Question 1.1

Is true. Here are two kinds of proofs, both correct:

Suppose on observation, i, is randomly selected from a population of size N, let this sample be denoted $S = \{i\}$. Then the treatment $D_i \in \{0, 1\}$ is randomly assigned to that individual with potential outcomes $(Y_{1,i}, Y_{0,i})$. We observe $Y_i = D_i Y_{1,i} + (1 - D_i)Y_{0,i}$. The sample average treatment effect (ATE) is $\hat{\rho} = 2D_i Y_i - 2(1 - D_i)Y_i$ which estimates population ATE, $\rho = \mathrm{E}(Y_{1,i} - Y_{0,i})$.

Part 1. True.

$$
\begin{aligned}
\mathrm{E}(\hat{\rho}) &= \mathrm{E}\left(\mathrm{E}(\hat{\rho}|S)\right) = \mathrm{E}\left(\mathrm{E}\left(2D_i\left(D_i Y_{1,i} + (1 - D_i)Y_{0,i}\right) - 2(1 - D_i)\left(D_i Y_{1,i} + (1 - D_i)Y_{0,i}\right)|S\right)\right) \\
&= \mathrm{E}\left(\mathrm{E}\left(4D_i^2 Y_{1,i} - 2D_i Y_{1,i} - 4D_i^2 Y_{0,i} + 6D_i Y_{0,i} - 2Y_{0,i}|S\right)\right) \\
&= \mathrm{E}\left(4\frac{1}{2}Y_{1,i} - 2\frac{1}{2}Y_{1,i} - 4\frac{1}{2}Y_{0,i} + 6\frac{1}{2}Y_{0,i} - 2Y_{0,i}\right) = \mathrm{E}(Y_{1,i} - Y_{0,i}) = \rho
\end{aligned}
$$

Thus $\hat{\rho}$ is unbiased due to the random sampling and random assignment.

Or

$\hat{\rho}$ is unbiased for ρ if $\mathrm{E}[\hat{\rho}] = \rho$.

$$
\begin{aligned}
\mathrm{E}[\hat{\rho}] &= \mathrm{E}[\mathrm{E}[\hat{\rho}|D_i]] && \text{(law of total exp.)} \\
&= \mathrm{E}[p(D_i = 0)\,\mathrm{E}[\hat{\rho}|D_i = 0] + p(D_i = 1)\,\mathrm{E}[\hat{\rho}|D_i = 1]] && \text{(def. of conditional exp.)} \\
&= \mathrm{E}\left[\frac{1}{2}\,\mathrm{E}[2Y_{1i}] + \frac{1}{2}\,\mathrm{E}[-2Y_{0i}]\right] && \text{(by linearity)} \\
&= \mathrm{E}[\mathrm{E}[Y_{1i}] - \mathrm{E}[Y_{0i}]] \\
&= \mathrm{E}[Y_{1i} - Y_{0i}] = \rho
\end{aligned}
$$

The statement is true.

Question 1.2

This is interesting; I thought I knew the answer, but some of the responses make it seem that the answer might depend on some finicky question about the definition of "consistent." I said the answer that I thought was correct in class, so I gave everyone full credit for this question. Even though everyone got credit, I sitll left comments if your logic was incorrect. Here are two responses that are each correct for a given interpretation of the question.

Part 2. True. Typically when discussing consistency, we look at $\text{Var}(\hat{\rho})$ as the sample size increases. We know with the randomization of treatment and random sampling the $\hat{\rho}$ is consistent for ρ. For sample size greater than 1, we can consider $\hat{\rho} = \frac{1}{n_1}\sum_{i:D_i=1} Y_i - \frac{1}{n_0}\sum_{i:D_i=0} Y_i = \bar{Y_1} - \bar{Y_0}$, the variance is

$$\text{Var}(\hat{\rho}) = \text{E}\left(\text{Var}(\bar{Y_1} - \bar{Y_0}|S)\right) + \text{Var}\left(\text{E}(\bar{Y_1} - \bar{Y_0}|S)\right) = \text{E}\left(\frac{s^2_{Y_1}}{n_1} + \frac{s^2_{Y_0}}{n_0} - \frac{s^2_{\rho}}{n}\right) + \frac{1}{n}\text{Var}(Y_1 - Y_0) = \frac{\sigma^2_{Y_1}}{n_1} + \frac{\sigma^2_{Y_0}}{n_0}$$

which converges to 0 as the sample size increases.

^assumes that “as the sample size increases” is the definition of “consistency,” aruing that the trick question is contrary to the definition.

In class, it was said that $\hat{\rho}$ is a consistent estimator of ρ as $N \to \inf$. For a fixed i and D_i, the population has no effect on the estimator (the outcome of the single unit does not depend on the size of the population–only on the treatment assignment in the switching equation). (My intuition prior to class was that this is inconsistent because the value should switch, depending on the treatment assignment, between $2Y_{1i}$ and $-2Y_{0i}$, so the estimator does not necessarily converge to a single value. Convergence is necessary for consistency.)

^assumes that convergence to a single value will never happen.

Question 2

The key thing is that you isolated the *causal* variables of interest, the identification strategy, and thought about potential threats to inference that could have been elimated in an ideal experiment. Many people chose to analyze Miller & Sutherland (2023), so here are some answers w my comments.

“The Effect of Gender on Interruptions at Congressional Hearings.” by Miller & Sutherland (2023)	
Causal Effect of Interest	Does gender affect frequency of interruption in Congress? Does being a woman speaking on women’s issues affect this frequency of interruption?
Identification Strategy brutal lol	While this is an empirical study, it does not attempt to make causal claims. Instead it models the probability of being interrupted, using logistic regression on “chunks” of text in transcripts with an interruption indicator as the response and speaker gender as the main predictor. The model includes covariates and their interaction with gender, such as an indicator for women’s issues discussions,seniority level, political party, etc. Fixed effects for a Senate indicator, the committee type, and the session are included.
Ideal Intervention and manipulability yes	The ideal intervention would be randomly assigning genders to incoming politicians. This is not well defined in terms of manipulability. Researchers would also need to randomly assign who speaks for a set amount of time at a given time and randomly assign whether the discussion was on a women’s issue. This more manipulable than assigning genders, but still not manipulable in practice as it would require specifying the nature of discussions in congress.

^ interprets the paper as not making causal claims. They clearly think they're making causal claims, but this is not well-defined: I suppose they're motivating CIA through the control variables and fixed effects

1)The main effect the authors are interested in is the average treatment effect of being female on the probability of getting interrupted in Congress. The authors are also interested in identifying conditional average treatment effects based on speech topic, speech timing, and the majority, party, and chairmanship statuses of the speaker.

---

2)The authors use congressional hearing transcripts as their source of data and logistic regression with Congress session and committee fixed effects and robust standard errors as their statistical technique.

3)The ideal intervention would be to randomly assign gender to members of Congress. However, gender is clearly not manipulable.

yes

^ interprets the question as causal but lacking a clear identification strategy.

This is an observational study, the authors extract interruptions in chucks of speech in Congressional hearings (binary), and correlate it with gender variable.

An ideal intervention is to randomly assign gender to Congressional speakers and see if they are more likely to be interrupted after the treatment. Yet, gender is one of the demographic variables we cannot manipulate at all.

yes

^ same here.

Question 3

1. The X distribution is normal with a long right tail, the Y distribution is left-censored and has a long right tail
2. Despite this difference, the bootstrapped sample means of each variable approach a normal distribution. This convergence happens “faster” in the sample size for the X distribution than for the Y distribution; that is the, the variance of the distribution is lower at each sample size. The mean of each variable is 6.
3. The random assignment here ensures that the true value of the means of the treatment and control will be equal in expectation across the replications. The distribution of “treatment effects” is thus normal at each sample size, but the variance (“width”) of the distribution varies significantly, decreasing in the sample size.