

תרגיל 3

הנחיות להגשה: יש להגיש קובץ קוד (do file / script), קובץ תוצאות (באמצעות log או sink) וקובץ נוסף (word או PDF) עם תשובות לשאלות הפתוחות והגרפים הנדרשים.

1. נניח שאנו מעוניינים לאמוד את ההשפעה של מספר שעות הלמידה שהתלמיד משקיע בהכנה למבחן (*hours*) על הציון במבחן (*score*).
 - a. כתבו מודל שמתאר זאת (באוכלוסייה).
 - b. אחת מהנחות המודל הפשוט היא שמתקיים $E(u|x) = 0$. ציינו לפחות שני גורמים שעשויים להיכלל בתוך ה- u . האם סביר שהגורמים האלו יהיו מתואמים עם *hours*?
 - c. בהתבסס על תשובתכם לסעיף הקודם, האם ניתן להסיק כי האפקט הסיבתי הוא אפס? האם אומד OLS יהיה מוטה?

2. בשאלה זו תתבקשו להראות איך משפיעים שינויים ביחידות המדידה על האומדים של ה- OLS, על סטיות התקן של האומדים ועל ה- R^2 . נתונים $\hat{\beta}_0, \hat{\beta}_1$ - אומדי השיפוע והחותך (בהתאמה) של רגרסיה של y על x :
$$y = \hat{\beta}_0 + \hat{\beta}_1 x$$
כמו כן נתונים שני קבועים c_1, c_2 ($c_2 \neq 0$). כעת נאמדת הרגרסיה:

$$c_1 y = \tilde{\beta}_0 + \tilde{\beta}_1 c_2 x$$

הראו בחישוב (ללא שימוש בתוכנות) מהם האומדים של $\tilde{\beta}_0, \tilde{\beta}_1$ סטיות התקן שלהם ו- R^2 של המודל החדש כפונקציה של $\hat{\beta}_0, \hat{\beta}_1, c_1, c_2$.

רמז: הציבו בנוסחת האומדים את הערכים החדשים במקום x ו- y .

האומדים הם:

$$\hat{\beta}_1 = \frac{\sum_1^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_1^n (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

3. לשאלה זו יש להשתמש בנתונים שבקובץ hprices.csv בתיקיית קבצי הנתונים.

קובץ זה מכיל נתונים על דירות כגון מחיר, שטח ועוד.

א. צרו משתנה חדש בשם centered_price ששווה למחיר הדירה פחות ממוצע

מחירי הדירות.

צרו משתנה נוסף בשם centered_sqrft באותו אופן עבור sqrft.

ב. אמדו את המודלים הבאים:

$$(1) \text{ price} = \beta_0 + \beta_1 \text{sqrft} + u$$

$$(2) \text{ centered_price} = \widetilde{\beta}_0 + \widetilde{\beta}_1 \text{centered_sqrft} + u$$

מהם האומדים להשפעת השטח על המחיר בשתי הרגרסיות ומהן סטיות התקן של

האומדים, מהו החותך ומהו ה- R^2 ?

ג. צרו גרף פיזור (scatter) של כל מודל ונסו להסביר את הקשר בין התוצאות

בסעיף הקודם. האם השפעת גודל הדירה על המחיר גדול יותר במודל (1) או

במודל (2)?

ד. צרו משתנה חדש בשם lprice2 ששווה ל:

$$\text{lprice2} = \log(\text{price} * 1000)$$

אמדו את המודלים הבאים:

$$(3) \text{ lprice} = \beta_0 + \beta_1 \text{sqrft} + u$$

$$(4) \text{ lprice2} = \widetilde{\beta}_0 + \widetilde{\beta}_1 \text{sqrft} + u$$

וחזרו על סעיף ג' עבור התוצאות החדשות.