Practice Quiz 2
G572
Spring 2023

1. To make a good passive prediction we must have which of the following:
    a. **Observational data**
    b. Experimental data
    c. An exogenous model
    d. An endogenous model

Treatments in an RCT are assigned differently than they would be in observational data and when we make passive predictions it is based on what we expect to happen when incentives/prices/trade-offs are as they would be in the real world.

2. To make a good active prediction we must have which of the following:
    a. Observational data
    b. Experimental data
    c. **An exogenous model**
    d. An endogenous model

It is true that experimental data is the easiest way to get this, but it isn't necessary (sufficient, but not necessary). What we really need is an exogenous model and technically we can get that even when we have observational data (not directly and note easily, but it is possible).

3. If gender in a data set is a binary variable defined as female = 1 and male = 0, which of the following is more likely to have a negative correlation for adults in the U.S. (this requires a little bit of outside knowledge, but not much)? (college degree = 1 for those that have a college degree)
    a. **Age and college degree (years 1920-1940)**
    b. Age and college degree (2010-2020)

Due to discrimination, women were less likely than men to have a college degree in the 1920-1940 period. That relationship in terms of likelihood of attending college is actually the opposite today.

4. Which kinds of models can have exposure to confounding factors?

    a. **Those used for passive prediction**
    b. Those used for active prediction
    c. Both of the above
    d. None of the above

We have no concerns about confounding factors when building a model for passive prediction. Confounding factors "mess up" our ability to estimate a causal effect of X on Y, but in a model to be used for passive prediction we have no need to estimate the causal impact of X on Y.

5. In the following model, the estimate of b_1 tells us how Y moves when X_1 increases by 1 unit:

$$Y = b_0 + b_1X_1 + b_2X_2 + U$$

    a. True
    b. False

Our estimate of b_1 tells us how Y moves when X_1 increases by 1 unit, holding X_2 constant/fixed.

6. Any prediction having to do with price as an X variable will always be an active prediction.
    a. True
    b. False

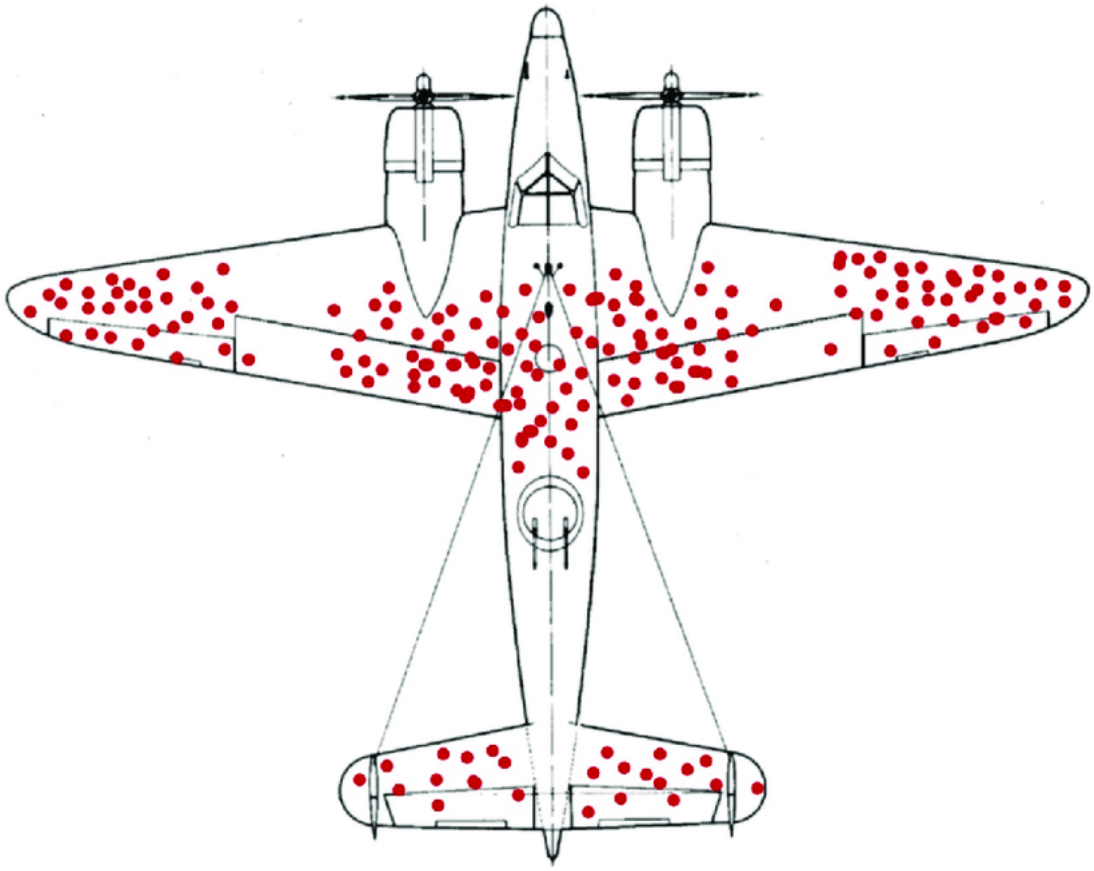
There are plenty of times where companies/suppliers/etc. are observing what their own managers do and/or what competitors are doing and simply want to passively predict what makes sense for them to have in terms of supply (or some other factor). It is true though that price can be a variable that could be used in active prediction if they were considering actively changing the price of a good or service they offered.

7. In an RCT, if we regress the outcome (Y) on the treatment (X), X will be exogenous.
    a. True
    b. False

True, the treatment will not be correlated with anything else.

8. Suppose Kroger has let the regional managers for each region in the U.S. set their own marketing budget for television ads. Which of the following is most likely to be a confounding factor in the impact of advertising on sales (assume you currently have no data on anything other than sales and TV ad spending) in the U.S.?
    a. Number of national sporting events occurring
    b. Population in region
    c. Both of the above
    d. None of the above

This is a challenging question – the key here is that a confounding factor must be correlated with X. Since X is regional advertising costs it would likely be correlated with something at the regional level (not to say that it wouldn't be correlated with anything else).



9. Use this image of a plane (as discussed in class) and where bullet holes were distributed to answer this question. "If we increase the armor over the fuselage, fewer planes will be shot down" is an example of a _____ prediction.
    a. active
    b. passive

    This question is also difficult and honestly a bit of a trick question (hence why it does not appear in the actual quiz). The key here is that although it would be unwise to use this information to make an active prediction (just like it is not advised to "go with

your gut feeling" rather than data when available), the statement is an active prediction.

10. Randomizing the assignment of the treatment ensures that the estimated coefficient will be a good estimate of the average treatment effect (the average of the causal effect for the entire population).
    a. True
    b. False

Again, a question that is harder than what would be on the actual quiz – We would need to also have a random sample of the population (or at worst a "representative" sample) to be able to extrapolate the results out to the entire population (which is what the ATE is).