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 5th of May, 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 2022: Two-tier experiment, with added coding example)
Imagine a government runs the following two-stage experiment. In the first stage, villages $j = 1, ..., J$ are randomized into a fiscal transfer $D_j \in \{0, 1\}$ from the central government. In the second stage, individuals in each village $i = 1, ..., N_j$ are randomized into a cash transfer $W_{ij} \in \{0, 1\}$. Let Y_{ij} be consumption of individual i in village j one year after the two treatments have been allocated. Let $Y_{ij}(d, w)$ be the potential outcome of (i, j) if their treatment status is $(d, w) \in \{0, 1\}^2$. So there are four potential outcomes $Y_{ij}(0, 0)$, $Y_{ij}(0, 1)$, $Y_{ij}(1, 0)$, and $Y_{ij}(1, 1)$.

1. How would you formulate the stable unit treatment value assumption (SUTVA) in this context? Is it satisfied?

2. Write down expressions for three average treatment effects (ATE) τ_w, τ_d, and τ_{dw}, corresponding to the ATE of the effect of being treated only by the fiscal transfer, only by the cash transfer, and by both, respectively. Are all three of these ATEs identified? Why or why not?

3. Construct a population estimator consisting of differences in CEFs that correspond to the identified ATEs. Write down a population regression equation that also estimates these ATEs.

4. Consider including village effects α_j instead of D_j in this regression. Propose statistical tests for whether villages matter for consumption based on both a random effects and a fixed effects perspective.

5. Create samples with $N_j = 10$ individuals per village and $J = 100$ villages. Assume that $\Pr(D_j = 1) = \Pr(W_{ij} = 1) = 0.5$, and i.i.d. distributions

$$\begin{aligned} Y_{ij}(0,0) &\sim N(0,1) \\ Y_{ij}(1,0) &\sim N(1,1) \\ Y_{ij}(0,1)|\mu_j &\sim N(\mu_j,1) \\ Y_{ij}(1,1) &= Y_{ij}(1,0) + Y_{ij}(0,1) + X_{ij} \end{aligned}$$

   with $\mu_j \sim N(2, 1)$ and $X_{ij} \sim N(\phi, 0.2)$ for $\phi = -1, -0.8, ..., 0, 0.2, ..., 1$. How would you interpret the parameter ϕ? What about μ_j?

6. Run two regression specifications of Y_{ij} on treatments for each value of ϕ, the first including only indicators for D_j and W_{ij}, and the second additionally including the interaction term $D_j \times W_{ij}$. Make a separate figure for each of the two specifications. In each figure, plot ϕ on the horizontal axis and the true values of τ_d and τ_w as well as the coefficients on D_j and W_{ij} on the vertical axis. Think about your choice of standard errors and plot confidence intervals for all coefficients using `graph twoway rcap`.

7. Fix $\phi = 0$ and run a regression of Y_{ij} on village fixed effects and W_{ij}. Compare your estimate of the coefficient on W_{ij} to the coefficient of a regression on D_j and W_{ij} as in the previous question. How do they compare? Would your answer change if $\phi = 1$?

8. Keep $\phi = 0$. Instead of estimating the ATE of W_{ij}, imagine we are interested in the village-specific effect of the cash transfer. Run a regression to estimate village-specific effects and call them $\widehat{\tau}_j$. Make a scatter plot to compare your estimates $\widehat{\tau}_j$ to the true values μ_j.

9. Use Empirical Bayes methods to shrink the estimates of $\widehat{\tau}_j$. Make a plot using the Stata command `graph twoway pcarrow` to show how your shrunk estimates $\widehat{\mathbb{E}}[\mu_j|\widehat{\tau}_j]$ move the noisy estimates of $\widehat{\tau}_j$ towards the true values μ_j.

Question 2. (Retake Exam 2022: Two workers and two machines, with coding example.) Two workers $i \in \{1, 2\}$ are assigned to two different machines $j \in \{1, 2\}$ over time $t \in \{1, 2\}$. Let $j : \{1, 2\} \times \{1, 2\} \to \{1, 2\}$ be the mapping from workers in periods to machines such that $j(1, 2)$ would index a worker that is assigned to machine 1 in period 2. Suppose that in period 1, we have $j(1, 1) = 1$ and $j(2, 1) = 2$, whereas in period 2 we have $j(1, 2) = 2$ and $j(2, 2) = 1$.

1. Let us assume that workers and machines are perfect substitutes, such that output is produced according to
$$Y_{it} = \alpha_i + \gamma_t + \psi_{j(i,t)} + \varepsilon_{it}.$$
Consider $\theta = (\alpha_1, \alpha_2, \gamma_1, \gamma_2, \psi_1, \psi_2)$. Explain why θ is not identified. What normalizations render θ identified, and why?

2. Let us now assume that workers and machines are perfect complements, such that output is produced according to
$$Y_{it} = \gamma_t + \min\left\{\alpha_i, \psi_{j(i,t)}\right\} + \varepsilon_{it}.$$
Is θ (with appropriate normalizations) now identified? Why or why not?

3. Imagine we can now observe a third worker $i = 3$ with assignment $j(3, 1) = j(3, 2) = 1$. Does this change your answer to part 1 of the question? What about part 2?

4. Consider the following data:

Y_{it}	$t = 1$	$t = 2$
$i = 1$	2	4
$i = 2$	3	3

Under the assumption of perfect substitutes, find the values of θ after normalizations $\gamma_1 = \psi_1 = 0$.

5. Simulate data for a firm for which workers and machines are perfect substitutes. There are $N = 10$ workers assigned over $T = 200$ periods to $J = 20$ machines. Worker productivity is distributed as $\alpha_i \sim \mathcal{N}(0, 1)$; machines as $\psi_j \sim \mathcal{N}(1, 1)$; periods as $\gamma_t \sim \mathcal{N}(0, 0.5)$; and noise as $\varepsilon_{it} \sim \mathcal{N}(0, 0.2)$; all independently and identically distributed. The worker-to-machine assignment process is random: specifically, a worker's assignment is determined by $j(i, t) \sim U\{1, ..., J\}$, i.e. each machine with equal probability, independently across periods.

6. State the conditions under which the worker and machine fixed effects are identified according to this DGP, and in particular given the random assignment process.

7. Estimate worker and machine effects simultaneously in a regression and save the estimated effects as $\widehat{\alpha}_i$ and $\widehat{\psi}_j$, respectively. Plot the true α_i against their estimated counterpart, and same for ψ_j. Comment on what you see in these plots.

8. Re-estimate the effects with $T = 5$ and plot the same figures as before. How did they change?

9. Imagine the firm realizes that its worker-to-machine assignment is not optimal, and so it changes the assignment process to maximize output. How will the assignment function $j(i, t)$ look like now? What is the implication for the identification of parameters in a connected-set fixed effects model?

Question 3. (Exam 2023: Differences-in-differences with transitory policy.)
Imagine we have data corresponding to the following treatment structure:

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

1. Set up a two-way fixed effects estimator of an outcome Y_{it} on the treatment $D_{it} \in \{0, 1\}$. Under what assumption(s) does the coefficient on the treatment correspond to some causal effect? Define the causal effect of interest.

2. Define
$$\tau_{gt} = N_{gt}^{-1} \sum_{i \in g} [Y_{it}(1) - Y_{it}(0)]$$
as the group-time specific average treatment effect, where N_{gt} is the number of observations in period t of cohort g and $Y_{it}(d)$ are potential outcomes for $d \in \{0, 1\}$. Find the weights of τ_{gt} associated with the four treated cells.

3. Consider a switching estimator in the spirit of de Chaisemartin and d'Haultfoeuille (2020). Which comparisons does it leverage, and which ones does it avoid? State any additional assumptions we need for any of these comparisons.

4. Consider the following CEF matrix:
$$\begin{bmatrix} 2 & 3 & 4 & 3 \\ 1 & 4 & 5 & 3 \end{bmatrix}.$$
Derive by hand estimates of any τ_{gt} that are identified. Imagine the weights across switching estimator comparisons are equal, what is the estimate of the causal effect define in the first part of this question?