Tidy Data

1. (short answer/numerical) Given the provided example data set, is it in a tidy format? Why or why not? If not, manipulate the data to a tidy format
2. (short answer/numerical) Given the two provided example data set, which variables are the keys and which are the values? Perform an inner/left/right/full join.

Regression

1. (numerical) Given the provided example data, what is $\bar{x}$, $\bar{y}$, $Cov(x, y)$, $V(x)$. Calculate $\hat{\beta}_0$ and $\hat{\beta}_1$.
2. (short answer) When should you use sample weights?

Directed Acyclic Graphs

1. (short answer) Given the provided variables provided and their described relationships, draw a DAG
2. (short answer) Given the provided DAG, list all of the paths and label what type of path they are
3. (short answer) Given the provided DAG, which variables are colliders?
4. (short answer) Given the provided DAG, on which variables should you condition?

Potential Outcomes

1. (numerical) Given the provided table and equations, calculate the ATE/ATT/ATU/SDO/selection bias/heterogeneous treatment effect bias.
2. (short answer) Intuitively, how does independence resolve selection bias?
3. (short answer) What is the Stable Unit Treatment Value Assumption and why is it important?

Randomized Experiments

1. (short answer) What does a balance table tell you? What should you do with statistically imbalanced variables?
2. (short answer) Given the provided example data set, which variables should be included as rows of a balance table?
3. (short answer) Intuitively, what is the conditional independence assumption and why is it important?
4. (short answer) Intuitively, what is the process for propensity score matching? Should you use all matches with all values of propensity scores from zero to one? Why are economists skeptical of propensity score matching?
5. (short answer) Experimental design: what is the relationship between the significance level and the power level? Ceteris paribus, how does a larger sample affect: the confidence level of a test; the power of a test; and detectable effect size?
6. (short answer) What is internal validity? What is external validity?

Instrumental Variables

1. (short answer) If ability is omitted from a return of education on wages regression, which way do you expect the OLS estimate to be biased. Assuming the 2SLS estimate from Card (1993) identifies a causal effect, why is the 2SLS estimate larger than the OLS estimate?

2. (short answer) Intuitively, list and explain the five assumptions required to estimate a LATE with heterogenous treatment effects. Which of the five assumptions are testable?

3. (short answer) List the three instrumental variable causal equations of interest by name and write them out using variables, coefficients and error terms appropriately.

4. (short answer) Which of the three instrumental variable causal equation of interest has a causal interpretation if monotonicity is violated?

5. (numerical) Given the provided estimates, show that the structural equation is the ratio of the reduced form and the first stage

6. (short answer) What is the problem with a weak instrument?

7. (short answer) Regarding compliance, what are the four sub populations? What is the empirical problem with the group that does not comply?

8. (short answer) If you are running a randomized experiment with one-sided non-compliance, what does your OLS estimate return, and what is the name of this instrumental variable causal equation? How can you get a version of a local average treatment effect.