1. What were the two major families of displays (diagrams for working with qualitative data) described in lecture?

2. You observe a positive and statistically significant association between a variable X and self-reported # of sexual partners among men. You have reason to believe that a substantial number of men exaggerate their number of sexual partners. Following our discussion of the Lewontin article, why can this specifically compromise your concluding from your data that a positive association (not necessarily causal) exists between X and one’s actual # of sexual partners?

3. I said that one way of thinking about what randomization does is that under randomization the mean difference between the error terms \((e_i - e_c)\) for the outcomes for treatment and control groups will be zero across the randomization set. What does the term “randomization set” refer to?

4. Explain the distinction between multiple causation and conjunctural causation.

5. What is the distinction between emic vs. etic categories?

6. You wish to test whether the implementation of a special reading program will improve student performance on subsequent reading exams. You get funding for your study from an agency especially concerned with helping disadvantaged students, who tend to have lower reading scores. Within a school, you are able to assign all students with reported family incomes below $30K to the program, while all students with family incomes above $30K are excluded from the program.

   (a) What is this design called?

   (b) Under the sometimes heroic assumptions discussed in class, you will get an unbiased estimate of the treatment effect with this design if you compute a regression in which post-program reading exam score is included as a dependent variable and what is/are included as independent variable(s)?

   (c) A critic comes along and smugly asks you how you can call this effect unbiased, given that (1) you don’t control for parent’s education, (2) parent’s education is plainly associated with the outcome, and (3) the treatment and control group will plainly differ in their mean levels of parental education if you base the assignment on family income. Put this critic in her/his place by briefly explaining why the estimated treatment effect will still be unbiased despite the existence of this pattern of associations with parental education.

7. What is meant by a “member check” in an ethnographic study?

8. I gave at least four reasons why being able to conduct multiple randomized experiments provide a much more powerful apparatus for causal inference than only being able to conduct a single experiment. Give two.

9. What was the distinction that your instructor drew between physical control and statistical control?

10. What is meant by the term “confirmation bias” as a source of invalidity of inferences drawn from qualitative data?

   (a) Your instructor highlighted two distinct ways that confirmation bias can cause problems. What were they?

11. To McCloskey and Ziliak, what is the “standard error of regressions”?

12. In a randomized experiment, X is a binary variable that represents whether or not the observation received the treatment, and Z is a pre-treatment characteristic of the observation. Say X and Z are completely uncorrelated. Even so, I have asserted that controlling for Z (covariance adjustment for Z) will improve the efficiency of the estimate of the causal (i.e., treatment) effect of X on Y. As a first matter, what does it mean to improve the efficiency of an estimate?
(a) If \( X \) and \( Z \) are not correlated, why would controlling for \( Z \) improve the efficiency of the estimated treatment effect of \( X \)?

13. What is the issue of reactivity in ethnographic research?

14. Draw a 2-by-2 table like the one below (you don’t have to label the rows and columns, as long as you have 2 of each) and put an A in all cells that represent conditions under which \( Z \) (uncontrolled) may be a source of spuriousness in the estimation of the treatment effect of \( X \) on \( Y \). Put a B in the cell that represents the condition presumably obtained when you conduct a properly randomized experiment where \( Z \) is some pre-assignment characteristic of the observations. (Note: a cell may contain both an A and B).

<table>
<thead>
<tr>
<th>( Z ) is a potential cause of variation in ( Y )</th>
<th>( Z ) is not correlated with ( X )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Z ) is correlated with ( X )</td>
<td>( Z ) is not a potential cause of variation in ( Y )</td>
</tr>
</tbody>
</table>

15. In his essay “Two Methods in Search of a Substance,” what methods were Coser’s two targets?

16. Drawing on Maxwell’s distinction, give a hypothetical example of an account (description) of an episode an ethnographer could provide that was descriptively valid but interpretatively invalid (not valid).

17. What does Holland think social scientists should be doing more of: (a) measuring the effects of causes or (b) measuring the causes of effects?

18. Provide an example, made-up or hypothetical as it may be, that would be transparently an example of a researcher committing Fischer’s “fallacy of responsibility for cause.”

19. Imagine a situation in which \( Z \) is the only potential source of spuriousness for the estimation of an effect of \( X \) on \( Y \). If \( X \) is positively associated with \( Y \) and \( Z \) is positively associated with \( X \), then under what association(s) between \( Z \) and \( Y \) (i.e., positive, negative, no) would the estimated causal effect of \( X \) on \( Y \) have a larger positive magnitude when you control for \( Z \) than when you do not control for \( Z \)?

(a) In this situation, where our estimate of the causal effect of \( X \) on \( Y \) is diminished relative to the true causal effect when we fail to control for \( Z \), what is the term used to describe \( Z \)?

20. When someone complains that the typical social psychology experiment, done in the laboratory with undergraduates as subjects, has low external validity, they could be complaining about either of two distinct things (or both). What are these?

21. What did Aronson et al. mean when they called randomization the “great equalizer”?

22. Your instructor characterized Devah Pager’s audit study as employing simultaneously a within-subjects and between-subjects design. What about Devah’s design was a within-subjects design, and what about it was a between-subjects design? (If you don’t know/recall how this pertains specifically to her experiment, you should at least explain what the distinction between a within- and between-subjects design is, although don’t expect anything like full credit for it.)

(a) What does it mean to say that Devah uses counterbalancing in the within-subject part of her design? (If you don’t know this specifically to her experiment, you should at least explain what counterbalancing is, although don’t expect more than partial credit for it.)

23. What is the causal transience assumption that Holland talks about?

24. Explain what it means to use blocking in a randomized experiment.

25. Can you sample on the value of the outcome in both prospective and retrospective studies? Explain why or why not.

26. Writers about qualitative methodology stress the importance of seeking out corroboration from multiple “sources” or explicable noncorroboration as a way of gaining confidence in descriptions or inferences. What is the general name for this strategy?
27. Your instructor gave out a handout in which a simulated example he showed how the effect of cultural capital on an outcome persisted when you controlled for a measure of SES, whether SES was measured as an ordinal categorical variable or measured continuously. Simply and briefly, what was the point of this handout?

28. You estimate $\beta_X$ in a regression of dependent variable $Y$ on independent variable $X$ in a model with no controls, and this estimate is statistically significantly different from zero. Then you control for $Z$ and your new estimate of $\beta_X = 0$. Explain how your interpretation of what the difference in $\beta_X$ between the two models implies about the relationship between $X$ and $Y$ depending on whether you conceive $Z$ as being a source of spuriousness or an intervening variable.

29. According to Katz, what is the paradox of the ethnographer’s warrant?

30. What does it mean to say that an experiment has a “2x3 factorial” design?

31. According to Holland, what is the fundamental problem of causal inference?

32. In our discussion of covariance adjustment in randomized experiments, it was noted that if you control for exogenous variable $Z$, your adjusted estimate of the treatment effect ($\hat{T}_{adjusted}$) of $X$ on $Y$ will probably differ from the estimate obtained without controlling for $Z$ ($\hat{T}$). What two things increase the magnitude of $|\hat{T} - \hat{T}_{adjusted}|$?

33. Duneier’s appendix to Sidewalk expresses the concern that Burawoy’s extended case method collapses two different tasks that Duneier suggests should be considered distinct issues. What are these?

34. What is Lieberson’s “formalistic fallacy”?

35. Why would one ever want to use a case-control sampling design for estimating the effect of an independent variable on an outcome rather than a design based on a simple random sample from a population? (Be specific; a one-word answer [or multi-word answers that can be reduced to a single word] is not enough, but a sentence is enough–your answer should indicate that you have some idea of what a case-control design is rather that you are simply guessing.)