension of random forests specifically designed to estimate heterogeneous treatment effects. They are particularly useful for understanding how

Causal Forests

Causal forests extend random forests to directly estimate heterogeneous treatment effects. The key ideas include:

1. **Honest trees**: Split the sample into a training set (used to determine splits) and an estimation set (used to estimate treatment effects within leaves)
2. **Orthogonalization**: Remove the effect of confounders by residualizing the outcome and treatment
3. **Adaptive sample splitting**: Focus on regions with high treatment effect heterogeneity

The algorithm builds many trees and averages the results to obtain the Conditional Average Treatment Effect (CATE) function:

Key advantages: - Directly targets heterogeneous treatment effects - Provides measures of variable importance for effect modification - Performs well with high-dimensio

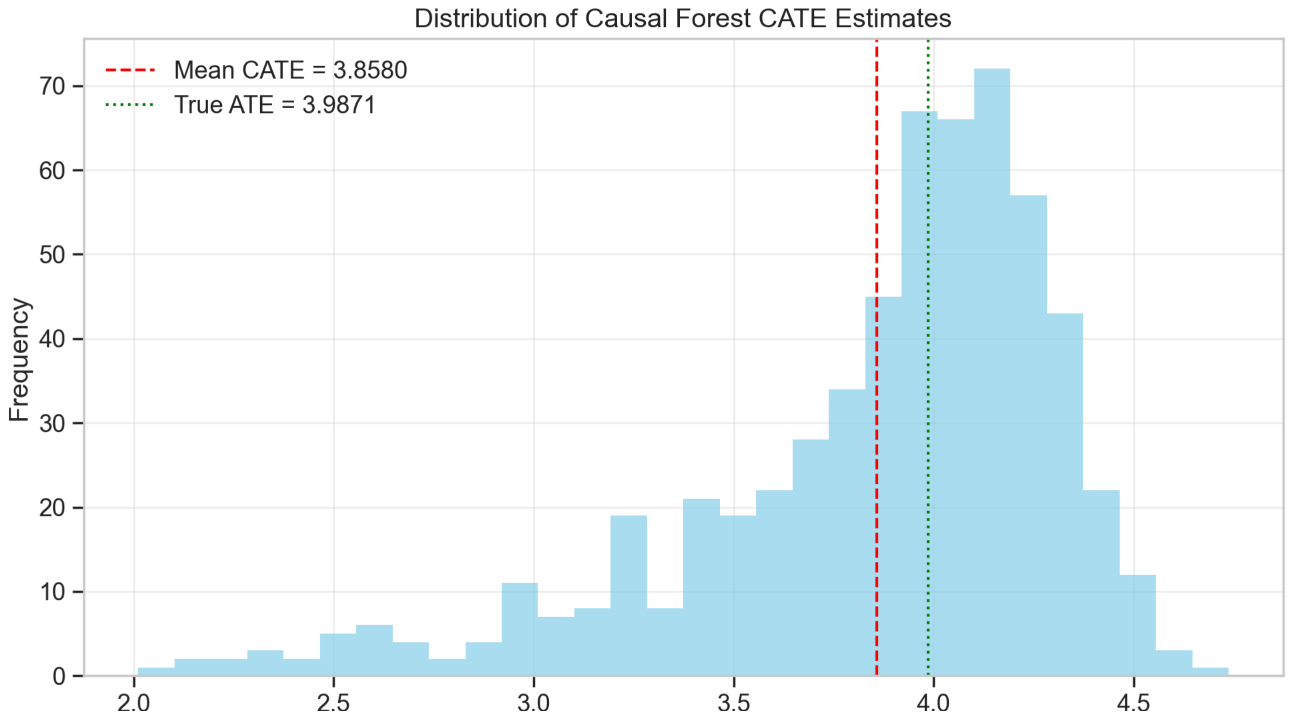
Implementation: Causal forests are available in packages like `econml` (Python) and `grf` (R)

Causal Forest Results

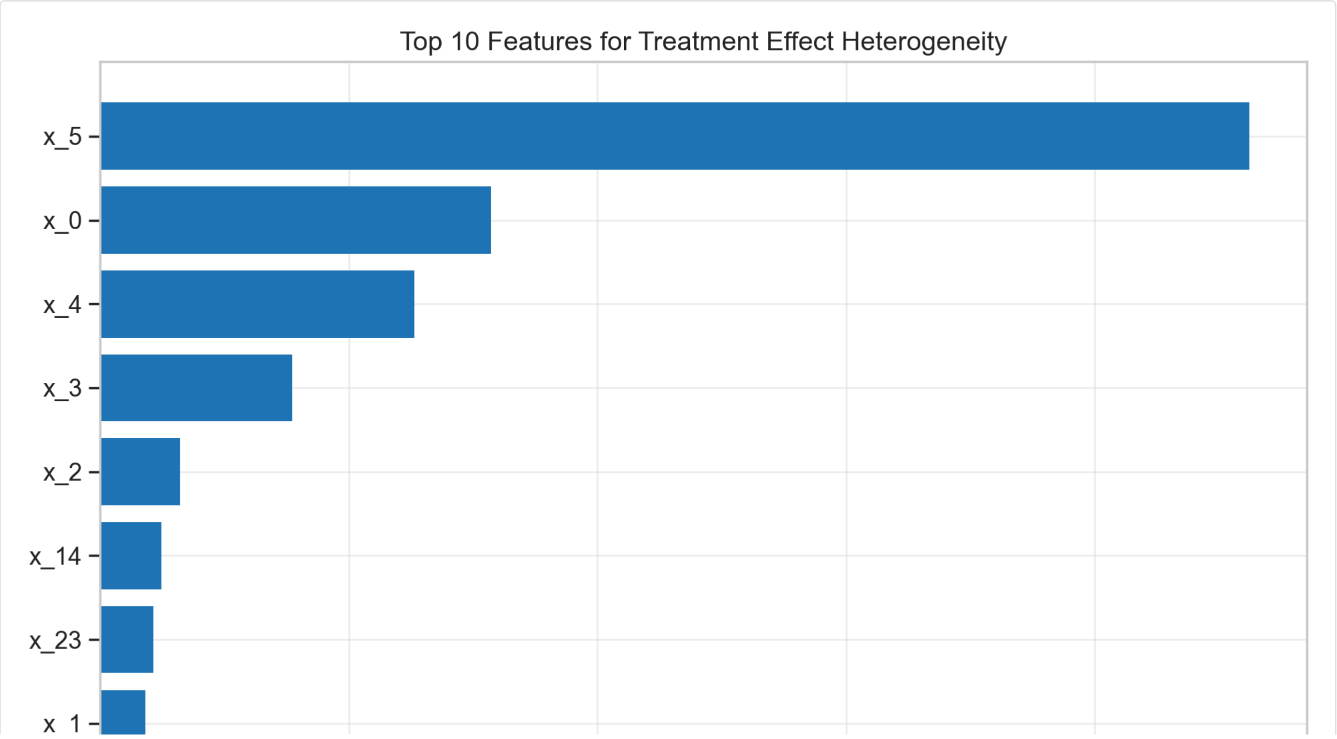
Causal Forest Results:

- **Estimated ATE**: 3.8580
- **True ATE**: 3.9871
- **Bias**: -0.1291
- **Absolute Bias**: 0.1291

Treatment Effect Heterogeneity



Feature Importance for Effect Modification

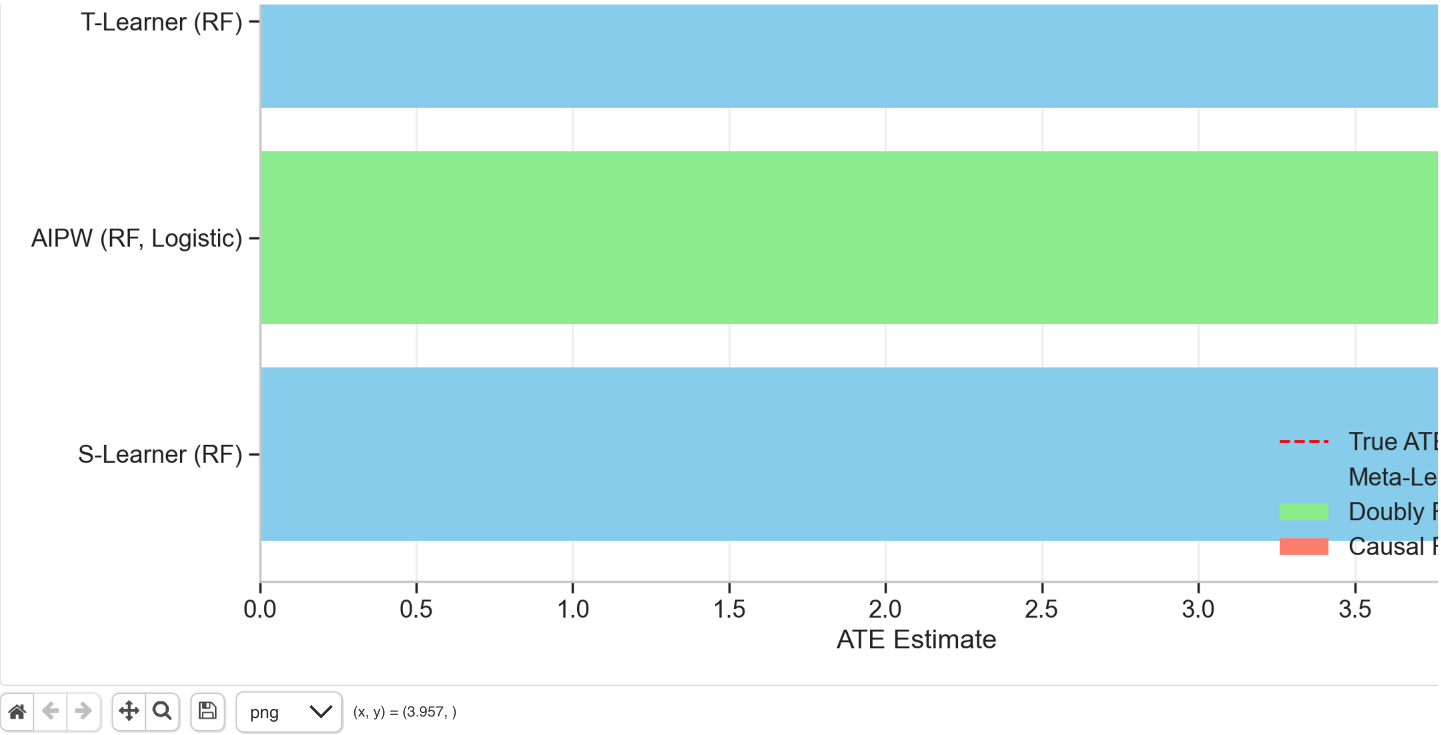


The feature importance plot shows which variables contribute most to treatment effect heterogeneity. These are the features that most strongly modify the effect of treatment,

6.3.4 Comparing All Advanced Methods {#compare-advanced}

Now let's compare all the advanced machine learning methods we've implemented to see which ones perform best in estimating causal effects in our dataset.

Comparison of Advanced Machine Learning Methods



Best Advanced Method: S-Learner (RF)

- **ATE Estimate**: 3.9831
- **True ATE**: 3.9871
- **Bias**: -0.0040

Advanced machine learning methods can significantly improve causal estimates by capturing complex relationships between variables without requiring strong paramet

Analysis of Advanced Methods:

1. **Meta-Learners** leverage flexible machine learning algorithms to model outcomes or treatment effects. They work well when relationships are complex but depend heavily o
2. **Doubly Robust Methods** combine outcome modeling and propensity score approaches, providing protection against model misspecification. They tend to have lower bias ar
3. **Causal Forests** excel at capturing treatment effect heterogeneity and provide valuable insights through feature importance. They're particularly useful for understanding wh

The best method depends on the specific context, data structure, and research question. In practice, it's valuable to implement multiple methods and compare their results, as w

□		Method object	ATE float64
□	**0**	S-Learner (RF)	3.9831111337
□	**2**	AIPW (RF, Logistic)	3.9978829665
□	**1**	T-Learner (RF)	3.9581051078
□	**3**	Causal Forest	3.8580102349