Causal inference HW2

1. Consider a completely randomized experiment with a binary outcome, with data summarized in the following 2 ×2 table:

	$Y = 1$	$Y = 0$	Total
$A = 1$	n_{11}	n_{10}	n_1
$A = 0$	n_{01}	n_{00}	n_0

Under the sharp null hypothesis, show that
(1) any test statistic is a function of n_{11} and other nonrandom fixed constants, and
(2) the exact distribution of n_{11} is Hypergeometric. Specify the parameters for the Hypergeometric distribution.

2.

This dataset used in this question is based on the Pennsylvania Reemployment Bonus Experiment, as analyzed in Koenker and Xiao (2002).

Koenker, R., & Xiao, Z. (2002). Inference on the Quantile Regression Process. *Econometrica*, 70(4), 1583–1612.

The experiment was designed to assess the effect of offering cash bonuses for early reemployment on the length of unemployment spells. Individuals who filed for unemployment benefits were randomly assigned to either a treatment group, eligible for a reemployment bonus, or a control group.

Importantly, the study employed a stratified randomized design, with randomization carried out within calendar quarters. That is, individuals were stratified by the quarter in which their unemployment spell began (denoted as **quarter**), and then randomized to treatment or control. This stratification helps control for seasonal variation in labor market conditions. The primary outcome variable is the duration of unemployment, denoted as **duration**, measured in weeks. Treatment assignment is indicated by the variable **treatment**. In addition to treatment status and quarter, the dataset includes a rich set of covariates:

- Gender (**female**)
- Race/ethnicity (**black**, **hispanic**)
- Number of dependents (**ndependents**)
- Expected job recall (**recall**)
- Age group indicators: **young** (under 35) and **old** (over 54)
- Industry background (**durable**): indicator for previous employment in the durable goods sector
- Region type (**lusd**): indicator for residence in a low-unemployment, short-duration area (Coatesville, Reading, or Lancaster)

Questions:

(1). Conduct analysis for the stratified randomized experiment (stratified by **quarter**) without covariate adjustment:

A. Perform a Fisher randomization test to evaluate the sharp null hypothesis that the treatment has no effect on any unit, and report the p-value of the test.

B. Conduct Neymanian repeated sampling inference to estimate the ATE. Report the point estimate, estimated standard error, and the 95% confidence interval.

(2). Perform regression adjustment within each stratum (i.e., within each level of **quarter**) using available covariates. Then aggregate the adjusted estimates across strata to estimate the overall ATE.

A. Report the point estimate, standard error, and 95% confidence interval of the adjusted estimator.

B. Compare the adjusted results with the unadjusted estimates from Question 1, and discuss whether regression adjustment improves the precision of the estimate or alters the substantive conclusions.

3.
Under the assumptions of exchangeability, positivity, and SUTVA, show

1. the distributional treatment/causal effect

$$\mathrm{DCE}_y = \Pr(Y(1) > y) - \Pr(Y(0) > y)$$

is nonparametrically identifiable for all y;

2. the quantile causal effect

$$\mathrm{QCE}_q = \mathrm{quantile}_q\{Y(1)\} - \mathrm{quantile}_q\{Y(0)\}$$

is nonparametrically identifiable for all q, where $\mathrm{quantile}_q\{\cdot\}$ is the qth quantile of a random variable.