המחלקה למנהל עסקים
האוניברסיטה העברית

אקונומטריקה למדע המידע
אלישבע צור (שורץ)

<u>תרגיל 7</u>
רגרסיה ליניארית ומשתנה עזר
יש לצרף את הקוד ששימש לפתרון
תאריך הגשה: 16 בדצמבר 2021
בהצלחה!

לשאלות 1ו-2 השתמשו בקובץ הנתונים wages, ייבאו את הנתונים ועשו בהם תיקון קטן:

```
data=read.csv("path/wages.csv",header=TRUE)
data$wage=data$ï..wage
```

1. . השמטת משתנה

השתמשו בקובץ הנתונים wages כדי לענות על שאלה זו. כמו כן, שימו לב שכל הרגרסיות הבאות כוללות חותך. בשאלה זו אנו רוצים לבחון את האפקט הסיבתי של השכלה על שכר. לשם כך נבחן את שתי הרגרסיות הבאות:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \epsilon$$

$$\log(\text{wage}) = \alpha_0 + \alpha_1 \text{educ} + \alpha_2 \text{IQ} + u$$

א. מצאו את האומדנים עבור β_1 α_1 ו-α_2

ב. על מי מבין המקדמים של educ - β_1 או α_1 הייתם "סומכים" יותר? הסבירו מדוע

ג. כתבו את הקשר המתמטי בין - β_1 ו α_1 והראו שהוא מתקיים

על מנת לכתוב את הקשר המתמטי - היעזרו גם ברגרסיה הבאה: $\text{IQ} = \pi_0 + \pi_1 \text{educ} + \epsilon$

2.משתנה עזר ושיטת 2SLS

מטרת תרגיל זה היא להשוות את האומדים וסטיות התקן המתקבלים משימוש נכון ב-2SLS לעומת אלה המתקבלים משימוש בפרוצדורות לא נכונות. השתמשו בקובץ הנתונים wages.

א. השתמשו בפרוצדורת 2SLS של R כדי לאמוד את המשוואה הבאה:

$$\log(\text{wage}) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + \beta_3 \text{tenure} + \beta_4 \text{black} + u$$

כאשר sibs (מספר האחים) הוא משתנה עזר ל – educ

הסבירו מדוע נדרש משתנה עזר עבור המשתנה educ וכן **מדוע** sibs יכול לשמש כמשתנה עזר עבורו

- על מנת להשתמש בפרוצדורת 2SLS של R עליכם קודם כל להתקין את החבילה המתאימה
  ```
  install.packages("ivreg", dependencies = TRUE)
  ```
  (זה יכול לקחת כמה דקות...)
- השתמשו בפקודה הבאה:

model1 <-ivreg::ivreg($y \sim x_1 + x_2 + x_3 + x_4 \mid z_1 + z_2 + z_3 + z_4$,data = data)

כאשר y הוא המשתנה התלוי

$x_1 - x_4$ הם המשתנים הבלתי תלויים

$z_1 - z_4$ הם המשתנים הבלתי תלויים **האקסוגניים** ומשתנה העזר

- השתמשו גם בפקודה summary על מנת לראות את סטיות התקן של האומדנים שלכם למקדמי הרגרסיה
- הגישו צילום של הפלט

ב. כעת בצעו את 2SLS באופן ידני. כלומר, אמדו רגרסיה של educ על sibs, exper, tenure ו-black וחשבו את התחזית $\widehat{educ_i}$ $(i = 1,2, \dots, n)$. אז אמדו את רגרסיית השלב השני log(wage) על $\widehat{educ}$, exper, tenure ו-black. בדקו שה-$\hat{\beta}_j$ המתקבלים זהים לאלה מחלק (1), אולם כי יש הבדל קטן בסטיות התקן. סטיות התקן המתקבלות מפרוצדורה ידנית של 2SLS אינן נכונות באופן כללי.

- בשלב הראשון אמדו את הרגרסיה הבאה:

step1=lm(x_1~ $x_2 + x_3 + x_4 + x_5$ data = data)

כאשר x_1 הוא המשתנה האנדוגני ו- $x_2 - x_5$ הם משתנה העזר ושאר המשתנים **האקסוגניים**

- כחלק מהשלב הראשון חשבו את $\widehat{educ}$ על ידי הפקודה

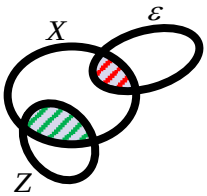
data$predict_educ=predict(step1)

- בשלב השני אמדו את הרגרסיה הרצויה תוך שימוש בתוצרים של השלב הראשון (כמו שלמדנו בכיתה)

step2=lm(y~x_1+x_2+x_3+x_4, data = data)

- השתמשו גם בפקודה summary על מנת לראות את סטיות התקן של האומדנים שלכם למקדמי הרגרסיה
- הגישו צילום של הפלט, שימו לב שאכן אומדני המקדמים יצאו או דבר בשתי השיטות, אך סטיות התקן מעט שונות

ג. בהתבסס על האילוסטרציה הבאה: (שראיתם בכיתה)



השלימו את שני התנאים שמשתנה העזר צריך לקיים, התייחסו גם לעוצמת הקשר שהייתם רוצים

3. **בניית מודל**

החוקר משהשאלה הקודמת רוצה לבנות מודל שיסביר את השכר על ידי השכלה בצורה הטובה ביותר. הוא חושב שאת התשואה להשכלה יש למדוד באחוזים – כלומר, יש למצוא את אחוז התשואה בשכר לכל שנת לימוד, בנוסף, הוא מאמין כי מסיבות שאינן קשורות לרמת השכלה – נשים מרוויחות פחות מגברים, והתשואה שלהן להשכלה – האחוז שבו יעלה שכרן לכל שנת לימוד – שונה משל גברים גם כן. כמו כן הוא מאמין כי קשרים חברתיים משפיעים גם הם על ההשכלה שאדם רוכש – אדם שיש לו קשרים חברתיים טובים ומגוונים – מכיר יותר מקומות ואפשרויות ומגלה יותר סקרנות, מצד שני – "**יותר מידי**" קשרים חברתיים מונעים ממנו להתרכז במטרה. לבסוף, הוא רוצה לפקח גם על "הכישרון" של אדם – אותו הוא מקווה לייצג על ידי ציון הIQ שלו

כתבו את המודל שהחוקר רוצה לאמוד על ידי שימוש במשתנים הבאים

Wage- שכר של פרט במדגם

Female – משתנה דמי ששווה 1 אם הפרט במדגם הוא אישה ו0 אחרת

Educ- מספר השנים שהפרט למד

IQ- ציון הIQ של הפרט

Friends - מספר החברים הטובים של הפרט