

✓ CMU 10-742 (Fall 2024) - Machine Learning in Healthcare

Assignment 4: Causal Modeling and Bayesian Trials

Out: Tues Oct 8 2024

Due: Tues Oct 22 2024

This assignment counts for 5 points out of the 35 total points allocated to the course problem sets.

This assignment covers material that you have not yet seen in lecture. This is by design. The lectures on causal modeling and Bayesian trials occur late in the semester, when you will be focused on your final projects. Working through this assignment now will give you a great background for these lectures. To succeed in this assignment, you may have to look up and learn various concepts, including:

- randomized control trials
- Bayesian trials (esp. multi-arm bandit trials)
- average treatment effect (ATE) and conditional average treatment effect (CATE)
- confounders
- propensity models

✓ Part 1: Designing a Bayesian Clinical Trial (1 point)

It is Feb 2025 and there's an outbreak of a new respiratory infectious disease, Covid-25. You are a clinical researcher and pulmonologist - and the CDC has tasked you with finding a reliable treatment as quickly as possible.

You are evaluating four candidate treatments:

Treatment	Description
A	2 tbsp Paxlovid, taken orally 1x day
B	2 tsp metformin taken orally 1x/day
C	1 tbsp Hydroxychloriquine, taken orally 2x/day
D	1 hour of gentle aerobic exercise (e.g. a walk) every day

You can measure the effect of each treatment on a patient using a Covid-detector, which accurately reports the change in disease state in 24 hours within the range 0 (no improvement) to 100

(complete improvement). (We assume, for simplicity, that none of the treatments can have a negative effect.)

At your health system, you have access to a population of 1000 new patients every day who are newly infected with Covid-25.

✓ 1.1

Briefly describe (in no more than a short paragraph) how you would design a double-blind randomized control trial (RCT) to assess the efficacy of treatments 1-4.

YOUR ANSWER HERE

✓ 1.2

Succinctly describe how you would implement a Bayesian multi-armed bandit (MAB) Bayesian trial using Thompson sampling to assess these treatments.

YOUR ANSWER HERE

✓ 1.3

Compare the RCT and MAB approach. What are the advantages and disadvantages of each?

YOUR ANSWER HERE

✓ Part 2: Causal Modeling (4 points)

Conratulations! You have just been appointed chief data scientist for Pied Piper Inc. (PPI).

PPI has been growing and just decided last year to start becoming self-insured, meaning that they will directly pay the healthcare costs for their employees, instead of relying on a commercial insurer like Aetna or Cigna. Not coincidentally, PPI's board is now looking for ways to reduce their overall spend on employee healthcare.

To that end, PPI just introduced a new program called Tofu For You ("Tofu") to all PPI employees. Employees who enroll in Tofu get daily video reminders on healthy eating. PPI hopes that Tofu will

reduce the overall healthcare costs for enrolled employees.

collecting (most) of the imports for this assignment in one place. You may not end
you may need others not listed here.

```
from google.colab import auth
import pandas as pd
import numpy as np
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, roc_auc_score, roc_curve, confusion_matr
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.feature_extraction.text import CountVectorizer
```

```
auth.authenticate_user()
```

```
# load data
!gsutil cp gs://10-742/assignment_4/tofu.csv ./
df = pd.read_csv('tofu.csv')
```

✓ 2.1

Report the mean and standard deviation for the next 12 months of healthcare spend.

YOUR ANSWER HERE

✓ 2.2

We'd like to figure out whether and how much Tofu affects healthcare spend. Using statistics terms, we're looking for the the Average Treatment Effect (ATE) of the Tofu program on healthcare spending.

For our first attempt to calculate ATE, we'll do the following:

- Split the data into two groups, one enrolled in Tofu (treatment group) and another group not enrolled (control group).
- Calculate the mean difference between `SpendLast12M` and `SpendNext12M` for each group.
- Calculate the ATE by subtracting the mean of the control group from the mean of the treatment group.

Based on this calculation, does the Tofu program increase or decrease future healthcare spend?

Can you think of any shortcomings with this approach to calculating ATE?

YOUR ANSWER HERE

✓ 2.3

A **confounder** is a feature (or "covariate") that affects whether the treatment is selected or not, and also affects the outcome. In this scenario, a confounder would affect (1) whether the person elects to enroll in Tofu or not, and (2) future healthcare spend.

Provide two examples of potential confounders in this trial.

YOUR ANSWER HERE

✓ 2.4

You might now ask whether gender has any affect on the ATE. To answer this question, we can calculate the *conditional average treatment effect* (CATE), conditioned on gender.

More specifically: calculate and report the ATE (as above) for the male and female cohort separately.

For which cohort (male or female) does Tofu appear to have a larger impact on future healthcare spend? (We are not asking you to test for statistical significance here.)

YOUR ANSWER HERE

✓ 2.5

Build a model, using logistic regression, to predict if a person will enroll in Tofu. You can use all features except `SpendNext12M`.

This is called a 'propensity model', because it models the *propensity* that an individual will self-select into a group (i.e. the group of Tofu participants).

Reserve 20% of the dataset for testing, and report your accuracy on that dataset. For comparison, we got `Accuracy=0.83`.

YOUR ANSWER HERE

✓ 2.6

Select 2000 employees who actually enrolled in Tofu. For each one, calculate their propensity score, and also find one employee who did not enroll, but who has a very similar propensity score. For this matched cohort, estimate the ATE.

Run this experiment ten times. Report the ATE for each experiment, and report the average ATE across all 10 trials.

How does this average ATE differ from the previously-calculated ATE?

This approach is called "Propensity Score Matching." While's not as statistically robust as a randomized control trial, this approach does attempt to estimate the effect of a treatment while accounting for the covariates.

YOUR ANSWER HERE

✓ 2.7

Let's fast forward three years. After the promising initial data, you convinced Pied Piper's board of directors to run a randomized control trial (RCT) to assess Tofu's impact. Pied Piper selected 10,000 employees and randomly assigned each to the treatment or control group.

The data for this RCT is in the `tofu_rct.csv` file.

What is the ATE for Tofu according to this trial? How close was your coarse estimate of ATE (from 2.2) and your propensity score matching estimate of this true ATE (from 2.6)?

```
!gsutil cp gs://10-742/assignment_4/tofu_rct.csv ./
df_rct = pd.read_csv('tofu_rct.csv')
```

YOUR ANSWER HERE

✓ 2.8

Are there any ethical considerations in performing an RCT here? How might you address those concerns?

YOUR ANSWER HERE

✓ 2.9

In general, an RCT is the best approach to removing confounding factors. Is there any opportunity for a confounder to remain in this RCT? If so, what would be an example of such a confounder?

YOUR ANSWER HERE