1. 1. Define the causal effect as the comparison of potential outcomes, for the same unit, at the same moment in time post-treatment: $\tau_i = Y_i(1) - Y_i(0)$. The fundamental problem of causal inference is therefore the problem that at most one of the potential outcomes can be realized and thus observed. Assessing the causal effect of a binary treatment requires observing more than a single unit, because we must have observations of potential outcomes under both treatments. To rule out the existence of a causal effect of a treatment relative to an alternative an exclusion restriction (such as SUTVA: potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes) is necessary.

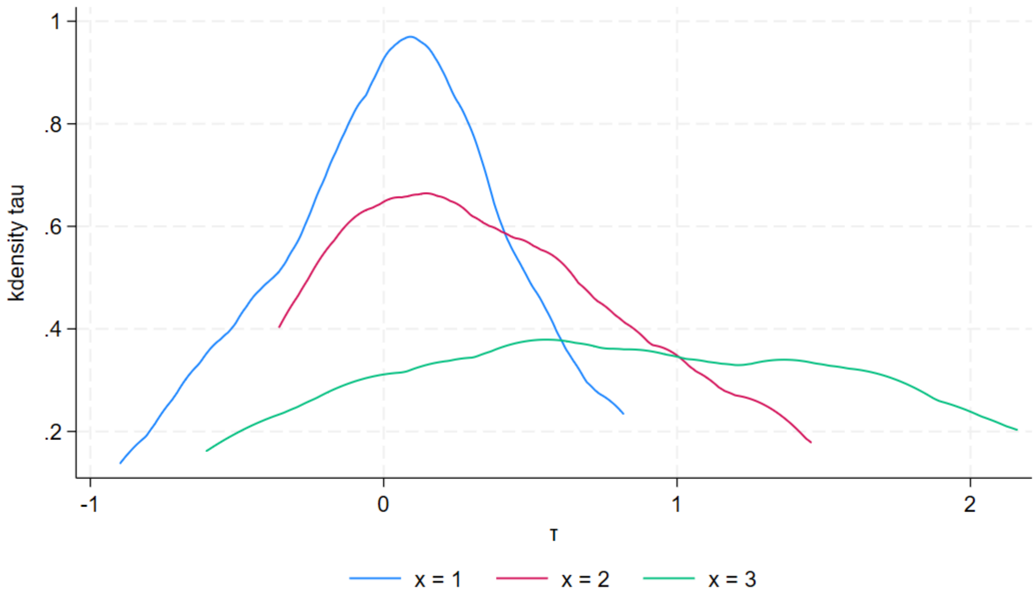
   2. The average treatment effect on the treated: $\tau = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1]$ is identified in the model $Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0), \quad D \in \{0,1\}$ if $\mathbb{E}[Y_i(d) \mid D_i] = \mathbb{E}[Y_i(d)] \quad \text{for } d \in \{0,1\}$ (mean indep. assumption/potential outcomes are missing at random: $D_i \perp \{Y_i(0), Y_i(1)\} \quad \forall i$) For suppose it doesn't: $\mathbb{E}[Y_i \mid D_i = 1] - \mathbb{E}[Y_i \mid D_i = 0] = \tau + (\mathbb{E}[Y_i(0) \mid D_i = 1] - \mathbb{E}[Y_i(0) \mid D_i = 0]) \neq \tau$.

   3. Similarly (taking conditional expectation), the CATTs $\tau(x)$ for each type $x \in \{1,2,3\}$: $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i = 1, X = x]$ are identified under the conditional Independence Assumption (CIA): $D_i \mid X_i \perp \{Y_i(0), Y_i(1)\}$ Additionally, $0 < \Pr(D_i = 1 \mid X_i = x) < 1 \quad \forall x \in \{1,2,3\}$ is necessary for the conditional expectation to be defined.

   4. Denote the observed outcomes Y_i^x and treatment status D_i^x for total number N_x observations with $X = x$. An expression based on observed data: $\hat{\tau}(x) = \frac{\sum_{i=1}^{N_x} Y_i^x D_i^x}{\sum_{i=1}^{N_x} D_i^x} - \frac{\sum_{i=1}^{N_x} Y_i^x (1 - D_i^x)}{\sum_{i=1}^{N_x} (1 - D_i^x)},$. The CATTs are defined for $x \in \{1,2\}$ and can be estimated as: $\hat{\tau}(1) = 2 - \frac{5+3}{2} = -2, \quad \hat{\tau}(2) = 4 - 5 = -1$.

   5.
```
set obs 100
gen x = cond(_n <= 50, 1, cond(_n <= 75, 2, 3))
gen y0 = runiform()
gen y1 = x * runiform()
gen d = runiform() < .5
gen tau = y1 - y0
```

   6. Individual causal effects $\tau_i(x)$:



[1] mach5689@student.su.se

7. ```
   gen y = y1 * d + y0 * (1 - d)
   reg y d
   scalar ate = _b[d]
   reg y d if x == 1
   scalar att1 = _b[d]
   reg y d if x == 2
   scalar att2 = _b[d]
   reg y d if x == 3
   scalar att3 = _b[d]

   gen tau_d = y1 - y0 if d==1
   gen tau1 = tau if x == 1
   gen tau2 = tau if x == 2
   gen tau3 = tau if x == 3

   egen tau_mean = mean(tau)
   scalar tau_mean = tau_mean
   egen tau1_mean = mean(tau1)
   scalar tau1_mean = tau1_mean
   egen tau2_mean = mean(tau2)
   scalar tau2_mean = tau2_mean
   egen tau3_mean = mean(tau3)
   scalar tau3_mean = tau3_mean
   egen tau_d_mean = mean(tau_d)
   scalar tau_d_mean = tau_d_mean

   matrix table = (ate, ate, att1, att2, att3) \
   (tau_mean, tau_d_mean, tau1_mean, tau2_mean, tau3_mean)
   matrix list table
                ATE          ATT        CATT(1)      CATT(2)      CATT(3)
   _b[d]   .27233104    .27233104    -.03207337      .389496    1.1113485
   mean    .32671812    .24986053     .03637347    .38980356    .84432203
   ```

   Estimates of the ATE/ATT/CATT for each type of X are reported in the first row of the table. Compared with the mean values of the unobserved individual treatment effects, shown in the second row we, the regression coefficient estimate of ATE/ATT is between the mean true values. The estimate of CATT(1) is close to the true value of zero (Y(1) and Y(0) iid). Respectively, the CATT(2) and CATT(3) estimates are close to and slightly higher than the mean values.
   ```

2. a. Adding zero to the ATT $\tau \equiv \mathbb{E}[\tau_i|D_i = 1] = \mathbb{E}[Y_i(1) - Y_i(0)|D_i = 1]$ shows is not point-identified: $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] - (\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0])$ because the known population quantity $\mathbb{E}[Y_i|D_i = 1] - \mathbb{E}[Y_i|D_i = 0] \neq \tau$ as in general $\mathbb{E}[Y_i(0)|D_i = 1] - \mathbb{E}[Y_i(0)|D_i = 0] \neq 0$.

b. Since $Y_i(1) = Y_i(0) + \tau$: $\mathbb{E}[\mathbb{E}[Y_i|D_i = 1, X_i = x] - \mathbb{E}[Y_i|D_i = 0, X_i = x]|D_i = 1] = \tau$ due to conditional mean indep., $\tau = \alpha_0 + \alpha_1 \cdot Pr(X_i = 1) = 1 - 2 \cdot 0.3 = 0.4$ if X_i is independent of D_i.

c. The CATT is identified as the difference-in-means between the D_i groups at $X_i = x : \tau(x) = \mathbb{E}[\tau_i|D_i = 1, X_i = x] = \mathbb{E}[Y_i|D_i = 1, X_i = x] - \mathbb{E}[Y_i|D_i = 0, X_i = x]$. Common support is necessary for identification as it ensures the conditional expectations are defined. Then $\tau(0) = \mathbb{E}[\tau_i|D_i = 1, X_i = 0] = \alpha_0 = 1, \quad \tau(1) = \mathbb{E}[\tau_i|D_i = 1, X_i = 1] = \alpha_0 + \alpha_1 = -1$.

d. There are $\binom{5}{3} = 10$ assignments

e. $\binom{3}{2}\binom{2}{1} = 6$ assignments are left, Completely Randomized Experiment: unit assignment probabilities are $n_1(x = 0)/N(x = 0) = 2/3, \quad n_1(x = 1)/N(x = 1) = 1/2$ for groups with $X_i = 0, X_i = 1$ respectively. Corresponding propensity scores are equal to the unit assignment probabilities since the unit assignment probability is constant within each group.

f.
```
set seed 42
set obs 200
gen y0 = rnormal()
gen x = runiform() < .3
gen d = runiform() < .5
gen tau_i = 1 + -2*x
gen y = y0 + d*tau_i
preserve
collapse y, by(d)
gen tau_hat = y - y[_n-1]
restore
reg y d
```

The ATT averaging and OLS estimates are the same: $\hat{\tau} = .6609471$ ($\neq 0.4$ from above)

g.
```
preserve
collapse y, by(d x)
sort x d
gen tau_x_hat = y - y[_n-1]
list tau_x_hat if x == 0 & d == 1
 1.058825
list tau_x_hat if x == 1 & d == 1
 -.435797 |
restore
gen dx0 = d*(1-x)
gen dx1 = d*x
reg y dx0 dx1 x
y         Coefficient
dx0       1.058825
dx1       -.435797
```

Both approaches result in CATTs estimates $\hat{\tau}(X = 0) = 1.058825$ and $\hat{\tau}(X = 1) = -.435797$.

h. The ATT estimates are centered around 0.4 in experiment 1 and slightly shifted to the left in experiment 2 (gender imbalance between control and treatment groups gave more weight to the treatment effect for female observations). The CATT estimates do not differ across experiments.

Figure 1: ATT

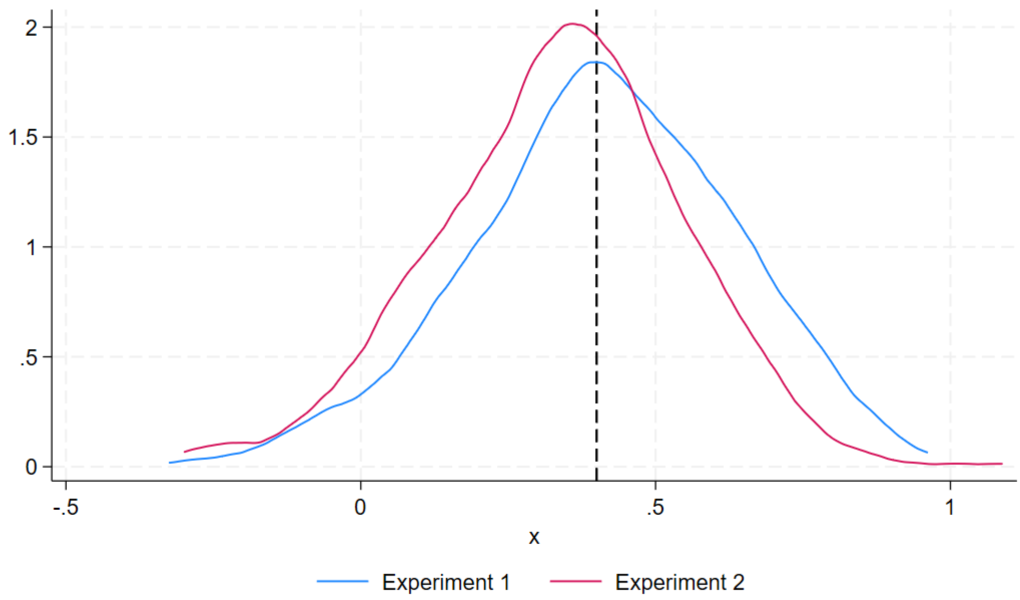
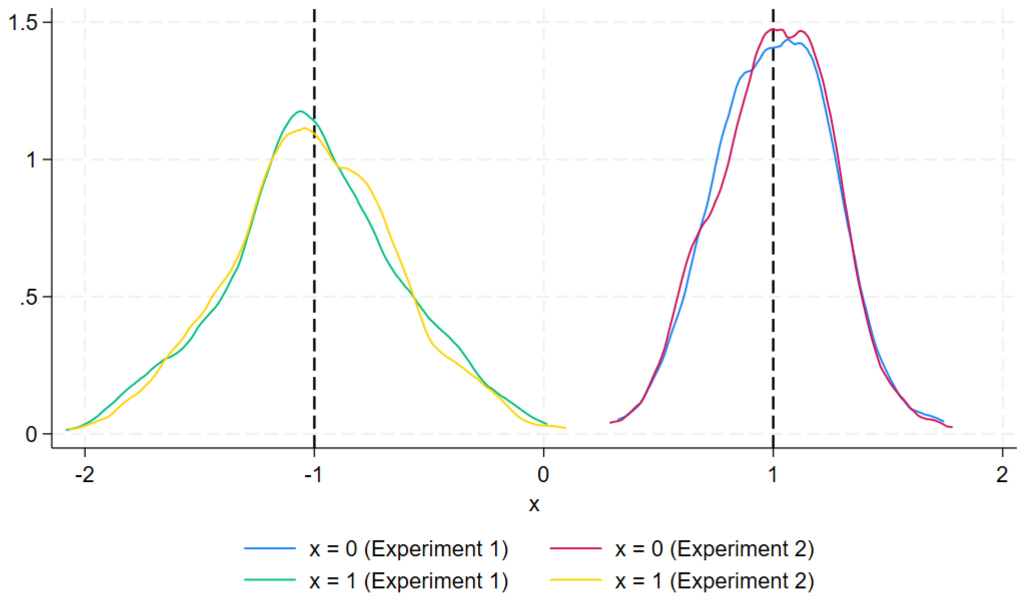


Figure 2: CATT



3. a.

$$\frac{\binom{N-1}{n_1-1}}{\binom{N}{n_1}} = \frac{(N-1)!}{(n_1-1)!(N-n_1)!} \cdot \frac{n_1!(N-n_1)!}{N!} = \frac{n_1}{N} = \frac{n_1}{5}$$

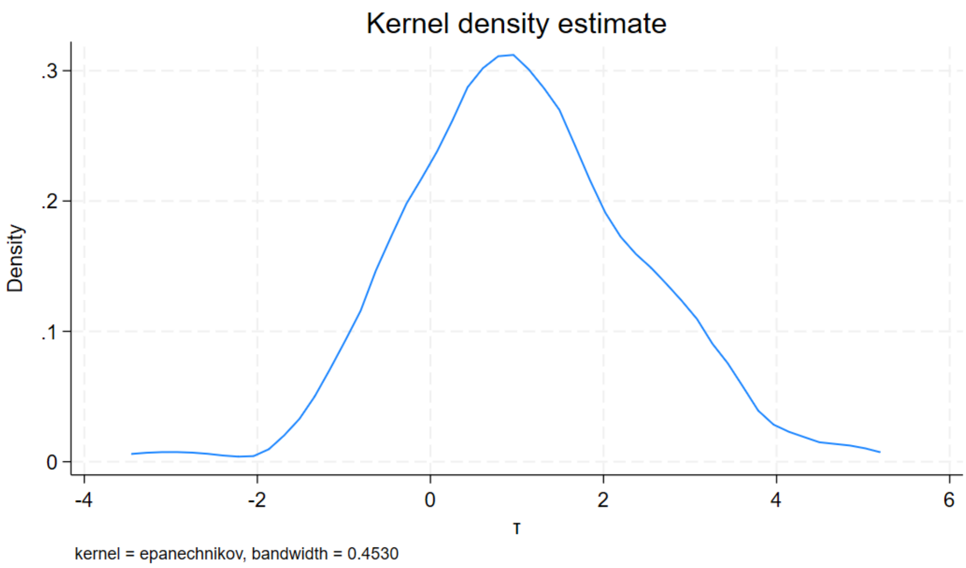
$$\frac{\binom{N-2}{n_1-2}}{\binom{N}{n_1}} = \frac{(N-2)!}{(n_1-2)!(N-n_1)!} \cdot \frac{n_1!(N-n_1)!}{N!} = \frac{n_1(n_1-1)}{N(N-1)} = \frac{n_1(n_1-1)}{20}$$

b. See (c) below: the value of τ can be determined from observable quantities, i.e. $\hat{\tau}$ averaged over repeated treatment assignments from the assignment mechanism identifies τ

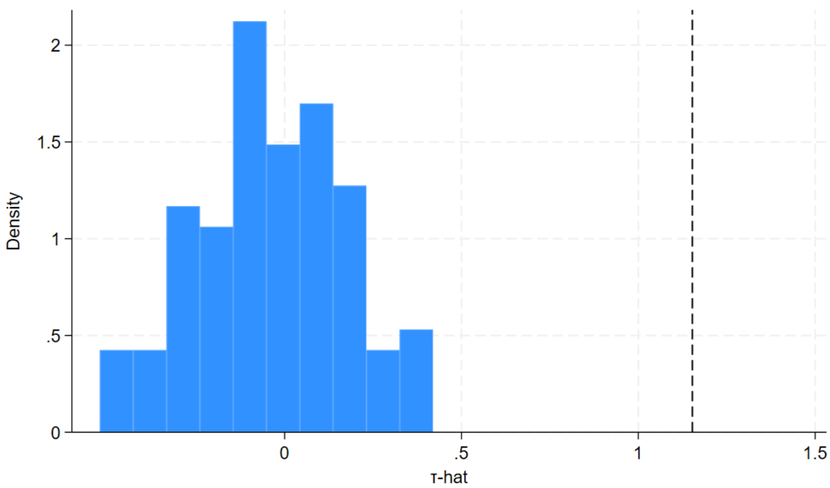
c. Note $Y_i D_i = Y_i(1)D_i, \quad Y_i(1-D_i) = Y_i(0)(1-D_i)$ and from (a) follows $E[D_i|(Y_i(1), Y_i(0))_{i=1}^N] = P(D_i = 1|(Y_i(1), Y_i(0))_{i=1}^N) = \frac{n_1}{N}$. Rewriting $\hat{\tau} = \frac{\sum_{i=1}^N Y_i D_i}{n_1} - \frac{\sum_{i=1}^N Y_i(1-D_i)}{N-n_1}$ we have

$$E[\hat{\tau}|(Y_i(1), Y_i(0))_{i=1}^N] = \frac{1}{N}\sum_{i=1}^N (Y_i(1) - Y_i(0)) = \tau.$$

d. Individual treatment effects τ_i



e. The fraction is 0.



f. The fraction is 0.