Please collect the answers to the questions below (including any tables and figures) in a `.pdf` and submit it together with the `.do` files that generated any tables and figures on Athena under the corresponding folder. If you write with pen/pencil, make sure to have clear photographs for submission. The deadline is 15th of April, 16:00.
Please remember to lay out your results as clearly as possible, and to comment on your code in a way that makes it easily accessible to others.

Question 1. (Exam 2023, extended with coding example: The CATT).
Individuals indexed by $i = 1, ..., N$ come from three types $X_i \in \{1, 2, 3\}$ with $0 < \Pr(X_i = x) < 1$ for $x \in \{1, 2, 3\}$ and $\sum_{i=1}^{N} \Pr(X_i = x) = 1$. Individuals are subject to a randomized treatment $D_i \in \{0, 1\}$ with corresponding potential outcomes defined as $Y_i(d)$ for $d \in \{0, 1\}$. Let $e(X_i)$ be the propensity score of an individual of type X_i and assume that $e(X_i) = X_i/3$.

1. Define the individual treatment effect of i according to this setup. Describe how it relates to the fundamental problem of causal inference. Under what assumptions is the individual causal effect well-defined? *Hint: Note the question asks when the individual causal effect is well-defined,* ***not*** *whether it is identified.*

2. Define the average treatment effect on the treated (ATT), which we denote by τ. Under what assumption(s) is it identified? Why?

3. Define conditional ATTs $\tau(x)$ for each type $x \in \{1, 2, 3\}$. Under what assumptions are they identified? Why?

4. How would you estimate $\tau(x)$? Write down an expression based on observed data $(Y_i, D_i.X_i)$ for $i = 1, ..., N$. For any $\tau(x)$ that is identified, estimate the CATT by hand using the following data:

i	Y_i	D_i	X_i
1	5	0	1
2	3	0	1
3	2	1	1
4	5	0	2
5	4	1	2
6	3	1	3
7	2	1	3

5. Simulate $N = 100$ observations with $\Pr(X_i = 1) = 1/2$ and $\Pr(X_i = 2) = \Pr(X_i = 3) = 1/4$. Let

$$Y_i(0) \sim U[0, 1]$$
$$Y_i(1) \sim X_i \times U[0, 1],$$

both i.i.d.

6. Generate individual causal effects τ_i. Make a single figure with three kernel density plots overlaid, one for each type of X_i of τ_i.

7. Estimate the ATE, the ATT, and the CATT for each type of X_i (if they are identified). Report these values in a table, in which you can compare them to their true value.

Question 2. (Treatment Effects by Gender).
Consider the following model:

$$Y_i = Y_i(0) + D_i\tau_i$$

where Y_i is the observed outcome, $D_i \in \{0,1\}$ is a binary treatment, $Y_i(d)$ for $d \in \{0,1\}$ are potential outcomes, and $\tau_i = Y_i(1) - Y_i(0)$ is the individual treatment effect. Assume that $X_i \in \{0,1\}$ (e.g. gender with $X_i = 1$ meaning female) affects treatment effects in the following way:

$$\tau_i = \alpha_0 + \alpha_1 X_i.$$

a. Show that, without further assumptions, the average treatment effect on the treated (ATT) $\tau \equiv \mathbb{E}[\tau_i | D_i = 1]$ is **not** (point-)identified.

b. Now assume that $\mathbb{E}[Y_i(0)|D_i = 0, X_i = x] = \mathbb{E}[Y_i(0)|D_i = 1, X_i = x]$. Show that τ is identified. Why is τ now identified? If $\Pr(X_i = 1) = 0.3$, $\alpha_0 = 1$ and $\alpha_1 = -2$, what is the true value of τ?

c. Show that, under the previous assumption and the assumption that

$$0 < \Pr(D_i = 1|X_i = x) < 1,$$

for $x \in \{0,1\}$, the conditional ATT (CATT) $\tau(X_i) \equiv \mathbb{E}[\tau_i|D_i = 1, X_i]$ is identified. Why is the assumption $0 < \Pr(D_i = 1|X_i = 1) < 1$ necessary for identification? What are the true values for $\tau(x)$, $x = 0, 1$?

d. Imagine we run an experiment with $N = 5$ individuals, with $X_i = 0$ for $i \leq 3$ and $X_i = 1$ for $i \geq 4$. We want $n_1 = 3$ treated individuals and $n_0 \equiv N - n_1 = 2$ controls. How many assignments are there for this setup?

e. Continuing part (d.), imagine we want to ensure that there are both women and men in the sample, so we restrict ourselves to assignments that have exactly two men and one woman (i.e. exactly two individuals with $X_i = 0$ and one individual with $X_i = 1$). How many assignments are left? What are the unit assignment probabilities of each of the N individuals? What are the propensity scores $e(X_i)$?

f. Generate a random sample of 200 individuals with $(Y_i(0), X_i) \sim \mathcal{N}(0,1) \times \text{Bernoulli}(p)$, $p = 0.3$, $\alpha_0 = 1$ and $\alpha_1 = -2$. You are running an experiment (i.e. D_i is randomly assigned, independently of other random variables in the model) with $\Pr(D_i = 1) = 0.5$, which we call Experiment 1. Estimate

$$\begin{aligned}\widehat{\tau} &= \widehat{\mathbb{E}}[Y_i|D_i = 1] - \widehat{\mathbb{E}}[Y_i|D_i = 0] \\ &= n_1^{-1} \sum_{i:D_i=1} Y_i - n_0^{-1} \sum_{i:D_i=0} Y_i\end{aligned}$$

simply by averaging ("collapsing") your data. Then compare this to the OLS estimate $\widehat{\tau}_{\text{OLS}}$. How do they compare?

g. Estimate the two CATTs

$$\widehat{\tau}(x) = n_1^{-1}(x) \sum_{i:D_i=1} 1[X_i = x] Y_i - n_0^{-1}(x) \sum_{i:D_i=0} 1[X_i = x] Y_i$$

for $x = 0, 1$ where $n_d(x)$ is the number of individuals with treatment status d and gender x, again by collapsing as well as via OLS.

h. Now imagine we want to run an experiment in which the assignment mechanism is such that $\Pr(D_i = 1)$ satisfies the fixed-number-of-women requirement from part (e.), keeping all other assumptions about the sample as in part (f.). Call this Experiment 2. To compare estimates from the two experiments, let's draw 500 datasets with $p = 0.3$ and $N = 100$ for each of the two experiments. In each of the datasets, produce estimates $\widehat{\tau}$ and $\widehat{\tau}(x)$ for $x = 0, 1$. Use a Kernel density estimator to plot the distributions of these three estimators. Specifically, make two plots: the one for $\widehat{\tau}$ has two density estimates, one for each experiment; the other plot has four density estimates, $\widehat{\tau}(0)$ and $\widehat{\tau}(1)$ for each experiment. Plot the true values of τ and $\tau(x)$ as well. How do the estimates differ across experiments?

Question 3. (Randomization Inference & Bootstrapping of the Sample ATE).
Consider data from an experiment with N units with $n_1 < N$ treated units and $n_0 = N - n_1$ control units. As the experiment is completely randomized, the assignment mechanism is given by

$$P(\mathbf{D}|\mathbf{Y}(\mathbf{0}), \mathbf{Y}(\mathbf{1})) = \begin{cases} \binom{N}{n_1}^{-1} & \text{if } \sum_{i=1}^{N} D_i = n_1 \\ 0 & \text{otherwise.} \end{cases}$$

a. Imagine $N = 5$. What is the probability that the first unit is treated? What is the probability that both of the first two units, $i = 1, 2$ are treated?

b. Our target parameter is the Sample Average Treatment Effect (SATE):

$$\tau = N^{-1} \sum_{i=1}^{N} \{Y_i(1) - Y_i(0)\}.$$

Show that τ is identified.

c. Show that the difference-in-means estimator $\hat{\tau}$ is unbiased conditional on potential outcomes for τ, where

$$\hat{\tau} = \frac{\sum_{i=1}^{N} Y_i D_i}{\sum_{j=1}^{N} D_j} - \frac{\sum_{i=1}^{N} Y_i (1 - D_i)}{\sum_{j=1}^{N} (1 - D_j)}.$$

d. Generate a dataset with $N = 100$ assuming the data generating process of a completely randomized experiment with $n_1 = 50$ and potential outcomes distributed as jointly Normal:

$$(Y_i(1), Y_i(0)) \sim N\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} \sigma & \rho \\ \rho & 1 \end{pmatrix}\right),$$

where $\sigma = 1$ is the variance of $Y_i(1)$ and $\rho = 0$ is the covariance. Use a nonparametric density estimator to plot the distribution of individual treatment effects $\tau_i = Y_i(1) - Y_i(0)$.

e. Keep only the observed outcome Y_i, and the treatment status D_i, and save the estimate $\hat{\tau}$ from a regression of Y_i on D_i and a constant. Also save the p-value of the coefficient of interest. Let us now conduct randomization inference under the sharp $H_0 : \tau = 0$. In principle, we would want to calculate the coefficient estimate for all potential alternative assignments, calculate their distribution, and compare how our realized treatment effect compares to that distribution of the estimator under the null, in particular in which percentile it is. That is typically infeasible. However, we can approximate that procedure by drawing large number of alternative assignments. Generate variables $Y_i^{H_0}(0)$ and $Y_i^{H_0}(1)$ under the sharp $H_0 : \tau = 0$. Generate 1000 alternative

assignments $(D^a; a = 1, ..., 1000)$ using the same randomization procedure used to generate the original assignment you saved in (d). For each alternative assignment, calculate $Y_i^{H_0,a} = Y_i^{H_0}(0) + (Y_i^{H_0}(1) - Y_i^{H_0}(0))D^a$, calculate $\hat{\tau}$ from a regression of $Y_i^{H_0,a}$ on the alternative assignment D^a and a constant, and save the coefficient estimate on D^a. Plot the distribution of those estimates and draw a line indicating the coefficient estimate from the true treatment assignment. Which fraction of coefficient estimates from alternative treatment assignments is larger than the coefficient estimate from the true treatment assignment?

f. Last we bootstrap the distribution of $\hat{\tau}$. Use the same dataset as in question 3.d and go through the following procedure 1000 times: each time draw from the original dataset 100 observations with replacement, and recalculate $\hat{\tau}$ given this alternative dataset. Plot the distribution of those estimates and draw a line indicating the coefficient estimate from the true sample. Which fraction of coefficient estimates from alternative treatment assignments is smaller than 0?