Causal Inference in Healthcare (PS-1)

Team Name: MangoDB

Team Members:
- Aryan Bhagat
- Chetan Borkar

1. Project Objective

The goal of this project is to estimate the effect of insulin treatment on the probability of developing diabetes using causal inference methods, specifically for a dataset from the Indian KIMA study.

2. Dataset Overview

The dataset contains demographic and medical information of patients, with a focus on insulin treatment and diabetes diagnosis rates. It provides variables such as:

- **Age**
- **Insulin use**
- **Diagnosis outcome** (whether the patient developed diabetes)
- **Other covariates** (e.g., BMI, blood pressure, gender)

3. Causal Inference Methods

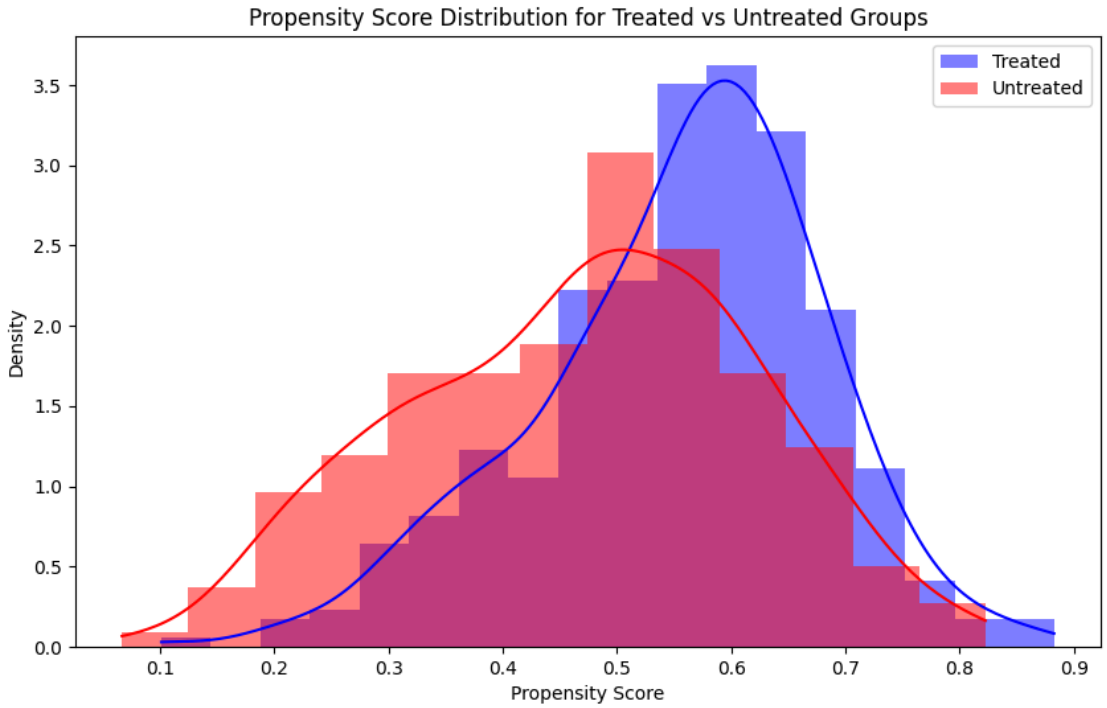
3.1 Propensity Score Matching (PSM)

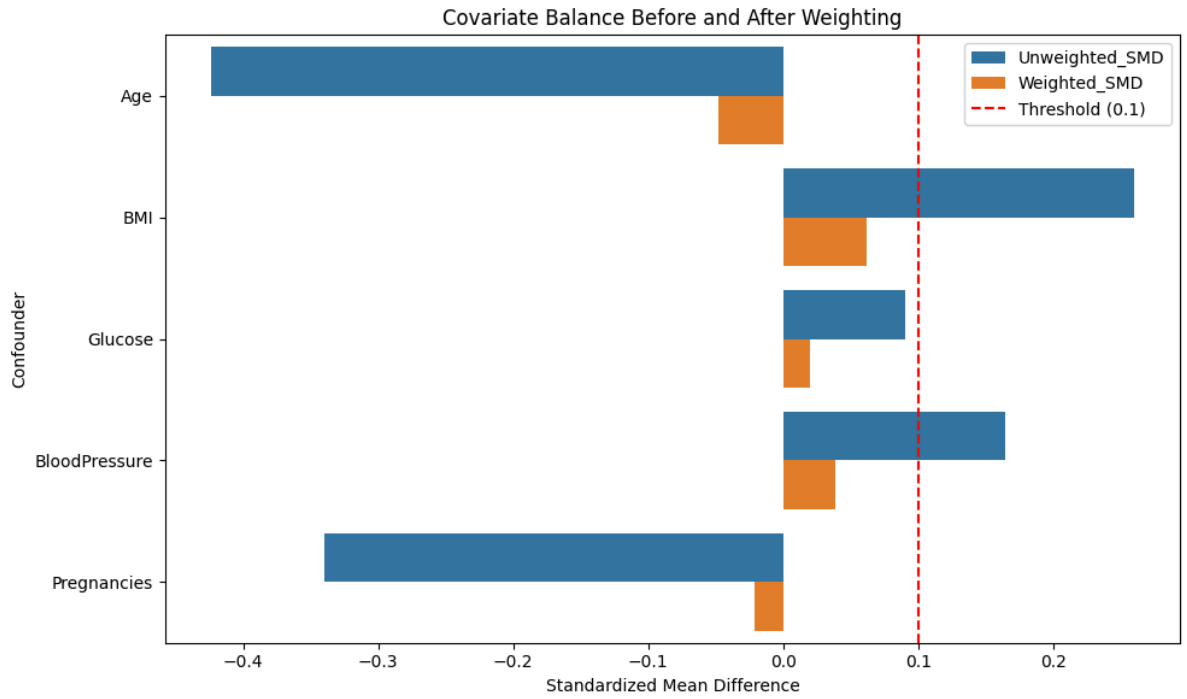
Steps Implemented:

1. Estimated propensity scores using logistic regression to predict the likelihood of receiving insulin based on covariates (age, gender, BMI, blood pressure).
2. Matched treated and untreated individuals with similar propensity scores, balancing the covariates across groups.
3. Used **Love plots** to check for balance between the treated and untreated groups after matching.

Key Findings:

- **Significant overlap in the propensity score distributions between treated and untreated groups, indicating that matching worked well.**





3.2 Inverse Probability Weighting (IPW)

Steps Implemented:

1. Calculated inverse weights for each individual based on their propensity scores.
2. Implemented weight trimming at the 95th percentile to avoid overly influential extreme weights.

Key Findings:

- **ATE (Average Treatment Effect) = -0.0307**: Suggests that insulin treatment is associated with a **3.07% reduction** in the likelihood of developing diabetes.
- **Doubly-Robust ATE (AIPW) = -0.0382**: A more reliable estimate combining IPW and outcome modeling.

3.3 Conditional Average Treatment Effect (CATE)

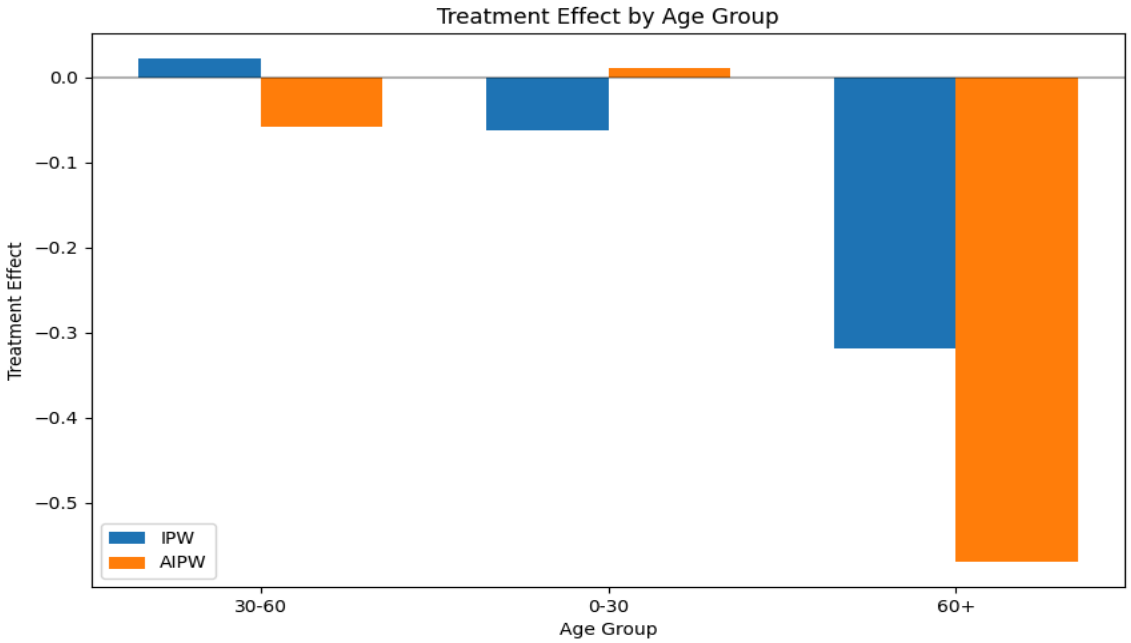
The **CATE** estimates the treatment effect for specific subgroups:

- **Age Group 30-60:** CATE = **0.0222** (slight positive treatment effect)
- **Age Group 0-30:** CATE = **-0.0625** (negative treatment effect)
- **Age Group 60+:** CATE = **-0.3192** (strong negative treatment effect in older individuals)

4. Augmented Inverse Probability Weighting (AIPW)

AIPW is a **doubly-robust estimator** that combines both the propensity score model (IPW) and an outcome model to estimate treatment effects.

- **Key Advantage:** AIPW provides robust estimates even if one of the models (propensity score or outcome model) is misspecified.
- **Key Result:** The doubly-robust **ATE using AIPW was -0.0382**, providing a more reliable estimate by correcting for potential model misspecification.



5. Summary and Clinical Implications

Summary

- The analysis suggests a **small but statistically significant negative treatment effect** of insulin on the likelihood of developing diabetes.
- The effect varies significantly across age groups:
    - **Older patients (60+) show a much stronger negative treatment effect.**
    - **Younger patients (0-30) experience a slight negative effect.**

Clinical Implications

- This finding suggests that **insulin treatment might be less effective in older individuals**.
- Further investigation is needed to understand the underlying causes of this variability.

Future Work

- Include **more covariates** such as lifestyle factors.
- Explore **potential interactions** between insulin and other treatments.
- Implement **advanced causal inference techniques** for better accuracy.