**EPI289 Spring 2024**

**Homework 5**

**TOTAL [34 points]**

**This homework is due on Wednesday, February 28, 2024 at 9:45 AM Eastern Time.**

**For multiple choice questions below, select the one best choice from the alternatives presented. For questions that may have multiple correct answers, the question will ask you to select all that apply.**

**For numerical answer questions, provide the number only (no letters, symbols, or other text) and keep in mind the units and number of decimal places requested in the question, if applicable. For short answer questions, please type your answers in the space provided.**

**You will need to use the NHEFS dataset to complete this homework.**

**PART 1 [10 points]**

**Suppose you conduct the study depicted in slide #33 of “Instrumental variable estimation (I).” You want to estimate the effect of type of chemotherapy A on 5-year mortality Y in the entire population, that is, E[Ya=1]-E[Ya=0]. You have collected data on A, Y, and the proposed instrument Z, but not on confounders U. The variable Z will be an instrument if it meets instrumental conditions (i), (ii), and (iii) that are described in slide #10.**

1. **[Multiple choice, 1 point] Suppose it is known that physicians who prefer a certain type of chemotherapy are also more likely to treat patients with adjuvant radiation therapy, which is known to lower 5-year mortality risk. Using only the information provided in this scenario, which of the following conditions, if any, is violated?**
   1. Instrumental condition 1
   2. Instrumental condition 2
   3. Instrumental condition 3
   4. Homogeneity
   5. Monotonicity
   6. **None of the conditions listed above are violated**
2. **[Essay question, 1 point] Explain your reasoning for Question 1. Please limit your response to one sentence.**

Physician preference is associated with treatment assignment, physician preference only affects the outcome of 5yr mortality risk through the treatment and there are no confounding on physician preference based on the scenario above.

1. **[Multiple choice, 1 point] Suppose it is known that younger patients are more likely to receive a certain type of chemotherapy, and younger patients also have a lower risk of 5-year mortality. Using only the information provided in this scenario, which of the following conditions, if any, is violated?**
   1. Instrumental condition 1
   2. Instrumental condition 2
   3. Instrumental condition 3
   4. Homogeneity
   5. Monotonicity
   6. **None of the conditions listed above are violated**
2. **[Essay question, 1 point] Explain your reasoning for Question 3. Please limit your response to one sentence.**

Younger patients would have an arrow into A and Y, meaning no confounder on  and it does not open any backdoor paths, thereby still satisfying the conditions.

1. **[Multiple choice, 1 point] Suppose patients cared for by physicians with more years of experience have a lower 5-year mortality risk, and physicians with more years of experience are more likely to prefer a certain type of chemotherapy. Using only the information provided in this scenario, which of the following conditions, if any, is violated?**
   1. Instrumental condition 1
   2. **Instrumental condition 2**
   3. Instrumental condition 3
   4. Homogeneity
   5. Monotonicity
   6. None of the conditions listed above are violated
2. **[Essay question, 1 point] Explain your reasoning for Question 5. Please limit your response to one sentence.**

Physician preference for chemotherapy treatment would have an arrow from physician experience into treatment and into 5-year mortality risk, thereby opening a backdoor path from physician preference to 5yr mortality risk and violating the condition that physician preference only affect 5yr mortality risk through treatment.

1. **[Multiple choice, 1 point] Suppose patients with a recent history of heart attack will never receive a certain type of chemotherapy. Using only the information provided in this scenario, which of the following conditions, if any, is violated?**
   1. **Instrumental condition 1**
   2. Instrumental condition 2
   3. Instrumental condition 3
   4. Homogeneity
   5. Monotonicity
   6. None of the conditions listed above are violated
2. **[Essay question, 1 point] Explain your reasoning for Question 7. Please limit your response to one sentence.**

If patients with a history of heart attack will never receive chemotherapy, physician preference on chemotherapy treatments are irrelevant and no longer associated with treatment A.

1. **[Multiple choice, 1 point] Suppose physicians are less likely to prescribe their preferred type of chemotherapy to patients with a history of liver disease, and a certain type of chemotherapy is more effective among patients with a history of liver disease. Using only the information provided in this scenario, which of the following conditions, if any, is violated?**
   1. Instrumental condition 1
   2. Instrumental condition 2
   3. **Instrumental condition 3**
   4. Homogeneity
   5. Monotonicity
   6. None of the conditions listed above are violated
2. **[Essay question, 1 point] Explain your reasoning for Question 9. Please limit your response to one sentence.**

Physician preference (Z) for chemotherapy treatment is confounded by history of liver disease and 5-year mortality risk is also affected by patients with a history of liver disease as chemotherapy may be more effective in improving their outcome, thereby violating the condition that Z has no shared common cause with Y.

**PART 2 [9.5 points]**

**Suppose you conduct the study depicted in slide #36 of “Instrumental variable estimation (I).” You want to estimate the effect of type of invasive procedures A on mortality Y in the entire population, that is, E[Ya=1]-E[Ya=0]. You have collected data on A, Y, and the proposed instrument Z, but not on confounders U. The variable Z will be an instrument if it meets instrumental conditions (i), (ii), and (iii) that are described in slide #10.**

1. **[Multiple choice, 0.5 point] Can you empirically verify (that is, using your data) the first instrumental condition?**
   1. **Yes**
   2. No
2. **[Essay question, 0.5 point] If you selected yes for Question 11, explain how you would empirically verify the first condition. If you selected no for Question 11, explain why you cannot verify this condition. Please limit your response to one sentence.**

To empirically verify the first condition, I could evaluate if the mean of invasive procedures varies across Z and see if Z and A are associated by calculating the risk difference.

1. **[Multiple choice, 0.5 point] Can you empirically verify (that is, using your data) the second instrumental condition?**
   1. Yes
   2. **No**
2. **[Essay question, 0.5 point] If you selected yes for Question 13, explain how you would empirically verify the second condition. If you selected no for Question 13, explain why you cannot verify this condition. Please limit your response to one sentence.**

If Z only had a path through A (invasive procedures) to Y (mortality), we would see an association between Z and Y. If condition 2 was violated, and there was a direct path from Z to Y, we would again see an association. Since there is an association observed under both scenarios, theres no way to empirically verify this condition.

1. **[Multiple choice, 0.5 point] Can you empirically verify (that is, using your data) the third instrumental condition?**
   1. Yes
   2. **No**
2. **[Essay question, 0.5 point] If you selected yes for Question 15, explain how you would empirically verify the third condition. If you selected no for Question 15, explain why you cannot verify this condition. Please limit your response to one sentence.**

Condition 3 requires that there is no common cause, aka no confounding, which cannot be verified as there may be some form of unmeasured confounding.

1. **[Essay question, 3 points] For the instrumental condition(s) you cannot verify, if any, use your subject-matter knowledge to describe a realistic scenario in which the condition(s) would be violated. Please limit your response to two sentences per instrumental condition.**

**-skip**

1. **[Essay question, 1 point] What other condition, besides the instrumental ones, is needed to estimate the average causal effect of *A* on *Y* in the population? You only need to provide one version of this condition. Please limit your response to one sentence.**

Homogeneity condition, which stipulates that the treatment is constant for all individuals (A on Y is the same for every individual).

1. **[Essay question, 1.5 point] Can you empirically verify the condition in Question 18? If yes, explain how. If not, explain why not. Please limit your response to one sentence.**

No, you cannot empirically verify the condition in Q18 because if it would require me to estimate the individual causal effect of A on Y for each participant, which is impossible as we cannot observe both counterfactual scenarios to estimate individual effects.

1. **[Essay question, 1 point] Describe a realistic scenario in which the condition described in Question 18 would be violated? Please limit your response to four sentences.**

Unless the treatment had no effect on the population, the assumption of a constant treatment effect would most likely be violated. For example, imagine an invasive procedure results in post-operative complications for some individuals but not for others, the treatment does not have the same effect on Y as everyone in the population.

**PART 3 [14.5 points]**

**Use the NHEFS data to estimate the average causal effect of smoking cessation *A* (1: yes, 0; no) on weight gain *Y* in the compliers had everybody’s outcome been measured (C=0). Use IV estimation based on the proposed instrument “Price of pack of cigarettes in 1982 greater or equal than $1.50” *Z* (1: yes, 0: no). Assume exchangeability between the censored and uncensored conditional on age (continuous variable; linear), sex (1: female, 0: male), race (1: nonwhite, 0: white), smoking intensity (continuous; linear), asthma diagnosis in 1971 (1: yes, 0: no) and smoking cessation *A* (1: yes, 0: no). Exclude subjects with missing values for the instrument.**

**Note: use non-stabilized inverse probability weighting to adjust for selection bias due to censoring.**

1. **[Multiple choice, 1 point] Which of the following expressions represents this effect on the additive scale using counterfactual notation?**
   1. E[Ya=1 − Ya=0|C=0, Az=1=1,Az=0=0]
   2. E[Ya=1,c=0 − Ya=0,c=0]
   3. **E[Ya=1,c=0 − Ya=0,c=0|Az=1=1,Az=0=1]**
   4. E[Ya=1,c=0 − Ya=0,c=0|Az=1=0,Az=0=1]
   5. E[Ya=1 − Ya=0|C=0, A=1, Z=0]
   6. E[Ya=1,c=0 − Ya=0,c=0|Az=1=1,Az=0=0]
2. **[Fill in the blank, 3 points] Provide a point estimate for this effect. Round your answer to the nearest tenth (i.e. 1 decimal place).**

**14kg**

1. **[Essay question, 1.5 points] Describe (in plain English so that your uncle can understand) the first instrumental variable assumption. Regardless of whether or not this assumption is verifiable, discuss why or why not you think this assumption is reasonable in this application. Please limit your response to two sentences.**
2. **[Essay question, 1.5 points] Describe (in plain English so that your uncle can understand) the second instrumental variable assumption. Regardless of whether or not this assumption is verifiable, discuss why or why not you think this assumption is reasonable in this application. Please limit your response to two sentences.**
3. **[Essay question, 1.5 points] Describe (in plain English so that your uncle can understand) the third instrumental variable assumption. Regardless of whether or not this assumption is verifiable, discuss why or why not you think this assumption is reasonable in this application. Please limit your response to two sentences.**
4. **[Essay question, 3 points] In addition to the three instrumental variable assumptions, are there any additional assumptions (specific to the setting of instrumental variable estimation) that are needed to endow the point estimate with a causal interpretation? If so, describe (in plain English so that your uncle can understand) this/these assumptions and discuss why or why not you think this/these assumption(s) is/are reasonable in this application. Please limit your response to two sentences per assumption.**
5. **[Multiple choice, 1 point] Who are the compliers?**
   1. People who would quit smoking if the price of cigarettes was >$1.50 and who would not quit smoking if the price of cigarettes was <$1.50.
   2. **People who in fact quit smoking when the price of cigarettes was >$1.50 and who in fact did not quit smoking when the price of cigarettes was <$1.50**
   3. People who would quit smoking if the price of cigarettes was <$1.50 and who would not quit smoking if the price of cigarettes was >$1.50.
   4. People who in fact quit smoking when the price of cigarettes was <$1.50 and who in fact did not quit smoking when the price of cigarettes was >$1.50
6. **[Essay question, 1 point] Who are the defiers in the context of the study variables? Please limit your response to one sentence.**

People who in fact did not quit smoking when the price of cigarettes was >$1.50 and who in fact did quit smoking when the price of cigarettes was <$1.50.

1. **[Numerical answer, 1 point] What would be the proportion of compliers had everybody’s outcome been measured (C=0)? Present your answer as a decimal (rather than a percentage) and round to the nearest thousandth (i.e. 3 decimal points).**

**0.251**

1. **[Essay question, ungraded] Provide your R code for Part 3.**

nhefs.iv<- nhefs[which(!is.na(nhefs$wt82) & !is.na(nhefs$price82)),] nhefs$censor <- ifelse(is.na(nhefs$wt82\_71) & !is.na(nhefs$price82), 1, 0)

nhefs.iv$highprice <- ifelse(nhefs.iv$price82>=1.5, 1, 0)

# Estimation of denominator of censoring weights

ipw.c.d <- glm(censor ~ sex + race + age + qsmk + asthma, family=binomial(), data=nhefs)

summary(ipw.c.d)

nhefs$pd.cens <- 1-predict(ipw.c.d, nhefs, type = "response")

nhefs$w <- ifelse(nhefs$censor==1, 1/nhefs$pd.cens, 1/(1-nhefs$pd.cens))

library(sem)

model1 <- tsls(wt82\_71 ~ qsmk, ~ highprice, weights=w, data = nhefs) summary(model1)

beta <- coef(model1)

SE <-sqrt(19.511)

# robust standard errors

lcl <- beta-1.96\*SE

ucl <- beta+1.96\*SE

cbind(beta, lcl, ucl)[2,]

#Those who took the treatment when assigned

compliers <- sum(nhefs.iv$highprice == 1 & nhefs.iv$qsmk == 1)

total <- nrow(nhefs.iv)

proportion <- compliers /total