# MA592 Introduction to Causal Inference

Homework 3: Estimating the effect of a treatment

Zhiwei Liang

10/13/2021

## 1. The effect of statins on MI

**(a) Estimate the risk of Myocardial Infarction in the entire group of 300,006 men, had none of them used statins, based on the data above, using Inverse Probability of Treatment Weighting.**

$$P(Y^{(0)}=1)=E(Y^{(0)})=E[\frac{YI_{A=0}}{P(A=0|X)}]$$

$$=\frac{nI_{A=0,X=1}}{N}\times\frac{1}{P(A=0|X=1)}+\frac{nI_{A=0,X=0}}{N}\times\frac{1}{P(A=0|X=0)}$$

$$=\frac{\frac{400}{50001/(50001+50002)}+\frac{360}{180002/(180002+20001)}}{300,006}$$

So the result is 0.004

**(b) Estimate the risk of Myocardial Infarction in the entire group of 300,006 men, had they all used statins, based on the data above, using Inverse Probability of Treatment Weighting.**

$$P(Y^{(1)}=1)=E(Y^{(1)})=E[\frac{YI_{A=1}}{P(A=1|X)}]$$

$$=\frac{nI_{A=1,X=1}}{N}\times\frac{1}{P(A=1|X=1)}+\frac{nI_{A=1,X=0}}{N}\times\frac{1}{P(A=1|X=0)}$$

$$=\frac{\frac{240}{50002/(50001+50002)}+\frac{30}{20001/(180002+20001)}}{300,006}$$

So the result is 0.0025999

**(c) Estimate the average effect of statins on Myocardial Infarction in the entire group of 300,006 men, based on the data above, using (a) and (b).**

$ATE=E(Y^{(1)}-Y^{(0)})=E(Y^{(1)})-E(Y^{(0)})$ = -0.0014001

**(d) Compare your results with the results you obtained in Questions 1(a), (b), and (c) of Homework 2. Are your answers the same? Can you explain this?**

**In HW2:** the results are 0.004, 0.0025999, -0.0014001

**In HW3:** the results are 0.003999952, 0.002599897, -0.001400055

The answer are the same (just keep different decimal).

**(e) Compare your results with the results you obtained in Question 9 of Homework 1. Can you say a few words about bias, and in this case, provide the numeric value of the bias? For now, just ignore random variation.**

**In HW1(9):** $Diff = P(MI|statins) - P(MI|nostatins) = 5.5267283 \times 10^{-4}$

The result indicates that the risk with statins is 0.06% higher than without statins.

**In 3(c):** The result is -0.0014001 in the entire group of men, and it indicates that the MI risk with statins is 0.14% lower than without statins.

The conclusions in HW1 and HW3 are totally contrast, the former indicates statins can prevent MI while the latter indicates the statins cannot. The reason of the bias is the given condition of X. In HW1, the probabilities of different risk status were ignored, we just calculate the $E(Y^{(i)}|A = i)$; In HW2, considering $P(X)$, we calculate $P(A = i|X)$.

**Bias** = HW1(9) - HW3(e) = $5.5267283 * 10^{-4} - (-0.0014) = 0.002$

**(f) Show that for this particular case, if we have No Unmeasured Confounding when including this cardiovascular risk variable, that your Inverse Probability of Treatment Weighting estimator leads to unbiased estimating equations. What are the benefits of such a result?**

IPTW eliminates bias by weighting the observed responses. So the observed responses can represent not only themselves but also the counterfactual(potential) responses in the same group(with the same confounder).

For example, when estimating the $E(Y^{(1)})$ in low risk group, men with statins are used to represent the men without statins in the group.

## 2.The effect of antibiotic medications on hospital death

The **notation** used in this question:

- $A$: treatment=1 for caz-avi, treatment=0 for colistin
- $X_1$: creatininehigh = 1 for patients with high creatinine levels, 0 otherwise
- $X_2$: infectiontype = 1 for bloodstream infections, 2 for urinary tract infections, and 3 for other types of infection
- $X_3$: pitt score is a measure of disease severity
- $Y$: hospitaldeath=1 for patients who died in the hospital and hospitaldeath=0 otherwis

```
mydata = readxl::read_xlsx("hw2_1000.xlsx")

colnames(mydata)[2] = "x3"
colnames(mydata)[3] = "x2"
colnames(mydata)[4] = "x1"
colnames(mydata)[5] = "a"
colnames(mydata)[6] = "y"

# make the Infection type as factor
mydata$x2 = as.factor(mydata$x2)
```

**(a) Estimate the probability of hospital death in the entire simulated population, had everyone been treated with caz-avi, using a conditioning argument. Please include all potential confounders in your analysis; "pre-testing" is often not recommended, we often try to decide which confounders to include based on subject matter knowledge.**

$$P(Y^{(1)} = 1) = E(Y^{(1)}) = E[\frac{nI_{A=1}}{P(A = 1|X)}]$$

$$= \frac{\sum_{i_1=0}^{1}\sum_{i_2=1}^{3}\sum_{i_3=0}^{1}\frac{nI_{A=1,X_1=i_1,X_2=i_2,X_3=i_3}}{P(A=1|X_1=i_1,X_2=i_2,X_3=i_3)}}{N}$$

```
library(dplyr)

# function to calculate P(X1,X2,X3)*E(Y^1|X1,X2,X3,A=1)
IPTW_1 = function(i1, i2, i3) {
    n = 1000
    n123 = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3))
    n123a = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3, a == 1))
    n123ay = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3, a == 1, y == 1))
    inverse_prob = n123a/n123
    p = n123ay/(inverse_prob * n)
    return(p)
}

# iteration to sum all the combinations of X1-X3
p1 = 0
for (i1 in 0:1) {
    for (i2 in 1:3) {
        for (i3 in 0:1) {
            p1 = p1 + sum(IPTW_1(i1, i2, i3))
        }
    }
}
p1
## [1] 0.08729967
```

The probability is 0.0872997.

**(b) Estimate the probability of hospital death in the entire simulated population, had everyone been treated with colistin, using a conditioning argument.**

$$P(Y^{(0)}=1) = E(Y^{(0)}) = E[\frac{YI_{A=0}}{P(A=0|X)}]$$

$$= \frac{\sum_{i_1=0}^{1}\sum_{i_2=1}^{3}\sum_{i_3=0}^{1}\frac{YI_{A=1}}{P(A=0|X_1=i_1,X_2=i_2,X_3=i_3)}}{N}$$

```
library(dplyr)

# function to calculate P(X1,X2,X3)*E(Y^1|X1,X2,X3,A=1)
IPTW_0 = function(i1, i2, i3) {
    n = 1000
    n123 = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3))
    n123a = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3, a == 0))
```

```
    n123ay = nrow(mydata %>%
        filter(x1 == i1, x2 == i2, x3 == i3, a == 0, y == 1))
    inverse_prob = n123a/n123
    p = n123ay/(inverse_prob * n)
    return(p)
}

# iteration to sum all the combinations of X1-X3
p0 = 0
for (i1 in 0:1) {
    for (i2 in 1:3) {
        for (i3 in 0:1) {
            p0 = p0 + sum(IPTW_0(i1, i2, i3))
        }
    }
}
p0
## [1] 0.35744
```

The probability is 0.35744.

**(c) What are the assumptions you made in 2(a) and 2(b)? Mention all assumptions, including potential modeling assumptions.**

1. $(Y^{(0)}, Y^{(1)})$ *independent of* $A$, *given* $X1, X2, X3$
2. Consistency

- If $A = 0$, we observe $Y = Y^{(0)}$.
- If $A = 1$, we observe $Y = Y^{(1)}$.

**(d) Estimate the average effect of caz-avi, as compared to colistin, on hospital death, in the entire simulated population.**

$ATE = E(Y^{(1)} - Y^{(0)}) = E(Y^{(1)}) - E(Y^{(0)})$ = -0.2701403

**(e) Compare your results from the conditioning arguments in Homework 2 with the results from Inverse Probability of Treatment Weighting you found here. What do you see? Can you explain why this is happening?**

The results are the same in HW2 and HW3.

**(f) Show that for this particular case, if we have No Unmeasured Confounding when including this cardiovascular risk variable, that your Inverse Probability of Treatment Weighting estimator leads to unbiased estimating equations. What are the benefits of such a result?**

IPTW eliminates bias by weighting the observed responses. So the observed responses can represent not only themselves but also the counterfactual(potential) responses in the same group(with the same confounder).

For example, when estimating the risk of death(had everyone been treated with caz-avi), in high creatinine levels/bloodstream infections/severer disease group, patients with caz-avi are used to represent the patients with colistin in this group.