

**מבוא לאקונומטריקה א'**

**החוג לכלכלה וניהול**

## נושאי הקורס:

|          |                                                        |
|----------|--------------------------------------------------------|
| 2.....   | הקדמה: תזכורת של סטטיסטיקה ומתמטיקה                    |
| 7.....   | פרק 1 - מה זו אקונומטריקה?                             |
| 12.....  | פרק 2 - אומדי הריבועים הפחותים (אר"פ) וההנחות הקלאסיות |
| 27.....  | פרק 3 - מודלים לא ליניאריים                            |
| 36.....  | פרק 4 - רגרסיה מרובה ומולטיקוליניאריות                 |
| 44.....  | פרק 5 - מבחן t                                         |
| 56.....  | פרק 6 - $R^2$ - מדד לטיב הרגרסיה ומבחן F               |
| 66.....  | פרק 7 - שינוי יחידות מדידה                             |
| 68.....  | פרק 8 - מבחן לדוגמא                                    |
| 72.....  | פרק 9 - מבחני ספציפיקציה- מבחן LM (כופלי לגרנג')       |
| 77.....  | פרק 10 - משתני דמי                                     |
| 99.....  | פרק 11 - הטרוסקדסטיות                                  |
| 112..... | פרק 12 - מתאם סידרתי                                   |
| 128..... | פרק 13 - סיכום בעיית המתאם הסדרתי והטרוסקדסטיות        |

**גול זה בול. בשבילך!**

## הקדמה: תזכורת של סטטיסטיקה ומתמטיקה

### הגדרות וסימונים

יש להבחין בין שני סוגים של משתנים: משתנים אמפיריים ("רגילים") לעומת משתנים מקריים.

משתנה אמפירי - משתנה שתוצאותיו ידועות מראש (כמו למשל המשתנה רמת הכנסה, גיל, מס' שנות לימוד במדגם מסוים).

משתנה מקרי - משתנה שתוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קוביה או בהטלת מטבע)

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (כמו למשל  $X_i$  או  $Y_i$ )

בנוסף לכך ישנם גם קבועים - המקבלים ערך קבוע ומסומנים באות לועזית ללא אינדקס (כמו למשל  $a$  או  $b$ )

באקונומטריקה נעסוק בעיקר במשתנים מקריים.

לכל משתנה מקרי  $X_i$  יש תוחלת המייצגת את מרכז ההתפלגות.

התוחלת מסומנת  $\mu_x$  או  $E(X)$ .

השונות מייצגת את מידת הפיזור של ההתפלגות.

השונות מסומנת  $\sigma_x^2$  או  $V(X)$ .

סטית התקן היא השורש של השונות והיא מסומנת  $\sigma_x$ .

לשני משתנים מקריים  $X$  ו- $Y$  יש שונות משותפת (covariance) המהווה מדד

להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם (יחס ישר או יחס הפוך).

השונות המשותפת מסומנת  $\text{cov}(x,y)$

כאשר:

$$Y, X \Leftrightarrow \text{Cov}(X, Y) = 0 \text{ בלתי מתואמים}$$

$$\text{Cov}(X, Y) > 0 \Leftrightarrow \text{מתאם חיובי בין המשתנים}$$

$$\text{Cov}(X, Y) < 0 \Leftrightarrow \text{מתאם שלילי בין המשתנים}$$

$$X, Y \Leftarrow \text{בלתי תלויים} \quad X, Y \Leftarrow \text{בלתי מתואמים}$$

מקדם מתאם של פירסון - מדד לכיוון ועוצמת הקשר הלניארי בין שני משתנים:

$\eta_{xy}$  מקדם המתאם מסומן

$$\eta_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$-1 \leq \eta \leq 1$$

כאשר:

$$\eta = 1 \text{ מתאם ליניארי חיובי מלא בין שני המשתנים}$$

$$\eta = -1 \text{ מתאם ליניארי שלילי מלא בין שני המשתנים}$$

$$\eta = 0 \text{ לא קיים מתאם ליניארי בין שני המשתנים}$$

### אמידה

פרמטר- ערך המשתנה הנחקר המתאר את כל האוכלוסיה

סטטיסטי/אומד- ערך המשתנה הנחקר המתאר את המדגם

| מדגם                                                                       | אוכלוסיה                                   |
|----------------------------------------------------------------------------|--------------------------------------------|
| $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$                                     | $E(X) = \mu$                               |
| $S_X^2 = \frac{S_{XX}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$      | $V(X) = \sigma^2 = E(X - E(X))^2$          |
| $\frac{S_{XY}}{n} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$ | $\text{cov}(X, Y) = E(X - E(X))(Y - E(Y))$ |

|                                                       |                                                               |
|-------------------------------------------------------|---------------------------------------------------------------|
| $r_{XY} = \frac{S_{XY}}{\sqrt{S_{XX}} \sqrt{S_{YY}}}$ | $\eta_{XY} = \frac{\text{cov}(X,Y)}{\sqrt{V(X)} \sqrt{V(Y)}}$ |
|-------------------------------------------------------|---------------------------------------------------------------|

### נוסחאות וחוקים בסטטיסטיקה

יהיו X ו-Y משתנים מקריים, ו-a, b קבועים:

**חוקי הסיגמה**

$$\sum_{t=1}^T X_t = X_1 + X_2 + X_3 + \dots + X_T \quad (1)$$

**(2) סכום של קבוע:**

$$\sum_{t=1}^T a = Ta$$

**(3) סכום של קבוע כפול משתנה = לקבוע כפול הסכום:**

$$\sum_{t=1}^T aX_t = a \sum_{t=1}^T X_t$$

**(4) סכום של סכום/הפרש = לסכום/הפרש הסכומים:**

$$\sum_{t=1}^T (X_t \pm Y_t) = \sum_{t=1}^T X_t \pm \sum_{t=1}^T Y_t$$

**(5) יש לשים לב כי:**

$$\sum_{t=1}^T X_t^2 \neq \left( \sum_{t=1}^T X_t \right)^2$$

$$\sum_{t=1}^T X_t Y_t \neq \sum_{t=1}^T X_t \sum_{t=1}^T Y_t$$

**הגדרות ופיתוחים**

**(1) סכום הסטיות מהממוצע = 0:**

$$\sum_{t=1}^T (X_t - \bar{X}) = 0$$

**(2) סכום הסטיות הריבועיות מהממוצע (מונה השונות):**

$$S_{XX} = \sum_{t=1}^T (X_t - \bar{X})^2 = \sum_{t=1}^T X_t^2 - T\bar{X}^2 = \sum_{t=1}^T (X_t - \bar{X})X_t$$

**(3) מונה של השונות המשותפת:**

$$S_{XY} = \sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y}) = \sum_{t=1}^T X_t Y_t - T\bar{X}\bar{Y} = \sum_{t=1}^T (X_t - \bar{X})Y_t = \sum_{t=1}^T (Y_t - \bar{Y})X_t$$

**חוקי התוחלת**

**(1) תוחלת של קבוע = קבוע:**

$$E(a) = a$$

**(2) תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:**

$$E(X \pm Y) = E(X) \pm E(Y)$$

$$E(\sum (X_i)) = \sum E(X_i)$$

**(3) תוחלת של כפל/חילוק  $\neq$  לכפל/חילוק התוחלות:**

$$E(X \cdot Y) \neq E(X) \cdot E(Y)$$

$$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

$$E(X^2) \neq [E(X)]^2$$

**(4) השפעת טרנספורמציה ליניארית על התוחלת:**

$$E(a + \frac{1}{a}X \pm b) = a + \frac{1}{a} \cdot E(X) \pm b$$

**חוקי השונות**

**(1) עבור X ו-Y בלתי תלויים/בלתי מתואמים מתקיים:**

**שונות של סכום/הפרש = סכום/הפרש השונות**

$$V(X \pm Y) = V(X) \pm V(Y)$$

$$V(\sum (X_i)) = \sum V(X_i)$$

**(2) עבור X ו-Y תלויים/מתואמים מתקיים:**

**שונות של סכום/הפרש  $\neq$  סכום/הפרש השונות**

$$V(X \pm Y) = V(X) + V(Y) \pm 2 \cdot Cov(X, Y)$$

**(3) שונות של קבוע = 0:**

$$V(a) = 0$$

$$V(a \pm x) = V(X)$$

**(4) השפעת טרנספורמציה ליניארית על השונות:**

$$V(aX + b) = a^2 V(X)$$

הערה חשובה:

חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).  
 חוקי הסכום מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

## חוקי השונות המשותפת

(1) שונות משותפת בין משתנה לקבוע  $= 0$ :

$$\text{cov}(X, a) = 0$$

(2) שונות משותפת של משתנים המוכפלים בקבוע:

$$\text{cov}(aX, bY) = ab \cdot \text{cov}(X, Y)$$

(3) שונות משותפת של משתנה עם עצמו = שונות המשתנה:

$$\text{cov}(X, X) = V(X)$$

$$\text{cov}(Y, Y) = V(Y)$$

## פרק 1 - מה זו אקונומטריקה?

אקונומטריקה היא שיטת מחקר שבאמצעותה אנחנו מוצאים קשר בין משתנים. למשל, אם נדע את הקשר בין שער הריבית לבין שער הדולר, נוכל לדעת איך שער הריבית משפיע על שער הדולר, ונוכל להשתמש בזה לתחזיות כלכליות.

במציאות קיימים חוקים הקושרים בין משתנים. חוקים אלה אינם ידועים לנו, אבל אנו יכולים לראות התוצאות שנובעות מהחוקים האלה. אנו משתמשים בתוצאות אלו כדי לשחזר את החוקים. השיחזור נקרא רגרסיה.

נתחיל בדוגמא:

מתווך דירות בתל אביב רצה לבדוק איך משפיע גודלה של דירה על המחיר שבו היא נמכרת.

הוא הניח 2 הנחות מקדימות:

- (1) רק גודל הדירה משפיע על מחיר הדירה באופן שיטתי. כל שאר הדברים המשפיעים על מחיר הדירה הם אקראיים ולא ניתנים לחיזוי.
- (2) ההשפעה של גודל הדירה על מחיר הדירה היא ליניארית.

שתי ההנחות האלה מאפיינות את הקשר. אם נסמן את גודל הדירה ב-  $X$  ואת מחיר הדירה ב-  $Y$ , נוכל לכתוב באופן מתמטי כי  $Y = \alpha + \beta X + u$ . זהו המודל של המתווך. ההנחות של המתווך נקראות הספציפיקציה של המודל.  $X$  ו-  $Y$  הם המשתנים של המודל.  $Y$  הוא המשתנה המוסבר של המודל.  $X$  הוא המשתנה המסביר של המודל (יכול להיות יותר ממשתנה מסביר אחד).  $\alpha$  ו-  $\beta$  הם הפרמטרים של המודל.  $\alpha$  נקרא חותך.  $\beta$ , או כל מקדם אחר של משתנה מסביר, נקרא שיפוע.  $u$  מכונה הפרעה האקראית (לעיתים מכונה בבדיחות הדעת קריזה).

**מודל:**

$$Y = \alpha + \beta X + u$$

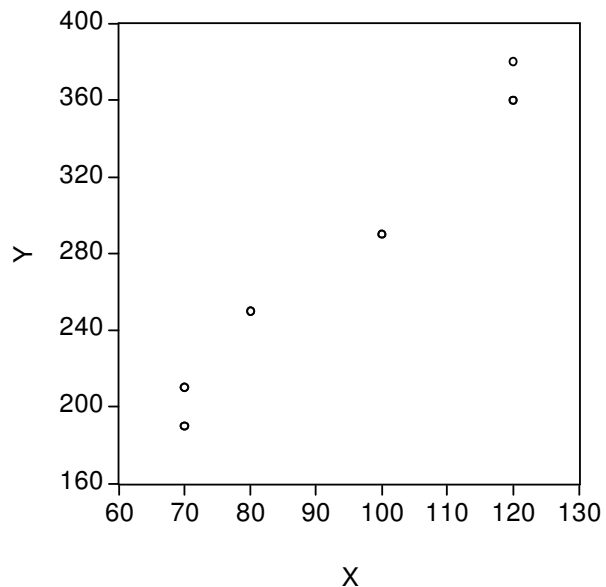
|                |                             |                |                         |
|----------------|-----------------------------|----------------|-------------------------|
| משתנה<br>מוסבר | פרמטרים:<br>$\alpha$ - חותך | משתנה<br>מסביר | הפרעה אקראית<br>"קריזה" |
|                | $\beta$ - שיפוע             |                |                         |

אחרי הגדרת המודל המתווך אסף נתונים על 6 דירות, שנמכרו בחודש האחרון באותו איזור.

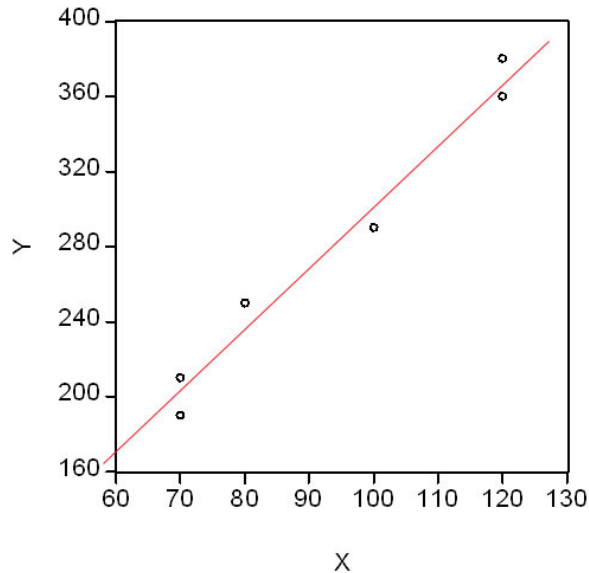
זהו המדגם של המתווך. במדגם יש 6 תצפיות. נוהגים להציג את המודל כאשר לכל משתנה נוסף אינדקס  $Y_t = \alpha + \beta X_t + u_t$ . האינדקס מייצג את מספר התצפית.

| מספר הדירה | גודל הדירה במ"ר | מחיר הדירה באלפי דולרים |
|------------|-----------------|-------------------------|
| 1          | $X_1 = 70$      | $Y_1 = 190$             |
| 2          | $X_2 = 70$      | $Y_2 = 210$             |
| 3          | $X_3 = 80$      | $Y_3 = 250$             |
| 4          | $X_4 = 100$     | $Y_4 = 290$             |
| 5          | $X_5 = 120$     | $Y_5 = 360$             |
| 6          | $X_6 = 120$     | $Y_6 = 380$             |

נציג את 6 התצפיות בגרף:



מהו הקו הישר המתאר את הקשר בין שני המשתנים בצורה הטובה ביותר? (הקו הוא ישר בגלל שהמתווך הניח לינאריות של המודל).  
 מסתבר שקו הרגרסיה הטוב ביותר הוא קו שחושב בשיטת הריבועים הפחותים  
 (השיטה תתואר במלואה בהמשך):



הנוסחה של הקו היא:  $\hat{Y}_i = -27.32 + 3.29 X_i$ .

זהו כנראה לא הקו האמיתי, אך ממילא את הקו האמיתי אף פעם אי אפשר לדעת.  
 סביר שקו זה הוא די קרוב לקו האמיתי.

לפי הנוסחה כל מ"ר נוסף שיש בדירה מעלה את מחירה ב-3,290 דולר.  
 מקו זה יודע המתווך להעריך מחירים של דירות. כשפנה אליו בעל דירה שגודלה 90 מ"ר ושאל אותו מה שווי הדירה, חישב המתווך לפי הנוסחה,  
 $-27.32 + 3.29 \cdot 90 = 268.78$ , והשיב לבעל הדירה: "המחיר שאתה יכול לקבל עליה הוא 268,780 דולר. אם יהיה לך מזל תקבל יותר, אבל יכול להיות שתצטרך למכור בפחות".

בשפה אקונומטרית נוכל לומר כי אם יהיה לו מזל אז ההפרעה האקראית תהיה חיובית, ואם לא – היא תהיה שלילית.

**לסיכום:**

(1) במודל  $Y_t = \alpha + \beta X_t + u_t$ ,  $\alpha$  ו- $\beta$  הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).

(2)  $\hat{\alpha}$  הוא האומד ל- $\alpha$ .  $\hat{\beta}$  הוא האומד ל- $\beta$ .

(3) אומדי ריבועים פחותים (אר"פ) הם אומדים שחושבו בשיטת הריבועים הפחותים. אומדי הריבועים הפחותים מסומנים בד"כ ע"י 'כובע' -  $\hat{\beta}$ .

אומדים אחרים מסומנים בד"כ ע"י 'תלתל' -  $\tilde{\beta}$ .

(4) בעוד  $\alpha$  ו- $\beta$  הם קבועים,  $\hat{\alpha}$  ו- $\hat{\beta}$  הם משתנים מקריים. מדוע? מפני שבכל מדגם מתקבלים  $\hat{\alpha}$  ו- $\hat{\beta}$  אחרים.

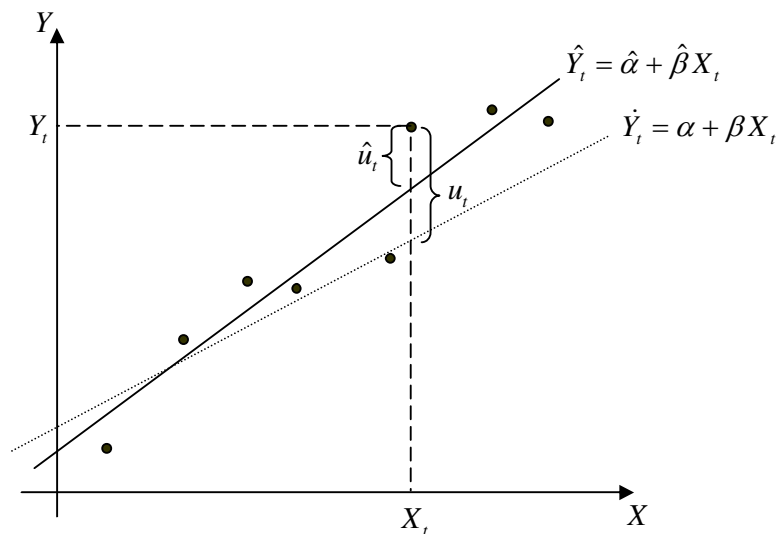
(5) את  $\alpha$  ו- $\beta$  אי אפשר לדעת, ולכן אי אפשר לדעת מהו הקו האמיתי, וכן אי אפשר לדעת את  $u_t$ .

(6) אפשר לדעת את  $\hat{u}_t$ , שהיא הסטיה מקו הרגרסיה. נגדיר זאת באופן הבא:

\* עבור  $X_t$ , הערך הצפוי של המשתנה המוסבר ( $\hat{Y}_t$ ) המתקבל לפי הרגרסיה

$$\text{הוא } \hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t.$$

\* הסטיה של התצפית ( $Y_t$ ) מהערך הצפוי לפי הרגרסיה ( $\hat{Y}_t$ ) היא  $\hat{u}_t = Y_t - \hat{Y}_t$ .



..... הקו האמיתי

———— קו הרגרסיה (הקו הנאמד)

• תצפית בודדת

## שלבבי התהליך האקונומטרי

### הגדרת מודל

כל הגורמים המשפיעים באופן שיטתי חייבים להופיע במודל. כל ההשפעות האקראיות באות לידי ביטוי בקריזה.

### איסוף נתונים

ככל שמספר התצפיות במדגם גדול יותר כן יהיו התוצאות טובות יותר.

### אמידה

יש שיטות אמידה רבות. אנחנו לומדים רק על אומדי הריבועים הפחותים.

### ניתוח סטטיסטי של התוצאות

מובהקות הרגרסיה (באמצעות מבחן  $F$ ), איכות הרגרסיה (באמצעות  $R^2$ ), מובהקות האומדים (באמצעות מבחן  $t$ ).

### ניתוח כלכלי של התוצאות

משמעות הקשר בין המשתנים וביצוע תחזיות אם יש צורך.

במבחן:

### הגדרת מודל

בד"כ איננו צריכים להגדיר את המודל אלא מגדירים אותו בשבילנו.

### איסוף נתונים

את הנתונים איננו צריכים לאסוף.

### אמידה

אנחנו צריכים לדעת באופן תאורטי איך אומדים וכן את תכונות האומדים. האמידה עצמה מבוצעת ע"י מחשב, ואנו מקבלים את תוצאותיה.

### ניתוח סטטיסטי של התוצאות

אנו צריכים לשלוט הן בתאוריה והן בפרקטיקה של הניתוח.

### ניתוח כלכלי של התוצאות

נדרש ברמה בסיסית.

## פרק 2 - אומדי הריבועים הפחותים (אר"פ) והנחות הקלאסיות

שיטת האמידה של  $\alpha$  ושל  $\beta$  נקראת שיטת הריבועים הפחותים

### Ordinary Least Squares (OLS)

השאלה הנשאלת בשיטת אמידה זו היא: איזה  $\hat{\alpha}$  ו- $\hat{\beta}$  יביאו למינימום את סכום

ריבועי טעויות האמידה.

ובתרגום מתימטי:

$$\min_{\hat{\alpha}\hat{\beta}} \sum \hat{u}_t^2 = \min_{\hat{\alpha}\hat{\beta}} \sum (y_t - \hat{y}_t)^2 = \min_{\hat{\alpha}\hat{\beta}} \sum [y_t - (\hat{\alpha} + \hat{\beta}x_t)]^2 = ?$$

מתוך גזירת הפונקציה הזו מתקבלים האומדים  $\hat{\alpha}$  ו- $\hat{\beta}$ .

| מודל עם רק עם<br>חותך<br>$Y_t = \alpha + u_t$ | מודל ללא חותך<br>$Y_t = \beta X_t + u_t$                        | מודל עם חותך ושיפוע<br>$Y_t = \alpha + \beta X_t + u_t$                                                                                                                                                                                                   |                  |
|-----------------------------------------------|-----------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------|
| $\hat{\alpha} = \bar{Y}$                      | $\hat{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}$ | $\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{t=1}^T (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^T (X_t - \bar{X})^2}$<br>$= \frac{\sum_{t=1}^T (X_t - \bar{X})Y_t}{\sum_{t=1}^T (X_t - \bar{X})^2}$<br>$\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$ | חישוב<br>האומדים |
| $E(\hat{\alpha}) = \alpha$                    | $E(\hat{\beta}) = \beta$                                        | $E(\hat{\beta}) = \beta$<br>$E(\hat{\alpha}) = \alpha$                                                                                                                                                                                                    | תוחלת<br>האומדים |
| $V(\hat{\alpha}) = \frac{\sigma_u^2}{T}$      | $V(\hat{\beta}) = \frac{\sigma_u^2}{\sum_{t=1}^T X_t^2}$        | $V(\hat{\beta}) = \frac{\sigma_u^2}{S_{XX}}$<br>$V(\hat{\alpha}) = \sigma_u^2 \left( \frac{1}{T} + \frac{\bar{X}^2}{S_{XX}} \right)$                                                                                                                      | שונות<br>האומדים |

**\*\*הערה חשובה:** בתהליך הגזירה של פונקציית הריבועים הפחותים מתקבלות

"המשוואות הנורמליות":

עבור המודל הקלאסי (עם חותך):

גזירה של  $\alpha$  מתקבלת המשוואה הנורמלית:  $\sum \hat{u}_i = 0$

גזירה של  $\beta$  מתקבלת המשוואה הנורמלית:  $\sum \hat{u}_i \cdot x_i = 0$

עבור מודל ללא חותך:

מתקבלת משוואה נורמלית אחת מגזירת  $\beta$  בלבד:  $\sum \hat{u}_i \cdot x_i = 0$

המשוואות הנורמליות צריכות להתקיים על מנת שפונקציית הריבועים הפחותים

תתקיים ( $\sum \hat{u}_i^2 = \min$ )

תירגול:

**?** כלכלן החליט לאמוד את המודל:  $y_i = \alpha + \beta x_i + u_i$ .

א. נסחו את בעיית ה-OLS.

ב. מצאו את תנאי סדר ראשון של בעיית ה-OLS (המשוואות הנורמליות).

ג. מצאו נוסחה לקבלת האומדים  $\hat{\alpha}, \hat{\beta}$ .

ד. הוכיחו כי קו הרגרסיה עובר דרך נקודת הממוצעים  $(\bar{X}, \bar{Y})$ .

ה. בהנחה והיינו בוחרים אומד אחר ל- $\beta$  שאינו אומד הריבועים

הפחותים, מה היה יחס הביטויים:  $\sum e_i$  ו- $\sum e_i^2$  של אומד זה ביחס

לאומד הריבועים הפחותים?

**?** כלכלן החליט לאמוד את המודל:  $y_i = \beta x_i + u_i$ .

א. נסחו את בעיית ה-OLS.

ב. מצאו את תנאי סדר ראשון של בעיית ה-OLS.

ג. מצאו נוסחה לקבלת  $\hat{\beta}$ .

ד. הוכיחו כי קו הרגרסיה אינו עובר דרך נקודת הממוצעים  $(\bar{X}, \bar{Y})$ .

ה. מהו התנאי שבו אומד הריבועים הפחותים שמצאתם בסעיף ג יהיה זהה לנוסחה של אומד הריבועים הפחותים שנמצא בשאלה הקודמת (במודל עם חותך)?

**?** חוקר רצה לחקור האם ציוני IQ משפיעים על הציון באקונומטריקה ולכן אסף

תצפיות מ-5 סטודנטים:

| SCORE | IQ  | $e_i$ |
|-------|-----|-------|
| 80    | 100 | 1     |
| 75    | 110 | -1    |
| 80    | 110 | 1     |
| 90    | 103 | 2     |
| 85    | 102 | -3    |

איזה מבין המודלים הבאים נאמד?

א.  $score_i = \hat{\alpha} + \hat{\beta} \cdot IQ_i$

ב.  $score_i = \hat{\beta} \cdot IQ_i$

ג.  $score_i = \hat{\alpha}$

ד.  $score_i = \bar{y}$

**?** נבדק הקשר שבין שכר לשעה שעובד מסוים מרוויח אצל מעסיק מסוים (X)

לבין כמות העובדים שמועסקים אצל אותו מעסיק (Y) (הניחו שכר שווה בין העובדים אצל אותו המעסיק).

לשם כך נדגמו 10 מעסיקים באופן מקרי ונתקבלו התוצאות הבאות:

$$\bar{x} = 35$$

$$\bar{y} = 5.8$$

$$\sum_{i=1}^{10} x_i^2 = 19,100$$

$$\sum_{i=1}^{10} y_i^2 = 440$$

$$\sum_{i=1}^{10} x_i y_i = 2858.85$$

מהי תחזית כמות העובדים המועסקים אצל מעסיק מסוים המשתכרים 25 ₪ לשעה?

## ההנחות הקלאסיות של מודל הרגרסיה:

כדי שהנוסחאות הנ"ל יהיו נכונות וכדי שתכונות האומדים (שיפורטו בהמשך) יתקיימו, צריכים להשמר מספר כללים. כללים אלו נקראים ההנחות הקלאסיות. קיימות 7 הנחות כאלה:

(1) קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.

$$u + \text{מסביר} = \alpha + \beta$$

מקדם  $\beta$ : שיפוע הקו המתאר את הקשר בין המסביר למוסבר.

כדי שהקשר יהיה ליניארי שיפוע  $\beta$  צריך להיות קבוע.

\*\* שימו לב כי ישנם מודלים בהם הקשר בין  $X$  ל- $Y$  הוא לא ליניארי אבל בין המסביר למוסבר כן נקבל קו ישר ששיפועו קבוע, כמו למשל במודל:

$$y = \alpha + \beta \ln x + u$$

$$(2) \text{ קיימים לפחות שני ערכי } X \text{ שונים זה מזה: } S_{xx} = \sum_{t=1}^T (X_t - \bar{X})^2 \neq 0$$

המשמעות הסטטיסטית של הנחה זו היא כי  $X$  הוא משתנה ולא קבוע. כלומר, יש לו פיזור או שונות השונה מ-0.

הבעיה ב- $X$  קבוע היא ששונותו שווה ל-0 וכאשר  $S_x^2 = 0$  הקשר בין ה- $X$  ל- $Y$  שווה גם הוא ל-0.

$$(3) \text{ תוחלת ההפרעה האקראית היא אפס לכל תצפית: } E(u_t) = 0 \text{ לכל } t$$

לכל ערך  $X$  באוכלוסיה יש פיזור מקרי של ערכי  $Y$  ושל טעויות או "קריזות" ( $u$ ), כל אחת מקריזות אלו איננה ניתנת לחיזוי אך בממוצע הן מתקזזות ומתאפסות ואנחנו פועלים לפי ההיגיון הכלכלי אותו ניתן לנבא על סמך הקו.

(4) ה- $X$  אינם משתנים מקריים.

אנו מניחים שהמשתנה המסביר הוא אקסוגני, כלומר ידוע מראש, משפיע על  $Y$  אבל לא מושפע ממנו בחזרה.

במילים אחרות, ניבוי Y על סמך X מסוים, מחייב את ה-X להיות משתנה אמפירי, ידוע מראש ולא אקראי ולהיות המשתנה המסביר, המשפיע במודל. למשל, אם נרצה לנבא את תצורות משפחה על סמך הכנסתה, כאשר נדגום משפחה ונשאל להכנסתה נצפה לקבל תשובה מסויימת (שההכנסה למשפחה לא תהיה אקראית) ולהניח כי זהו המשתנה המשפיע על התצורות ולא להיפך במודל הניבוי הנוכחי בו אנו משתמשים.

**\*\* שימו לב כי מהנחה זו משתמע גם כי המתאם בין הטעויות לערכי X שווה ל-0:**  
 $\text{cov}(X_t, u_t) = 0$  (שכן המתאם בין X לבין U שווה ל-0 עבור כל t).

**(5) הומוסקדסטיות: השונות של ההפרעה האקראית זהה לכל תצפית ותצפית:**

$$V(u_t) = \sigma_u^2 \text{ לכל } t$$

הפיזור סביב קו הרגרסיה הוא אחיד.

**(6) אין מתאם בין הפרעות אקראיות:  $\text{cov}(u_t, u_s) = 0$  לכל  $t \neq s$**

"הקריזות" של תצפיות שונות אינן תלויות אחת בשניה.  
הדבר תלוי בדגימה האקראית של התצפיות.

למשל, אם אנו בוחנים השפעה של ההכנסה על התצורות של משפחות, אם דגמנו באופן אקראי את המשפחות, לא יהיה קשר בין הטעות בניבוי של תצורות משפחה מסוימת ( $u_t$ ) לטעות בניבוי התצורות של משפחה אחרת ( $u_s$ ).

**(7) ההפרעות האקראיות מתפלגות נורמלית:  $u_t \approx N$**

התפלגות נורמלית של טעויות סביב התוחלת (ששווה כאמור ל-0) משמעה שרוב הטעויות בניבוי הן קטנות ולא מאוד משמעותיות.

### לסיכום:

**(1) קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.**

$$\text{X איננו קבוע: } S_{XX} = \sum_{t=1}^T (X_t - \bar{X})^2 \neq 0 \quad (2)$$

**(3) תוחלת ההפרעה האקראית היא אפס לכל תצפית:  $E(u_t) = 0$  לכל t**

(4)  $X_t$  אינם משתנים מקריים  $\Leftrightarrow$  ניתן להוציא אותם מחוץ לתוחלת

$$\text{cov}(X_t, u_t) = 0 \Leftrightarrow \text{ולשונויות}$$

(5) הומוסקדסטיות: שונות ההפרעה האקראית קבועה לכל תצפית:

$$V(u_t) = \sigma_u^2 \text{ לכל } t$$

$$\text{cov}(u_t, u_s) = 0 \text{ ב"ת: } t \neq s \text{ לכל}$$

(7) ההפרעות האקראיות מתפלגות נורמלית:  $u_t \approx N$

? שכר של עובדים מנובא על ידי השכלתם במודל הבא:  $w_i = \alpha + \beta \cdot s_i + u_i$

א. כתבו את ההנחות הקלאסיות במונחי המשתנים של המודל הנתון והסבירו אותן.

ב. התייחסו לכל אחת מהטענות הבאות וקבעו האם היא:

הנחה קלאסית/משוואה נורמאלית (או תוצאה הנובעת ממשוואה נורמאלית)/אף אחד מהשניים:

$$1. \text{cov}(s_i, u_i) = 0$$

$$2. \text{cov}(s_i, e_i) = 0$$

$$3. E(u_i) = 0$$

$$4. \text{cov}(u_i, u_j) = 0$$

$$5. \bar{e} = 0$$

$$6. \bar{w} = \bar{\hat{w}}$$

$$7. \sum u_i = 0$$

$$8. V(u_i) = \sigma_i$$

$$9. S_s^2 \neq 0$$

## תכונות האומדים

אומדי הריבועים הפחותים הם לינאריים, חסרי הטיה, יעילים ועקיבים.

### 1) לינאריות

אר"פ ניתנים להצגה כקומבינציה לינארית של  $Y_t$ .

במילים אחרות, כדי ש- $\hat{\beta}$  למשל, תהיה אומד לינארי צריך להתקיים:

$$\hat{\beta} = \sum W_t \cdot Y_t \quad \text{כאשר } W_t \text{ היא קומבינציה של ערכי } X.$$

$$\text{למשל: } \hat{\beta} = \frac{\sum X_t \cdot Y_t}{\sum X_t^2}$$

אומד זה ניתן להצגה בצורה הבאה:

$$\hat{\beta} = \frac{X_1}{\sum X^2} \cdot Y_1 + \frac{X_2}{\sum X^2} \cdot Y_2 + \dots + \frac{X_T}{\sum X^2} \cdot Y_T$$

$$\hat{\beta} = \frac{W_t}{a} \cdot Y_t$$

לפיכך מדובר באומד לינארי.

הוכחת לינאריות עבור האומדים של המודל הקלאסי:

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} = \sum w_t Y_t, \quad w_t = \frac{X_t - \bar{X}}{S_{xx}}$$

$$\hat{\alpha} = \sum v_t Y_t, \quad v_t = \frac{1}{T} - w_t \bar{X}$$

### כלל אצבע-כיצד יודעים אם אומד הוא לינארי?

הלכה למעשה יש לבדוק האם מתקיימים 3 התנאים הבאים:

1) המשתנים המקריים (ה- $y_t$ ) הם ממעלה ראשונה (כלומר לא יהיו נתונים

בחזקה או בשורש).

2) בין המשתנים המקריים (ה- $y_t$ ) יש סכום או הפרש (ולא כפל או חילוק).

3) כל שאר הגורמים פרט ל- $y_t$  אינם משתנים מקריים (בהתאם להנחות,

כזכור,  $x_t$  איננו משתנה מקרי).

כלל אצבע: אם בנוסחה של האומד לא מופיעים סימני כפל בין  $Y_t$ -ים או העלאה בחזקה/שורש של  $Y_t$  וכן ה-  $Y_t$ -ים לא מופיעים במכנה, אז סביר להניח שהאומד לינארי.

? האם האומד:  $\hat{\beta} = \frac{\sum_{t=1}^T (X_t - \bar{X}) Y_t}{\sum_{t=1}^T (X_t - \bar{X})^2}$  הוא לינארי?

? האם האומד  $\tilde{\beta} = \frac{\sum_{t=1}^T Y_t^3 \sum_{t=1}^T X_t (Z_t + Y_t)}{\sum_{t=1}^T X_t^2}$  הוא לינארי?

? כלכלן החליט לאמוד את המודל:  $y_i = \alpha + \beta x_i + u_i$ .

1.  $\tilde{\beta} = \frac{\sum x_i y_i}{\sum x_i^2}$

2.  $\tilde{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$

3.  $\tilde{\beta} = \sum \left( \frac{y_i}{x_i} \right)^2$

4.  $\tilde{\beta} = \frac{y_N - y_1}{x_N - x_1}$

5.  $\tilde{\beta} = \frac{\bar{y}}{\bar{x}}$

6.  $\tilde{\beta} = \frac{\sum x_i y_i}{\sum (x_i - \bar{x})^2}$

מי מהאומדים הוא לינארי ומהן המשקולות?

2) חוסר הטיה

התוחלת של אר"פ שווה לערך האמיתי של הפרמטר. כלומר, אומד  $\hat{\theta}$  מסוים יהווה אח"ה לפרמטר  $\theta$  אותו הוא אומד באוכלוסיה אם מתקיים:

$$E(\hat{\theta}) = \theta$$

זהו מושג תאורטי (ולא קונקרטי) שאומר כי ממוצע כל האומדים ( $\hat{\theta}$ ) של אינסוף המדגמים האפשריים בגודל מסוים שווה לפרמטר ( $\theta$ ).

עבור מדגם מקרי אחד האומד איננו שווה לפרמטר ( $\hat{\theta} \neq \theta$ ) אבל על פני אינסוף המדגמים האפשריים, ממוצע האומדים ( $E(\hat{\theta})$ ) צריך להיות שווה לפרמטר ( $\theta$ ) כדי שהאומד יהיה אח"ה.

כיצד יודעים אם אומד הוא חסר הטיה?

בשלב הראשון יש לבצע עבודת הכנה –

מבטאים את האומד באמצעות הפרמטר האמיתי: מתחילים מהאומד המוצע, מציבים במקום ה- $Y_i$  את המודל ומפתחים אלגברית.

**\*\* יש לזכור כי:**

$\frac{u_i}{y_i}$  מהווים משתנים מקריים  $\Leftarrow$  נשארים בתוך התוחלת, השונות וה- $\sum$ .

$x_i$  איננו משתנה מקרי (על פי הנחה מס' 4)  $\Leftarrow$  יוצא מחוץ לתוחלת ולשונות אך

נשאר בתוך ה- $\sum$

$\frac{\alpha}{\beta}$  קבועים  $\Leftarrow$  יוצאים מחוץ לתוחלת, לשונות ול- $\sum$

**דוגמא:**

עבור המודל  $Y_i = \beta X_i + u_i$  והאומד המתאים לו  $\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2}$ :

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i (\beta X_i + u_i)}{\sum X_i^2} = \frac{\beta \sum X_i^2}{\sum X_i^2} + \frac{\sum X_i u_i}{\sum X_i^2} = \beta + \frac{\sum X_i u_i}{\sum X_i^2}$$

שלב מקדים זה יעשה לפני בדיקת חוסר הטייה, יעילות ועקיבות.  
הוכחת חוסר הטייה – מפעילים תוחלת על האומד, ואם התוחלת שווה לפרמטר  
האמיתי אז האומד חסר הטייה.

בשפה מתמטית: אם  $E(\hat{\beta}) = \beta$ , אז  $\hat{\beta}$  הוא אומד חסר הטייה ל- $\beta$ .  
כדי שהדבר יתקיים הנחות (3) ו-(4) חייבות להתקיים.

המשך הדוגמא שלעיל:

$$E(\hat{\beta}) = E\left(\beta + \frac{\sum X_i u_i}{\sum X_i^2}\right) = E(\beta) + E\left(\frac{\sum X_i u_i}{\sum X_i^2}\right) = \beta + \frac{\sum X_i E(u_i)}{\sum X_i^2} = \beta$$

מסקנה: האומד חסר הטייה!

כלל אצבע:

אם בעבודת ההכנה נשארים בסוף הפיתוח רק שני סוגי איברים:

(1) הפרמטר האמיתי

(2) איבר או כמה איברים שמכילים את  $u_i$  (קומבינציה ליניארית של  $u_i$ )

אז האומד חסר הטייה.

למשל, בעבודת ההכנה שלעיל:

$$\hat{\beta} = \beta + \frac{\sum X_i u_i}{\sum X_i^2}$$

הפרמטר  
האמיתי
איבר המכיל  
את  $u_i$

**? נתון האומד הבא:**

$$\tilde{\beta} = \frac{\sum_{t=1}^T X_t Y_t}{\sum_{t=1}^T X_t^2}$$

**האם האומד הנ"ל הוא חסר הטיה?**

1. בדוק במודל עם חותך
2. בדוק במודל ללא חותך

### (3) יעילות

יעילות פירושה השונות הקטנה ביותר. ככל שהשונות של האומד קטנה יותר, כך יש הסתברות גבוהה יותר שהוא יהיה קרוב יותר לפרמטר האמיתי באוכלוסייה אותו הוא אומד.

$\hat{\theta}_1$  יקרא אומד יעיל יותר מ- $\hat{\theta}_2$  אם מתקיים שהשונות שלו קטנה יותר:

$$V(\hat{\theta}_1) < V(\hat{\theta}_2)$$

**משפט גאוס מרקוב:**

יעילות היא תמיד מושג השוואתי. לכן בכדי לדעת האם השונות של האומד היא המינימאלית האפשרית נשתמש במשפט גאוס מרקוב.  
לפי משפט גאוס-מרקוב אר"פ הם בעלי השונות הנמוכה ביותר בקבוצה שלהם (קבוצת האומדים הלינאריים חסרי ההטיה), והם נקראים B.L.U.E (Best Linear Unbiased Estimation).

**כלומר:**

אם האומד שלנו הוא ליניארי וחסר הטיה  $\Leftarrow$  מבלי לחשב את שונותו נדע לפי משפט גאוס-מרקוב שהיא גדולה יותר משל אומד הריבועים הפחותים.  
אם האומד איננו ליניארי ו/או חסר הטיה  $\Leftarrow$  לא ניתן להשתמש במשפט גאוס-מרקוב ואז היחס בין שונות האומד לשונות אומד הריבועים הפחותים המקביל איננו ידוע.

**כיצד מחשבים שונות של אומד?**

ראשית כל, הנחות (4), (5) ו-(6) חייבות להתקיים. אם הן מתקיימות, מחשבים את השונות של האיברים המכילים את  $u_t$  מהפיתוח הקודם.

נדגים על ידי חישוב שונות אר"פ  $\hat{\beta}$  :

**1. במודל ללא חותך**

$$\begin{aligned} V(\hat{\beta}) &= V\left(\frac{\sum X_t u_t}{\sum X_t^2}\right) = \frac{V(\sum X_t u_t)}{(\sum X_t^2)^2} = \frac{\sum V(X_t u_t)}{(\sum X_t^2)^2} = \\ &= \frac{\sum X_t^2 V(u_t)}{(\sum X_t^2)^2} = \frac{\sum X_t^2 \sigma_u^2}{(\sum X_t^2)^2} = \frac{\sigma_u^2 \sum X_t^2}{(\sum X_t^2)^2} = \frac{\sigma_u^2}{\sum X_t^2} \end{aligned}$$

**2. במודל עם חותך**

$$\begin{aligned} V(\hat{\beta}) &= V\left(\frac{\sum (X_t - \bar{X}) u_t}{S_{XX}}\right) = \frac{V(\sum (X_t - \bar{X}) u_t)}{S_{XX}^2} = \frac{\sum V((X_t - \bar{X}) u_t)}{S_{XX}^2} = \\ &= \frac{\sum [(X_t - \bar{X})^2 V(u_t)]}{S_{XX}^2} = \frac{\sum [(X_t - \bar{X})^2 \sigma_u^2]}{S_{XX}^2} = \frac{\sigma_u^2 \sum (X_t - \bar{X})^2}{S_{XX}^2} = \frac{\sigma_u^2 S_{XX}}{S_{XX}^2} = \frac{\sigma_u^2}{S_{XX}} \end{aligned}$$

**(4) עקיבות**

ככל שהמדגם יגדל כן יתקרב האומד לערך האמיתי של הפרמטר. אם נגדיל את המדגם לאינסוף תצפיות ונחשב את האומד, הוא יהיה שווה לפרמטר

$$\begin{aligned} &(\hat{\theta} \rightarrow \theta) \\ &T \rightarrow \infty \end{aligned}$$

**תנאי הכרחי לעקיבות:** האומד חייב להיות פונקציה של גודל המדגם. במילים אחרות, האומד צריך להיות מושפע מגודל המדגם. ברגע שהאומד עונה על תנאי זה הוא יהיה עקיב. אומד המחושב במדגם סופי בהגדרה לא יוכל להיות עקיב לפרמטר באוכלוסיה.

**סיכום: השלבים להוכחת התכונות****(1) הוכחת ליניאריות****(2) הכנת האומד  $\Leftarrow$  להציב במקום  $Y_t$  את המודל האמיתי.**

$$Y_t = \alpha + \beta X_t + u_t \quad \text{במודל עם חותך:}$$

במודל ללא חותך:  $Y_t = \beta X_t + u_t$

(3) פיתוח האלגברה

(4) חישוב תוחלת, שונות, עקיבות.

- ליניאריות מהווה תנאי הכרחי לחוסר הטיה.
- ליניאריות וחוסר הטיה מהוות תנאי הכרחי לבחינת היעילות של האומד לפי משפט גאוס-מרקוב.
- עקיבות איננה תלויה בתכונות האחרות, אלא רק בהיותו של האומד פונקציה של גודל המדגם (לא מחושב על מדגם סופי). כך שאומד לא חייב להיות ליניארי או חסר הטיה כדי להיות עקיב.
- העקיבות משפיעה על היעילות של האומד. עבור אומדים התלויים בגודל המדגם: ככל שגודל המדגם גדול יותר כך שונות האומד קטנה והאומד יהיה יעיל יותר לפרמטר באוכ'.

### תרגול ממבחנים

**? תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 25 נקודות)**

נתון המודל  $Y_t = \alpha + \beta X_t + u_t$ ,  $T = 100$ , כאשר מתקיימות כל ההנחות הקלאסיות.

$$\tilde{\beta} = \frac{\sum_{t=51}^{100} Y_t - \sum_{t=1}^{50} Y_t}{\sum_{t=51}^{100} X_t - \sum_{t=1}^{50} X_t} \quad \text{נתון האומד}$$

- |                                                        |                |
|--------------------------------------------------------|----------------|
| א. האומד $\tilde{\beta}$ הינו אומד חסר הטיה ל- $\beta$ | נכון / לא נכון |
| ב. האומד $\tilde{\beta}$ הינו אומד עקיב ל- $\beta$     | נכון / לא נכון |
| ג. האומד $\tilde{\beta}$ הינו אומד לינארי ל- $\beta$   | נכון / לא נכון |
| ד. האומד $\tilde{\beta}$ הינו אומד יעיל ל- $\beta$     | נכון / לא נכון |
| ה. השונות האמיתית של $\tilde{\beta}$ היא:              |                |

**? תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 14 נקודות)**

נתון המודל  $Y_t = \beta X_t + u_t$ , כאשר כל ההנחות הקלאסיות מתקיימות.

(יש לשים לב המודל ללא חותך)

$$\tilde{\beta} = \frac{\sum Y_t}{\sum X_t} \quad \text{נתון האומד:}$$

א. האומד  $\tilde{\beta}$  הינו אומד מוטה ל-  $\beta$  : נכון / לא נכון / אי אפשר לדעת

ב. על סמך משפט גאוס-מרקוב ניתן להסיק כי  $\tilde{\beta}$  איננו אומד יעיל יותר מאומד הריבועים הפחותים : נכון / לא נכון / לא ניתן לדעת

ג. מהי השונות האמיתית של  $\tilde{\beta}$  ?

**? תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 16 נקודות)**

נתון המודל  $Y_t = \beta X_t + u_t$ , כאשר כל ההנחות הקלאסיות מתקיימות.

(יש לשים לב המודל ללא חותך)

$$\tilde{\beta} = \frac{\sum X_t Y_t}{\sum (X_t - \bar{X})^2} \quad \text{נתון האומד:}$$

א. מהי התוחלת של  $\tilde{\beta}$  ?

ב.  $E(\tilde{\beta}) < \beta$  . נכון / לא נכון / אי אפשר לדעת

ג. על סמך משפט גאוס-מרקוב ניתן להסיק כי אומד הריבועים הפחותים הינו אומד יעיל יותר מ-  $\tilde{\beta}$  . נכון / לא נכון / לא ניתן לדעת

ד. מהי השונות האמיתית של האומד  $\tilde{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$  ?

## ? שאלות נוספות מתוך מבחנים

בכל השאלות ההנחות הקלאסיות מתקיימות.

האומדים הם אר"פ, והמודל הוא  $Y_t = \alpha + \beta X_t + u_t$ .

1.  $E(Y_t) = E(\hat{Y}_t)$  נכון / לא נכון / אי אפשר לדעת

2.  $\sum_{t=1}^T (X_t - \bar{X}) \bar{Y} \neq 0$  נכון / לא נכון / אי אפשר לדעת

3. אמידת המודל בשיטת הריבועים הפחותים תתן את התוצאה:  $\sum_{t=1}^T u_t = 0$

נכון / לא נכון / אי אפשר לדעת

4. אם נתון ש-  $r_{xy} = 0.57$ , אזי  $\hat{\beta}$ :

א. הוא בהכרח שלילי

ב. הוא בהכרח חיובי

ג. הוא בהכרח שווה לאפס

ד. לא ניתן לקבוע את סימנו על סמך הנתונים הקיימים

5. סמן את הטענה הנכונה בהכרח:

א.  $\sum_{t=1}^T (X_t - \bar{Y}) \hat{u}_t = 0$

ב.  $S_{xx} = \sum_{t=1}^T X_t^2 - (T\bar{X})^2$

ג.  $\sum_{t=1}^T X_t u_t = 0$

ד. אף אחת מהטענות הנ"ל אינה נכונה בהכרח.

6. אומדי הריבועים הפחותים אינם חסרי הטיה, אם נתון שהשונות של  $u_t$  אינה

קבועה. נכון / לא נכון / אי אפשר לדעת

7. אומד חסר הטיה הוא אינו בהכרח גם אומד עקיב.

נכון / לא נכון / אי אפשר לדעת

### פרק 3 - מודלים לא ליניאריים

עד עכשיו דיברנו רק על מודלים ליניאריים (linear-linear). בפרק זה נלמד גם על מודלים שאינם ליניאריים: מודל חצי לוגריתמי (semi-log), מודל לוגריתמי כפול (double-log) ומודל לוג ליניארי (linear-log).

נשאלת השאלה-מתי מודל מוגדר כליניארי?

מודל מוגדר כליניארי כאשר הוא מתאר קשר קוי בין המשתנים- המסביר והמוסבר שלו.

למשל המודל הליניארי הקלאסי:  $Y = \alpha + \beta X + u$  מתאר קשר קוי בין  $x$  ל- $y$ .

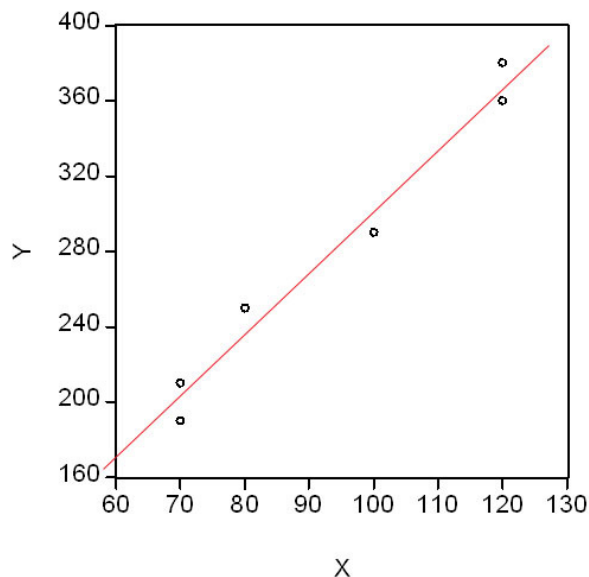
המשמעות של קשר ליניארי היא שהנגזרת-  $\frac{\partial Y}{\partial X}$  היא קבועה.

נגזרת זו מתארת את השינוי השולי (השיפוע של הגרף): אם מגדילים את  $x$

ביחידה אחת, בכמה יחידות משתנה  $y$ .

במודל הליניארי- שינוי זה הוא קבוע ושווה ל-  $\beta$ .

גרף המתאר את הקשר יראה כך:



בניגוד למודל הליניארי, שלושת המודלים האחרים (המודלים הלוגריתמיים) מתארים קשרים שאינם ליניאריים בין  $X$  ל- $Y$ . במודלים אלו השינוי השולי (השיפוע) לא יהיה קבוע, אלא תלוי במשתנים -  $x$  או  $y$  או בשניהם:

(1) במודל החצי לוגריתמי הקשר בין  $x$  ל-  $y$  מתואר על ידי הפונקציה הבאה:

$$Y = e^{\alpha + \beta x + u}$$

גרף המתאר את הקשר יראה כך:

כפי שניתן לראות, השינוי השולי איננו קבוע אלא תלוי ב- $y$ .

$$\frac{\partial Y}{\partial X} = \beta \cdot Y \text{ : השינוי השולי}$$

ככל ש- $Y$  גדל כך השיפוע ( $\beta$ ) גדל.

$$\beta = \frac{\frac{\partial Y}{\partial X}}{Y} \text{ : שיעור השינוי השולי}$$

שיעור שינוי שולי אומר: אם מגדילים את  $X$  ביחידה, בכמה % ישתנה  $Y$ .

במודל החצי לוגריתמי עבור עליה ביחידה אחת של  $X$ ,  $Y$  ישתנה ב-

$$100 \cdot \beta \%$$

במודלים אלו, המתארים שיעורי תשואה, השינוי באחוזים הוא קבוע למרות שהשינוי השולי איננו קבוע.

(2) במודל הלוגריתמי הכפול הקשר בין  $x$  ל-  $y$  מתואר על ידי הפונקציה

$$Y = e^\alpha \cdot X^\beta \cdot e^u$$

הבאה:  $Y = e^\alpha \cdot X^\beta \cdot e^u$   
גרף המתאר את הקשר יראה כך:

כפי שניתן לראות, השינוי השולי איננו קבוע אלא תלוי ב-  $x$  וב-  $y$ .

$$\frac{\partial Y}{\partial X} = \beta \cdot \frac{Y}{X} \quad \text{השינוי השולי:}$$

$$\beta = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y} = \frac{\frac{\partial Y}{\partial X}}{\frac{Y}{X}} \quad \text{משמעות ה- } \beta \text{ במודל כזה היא הגמישות:}$$

משמעות הגמישות היא שינוי שולי באחוזים: אם מגדילים את  $X$  ב- %

אחד, בכמה % ישתנה  $Y$ .

במודל הלוגריתמי הכפול ה-  $\beta$  מייצגת את הגמישות, כלומר אם נגדיל את

$X$  ב- % אחד,  $Y$  ישתנה ב-  $\beta$  %.

במודלים אלו הגמישות היא קבועה למרות שהשינוי השולי איננו קבוע.

(3) במודל הלוג-ליניארי הקשר בין  $x$  ל-  $y$  מתואר על ידי הפונקציה הבאה:

$$e^y = e^\alpha \cdot X^\beta \cdot e^u$$

גרף המתאר את הקשר יראה כך:

כפי שניתן לראות, השינוי השולי איננו קבוע אלא תלוי ב-X. ככל ש-X עולה כך פוחת השינוי השולי.

$$\text{השינוי השולי: } \frac{\partial Y}{\partial X} = \frac{\beta}{X}.$$

$$\beta = \frac{\frac{\partial Y}{\partial X}}{\frac{X}{Y}} : \beta \text{ עולה ב-} Y \text{ עולה ב-} \beta$$

ל- $\beta$  אין משמעות כלכלית במודל זה.

### גמישות

בנוסף למשמעות ה- $\beta$  בכל אחד מהמודלים, מושג נוסף שיש להכיר הוא מושג הגמישות.

כאמור, גמישות משמעה: שינוי שולי באחוזים. כלומר בכמה % ישתנה Y אם X יגדל ב-% אחד.

הביטוי המתימטי לגמישות:

$$\frac{\frac{\partial Y}{Y}}{\frac{\partial X}{X}} = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y}$$

כלומר, כדי לחשב גמישות יש להכפיל את השינוי השולי  $\left(\frac{\partial Y}{\partial X}\right)$  ב- $\frac{X}{Y}$

$$(1) \text{ במודל הליניארי- הגמישות: } \frac{\beta X}{Y}$$

$$(2) \text{ במודל החצי לוגריתמי- הגמישות: } \beta Y \cdot \frac{X}{Y} = \beta X$$

$$(3) \text{ במודל הלוגריתמי הכפול- הגמישות: } \beta \frac{Y}{X} \cdot \frac{X}{Y} = \beta$$

$$(4) \text{ במודל הלוג-ליניארי- הגמישות: } \frac{\beta}{X} \cdot \frac{X}{Y} = \frac{\beta}{Y}$$

ניתן לראות כי פרט למודל הלוגריתמי הכפול שבו הגמישות היא קבועה, הגמישות

של המודלים האחרים משתנה כפונקציה של X או של Y או של שניהם. כלומר

ניתן לחשבה עבור נקודה ספציפית על הגרף  $(X_t, Y_t)$  בלבד.

### טרנספורמציות של המודלים הלא ליניאריים לקו ישר:

בכדי שניתן יהיה לאמוד את המודלים הלא ליניאריים בשיטת OLS, עליהם לעבור טרנספורמציה לקו ישר.

טרנספורמציה של המודלים לקו ישר תאפשר לתאר את הקשר בין המשתנה המסביר למשתנה המוסבר באופן ליניארי.

טרנספורמציה זו תתבצע על ידי הוצאת  $\ln$  (לוג טבעי) משתי צידי המשוואה בכדי לבטל את ה-  $e$ .

תזכורת של חוקי לוגים:

$$LN(e^x) = X$$

$$LN(X^Y) = Y \cdot LN(X)$$

$$LN(X \cdot Y) = LN(X) + LN(Y)$$

$$LN\left(\frac{X}{Y}\right) = LN(X) - LN(Y)$$

| המודל             | לפני הטרנספורמציה                        | אחרי הטרנספורמציה                  |
|-------------------|------------------------------------------|------------------------------------|
| (1) לוג-ליניארי   | $e^y = e^\alpha \cdot X^\beta \cdot e^u$ | $Y = \alpha + \beta \ln X + u$     |
| (2) לוגריתמי כפול | $Y = e^\alpha \cdot X^\beta \cdot e^u$   | $\ln Y = \alpha + \beta \ln X + u$ |
| (3) חצי לוגריתמי  | $Y = e^{\alpha + \beta x + u}$           | $\ln Y = \alpha + \beta X + u$     |

אם נתייחס למשתנה המסביר או המוסבר בתוספת הלוג, ניתן יהיה לתאר את הקשר ביניהם באופן ליניארי.

**סיכום:**

| המודל                                                                                                    | משמעות ה- $\beta$                                                            | השינוי השולי<br>$(\frac{\partial Y}{\partial X})$<br>בכמה ישתנה Y<br>אם נגדיל את X<br>ביחידה? | הגמישות<br>$(\frac{\partial Y}{\partial X} \cdot \frac{X}{Y})$<br>בכמה % ישתנה Y אם<br>נגדיל את X ב-1%? |
|----------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|
| <b>ליניארי</b><br>$Y = \alpha + \beta X + u$                                                             | השינוי השולי<br>אם נגדיל את X ביחידה<br>Y ישתנה ב- $\beta$ יחידות            | $\beta$                                                                                       | $\frac{\beta X}{Y}$                                                                                     |
| <b>חצי לוגריתמי</b><br>$\ln Y = \alpha + \beta X + u$<br>( $Y = e^{\alpha + \beta X + u}$ )              | שיעור השינוי השולי<br>אם נגדיל את X ביחידה<br>Y ישתנה ב- $100 \cdot \beta\%$ | $\beta Y$                                                                                     | $\beta X$                                                                                               |
| <b>לוגריתמי כפול</b><br>$\ln Y = \alpha + \beta \ln X + u$<br>( $Y = e^\alpha \cdot X^\beta \cdot e^u$ ) | הגמישות<br>אם נגדיל את X ב-1%<br>Y ישתנה ב- $\beta\%$                        | $\frac{\beta Y}{X}$                                                                           | $\beta$                                                                                                 |
| <b>לוג ליניארי</b><br>$Y = \alpha + \beta \ln X + u$<br>( $e^Y = e^\alpha \cdot X^\beta \cdot e^u$ )     | אין משמעות כלכלית<br>אם נגדיל את X ב-1%<br>Y ישתנה ב- $\beta$                | $\frac{\beta}{X}$                                                                             | $\frac{\beta}{Y}$                                                                                       |

**\*\*המשתנה שיש בו LN השינוי בו יהיה באחוזים**

## תירגול

**?** על מנת לאמד את התשואה להשכלה בישראל בשנים 1948-1990 נאמדו המודלים

**הבאים :**

$$MWAGE_t = 139.547 + 118.628 \cdot SCL_t \quad (1)$$

$$MWAGE_t = -1445.08 + 1239.60 \cdot LN(SCL)_t \quad (2)$$

$$LN(MWAGE)_t = 5.244 + 0.778 \cdot LN(SCL)_t \quad (3)$$

$$LN(MWAGE)_t = 6.292 + 0.070 \cdot SCL_t \quad (4)$$

א. הסבירו את המשמעות של  $\beta$  בכל אחד מהמודלים

ב. חשבו את הגמישות בנקודת הממוצעים : (12.311,1600.01) עבור כל אחד מהמודלים.

**?** נתונים תוצאות האמידה של המודלים הבאים :

$$\hat{Y} = e^{4.5} \cdot X^{0.05} \quad (1)$$

$$\hat{Y} = e^{4.5+0.05X} \quad (2)$$

$$\hat{Y} = 4.5 + \frac{0.05}{X} \quad (3)$$

$$\hat{Y} = \frac{1}{1 + e^{4.5+0.05X}} \quad (4)$$

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

ב. עבור כל אחד מהמודלים ערכו תחזית נקודתית עבור  $X=6$

**? נתונים המודלים הבאים עבור התוצר במשק:**

$$1. \quad Q_i = AK_i^{\beta_1} e^{u_i}$$

$$2. \quad Q_i = Ae^{\beta_1 L_i + u_i}$$

$$3. \quad Q_i = A + K_i^{\beta_1} + e^{u_i}$$

$$4. \quad Q_i = A + \frac{\beta_1}{L_i} + u_i$$

$$5. \quad Q_i = A + \beta_1 \sqrt{K_i} + u_i$$

$$6. \quad Q_i = e^{A + \beta_1 K_i + u_i}$$

$$7. \quad Q_i = A\left(\frac{K_i}{2} + 7\right)^{\beta_1} e^{u_i}$$

$$8. \quad Q_i = A + \beta_1 L_i + u_i$$

$$9. \quad Q_i = A + \beta_1 \left(\frac{K_i}{L_i}\right) + u_i$$

**כאשר:**

**Q** - הוצאות צריכה על מוצר מסויים על ידי פרט מסויים.

**A** - הוצאות צריכה על המוצר בהינתן רמת הכנסה אפסית.

**K** - הכנסת הפרט.

**L** - שנות לימוד.

**א.** מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?

- ב. מי מבין המודלים שלא ניתנים לאמידה בשיטת OLS ניתן להביא למודל ליניארי בפרמטרים ועל כן לאמוד את הפרמטרים שלו?
- ג. עבור כל אחד מהמודלים קבעו מיהו המשתנה המוסבר ומיהו המסביר במשוואת הרגרסיה הליניארית.
- ד. עקומת אנג'ל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנג'ל?
- ? נתון המודל הבא:

$$Q_i = \frac{A}{K_i^{\beta_1}} e^{u_i}$$

- א. האם ניתן לאמוד את המודל בשיטת OLS?
- ב. מה המשוואה שצריך לאמוד על מנת לקבל את הפרמטרים למודל זה (כלומר כיצד הופכים את המודל לליניארי בפרמטרים)?
- ג. נאמד המודל הבא:

$$\ln(Q_i) = \alpha_0 + \alpha_1 \ln(K_i) + u_i$$

והתקבלו התוצאות הבאות:  $\hat{\alpha}_0 = 3, \hat{\alpha}_1 = 0.8$

מהם האומדנים עבור  $A, \beta_1$  ?

### פרק 4 - רגרסיה מרובה ומולטיקוליניאריות

עד עתה דיברנו על מודלים שכללו משתנה מסביר אחד בלבד. אולם במציאות קיימים משתנים רבים שמסבירים משתנה תלוי מסויים. למשל, הכנסה של עובד יכולה להיות מוסברת על ידי משתנים שונים כגון: ותק במקום העבודה, השכלה, גיל וכו'.

מודל רגרסיה מרובה נראה כך:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_j X_{ji} + U_i$$

כאשר:

$$Y_i = \text{משתנה תלוי}$$

$$X_{1i} \dots X_{ji} = \text{משתנים ב"ת}$$

$$U_i = \text{טעות מקרית המקיימת את כל ההנחות הקלאסיות.}$$

#### מקדמי מודל הרגרסיה המרובה:

$$\alpha = \text{חותך אחד שמשמעותו: הציון המנובא כאשר כל המשתנים הב"ת} = 0.$$

$$\beta_1 \dots \beta_j = \text{מקדמי השיפוע. מס' הבטות} = \text{למספר המשתנים הב"ת במודל.}$$

משמעות מקדם השיפוע  $\beta_j$ : ההשפעה הייחודית של המשתנה הב"ת המסוים לניבוי המשתנה התלוי, בניכוי השפעתם של כל יתר המשתנים הב"ת האחרים המצויים במשוואת הרגרסיה.

#### אמידת מודל הרגרסיה המרובה:

ברגרסיה מרובה, כמו ברגרסיה פשוטה, שיטת האמידה הטובה ביותר היא שיטת הריבועים הפחותים.

כלומר, נרצה להביא את סכום הטעויות בניבוי למינימום:

$$\text{Min} \sum e_i^2 = \text{Min} \sum (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_j X_{ji})^2$$

מפיתרון פונקציית הריבועים הפחותים נקבל את אומדי הרגרסיה:  $\hat{\alpha}, \hat{\beta}_1 \dots \hat{\beta}_j$

**הפיתרון יהיה על פי נגזרות חלקיות לפי כל אומד:**

$$\frac{\partial}{\partial \hat{\alpha}} = -2 \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}) = 0$$

$$\frac{\partial}{\partial \hat{\beta}_1} = -2 \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}) X_{1i} = 0$$

$$\frac{\partial}{\partial \hat{\beta}_2} = -2 \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}) X_{2i} = 0$$

$$\frac{\partial}{\partial \hat{\beta}_3} = -2 \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}) X_{3i} = 0$$

•  
•  
•

$$\frac{\partial}{\partial \hat{\beta}_J} = -2 \sum_1^n (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}) X_{Ji} = 0$$

**אם נשתמש בזהות:  $e_i = Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \dots - \hat{\beta}_J X_{Ji}$  נקבל את**

**המשוואות הנורמאליות:**

$$\sum_{i=1}^n e_i = 0 \quad \text{בגלל שיש חותך}$$

$$\sum_{i=1}^n e_i X_{1i} = 0 \quad \text{בגלל שיש את } X_{1i}$$

$$\sum_{i=1}^n e_i X_{2i} = 0 \quad \text{בגלל שיש את } X_{2i}$$

$$\sum_{i=1}^n e_i X_{3i} = 0 \quad \text{בגלל שיש את } X_{3i}$$

•  
•  
•

$$\sum_{i=1}^n e_i X_{Ji} = 0 \quad \text{בגלל שיש את } X_{Ji}$$

**אם נפתור את המשוואות נקבל את הנוסחאות למציאת האומדים, אך זה מעבר לדרישות הקורס (הפיתרון הוא מטריציוני).**

דוגמא: מקרה פרטי, מודל עם שני משתנים מסבירים:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$

הנוסחאות הנורמאליות הן:

$$\sum_{i=1}^n e_i = 0$$

$$\sum_{i=1}^n e_i X_{1i} = 0$$

$$\sum_{i=1}^n e_i X_{2i} = 0$$

מפיתרון מערכת המשוואות נקבל את הנוסחאות הבאות לחישוב האומדים:

$$\hat{\alpha} = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$\hat{\beta}_1 = \frac{(r_{y1} - r_{y2} * r_{12}) * \hat{s}_y}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x1}}$$

$$\hat{\beta}_2 = \frac{(r_{y2} - r_{y1} * r_{12}) * \hat{s}_y}{1 - r_{12}^2} * \frac{\hat{s}_y}{\hat{s}_{x2}}$$

**\*\*הערה:** ניתן לראות כי אם לא קיים מתאם בין המשתנים הב"ת  $r_{12} = 0$ , שיפועי

הרגרסיה המרובה זהים לשיפועי הרגרסיה הפשוטה:

$$\hat{\beta}_1 = r_{y1} \cdot \frac{\hat{s}_y}{\hat{s}_{x1}} = \frac{\text{cov}(y, x_1)}{\text{var}(x_1)}$$

$$\hat{\beta}_2 = r_{y2} \cdot \frac{\hat{s}_y}{\hat{s}_{x2}} = \frac{\text{cov}(y, x_2)}{\text{var}(x_2)}$$

**?** כלכלן החליט לאמוד מודל ליניארי עם שלושה משתנים מסבירים  $x_1, x_2, x_3$

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

א. מהי בעיית ה-OLS שעליו לפתור?

ב. מצאו את תנאי סדר ראשון של הבעיה.

**?** כלכלן החליט לבחון מה משפיע על שער הדולר בישראל.

לכן אסף מדגם בין ארבע תצפיות חודשיות.

להלן טבלה מסכמת:

| חודש    | X2<br>השקעות זרים בישראל<br>(במיליוני דולרים) | X1<br>שער הריבית | Y<br>דולר | טעות (e <sub>i</sub> ) |
|---------|-----------------------------------------------|------------------|-----------|------------------------|
| אוגוסט  | 100                                           | 3                | 3.2       | -5                     |
| ספטמבר  | 95                                            | 3.5              | 3.6       | 6                      |
| אוקטובר | 90                                            | 3.5              | 3.8       | 0                      |
| נובמבר  | 100                                           | 3                | 3.5       | -2                     |

מהו המודל אשר אותו אמד הכלכלן?

**?** הניחו כי הקשר באוכלוסייה בין X ל-Y נתון ע"י המשוואה הבאה:

$$Y_i = 2 + \beta_1 X_{1i} + 5X_{2i} + u_i \text{ וכל ההנחות הקלאסיות מתקיימות.}$$

נתון האומד:

$$\tilde{\beta}_1 = \frac{\sum X_{1i}((Y_i - 8X_{2i} - 2) - (\bar{y} - 8\bar{X}_2 - 2))}{\sum X_{1i}^2}$$

א. חשבו את תוחלת האומד

ב. חשבו את שונות האומד

ג. מהו היחס בין שונות האומד הנ"ל, לבין שונות אומד הריבועים הפחותים?

**?** הניחו כי הקשר באוכלוסייה בין X ל-Y נתון ע"י המשוואה הבאה:

$$y_i = \alpha + \beta_1 x_{1i} + 8x_{2i} + u_i$$

כל ההנחות הקלאסיות מתקיימות וכן:  $\sum x_{1i} = 0$

אומדים את  $\beta_1$  באופן הבא:

$$b_1 = \frac{\sum (x_{1i} - \bar{x})(y_i - 8x_{2i} - (\bar{y} - 8\bar{x}_2))}{\sum (x_{1i} - \bar{x})^2}$$

א. האם האומד חסר הטיה?

ב. מהי שונות האומד?

## מולטיקוליניאריות

אחת הבעיות העיקריות היכולות לצוץ בגרסיה רבת משתנים נקראת "מולטיקוליניאריות".

מולטיקוליניאריות מתייחסת למתאם בין המשתנים המסבירים במודל. נבחין בין מולטיקוליניאריות מלאה לחלקית.

מולטיקוליניאריות מלאה

מתאם מלא בין המשתנים המסבירים במודל.

הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני:

$$x_1 = a + bx_2 \quad (x_1 \text{ הוא קומבינציה ליניארית מלאה של } x_2) \quad \text{מכאן ש: } r_{12} = 1$$

שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל

$$x_1 = x_2^2), \text{ אז בהכרח } r_{12} \neq 1.$$

**\*\*הערה:** מולטיקוליניאריות מלאה יכולה להיווצר גם כאשר קבוצה של משתנים מסבירים מהווה קומבינציה ליניארית מלאה של אחד המשתנים המסבירים:

$$x_1 + x_2 = a + bx_3$$

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני.

**לדוגמא:**

נניח שאנו רוצים לאמוד את הביקוש לדירות בתל אביב כפונקציה של מחירן

בשקלים ( $x_1$ ) ובדולרים ( $x_2$ ):

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + u_t$$

בהנחה ששער הדולר נותר קבוע, המחיר בדולרים מהווה קומבינציה ליניארית מלאה של המחיר בשקלים (המחיר בשקלים \* שער הדולר). במקרה כזה לא ניתן להפריד את ההשפעה של שני המשתנים הב"ת זה מזה ומדובר בעצם באותו המשתנה.

מדוע זה בעייתי?

כיוון שלא ניתן לאמוד את המודל שכן אר"פ אינם מוגדרים.

הסבר: בגזירת אר"פ, נוצר מצב של תלות ליניארית בין המשוואות הנורמאליות וחלקן יתבטלו. ניוותר עם יותר נעלמים ממשוואות ועם אינסוף פתרונות, כך שלא נוכל להגדיר את האומדים.

**בדוגמא שלנו:**

$$Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

$$\sum e_i = 0 : \hat{\alpha}$$

$$\sum e_i x_{1i} = 0 : \hat{\beta}_1$$

$$\sum e_i x_{2i} = 0 : \hat{\beta}_2$$

$$\sum e_i (bx_{2i}) = 0$$

$$\text{מכיון ש: } x_1 = bx_2 \text{ (שער הדולר) אז גזירת } \hat{\beta}_1 : b \sum e_i x_{2i} = 0 \text{ והמשוואה}$$

$$0 = 0$$

מתבטלת.

פיתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקוליניאריות חלקית

כאשר יש מתאם גבוה מאוד (אך לא מושלם) בין 2 משתנים מסבירים במודל או בין קבוצה של משתנים מסבירים עלולה להיווצר בעיה של מולטיקוליניאריות חלקית:

$$x_1 = a + bx_2 + u_i$$

$$x_1 + x_2 = a + bx_3 + u_i$$

**לדוגמא:**

נניח שאנו רוצים לאמוד את הציונים בתואר ראשון ע"י ציוני הפסיכומטרי ( $x_1$ )

וציוני הבגרות ( $x_2$ ):

$$Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$$

מכיון שיש מתאם גבוה בין ציוני הפסיכומטרי וציוני הבגרות לא נוכל לבדוד באופן

מלא את ההשפעה המדויקת של כל אחד מהם על ציוני ה-B.A.

כל אחד מהמשתנים הב"ת "יגזול" מן ההשפעה הייחודית שיש למשתנה הב"ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב"ת יהיה מובהק, התרומה הייחודית של כל משתנה ב"ת לניבוי התלוי לא תהיה מובהקת.

**? נתון המודל :** 
$$Y_i = \alpha + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + U_i$$

חוו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

א. בהנחה כי מתקיים:  $X_{1i} - 2X_{2i} = 1$

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

ב. בהנחה כי מתקיים:  $x_{1i} = x_{2i}^2$

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

ג. הוכיחו תשובותיכם לסעיפים ב ו- ג.

ד. בהנחה כי מתקיים:  $r_{12} = 0.98$

1. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

2. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

**? כלכלן אמד את המודל** 
$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$$

בשל החשש ממולטיקוליניאריות בחן הכלכלן את המתאם בין כל זוג של משתנים מסבירים וקיבל:

$$r_{x1,x2} = 0.9, \quad r_{x1,x3} = 0.99, \quad r_{x3,x2} = 0.5$$

לכן הסיק כי אין בעיה של מולטיקוליניאריות מושלמת במודל. האם הוא צודק?

**? כלכלן אמד את המודל הבא:**

$$\ln(Q_i) = \alpha + \beta_1 \ln(K_i) + \beta_2 \ln(K_i^2) + \beta_3 L_i^{0.5} + u_i$$

האם קיימת בעיה של מולטיקוליניאריות במודל?

? להלן מודל של שכר  $W_i$ , כפונקציה של שנות לימוד  $S_i$  ושל גיל  $A_i$ :

$$W_i = \alpha + \beta_1 \cdot S_i + \beta_2 \cdot A_i + u_i \quad (1)$$

בנוסף למשתנים במשוואה, החליט החוקר להוסיף גם את משתנה הוותק:  $EXP_i$ . מכיוון שלא היו בידו נתונים על הוותק, החליט החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים (מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).

להלן משוואה מס' 2:

$$W_i = \alpha + \beta_1 \cdot S_i + \beta_2 \cdot A_i + \beta_3 \cdot EXP_i + w_i \quad (2)$$

חוזה דעתך על המשוואה השנייה.

## פרק 5 - מבחן t

עד מעתה למדנו לאמוד את מקדמי הרגרסיה ( $\alpha$ -ה- ו- $\beta$ ) על סמך מדגם מקרי יחיד. כעת נלמד כיצד לבדוק האם המקדמים שאמדנו מובהקים באוכלוסיה.

המבחן הסטטיסטי שנבצע למובהקות מקדמי הרגרסיה נקרא מבחן t. נלמד כיצד לבצע מבחן t למובהקות ה- $\beta$ , ה- $\alpha$  כמו גם למובהקות קשרים ליניאריים בין המקדמים (t מורכב).

### מבחן למובהקות ה- $\beta$ (מקדם השיפוע)

מבחן העונה לשאלה: האם משתנה מסביר מסוים רלוונטי למודל/ משפיע על המשתנה התלוי (מובהק)?

השערות:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

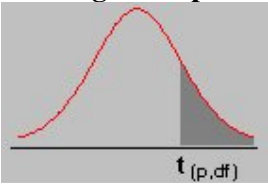
סטטיסטי המבחן:

$$t_{\beta=0} = \frac{\hat{\beta} - \beta_0}{S_{\hat{\beta}}}$$

כלל הכרעה ומסקנה:

נדחה את  $H_0$  אם:

שימוש בטבלת T:  $|t_{\beta=0}| > t_{(n-K, \frac{\alpha}{2})}$

| t table with right tail probabilities                                               |          |          |          |          |          |          |          |          |
|-------------------------------------------------------------------------------------|----------|----------|----------|----------|----------|----------|----------|----------|
|  |          |          |          |          |          |          |          |          |
| df\p                                                                                | 0.40     | 0.25     | 0.10     | 0.05     | 0.025    | 0.01     | 0.005    | 0.0005   |
| 1                                                                                   | 0.324920 | 1.000000 | 3.077684 | 6.313752 | 12.70620 | 31.82052 | 63.65674 | 636.6192 |
| 2                                                                                   | 0.288675 | 0.816497 | 1.885618 | 2.919986 | 4.30265  | 6.96456  | 9.92484  | 31.5991  |
| 3                                                                                   | 0.276671 | 0.764892 | 1.637744 | 2.353363 | 3.18245  | 4.54070  | 5.84091  | 12.9240  |

|    |          |          |          |          |         |         |         |        |
|----|----------|----------|----------|----------|---------|---------|---------|--------|
| 4  | 0.270722 | 0.740697 | 1.533206 | 2.131847 | 2.77645 | 3.74695 | 4.60409 | 8.6103 |
| 5  | 0.267181 | 0.726687 | 1.475884 | 2.015048 | 2.57058 | 3.36493 | 4.03214 | 6.8688 |
| 6  | 0.264835 | 0.717558 | 1.439756 | 1.943180 | 2.44691 | 3.14267 | 3.70743 | 5.9588 |
| 7  | 0.263167 | 0.711142 | 1.414924 | 1.894579 | 2.36462 | 2.99795 | 3.49948 | 5.4079 |
| 8  | 0.261921 | 0.706387 | 1.396815 | 1.859548 | 2.30600 | 2.89646 | 3.35539 | 5.0413 |
| 9  | 0.260955 | 0.702722 | 1.383029 | 1.833113 | 2.26216 | 2.82144 | 3.24984 | 4.7809 |
| 10 | 0.260185 | 0.699812 | 1.372184 | 1.812461 | 2.22814 | 2.76377 | 3.16927 | 4.5869 |
| 11 | 0.259556 | 0.697445 | 1.363430 | 1.795885 | 2.20099 | 2.71808 | 3.10581 | 4.4370 |
| 12 | 0.259033 | 0.695483 | 1.356217 | 1.782288 | 2.17881 | 2.68100 | 3.05454 | 4.3178 |
| 13 | 0.258591 | 0.693829 | 1.350171 | 1.770933 | 2.16037 | 2.65031 | 3.01228 | 4.2208 |
| 14 | 0.258213 | 0.692417 | 1.345030 | 1.761310 | 2.14479 | 2.62449 | 2.97684 | 4.1405 |
| 15 | 0.257885 | 0.691197 | 1.340606 | 1.753050 | 2.13145 | 2.60248 | 2.94671 | 4.0728 |
| 16 | 0.257599 | 0.690132 | 1.336757 | 1.745884 | 2.11991 | 2.58349 | 2.92078 | 4.0150 |
| 17 | 0.257347 | 0.689195 | 1.333379 | 1.739607 | 2.10982 | 2.56693 | 2.89823 | 3.9651 |
| 18 | 0.257123 | 0.688364 | 1.330391 | 1.734064 | 2.10092 | 2.55238 | 2.87844 | 3.9216 |
| 19 | 0.256923 | 0.687621 | 1.327728 | 1.729133 | 2.09302 | 2.53948 | 2.86093 | 3.8834 |
| 20 | 0.256743 | 0.686954 | 1.325341 | 1.724718 | 2.08596 | 2.52798 | 2.84534 | 3.8495 |
| 21 | 0.256580 | 0.686352 | 1.323188 | 1.720743 | 2.07961 | 2.51765 | 2.83136 | 3.8193 |
| 22 | 0.256432 | 0.685805 | 1.321237 | 1.717144 | 2.07387 | 2.50832 | 2.81876 | 3.7921 |
| 23 | 0.256297 | 0.685306 | 1.319460 | 1.713872 | 2.06866 | 2.49987 | 2.80734 | 3.7676 |
| 24 | 0.256173 | 0.684850 | 1.317836 | 1.710882 | 2.06390 | 2.49216 | 2.79694 | 3.7454 |
| 25 | 0.256060 | 0.684430 | 1.316345 | 1.708141 | 2.05954 | 2.48511 | 2.78744 | 3.7251 |
| 26 | 0.255955 | 0.684043 | 1.314972 | 1.705618 | 2.05553 | 2.47863 | 2.77871 | 3.7066 |
| 27 | 0.255858 | 0.683685 | 1.313703 | 1.703288 | 2.05183 | 2.47266 | 2.77068 | 3.6896 |
| 28 | 0.255768 | 0.683353 | 1.312527 | 1.701131 | 2.04841 | 2.46714 | 2.76326 | 3.6739 |
| 29 | 0.255684 | 0.683044 | 1.311434 | 1.699127 | 2.04523 | 2.46202 | 2.75639 | 3.6594 |
| 30 | 0.255605 | 0.682756 | 1.310415 | 1.697261 | 2.04227 | 2.45726 | 2.75000 | 3.6460 |

|     |          |          |          |          |         |         |         |        |
|-----|----------|----------|----------|----------|---------|---------|---------|--------|
| inf | 0.253347 | 0.674490 | 1.281552 | 1.644854 | 1.95996 | 2.32635 | 2.57583 | 3.2905 |
|-----|----------|----------|----------|----------|---------|---------|---------|--------|

**מסקנה:** יש/אין עדות לכך שהמשתנה הב"ת המסוים מובהק באוכ' (ולכן רלוונטי למודל).

**הערות:**

- ניתן גם לבצע מבחן מובהקות חד צדדי ל-  $\beta$  הנותן מענה על השאלה: האם מקדם השיפוע או הקשר בין המשתנים הוא חיובי או שלילי באוכ'?

**ההשערות:**

$$H_0 : \beta = 0$$

$$H_1 : \beta > / < 0$$

**כלל ההכרעה:**

**שימוש בטבלת T:**

$$t_{\beta=0} > t_{(n-K, \alpha)}$$

$$t_{\beta=0} < -t_{(n-K, \alpha)}$$

- ניתן לבדוק בנוסף האם ה- $\beta$  (השינוי השולי) שווה לערך מסוים באוכ'.

**השערות לדוגמא:**

$$H_0 : \beta = 2$$

$$H_1 : \beta \neq 2$$

**סטטיסטי המבחן:**

$$t_{\beta=2} = \frac{\hat{\beta} - 2}{S_{\hat{\beta}}}$$

**כלל הכרעה:**

**על סמך טבלת T**

- רב"ס ל- $\beta$ :**

$$P(\hat{\beta} - t_{(n-K, \frac{\alpha}{2})} \cdot S_{\hat{\beta}} \leq \beta \leq \hat{\beta} + t_{(n-K, \frac{\alpha}{2})} \cdot S_{\hat{\beta}}) = 1 - \alpha$$

ניתן לבדוק השערות באמצעות הרב"ס. צריך לבדוק האם הרב"ס מכיל את הערך המבוקש (את  $\beta_0$ ) אם כן- נקבל את  $H_0$  ואם לא-נדחה אותה.

### מבחן למובהקות ה- $\alpha$ (החותך)

עונה על השאלה- האם קו הרגרסיה יוצא מראשית הצירים (החותך של קו

הרגרסיה=0):

השערות:

$$H_0 : \alpha = 0$$

$$H_1 : \alpha \neq 0$$

סטטיסטי המבחן:

$$t_{\alpha=0} = \frac{\hat{\alpha} - \alpha_0}{S_{\hat{\alpha}}}$$

כלל הכרעה ומסקנה:

נדחה את  $H_0$  אם:

$$|t_{\alpha=0}| > t_{(n-K, \frac{\alpha}{2})} : T$$

מסקנה: יש/אין עדות לכך שקו הרגרסיה עובר דרך ראשית הצירים

- ניתן גם לבצע מבחן מובהקות חד צדדי ל- $\alpha$  הנותן מענה על השאלה: האם החותך הוא חיובי או שלילי באוכ'?

ההשערות:

$$H_0 : \alpha = 0$$

$$H_1 : \alpha > / < 0$$

כלל ההכרעה:

שימוש בטבלת T:

$$t_{\alpha=0} > t_{(n-K, \alpha)}$$

$$t_{\alpha=0} < -t_{(n-K, \alpha)}$$

- ניתן לבדוק בנוסף האם ה- $\alpha$  (החותך) שווה לערך מסוים באוכ'.

השערות לדוגמא:

$$H_0 : \alpha = 2$$

$$H_1 : \alpha \neq 2$$

סטטיסטי המבחן:

$$t_{\alpha=2} = \frac{\hat{\alpha} - 2}{S_{\hat{\alpha}}}$$

כלל הכרעה:

על סמך טבלת T :

נדחה את H0 אם:

$$|t_{\alpha=2}| > t_{(n-K, \frac{\alpha}{2})}$$

• רב"ס ל- $\alpha$ :

$$P(\hat{\alpha} - t_{(n-K, \frac{\alpha}{2})} \cdot S_{\hat{\alpha}} \leq \alpha \leq \hat{\alpha} + t_{(n-K, \frac{\alpha}{2})} \cdot S_{\hat{\alpha}}) = 1 - \alpha$$

ניתן לבדוק השערות באמצעות הרב"ס. צריך לבדוק האם הרב"ס מכיל את הערך המבוקש (את  $\alpha_0$ ) אם כן- נקבל את H0 ואם לא-נדחה אותה.

**?** חוקר רצה לבחון את השפעת ההכנסה (INCOME) על גובה המס (TAX)

(במיליארדי \$) שגובה מדינה במערב לפי המודל :  $TAX_i = \alpha + \beta \cdot INCOME_i + u_i$

לשם כך אסף נתונים מ-51 מדינות . להלן התוצאות:

$$TAX_i = -0.086912 + 0.152232 \cdot INCOME_i$$

(0.08953) (0.01622)

סטיות התקן של האומדים נתונות בסוגריים.

א. מהי המשמעות הכלכלית של  $\beta$  ושל  $\alpha$ ?

ב. האם ההכנסה משפיעה על גודל המס? בדקו ברמת מובהקות של 5%.

ג. בדקו את ההשערה כי כאשר ההכנסה אפסית, גודל המס שונה מ-0 באוכלוסיה.

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס ברמת

מובהקות של 5% וברמת מובהקות של 1%

ה. בנו רווח-סמך לשיפוע הרגרסיה ברמת ביטחון של 95%.

ו. בדקו את ההשערה שתוספת של מיליארד \$ להכנסה תגדיל את המס ב-

0.2 מיליארד \$, ברמת מובהקות של 0.05.

**? חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר (SALARY) לפי**

**המודל:**  $\ln(\text{SALARY}_i) = \alpha + \beta \cdot \text{EXP}_i + u_i$ . הוא אסף 403 תצפיות, ואמד את

**הפרמטרים. להלן תוצאות האמידה:**

$$\ln(\text{SALARY})_i = 7.334 - 0.0087 \cdot \text{EXP}_i$$

(0.068) (0.0026)

**א. האם קיים קשר חיובי מובהק בין ותק ללוג השכר?**

**ב. בדוק את ההשערה כי שיעור התשואה בשכר לשנת ותק קטנה מ: -0.9**

**ג. מהי תחזית השכר עבור אדם בעל 10 שנות ותק?**

**? נאמד המודל  $Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i + \beta_3 W_i + \beta_4 S_i + u_i$  והתקבלו התוצאות הבאות:**

$$\hat{y}_i = 5.06 + 0.97x_i + 3z_i - 5.02w_i + 8.97s_i$$

(0.456) (0.42) (0.08) (0.7) (0.29)

**א. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01**

**ב. בנו רווח בר סמך להשפעת X על Y**

### תחזית

המטרה של קו הרגרסיה הוא ביצוע תחזיות. תחזית נקודתית מחושבת על פי קו

הרגרסיה שאמדנו. נציב במקום ה-Xים ערכים נתונים ונקבל למה שווה ה-Y

המנובא.

**לדוגמא:**

**נתונה משוואת הרגרסיה הבאה:**

$$\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i}$$

כאשר  $y_i$  הינו סה"כ הוצאות משק בית i לחינוך לשבוע,  $x_{ji}$  הינו גילו של הילד j

מה יהיה סה"כ הוצאות משק הבית אם גיל הילד הראשון הוא 2 שנים, של השני

4.5 שנים, השלישי הוא בן 5 ואילו הרביעי בן 8?

**תשובה:**

**נציב במשוואת הרגרסיה:**

$$\hat{y}_i = 13 + 8 \cdot 2 + 7 \cdot 4.5 + 2 \cdot 5 + 9 \cdot 8 = 142.5$$

כלומר תחזית הוצאות משק הבית לחינוך עבור גילאי הילדים הנ"ל תהיה 142.5 ₪ לשבוע.

ברגרסיה פשוטה, כאשר המודל הינו:  $Y_i = \alpha + \beta X_i + u_i$  אנו יכולים גם לאמוד את

התחזית באוכלוסיה עבור ערך מסוים של  $X$ , באמצעות רווח בר סמך:

• רווח בר סמך לתחזית עבור  $X_f$  מסוים:

$$\hat{Y} \pm t_{(n-2, \frac{\alpha}{2})} S_u \sqrt{1 + \frac{1}{n} + \frac{(X_f - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

$$S_u^2 = MSE = \frac{SSE}{n-2} \quad \sum (X_i - \bar{X})^2 = S_{xx} = (n-1)S_x^2$$

רישום הרב"ס:  $p(\text{---} \leq Y \leq \text{---}) = 1 - \alpha$

התחזית מדויקת יותר (שונות התחזית קטנה יותר) כאשר:

1.  $n$  (גודל המדגם) גדול יותר
2. שונות המשתנה המסביר  $X$  גדולה יותר
3.  $X_f$  קרוב יותר ל  $\bar{X}$
4. האומד לשונות הטעויות  $S_u$ , קטן יותר.

**לדוגמא:**

במדגם של 30 דירות המושכרות לסטודנטים ברדיוס של עד 2 ק"מ מסביב למכללה נחקר הקשר בין שכר דירה למספר הסטודנטים הגרים בדירה. להלן תוצאות האמידה:

$$\hat{Y}_i = 686.207 + 233.52 \cdot X_i$$

נתון בנוסף כי :

$$S_x^2 = 1.313^2$$

$$S_u^2 = 414.055^2$$

$$\bar{x} = 3$$

1. חשבו אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.

תשובה:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

יש לחשב תחזית נקודתית עבור  $X=2$ :

$$\hat{Y}_{x=2} = 686.207 + 233.52 \cdot 2 = 1153.247$$

2. אמוד את שכר הדירה שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.

**תשובה:**

יש לחשב רב"ס לתחזית עבור  $X=2$ ,  $1 - \alpha = 0.95$ :

$$\hat{Y} \pm t_{n-2, \frac{\alpha}{2}} S_u \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$

**חישובי עזר:**

$$\sum (X_i - \bar{X})^2 = S_{xx} = (n-1)S_x^2 = (30-1) \cdot 1.31306^2 = 49.99$$

$$t_{28, 0.025} = 2.048$$

$$1153.247 \pm 2.048 \cdot 414.055 \sqrt{1 + \frac{1}{30} + \frac{(2-3)^2}{49.99}}$$

$$1153.247 \pm 870.305$$

$$p(282.94 \leq Y_{x=2} \leq 2023.55) = 0.95$$

### מבחן t מורכב (בחינת קשרים ליניאריים בין הפרמטרים)

לעיתים אנחנו מתבקשים לבדוק השערות העוסקות בקשרים בין הפרמטרים. כמו

למשל:  $H_0: \alpha = 5\beta$  או  $H_0: \beta_1 = 2 \cdot \beta_2$ .

במקרים אלו נרשום את השערות האפס כך:  $H_0: \alpha - 5\beta = 0$  ו-  $H_0: \beta_1 - 2 \cdot \beta_2 = 0$

ונחשב את סטטיסטי המבחן t:

$$t_{\hat{\beta}_1 - 2\hat{\beta}_2} = \frac{(\hat{\beta}_1 - 2\hat{\beta}_2) - 0}{S_{\hat{\beta}_1 - 2\hat{\beta}_2}} \quad \text{או} \quad t_{\hat{\alpha} - 5\hat{\beta}} = \frac{(\hat{\alpha} - 5\hat{\beta}) - 0}{S_{\hat{\alpha} - 5\hat{\beta}}}$$

כאשר את טעות התקן של המבחן מחשבים תוך שימוש בנוסחאות:

$$V(X \pm Y) = V(X) + V(Y) \pm 2\text{cov}(X, Y)$$

$$V(aX) = a^2 V(X)$$

$$\text{cov}(aX, bY) = a \cdot b \cdot \text{cov}(X, Y)$$

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.

לשם כך יש לקבל נתונים על השונות המשותפות של הפרמטרים (cov).

**דוגמא:**

$$Y_t = \alpha + \beta X_t + u_t \quad \text{נתון המודל}$$

שלצורך אמידתו נאספו 240 תצפיות ונתקבל ש:

$$\hat{Y}_i = 5.25 + 0.96 X_i$$

$$(0.25) \quad (0.12)$$

$$\text{cov}(\hat{\alpha}, \hat{\beta}) = -0.003 \quad \text{נתון בנוסף כי:}$$

$$H_0: \alpha = 5\beta \quad \text{יש לבדוק את ההשערה:}$$

**תשובה:**

$$H_0: \alpha - 5\beta = 0$$

$$H_1: \alpha - 5\beta \neq 0$$

**סטטיסטי המבחן:**

$$t = \frac{(\hat{\alpha} - 5\hat{\beta}) - 0}{S_{\hat{\alpha} - 5\hat{\beta}}} = \frac{0.45}{0.6726} = 0.669$$

**חישובי עזר:**

$$\hat{\alpha} - 5\hat{\beta} = 5.25 - 5 \cdot 0.96 = 0.45 \quad \text{נציב את האומדים:}$$

$$(\hat{S}_{\hat{\alpha} - 5\hat{\beta}}^2) \quad \text{חישוב השונות:}$$

$$V(\hat{\alpha} - 5\hat{\beta}) = V(\hat{\alpha}) + V(5\hat{\beta}) - 2\text{cov}(\hat{\alpha}, 5\hat{\beta}) =$$

$$V(\hat{\alpha}) + 5^2 V(\hat{\beta}) - 2 \cdot 5 \text{cov}(\hat{\alpha}, \hat{\beta}) =$$

$$0.0625 + 25 \cdot 0.0144 - 10 \cdot (-0.003) = 0.4525$$

$$S_{\hat{\alpha} - 5\hat{\beta}} = \sqrt{0.4525} = 0.6726 \quad \text{סטית התקן היא:}$$

$$\text{כלל הכרעה: } t_{\hat{\alpha} - 5\hat{\beta}} = 0.66 < t_{(238, 0.025)} = 1.96, \quad \text{לכן אין סיבה מספקת לדחות את}$$

**השערת האפס.**

**מסקנה:** אין עדות לכך שה-  $\alpha \neq 5\beta$

**?** על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת

2007 ונאמדה המשוואה הבאה:

$$C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$$

להלן תוצאות האמידה של המשוואה הנ"ל:

$$C_i = -107.226 + 0.743W_i + 0.561P_i$$

$$(0.0678) \quad (0.4)$$

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.009 \quad \text{ש: גמ}$$

יש לבדוק את ההשערה שהנטיה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטיה השולית לצרוך מתוך ההון.

תשובה:

$$\beta_1 = \beta_2 \quad \text{יש לבדוק את ההשערה כי:}$$

השערות:

$$H_0: \beta_1 - \beta_2 = 0$$

$$H_1: \beta_1 - \beta_2 \neq 0$$

חישוב סטטיסטי המבחן:

$$t_{\hat{\beta}_1 - \hat{\beta}_2} = \frac{(\hat{\beta}_1 - \hat{\beta}_2) - 0}{S_{\hat{\beta}_1 - \hat{\beta}_2}} = \frac{0.182}{0.194} = 0.938$$

חישובי עזר:

$$\hat{\beta}_1 - \hat{\beta}_2 = 0.743 - 0.561 = 0.182$$

$$\begin{aligned} S_{\hat{\beta}_1 - \hat{\beta}_2}^2 &= S_{\hat{\beta}_1}^2 + S_{\hat{\beta}_2}^2 - 2 \cdot S_{\hat{\beta}_1} \cdot S_{\hat{\beta}_2} \cdot \text{corr}(\hat{\beta}_1, \hat{\beta}_2) \\ &= 0.0046 + 0.016 - 2 \cdot (-0.009) = 0.038 \end{aligned}$$

$$S_{\hat{\beta}_1 - \hat{\beta}_2} = \sqrt{0.038} = 0.194$$

כלל הכרעה:

$$t_{\hat{\beta}_1 - \hat{\beta}_2} = 0.938 < t_{(39, 0.025)} = 2$$

מסקנה: אין עדות לכך שהנטיה השולית לצריכה הן שונות.

## תירגול מסכם

**? כלכלן בנה עבור מכבי ת"א מודל החוזה את השכר שיש לשלם לשחקן כדורסל**

**לחוזה של שנה:**

$$\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + u_i$$

**כאשר:**

**Y: שכר השחקן באלפי \$**

**X1: מס' נקודות שקולע השחקן בממוצע למשחק**

**X2: מס' האסיסטים שיש לשחקן בממוצע למשחק**

**X3: מס' הדקות שיושב שחקן על הספסל בממוצע למשחק.**

**הכלכלן דגם 34 משחקים וקיבל את התוצאות הבאות:**

$$\hat{Y}_i = 120 + 18X_{1i} + 8X_{2i} - 22X_{3i}$$

$$(2.2) \quad (3) \quad (4.4) \quad (-5)$$

**\*\*הערכים שבסוגריים הם ערכי t.**

**התקבל בנוסף כי:**

$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = 4, \text{cov}(\hat{\beta}_1, \hat{\beta}_3) = -3, \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -6$$

**א. תנו פירוש למקדמי הרגרסיה.**

**ב. איזה מהמשתנים הב"ת רלוונטי למודל?**

**ג. בנו רב"ס למשתנים המובהקים.**

**ד. מייקל ג'ורדן הצטרף למכבי והוא דורש 2 מיליון \$ לעונה.**

**ידוע כי מייקל קולע 45 נקודות בממוצע למשחק, מוסר 15 אסיסטים**

**בממוצע למשחק ויושב 5 דקות בממוצע על הספסל. כמה צריך לשלם לו?**

**ה. לטענת שימרון מזרחי מס' הנקודות הממוצע שקולע שחקן למשחק צריך**

**להשפיע פי 4 ממספר האסיסטים הממוצע שלו. האם הוא צודק?**

$$\ln(Q_i) = \alpha + \beta_1 \ln(K_i) + u_i \quad \text{? כלכלן אמד את המודל הבא:}$$

**שמתאר את הקשר שבין צריכת מוצר מסוים להכנסת הפרט (עקומת אנג'ל):**

**K = הכנסה חודשית באלפי שקלים**

**Q = צריכה שנתית באלפי שקלים**

לשם כך אסף 60 נתונים והריץ רגרסיה . התוצאות אשר קיבל הן:

$$S_K = 1.5, S_Q = 0.05, \text{cov}(\hat{\alpha}, \hat{\beta}) = -0.05, t_{\hat{\alpha}} = 3, t_{\hat{\beta}} = -7, \hat{\alpha} = 4, \hat{\beta} = -2$$

הממוצעים הינה: (6.7, 0.4)

א. הכלכלן ביקש לבדוק את ההשערה כי הגמישות במודל יחידתית ושווה

ל:1-

ב. בידקו את ההשערה כי מקדם החיתוך של קו הרגרסיה הוא כפול ממקדם

השיפוע.

ג. חיים משתכר בממוצע לחודש 10,000 ₪ , כמה ישקיע בצריכת המוצר

בשנה ?

ד. בנו רב"ס לתחזית הצריכה של חיים באוכלוסיה.

## פרק 6 - $R^2$ - מדד לטיב הרגרסיה ומבחן F

בפרק זה נלמד כיצד לבדוק האם הרגרסיה שווה משהו והאם ניתן לסמוך עליה בתחזיות.

### מדד $R^2$ לטיב הרגרסיה

מדד  $R^2$  לטיב ההתאמה או לטיב הרגרסיה עונה על השאלה: איזה אחוז מהשונות של המשתנה התלוי ( $Y$ ) מוסבר על ידי קו הרגרסיה (ההיגיון הכלכלי ה- $X$ ים)?  
מדד לפרופורציית השונות המוסברת:

$$R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS}$$

ניתן לכתוב את הנוסחה גם כך:

$$R^2 = \frac{\sum (\hat{Y}_i - \bar{Y})^2}{\sum (Y_i - \bar{Y})^2} = 1 - \frac{\sum e_i^2}{\sum (Y_i - \bar{Y})^2}$$

נוסחה זו מתבססת על הנוסחה לפירוק השונות של קו הרגרסיה:

$$TSS = RSS + ESS$$

$$\sum (Y_i - \bar{Y})^2 = \sum (\hat{Y}_i - \bar{Y})^2 + \sum e_i^2$$

תכונות  $R^2$ :

- מדד  $R^2$  (פרופורציית השונות המוסברת) נע בין 0 ל-1:

$$0 \leq R^2 \leq 1$$

ככל שקרוב יותר ל-1 ההתאמה טובה יותר ולהיפך, ככל שקרוב יותר ל-0 ההתאמה גרועה יותר.

כאשר  $R^2 = 1$  ההתאמה מושלמת ואין שום טעויות בניבוי במודל ואילו

כאשר  $R^2 = 0$  הכל טעות ואין שום הסבר במודל (זה קורה כשאין שום

משתנה מסביר במודל ויש רק חותך:  $y_i = \alpha + u_i$ ).

- אר"פ מביא למקסימום את  $R^2$
- לא ניתן להשוות בין מודלים שבהם אין את אותו משתנה מוסבר.
- בהוספת משתנים מסבירים נוספים למודל,  $R^2$  יכול רק לעלות או "להתנפח" או לכל היותר להשאר ללא שינוי (כתוצאה מהירידה בשונות הלא מוסברת  $\sum e_i^2$ ). זהו למעשה החיסרון הגדול של המדד.

כדי להתגבר על חיסרון זה קיים מדד נוסף והוא:  $R^2_{adj}$  ( $R^2$  מתוקן):

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}$$

$K = \text{מס' הפרמטרים במודל (כולל החותך)}$

ככלל מתקיים תמיד ש:  $\bar{R}^2 < R^2$

המדד המתוקן לפרופורצית השונות המוסברת, לוקח את  $R^2$  הרגיל "ומעניש" אותו על מספר המשתנים הב"ת שיש במודל. לכן, בניגוד ל-  $R^2$ , ה-  $\bar{R}^2$  יכול לרדת בהוספת משתנים למודל ולא רק לעלות.

משום כך, המדד המתוקן-  $\bar{R}^2$  עדיף על המדד של  $R^2$  בכדי לבחון האם כדאי לנו להוסיף משתנים ב"ת למודל.

זהויות שכדאי לדעת לגבי  $R^2$ :

במודל רגרסיה פשוטה:  $y_i = \alpha + \beta x_i + u_i$  מתקיים:

$$1. R^2 = r_{yx}^2$$

$$2. r_{yx} = \hat{\beta} \frac{S_x}{S_y}$$

$$3. R^2 = \hat{\beta}^2 \frac{S_x^2}{S_y^2}$$

במודלים :  $y_i = \alpha_1 + \beta_1 x_i + u_i$  מתקיים:  
 $x_i = \alpha_2 + \beta_2 y_i + \varepsilon_i$

1. הם בעלי אותו  $R^2$

2.  $R^2 = \beta_1 \cdot \beta_2$

? דרגו את המודלים הבאים (לפי קריטריון  $R^2$ ):

1.  $y_i = \alpha + \beta x_{1i} + u_i \quad R^2 = 0.15$
2.  $y_i = \alpha + u_i$
3.  $y_i = \beta x_{1i} + u_i$
4.  $y_i = \alpha + \beta x_{1i} + \beta_2 x_{2i} + \varepsilon_i$
5.  $y_i = \alpha + \beta_2 x_{2i} + u_i \quad R^2 = 0.20$

? על סמך מדגם של 100 תצפיות נאמדו המודלים הבאים:

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \quad (1)$$

$$R^2 = 0.70 \quad \hat{y}_i = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i} \quad (2)$$

$$R^2 = 0.65 \quad \hat{y}_i = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2i} \quad (3)$$

1) שלושה חוקרים העלו טענה לגבי מקדם  $R^2$  של משוואה מס' (1):

א. אי אפשר לדעת מהנתונים המובאים לעיל אם  $R^2$  של משוואה (1) הוא גדול או קטן מ-0.70.

ב. אי אפשר לדעת מהנתונים המובאים לעיל אם  $R^2$  של משוואה (1) הוא גדול או קטן מ-0.65.

ג. ניתן לצפות כי  $R^2$  של משוואה (1) יהיה גדול מ-0.70.

בהתייחס לטענות החוקרים ניתן לומר:

1. רק הטענה של חוקר א נכונה
2. רק הטענה של חוקר ב נכונה
3. רק הטענה של חוקר ג נכונה
4. כל הטענות שגויות

2) חוו דעתכם על הטענות הבאות המתייחסות ל- $\bar{R}^2$ :

א. ניתן לצפות ש- $\bar{R}^2$  של משוואה (1) יהיה גדול מ-0.7 :

נכון/לא נכון/ לא ניתן לדעת

ב. ניתן לצפות כי  $\bar{R}^2$  של משוואה (2) יהיה קטן מ-0.7:

נכון/לא נכון/ לא ניתן לדעת

ג. ניתן לצפות כי  $\bar{R}^2$  של משוואה (3) יהיה קטן מ-0.7:

נכון/לא נכון/ לא ניתן לדעת

**?** על סמך מדגם של 80 משפחות המונות כל אחת 4 ילדים, נאמדו המשוואות

הבאות:

$$R^2 = 0.77 \quad \hat{y}_i = 5 + 2x_{1i} + 2x_{2i} \quad (1)$$

$$R^2 = 0.62 \quad \hat{y}_i = 24 + 0.8x_{1i} \quad (2)$$

$$R^2 = 0.25 \quad \hat{y}_i = 14 + 0.7x_{2i} \quad (3)$$

$$R^2 = 0.30 \quad \hat{y}_i = 4 + 0.5w_i \quad (4)$$

$$R^2 = 0.45 \quad \ln(y)_i = 7 + 0.9x_{1i} + 0.6x_{2i} \quad (5)$$

$$\ln(y)_i = 11 + 0.7x_{1i} + 0.9x_{2i} + 0.6x_{3i} \quad (6)$$

$$\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i} \quad (7)$$

כאשר  $y_i$  הינו סה"כ הוצאות משק בית  $i$ ,  $x_{ji}$  הינו גילו של הילד  $j$ , ונתון כי

$$w_i = 2x_{3i} + x_{1i} - x_{2i}$$

דרגו את הרגרסיות לפי קריטריון  $R^2$  (הימני עדיף על השמאלי)

**?** נתונות שתי המשוואות הבאות:

$$y_i = 58 + b_1x_i + e_{1i} \quad \text{ו-} \quad x_i = a_2 - 0.2y_i + e_{2i} \quad \text{כאשר} \quad \bar{y} = \bar{x} = 40$$

למה שווה מקדם המתאם של פירסון בין X ל-Y?

1. 0.09

2. 0.69

3. 0.3

4. 0.72

5. אף תשובה לא נכונה.

### מבחן F

מבחן F משמש אותנו לבדיקת מובהקות מודל הרגרסיה כולו כמו גם לבדיקת הגבלות שונות שאנו רוצים לבדוק האם מתקיימות במודל (מבחן WALD).

מבחן t הוא למעשה מקרה פרטי של מבחן F כאשר קיימת מגבלה אחת (או שיוויון אחד) ב-H0.

אז קיים הקשר בין המבחנים:

$$F = t^2$$

אם קיימת יותר ממגבלה אחת (יותר משיוויון אחד) ב-H0 רק מבחן F יתאים.

### ביצוע מבחן F (מבחן WALD)

אם רוצים לבדוק השערת אפס שיש בה מספר שוויונים משתמשים במבחן F (במבחן t היה רק שוויון אחד בהשערת האפס).  
איך עושים זאת?

1) אומדים את המודל המקורי. מודל זה נקרא המודל החופשי או המודל הלא-מוגבל, ובאנגלית Unrestricted. בתהליך האמידה של המודל הלא-מוגבל

מקבלים את סכום ריבועי הסטיות של הטעויות, נסמן אותו ב-  $\sum e_{iUR}^2$ .

2) מגדירים את כל השוויונים של השערת האפס.

3) מציבים את השוויונים של השערת האפס במודל המקורי. באופן הזה הופכים אותו למודל כפוי (כופים עליו את השערת האפס), או מודל מוגבל, או באנגלית Restricted.

4) אומדים את המודל המוגבל. בתהליך האמידה של המודל המוגבל מקבלים

את סכום ריבועי הסטיות של הטעויות, נסמן אותו ב-  $\sum e_{iR}^2$ .

5) אם יודעים את מספר התצפיות, n, מספר הפרמטרים במודל הלא-מוגבל, k, ומספר השוויונים או ההגבלות בהשערת האפס, m, אפשר לחשב את

הסטטיסטי:

$$\frac{(\sum e_{\gamma}^2 - \sum e_{\gamma\gamma}^2)/m}{\sum e_{\gamma\gamma}^2/(n-k)} \sim F_{(m, n-k, 1-\alpha)}$$

(כש-  $m$  מספר המגבלות)

כאשר לשתי הרגרסיות (המוגבלת והלא מוגבלת) אותו משתנה מוסבר אפשר גם :

$$\frac{(R_{\gamma\gamma}^2 - R_{\gamma}^2)/m}{(1 - R_{\gamma\gamma}^2)/(n-k)} \sim F_{(m, n-k, 1-\alpha)}$$

כלל הכרעה לדחיית  $H_0$ :

$$F_{stat} > F_{(m, n-k; 1-\alpha)}$$

אם דוחים את  $H_0$  המסקנה היא שהמודל המקורי (הלא-מוגבל) הוא הרלוונטי ולהיפך.

**?** נאמד המודל  $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$  והתקבל כי  $\sum e^2 = 620.1683$

וכי  $R^2 = 0.99$

הועלתה ההשערה כי ההשפעה על  $Y$  של משתנה  $S$  היא פי 3 מזו של משתנה  $Z$ , וכן כי החותך הוא 5.

א. מהי השערת האפס?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

המשך השאלה

מאמידת המודל המוגבל התקבל כי  $\sum e^2 = 623.99$  וכי  $R^2 = 0.99$

ג. חשב את הסטטיסטי של WALD.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

**? במדגם של 82 תצפיות התקבל:**

$$y_i = 12 + 3x_{1i} + 4x_{2i} + e_i \quad R^2 = 0.73$$

**א. בחנו את ההשערה כי :**

$$H_0: \beta_2 = 0$$

$$H_1: \beta_2 \neq 0$$

**כאשר נתון כי לאחר אמידת המודל המוגבל התקבל כי:  $R^2 = 0.6$**

**ב. חשבו את  $S_{\hat{\beta}_2}$**

**? על מנת לאמוד את פונקצית התצרוכת נאספו נתונים על 42 משקי בית בשנת**

**2007 ונאמדה המשוואה הבאה:**

$$C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$$

**מתוצאות האמידה של המשוואה הנ"ל התקבל כי:  $\sum e^2 = 52968$**

**על מנת לבדוק את ההשערה שהנטיה השולית לצורך (נש"צ) מתוך ההכנסה זהה**

**לנטיה השולית לצורך מתוך ההון נאמדה גם המשוואה הבאה:**

$$C_i = \alpha + \beta_1 \cdot Y_i + u_i \quad \text{כאשר: } Y_i = \text{סה"כ ההכנסה של משק בית } i \text{ (} W_i + P_i \text{)}$$

**התקבל:  $\sum e^2 = 54156$**

**א. בדקו את ההשערה**

**ב. חשבו את סטטיסטי t לבדיקת ההשערה**

**מבחן F למובהקות המודל**

**מבחן מובהקות העונה לשאלה: האם מודל הרגרסיה שלנו לניבוי משתנה תלוי**

**מסויים על ידי המשתנים ב"ת, מובהק באוכלוסיה?**

**השערות:**

$$H_0: \beta_1 = \beta_2 = \dots = \beta_K = 0$$

$$H_1: \text{OTHERWISE}$$

**המודל הלא מוגבל יהיה:**

$$U: \quad Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_K X_K + u_t$$

**המודל המוגבל יהיה:**

$$R: \quad Y_t = \alpha + u_t$$

מאחר ו-  $R_y^2 = 0$  ו-  $m = k - 1$  :

$$F = \frac{\frac{R_U^2 - R_R^2}{1 - R_U^2}}{\frac{m}{n - k}} = \frac{\frac{R_U^2}{1 - R_U^2}}{\frac{k - 1}{n - k}}$$

**הערה:** בדיקת מובהקות המודל ברגרסיה מרובה ניתנת לביצוע רק על ידי מבחן F

מאחר ויש לי יותר ממגבלה אחת בהשערת האפס.

לעומת זאת בדיקת מובהקות המודל ברגרסיה חד משתנית ניתנת לביצוע גם על ידי

מבחן t שכן יש לי רק מגבלה אחת בהשערת האפס:  $F = t^2$

**? נתון המודל :**

$$y = A \frac{x_{1i}^{\beta_1}}{x_{3i}^{\beta_3}} e^{\beta_2 x_2} e^{u_i}$$

באמידת מדגם של 58 נבדקים התקבל  $R^2 = 0.56$

האם המודל מובהק?

**לסיכום:** מתי נשתמש במבחן-t ומתי במבחן-F ?

- כאשר בהשערות ישנם סימני אי שוויון (השערות חד צדדיות), נשתמש בהתפלגות t
- כאשר יש סימן שוויון אחד בהשערת האפס, ניתן להשתמש ב-t או ב-F (תלוי מה יותר נוח ואילו נתונים זמינים לנו).
- כאשר יש בהשערת האפס יותר מסימן שוויון אחד, נשתמש בהתפלגות F.

מולטיקוליניאריות חלקית-סתירה בין מבחן F למבחן t

לעיתים נוצר מצב שבו המודל הוא מובהק אולם אף אחד מהשיפועים לא יוצא מובהק. מצב זה מעיד על בעיה של מולטיקוליניאריות חלקית במודל (מתאמים גבוהים בין המשתנים הב"ת). כיוון שמולטיקוליניאריות מגדילה את טעות התקן של האומדים היא מקטינה את סטטיסטי t ולכן גורמת לכך שנקבל את  $H_0$ .

### תירגול מסכם

**? נאמדו חמשת המודלים הבאים על 70 תצפיות:**

$$1. I_i = 12 + 0.13 \cdot \exp_i + 0.08 \cdot scl_i + 2 \cdot workh_i + u_i \quad ESS = 130$$

$$2. I_i = 11 + 0.1 \cdot scl + 0.1 \cdot workh_i + u_i \quad ESS = 150$$

$$3. I_i = 9 + 0.22 \cdot scl + u_i \quad ESS = 151$$

$$4. I_i = 15 + 0.15 \cdot workh_i + u_i \quad ESS = 152$$

$$5. I_i = 25 + u_i \quad ESS = 200$$

**(I) המשתנה המוסבר הוא הכנסה מעבודה**

**והמשתנים המסבירים שבחנו הם מספר שנות הלימוד (scl), מספר שעות עבודה**

**(workh) וותק בעבודה (exp)**

**הערה: הניחו כי ערך F הקריטי הוא 4.**

**א. האם לשעות עבודה (workh) ישנה השפעה מובהקת על ההכנסה**

**במשוואה 2?**

**ב. האם לשנות לימוד ישנה השפעה מובהקת על ההכנסה במשוואה 2?**

**ג. האם רגרסיה 2 מובהקת? (בחנו האם יש הסבר במודל 2),**

**כיצד זה מסתדר עם תשובתכם ל-א ו-ב.**

**ד. האם השפעת הוותק יכול להיות 0.15?**

**ה. כלכלן נוסף הציע להריץ את המודל:**

$$I_i + \exp_i = 2 - 3(scl_i - workh_i) + u_i \quad ESS = 145$$

**איזו השערה ניתן לבחון באמצעות מודל זה? כמה דרגות חופש יש לסטטיסטי**

**שנקבל? בחנו אותה.**

**? על סמך מדגם של 40 תצפיות נאמדו המשוואות הבאות:**

$$R^2 = 0.76 \quad y_i = 2 + 3X_{1i} + 4X_{2i} + e_i \quad (1)$$

$$y_i = 3 + 5D_i + e_i \quad R^2 = 0.60 \quad (2)$$

$$D_i = 0.2X_{1i} + X_{2i} \quad (3)$$

**כאשר Y הינו הציון בתואר ראשון, X1 ציוני הבגרות ו-X2 ציוני הפסיכומטרי**

א. בדקו את ההשערה כי ציוני הבגרות וציוני הפסיכומטרי ביחד לא משפיעים על ציוני תואר ראשון.

ב. בדקו את ההשערה כי רגרסיה 2 מובהקת

ג. איזה השערה ניתן לבדוק באמצעות רגרסיה 1 ו-2?

**?** על סמך מדגם של 80 משפחות המונות כל אחת 4 ילדים, נאמדו המשוואות

הבאות:

$$R^2 = 0.6 \quad \hat{y}_i = 5 + 2x_{1i} + 2x_{2i} \quad (1)$$

$$R^2 = 0.45 \quad \hat{y}_i = 11 + 0.9x_{2i} + 0.6x_{3i} \quad (2)$$

$$R^2 = 0.78 \quad \hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} \quad (3)$$

כאשר  $y_i$  הינו סה"כ הוצאות משק בית  $i$ ,  $x_{ji}$  הינו גילו של הילד  $j$ .

חשבו את האומדן לסטית התקן של המקדם  $X_3$  ברגרסיה 3

**?** נתון המודל:  $y_i = AX_{1i}^{\beta_1} e^{\beta_2 X_{2i} + \beta_3 X_{3i}} e^{u_i}$

א. מהי המשוואה לאמידת המקדמים של המודל?

ב. מה המודל המוגבל עבור ההשערה:  $\beta_1 = 2\beta_3; \beta_2 = 3\beta_3$

ג. מהן דרגות החופש במונה ובמכנה?

ד. רשמו את הנוסחה לחישוב סטטיסטי המבחן

**?** המודל הבא מתאר את פונקציית הייצור של מוצר  $P$ :

$$\ln(P_i) = \alpha + \beta_S \ln(S_i) + \beta_J \ln(J_i) + \varepsilon_i$$

כאשר  $S$  ו- $J$  הן שתי התשומות בייצור ( $S$ =תשומת ההון ו- $J$ =תשומת העבודה).

מהו המודל המוגבל המתאים לבדיקת ההשערה כי פונקציית הייצור מקיימת

תק"ל (תשואה קבועה לגודל)?

## פרק 7 - שינוי יחידות מדידה

שינוי ליניארי (טרנספורמציה ליניארית) שנעשה במשתנה המוסבר או במשתנה המסביר במודל.  
שינוי ליניארי משמעו: הוספה/החסרה של קבוע ו/או הכפלה/חילוק של קבוע של אחד או שני המשתנים.

- טרנספורמציה ליניארית של המשתנים לא תשפיע על  $R^2$ ,  $F$ ,  $t_{\hat{\beta}}$  ו-  $PF$ .
- האומדים ( $\hat{\alpha}$  ו-  $\hat{\beta}$ ) וסטיות התקן שלהם ( $S_{\hat{\alpha}}$  ו-  $S_{\hat{\beta}}$ ) עשויים להשתנות וכך גם  $t_{\hat{\alpha}}$  (נסמן את הפרמטרים שלאחר השינוי ב-  $\alpha'$  ו-  $\beta'$  ואת האומדים ב-  $\hat{\alpha}'$  ו-  $\hat{\beta}'$ ).

השינויים מסוכמים בטבלה הבאה:

| $S_{\hat{\alpha}}$                        | $S_{\hat{\beta}}$                              | $\hat{\alpha}'$                               | $\hat{\beta}'$                         |                                                                           |
|-------------------------------------------|------------------------------------------------|-----------------------------------------------|----------------------------------------|---------------------------------------------------------------------------|
| $s_{\hat{\alpha}'} \neq s_{\hat{\alpha}}$ | $s_{\hat{\beta}'} = s_{\hat{\beta}}$           | $\hat{\alpha}' = \hat{\alpha} - \hat{\beta}d$ | $\hat{\beta}' = \hat{\beta}$           | <b>הוספת קבוע ל- <math>X</math>:</b><br>$Y = \alpha' + \beta'(X + d) + v$ |
| $s_{\hat{\alpha}'} = s_{\hat{\alpha}}$    | $s_{\hat{\beta}'} = s_{\hat{\beta}}$           | $\hat{\alpha}' = \hat{\alpha} + d$            | $\hat{\beta}' = \hat{\beta}$           | <b>הוספת קבוע ל- <math>Y</math>:</b><br>$Y + d = \alpha' + \beta'X + v$   |
| $s_{\hat{\alpha}'} = s_{\hat{\alpha}}$    | $s_{\hat{\beta}'} = \frac{s_{\hat{\beta}}}{d}$ | $\hat{\alpha}' = \hat{\alpha}$                | $\hat{\beta}' = \frac{\hat{\beta}}{d}$ | <b>הכפלת <math>X</math> פי קבוע:</b><br>$Y = \alpha' + \beta'(dX) + v$    |
| $s_{\hat{\alpha}'} = ds_{\hat{\alpha}}$   | $s_{\hat{\beta}'} = ds_{\hat{\beta}}$          | $\hat{\alpha}' = d\hat{\alpha}$               | $\hat{\beta}' = d\hat{\beta}$          | <b>הכפלת <math>Y</math> פי קבוע:</b><br>$dY = \alpha' + \beta'X + v$      |

מסקנות מהטבלה:

$$t_{(\hat{\beta}'=0)} = t_{(\hat{\beta}=0)} \text{ תמיד.}$$

$$t_{(\hat{\alpha}'=0)} = t_{(\hat{\alpha}=0)} \text{ רק בהכפלות.}$$

**? חוקר ביקש לאמוד את הקשר בין שכר בש"ח (MWAGE) לבין שנות**

**לימוד (SCL) באמצעות 2 מודלים שונים.**

**להלן תוצאות האמידה :**

$$MWAGE_t = 139.54 + 118.62 \cdot SCL_t \quad (1)$$

$$MWAGE_t = -1445.08 + 1239.60 \cdot \ln(SCL)_t \quad (2)$$

**חשבו מחדש את מקדמי הרגרסיה וסטטיסטי המבחן F בכל אחד מהמודלים כתוצאה:**

1. התברר כי נעשה טעות בחישוב מספר שנות הלימוד, ויש צורך להוסיף 20% למשתנה המקורי.
2. התברר כי הקשר בין שכר לשנות לימוד הוא ריבועי ולכן יש צורך להעלות את המשתנה המקורי של מספר שנות הלימוד בריבוע.

**תזכורת של חוקי לוגים:**

$$\ln(X \cdot Y) = \ln(X) + \ln(Y)$$

$$\ln\left(\frac{X}{Y}\right) = \ln(X) - \ln(Y)$$

**? בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:**

**החוקר טען כי יש לבדוק את הקשר בין שכר לותק ע"י שימוש בשכר נטו (NET)**

**ולא בשכר ברוטו (SALARY). (קיים שיעור מס קבוע של 20%). המודל הוא:**

$$\ln(\text{NET}_t) = \alpha' + \beta' \cdot \text{EXP}_t + v_t$$

**מה יהיו ערכי האומדים, סטיות התקן שלהם וטיב ההתאמה באמידת מודל זה?**

## מבחן לדוגמא - שאלות כלולות עד פרק 7

### פרק 8 - מבחן לדוגמא

לשם חישובים הנח כי ערך  $t$  הינו 2 וערך  $F$  הינו 4.

(1) הנח כי הקשר באוכל' בין  $X$  ל- $Y$  נתון על ידי המשוואה

הבאה:  $Y_t = \beta \cdot X_t + U_t$ , כאשר כל ההנחות הקלאסיות מתקיימות.

$$\tilde{\beta} = \frac{\sum Y_t}{S_{xx}} \quad \text{נתון האומד:}$$

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

ג. אומד זה יעיל פחות מאומד הריבועים הפחותים:

נכון/ לא נכון/ אי אפשר לדעת

ד. האם אומד זה הוא blue?

ה. אומד  $\tilde{\beta}$  מוגדר רק כאשר  $S_x^2 \neq 0$ : נכון/ לא נכון/ אי אפשר לדעת

ו. חשבו את השונות של  $\tilde{\beta}$  עבור מודל שבו  $\alpha \neq 0$

ז. שונות האומד (שחושבה בסעיף הקודם) הינה גדולה משונות המודל

הנתון: נכון/ לא נכון/ אי אפשר לדעת

(2) על סמך מדגם של 60 משפחות שלכל אחת 3 ילדים נאמדו המשוואות

הבאות:

$$\hat{y}_i = 15 + 0.7x_{1i} + 0.35x_{2i} + 0.20x_{3i} \quad R^2 = 0.85 \quad (1)$$

$$y_i = 2 + 0.1z_i \quad R^2 = 0.25 \quad (2)$$

$$z_i = x_{1i} - x_{2i} + 2x_{3i} \quad (3)$$

כאשר  $y_i$  הינן הוצאות משק הבית על חינוך הילדים ואילו  $x_{ji}$  הינו גילו של

הילד  $j$

א. ההשערה שניתן לבדוק באמצעות המשוואות הנתונות הינה:

$$1. \quad HO: \beta_1 = \beta_2; \beta_1 = 2\beta_3$$

$$2. \quad HO: \beta_1 = -\beta_2 = 2\beta_3$$

$$3. \quad HO: \beta_2 = -\beta_1; \beta_3 = 2\beta_1$$

4. לא ניתן לדעת.

ב. סטטיסטי המבחן שניתן לבדוק באמצעות המשוואות הנתונות שווה בקירוב ל:

1. 56

2. 57

3. 112

4. 74.66

3) כלכלן הציע את המודלים הבאים:

$$(1) y_i = \beta_1 \ln(x_i) + \beta_2 \ln(0.5x_i) + u_i$$

$$(2) y_i = \beta_1 \ln(x_i) + \beta_2 \ln(x_i^{0.5}) + u_i$$

האם ניתן לאמוד את המודלים בשיטת OLS?

א. אין בעיה לאמוד את שני המודלים

ב. לא ניתן לאמוד את המודל הראשון בלבד

ג. לא ניתן לאמוד את המודל השני בלבד

ד. לא ניתן לאמוד את שני המודלים

4) כלכלן אמד את המודל הבא:  $y_i = \alpha_0 + \alpha_1 \ln(x_i) + u_i$

וקיבל את האומדנים:  $\hat{\alpha}_0 = 10$  ו-  $\hat{\alpha}_1 = 6$

על אותו המדגם אמד חברו את המודל הבא:  $y_i = \beta_0 + \beta_1 \ln(x_i^2) + u_i$

מכאן ש:

א.  $\hat{\beta}_0 = 5$  ו-  $\hat{\beta}_1 = 3$

ב.  $\hat{\beta}_0 = 10$  ו-  $\hat{\beta}_1 = 3$

ג.  $\hat{\beta}_0 = 5$  ו-  $\hat{\beta}_1 = 6$

ד. כל התשובות שגויות

5) על סמך מדגם של 95 תצפיות נאמד המודל הבא:

$$y_i = 2 + 0.5x_{1i} + 0.3x_{2i} \quad R^2 = 0.73$$

(1) (2)

הערכים שבסוגריים הם סטיות התקן של המקדמים.

א. בדוק האם המודל מובהק

ב. בדוק האם מקדמי השיפוע מובהקים

ג. מה תוכל להסיק מסעיפים א ו- ב?

6) על סמך מדגם של 52 תצפיות נאמדו המשוואות הבאות:

$$R^2 = 0.84 \quad y_i = 4 + 0.1x_{1i} + 0.8x_{2i} \quad (1)$$

$$R^2 = 0.7 \quad y_i = 2 + 0.8x_{1i} \quad (2)$$

$$R^2 = 0.25 \quad y_i = 7 + 0.23x_{2i} \quad (3)$$

$$R^2 = 0.55 \quad y_i = 3 + 0.23z_i \quad (4)$$

כאשר  $x_{1i}$  ו- $x_{2i}$  הם השכלת הבעל והאשה בהתאמה במשפחה i

ו- $y_i$  הכנסת משק בית i

$$z_i = x_{1i} + 2.2x_{2i} \quad \text{כמו כן נתון כי:}$$

א. בדוק את ההשערה כי להשכלה אין השפעה על הכנסות המשפחה

ב. איזה השערה ניתן לבדוק תוך שימוש במשוואות (1) ו- (4)? בדוק אותה.

ג. חשב את סטית התקן של המקדם  $x_{1i}$  ברגרסיה (1).

7) חוקר מעוניין לאמוד את המודל:  $y_i = \alpha + u_i$

א. חשב את נוסחת אומד הריבועים הפחותים ל- $\alpha$  על ידי פיתרון בעיית

המינימיזציה של סכום ריבועי הסטיות.

ב. חשב את נוסחת שונותו של האומד.

8) על סמך מדגם של 45 תצפיות נאמדו המודלים הבאים:

$$R^2 = 0.75 \quad y_i = 5.4 + 1.2x_{2i} + 4.4x_{3i} + u_i \quad (1)$$

$$R^2 = 0.65 \quad y_i = 6.3 + 5.8x_{3i} + u_i \quad (2)$$

$$R^2 = 0.70 \quad y_i = 5.7 + 1.2x_{2i} + u_i \quad (3)$$

$$R^2 = 0.56 \quad y_i = 3.9 + 3.4\ln(x_{2i}) + u_i \quad (4)$$

$$\ln(y_i) = 2.4 + 1.8x_{2i} + 2.7x_{3i}^2 + 4.2x_{4i}^2 + u_i \quad (5)$$

$$y_i = 1.3 + 3.1x_{2i} + 0.5x_{3i} + 4.8x_{4i}^2 + 1.5x_{5i}^2 + u_i \quad (6)$$

א. דרג את הרגרסיות על פי מדד ההסבר (מהנמוך לגבוה)

ב. בדוק את ההשערות שלמשתנים  $X_2$  ו- $X_3$  ביחד אין השפעה על Y

במודל (1).

ג. בדוק בהסתמך על מודל (2) האם המשתנה  $X_2$  מובהק ברגרסיה (1).

- ד. ברגרסיה (1) נתונים כעת אומדי הטעויות הסטנדרטיות (סטיות התקן) של מקדמי  $X_2$  ו- $X_3$  0.5 ו-2.5 בהתאמה. בדוק עבור כל אחד מהמקדמים הנ"ל האם מובהק ומה אפשר ללמוד מרגרסיה (1).
- ה. איזו השערה ניתן לבדוק תוך שימוש במשוואות (6) ו-(3)?

## פרק 9 - מבחני ספציפיקציה - מבחן LM (כופלי לגרנג')

במבחן כופלי לגרנג' (LM) אנו בודקים האם משתנה או משתנים מסבירים מסוימים רלוונטיים למודל.

**לדוגמא:**

נניח שיש לנו מודל הכולל 4 משתנים מסבירים (UNRESTRICTED):

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t}$$

לגבי השניים הראשונים אנו בטוחים כי הם רלוונטיים וחייבים להופיע במודל. לגבי השניים האחרונים אנחנו לא בטוחים.

השערות:

$$H_0: \beta_3 = \beta_4 = 0$$

$$H_1: \text{OTHERWISE}$$

המודל המוגבל (RESTRICTED) הינו:

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t}$$

במבחן LM אומדים את המודל המוגבל ומקבלים עבור כל תצפית את הסטייה מקו

$$\text{הרגרסיה} - Y_t - \hat{Y}_t = \hat{u}_t$$

כעת אומדים את רגרסיית העזר שבה מנסים לנבא את הסטייה מקו הרגרסיה עבור כל תצפית:

$$\hat{u}_t = \delta_0 + \delta_1 x_{1t} + \delta_2 x_{2t} + \delta_3 x_{3t} + \delta_4 x_{4t} + \omega_t$$

הרציונאל של המבחן הוא:

אחת מן ההנחות הקלאסיות של מודל הרגרסיה היא כי המשתנים הב"ת אינם מתואמים עם טעות האמידה:  $\text{cov}(X_t, u_t) = 0$ .

המשתנים המצויים בתוך המודל (השניים הראשונים), לא מתואמים עם טעות האמידה ( $R_{u12}^2 = 0$ ).

לגבי שני המשתנים שיש לנו ספק לגביהם:

אם הם יהיו מתואמים עם טעות האמידה ( $R_{u34}^2 \neq 0$ ) – זו אינדיקציה ש"שכחנו" להוסיף למודל וכי הם רלוונטיים (דוחים  $H_0$ ).

אם הם אינם מתואמים עם טעות האמידה ( $R_{u34}^2 = 0$ ), המודל מושלם כמו שהוא (מקבלים את  $H_0$ ).

• הסבר באמצעות מעגלי וואן:

#### חישוב הסטטיסטי:

הכפלת ה-  $R^2$  של רגרסיית העזר במספר התצפיות (T).

$$LM_{stat} = R^2 \cdot T$$

#### כלל הכרעה של $H_0$ :

אם  $LM_{stat} > \chi_m^2$  נדחה את  $H_0$ .

m - מס' ההגבלות ב-  $H_0$ .

UNRESTRICTED

Dependent variable: Y

| Analysis of Variance |       |                |             |         |        |
|----------------------|-------|----------------|-------------|---------|--------|
| Source               | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
| Model                | ----- | 646169.84      | -----       | -----   | 0.0000 |
| Error                | ----- | 620.17         |             |         |        |
| C Total              | 203   | 646790.01      |             |         |        |
|                      |       |                |             |         |        |
| Root MSE             | ----- | R-square       | -----       |         |        |
| Dep Mean             | ----- | Adj R-sq       | -----       |         |        |
| C.V.                 | ----- |                |             |         |        |

| Parameter Estimates |    |                    |                |                       |         |
|---------------------|----|--------------------|----------------|-----------------------|---------|
| Variable            | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob> T |
| INTERCEP            | 1  | 5.067731           | 0.456604       | 11.09874              | 0.0000  |
| X1                  | 1  | 0.975726           | 0.042711       | 22.84485              | 0.0000  |
| X2                  | 1  | 3.005385           | 0.008679       | 346.2721              | 0.0000  |
| X3                  | 1  | -5.029101          | 0.073149       | -68.75146             | 0.0000  |
| X4                  | 1  | 8.974106           | 0.029075       | 308.6485              | 0.0000  |

RESTRICTED

Dependent variable: Y

| Analysis of Variance                                                                                                         |  |        |      |  |  |
|------------------------------------------------------------------------------------------------------------------------------|--|--------|------|--|--|
|                                                                                                                              |  | Sum of | Mean |  |  |
| <p>לפתרון מלא בסרטון וידאו היכנסו ל- <a href="http://www.GooL.co.il">www.GooL.co.il</a></p> <p>כתבה ופתרה - קרן ברוסרד ©</p> |  |        |      |  |  |

| Source  | DF    | Squares   | Square | F Value | Prob>F |
|---------|-------|-----------|--------|---------|--------|
| Model   | ----- | 646001.81 |        |         |        |
| Error   | ----- | 788.2     |        |         |        |
| C Total | 203   | 646790.01 |        |         |        |

|          |       |          |       |
|----------|-------|----------|-------|
| Root MSE | ----- | R-square | ----- |
| Dep Mean | ----- | Adj R-sq | ----- |
| C.V.     | ----- |          |       |

#### Parameter Estimates

| Variable | DF | Parameter | Standard | T for H0:   |         |
|----------|----|-----------|----------|-------------|---------|
|          |    | Estimate  | Error    | Parameter=0 | Prob> T |
| INTERCEP | 1  | 7.067731  | 0.656604 | 10.76406    | 0.0000  |
| X1       | 1  | 26.36455  | 0.756627 | 34.84485    | 0.0000  |
| X2       | 1  | 29.58626  | 0.076993 | 384.2721    | 0.0000  |

#### רגרסיית עזר

Dependent variable :RES

#### Analysis of Variance

| Source  | DF    | Sum of    | Mean   | F Value | Prob>F |
|---------|-------|-----------|--------|---------|--------|
|         |       | Squares   | Square |         |        |
| Model   | ----- | 646169.84 | -----  | -----   | 0.0000 |
| Error   | ----- | 620.17    |        |         |        |
| C Total | 203   | 646790.01 |        |         |        |

|          |       |          |       |
|----------|-------|----------|-------|
| Root MSE | ----- | R-square | 0.213 |
| Dep Mean | ----- | Adj R-sq | ----- |
| C.V.     | ----- |          |       |

#### Parameter Estimates

| Parameter | Standard | T for H0: |
|-----------|----------|-----------|
|-----------|----------|-----------|

| Variable | DF | Estimate  | Error    | Parameter=0 | Prob> T |
|----------|----|-----------|----------|-------------|---------|
| INTERCEP | 1  | 5.9608892 | 0.776604 | 7.675584    | 0.0000  |
| X1       | 1  | 1.2077723 | 0.978845 | 1.233875    | 0.8455  |
| X2       | 1  | 0.4840697 | 0.886754 | 0.545889    | 0.9976  |
| X3       | 1  | -5.029101 | 0.073149 | -68.75146   | 0.0000  |
| X4       | 1  | 8.974106  | 0.029075 | 308.6485    | 0.0000  |

(1) בדקו את הטענה כי לפחות אחד מן המשתנים הנוספים רלוונטי למודל בשתי דרכים.

(2) איזה מן המשתנים הנוספים רלוונטי למודל?

**\*\*שימו לב כי:**

עבור המשתנים הנוספים למודל - כל המדדים (הבטות, ערכי t וה-Pvalu) ברגרסיית העזר שווים לאלו של הרגרסיה הלא מוגבלת.  
עבור המשתנים הקיימים במודל - המדדים אינם שווים בין שתי הרגרסיות.

(3) הסבירו את הקשרים בין שלוש המשוואות: (R) (U) ו (עזר) ואת הקשר בין מבחן WALS ומבחן LM.

הקשר בין משוואות ה-U, ה-R והעזר:

א.  $U = R + \text{עזר}$

ב.  $ESS_U = ESS_y$

ג.  $ESS_R = TSS_y$

ד.  $R^2 = 1 - \frac{ESS_y}{TSS_y} = 1 - \frac{ESS_U}{ESS_R} = \frac{ESS_R - ESS_U}{ESS_R}$

(4) שחזרו בעזרת שתי המשוואות הראשונות (U ו-R) את  $LM_{stat}$

(5) שחזרו בעזרת המשוואה האחרונה (רגרסיית העזר) את  $WALS_{stat}$

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

## פרק 10 - משתני דמי

הנושא של משתני דמי מטפל בהכנסת משתנים ב"ת איכותיים למודל הרגרסיה.

עד כה כל המשתנים הב"ת שהכנסנו למודל היו כמותיים, כלומר קיבלו ערכים מספריים.

למשל, נניח שאנו סבורים שמש' שנות הלימוד של אדם משפיעות על שכרו:

$$W_t = \text{השכר (המשתנה התלוי)}$$

$$S_t = \text{שנות לימוד (המשתנה הב"ת)}$$

$$\text{משוואת הרגרסיה: } W_t = \alpha + \beta \cdot S_t$$

במקרה זה המשתנה המסביר (כמו גם המוסבר) הוא כמותי.

נניח שאנו סבורים שגם משתנה המגדר משפיע על השכר. משתנה זה איננו כמותי כמו שנות לימוד אלא איכותי שכן הוא לא מקבל ערכים מספריים אלא ערכים קטגוריאליים כ"גבר" או "אישה".

נשאלת השאלה כיצד נכניס אותו לתוך משוואת הרגרסיה?

נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר".

משתנה כזה נקרא משתנה דמי (dummy variable).

לעומת משתנה רגיל ש"פועל" תמיד, משתנה זה "יפעל" רק אם מדובר בגבר.

ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

(1) משתנה דמי לחותך- המין משפיע על השכר ההתחלתי בלבד

(2) משתנה דמי לשיפוע- המין משפיע על התוספת לשכר בגין שנות הלימוד

(3) משתנה דמי לכל הפונקציה- המין משפיע גם על החותך וגם על השיפוע

(1) *משתנה דמי לחותך*

המין משפיע על השכר ההתחלתי בלבד.

$$W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t \quad \text{המודל:}$$

החותך מייצג כאן את השכר ההתחלתי.

שכר ההתחלתי של אישה:  $\alpha_0$

שכר ההתחלתי של גבר:  $\alpha_0 + \alpha_1$

הבדל בשכר בין נשים וגברים:  $\alpha_1$  (הפרש בין החותכים)

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש

$$H_0: \alpha_1 = 0 \quad \text{החותכים:}$$

**\*\* השיפוע מייצג את התוספת בשכר כפונקציה של מס' שנות הלימוד והוא זהה עבור נשים וגברים.**

**? על בסיס מדגם של 50 איש העובדים בחברה מסוימת התקבלו התוצאות**

**הבאות:**

$$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$$

(24) (56) (134) (S.E)

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

א. מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?

ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?

ג. האם הבדל זה מובהק באוכלוסיה?

ד. בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ-500 ₪ מזה של נשים.

ה. בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.

פונקצית רגרסיה המכילה משתנים איכותיים בלבד

המגדר הוא המשתנה היחיד במשוואה:

$$W_t = \alpha_0 + \alpha_1 D + u_t$$

החותך מייצג כאן את השכר הממוצע עבור כל קטגוריה.

שכר הממוצע של אישה:  $\alpha_0$

שכר הממוצע של גבר:  $\alpha_0 + \alpha_1$

הבדל בשכר הממוצע בין נשים וגברים:  $\alpha_1$  (הפרש בין החותכים)

בדיקת השערות על משתנה הדמי: מבחן  $t$ :  $H_0: \alpha_1 = 0$  (מבחן זה למבחן  $t$  להבדל בין ממוצעים).

**?** על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים. תוצאות האמידה:

$$W_t = 5200 + 1120 \cdot D$$

$$S_{\hat{\alpha}_1} = 63 \text{ נתון:}$$

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

**(2) משתנה דמי לשיפוע**

המגדר משפיע על התוספת לשכר בגין שנות הלימוד.

$$W_t = \alpha + \beta_0 S_t + \beta_1 D S_t + u_t$$

השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.

אצל אישה: התוספת לשכר בגין שנות לימוד -  $\beta_0$

אצל גבר: התוספת לשכר בגין שנות לימוד -  $\beta_0 + \beta_1$

הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד:  $\beta_1$  (הפרש השיפועים)

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש

$$H_0: \beta_1 = 0$$

\*\* החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.

? על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים

לנשים בתוספת לשכר בגין שנות הלימוד. תוצאות האמידה נתונות להלן:

$$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$

$$(68) \quad (23) \quad (25)$$

בדוק את ההשערה.

(3) משתנה דמי לכל הפונקציה

המין משפיע גם על החותך וגם על השיפוע. הווה אומר, גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.

$$W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 D S_t + u_t$$

השכר ההתחלתי של אישה:  $\alpha_0$

השכר ההתחלתי של גבר:  $\alpha_0 + \alpha_1$

הבדל בשכר ההתחלתי בין המינים:  $\alpha_1$  (הבדל בחותכים)

אצל אישה-התוספת לשכר בגין שנות הלימוד:  $\beta_0$

אצל גבר-התוספת לשכר בגין שנות הלימוד:  $\beta_0 + \beta_1$

הבדל בין המינים בתוספת לשכר בגין שנות הלימוד:  $\beta_1$  (הבדל בשיפועים)

בדיקת השערות למשתני הדמי:

$$H_0: \alpha_1 = \beta_1 = 0$$

באמצעות מבחן WALD יש לבדוק:

לפחות אחד הפרמטרים שונה מ-0:  $H1$

אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים  
בנפרד:

$$H0: \beta_1 = 0 \quad \text{ו-} \quad H0: \alpha_1 = 0$$

### מבחן CHOW

דרך נוספת לבדיקת ההבדל בין הקטגוריות, בלא יצירת משתני דמי:

חלוקת המדגם לפי הקטגוריות של המשתנה האיכותי. מדגם של גברים ( $T_m$ ) ושל נשים ( $T_f$ ).

עבור כל קבוצה לאמוד משוואות רגרסיה לניבוי שכר על ידי שנות לימוד .

$$\text{נשים: } W_t = \alpha_f + \beta_f X_t + u_t$$

$$\text{גברים: } W_t = \alpha_m + \beta_m X_t + u_t$$

$$\text{השערות: } H_0: \alpha_f = \alpha_m; \beta_f = \beta_m$$

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALT שהשתמשנו בו מקודם):

המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן לא יכול את המדגם המאוחד כי אין צורך בשתי רגרסיות נפרדות:

$$W_t = \alpha + \beta X_t + u_t$$

$$\begin{aligned} ESS_U &= ESS_f + ESS_m \\ DF_U &= DF_f + DF_m \end{aligned} \quad \text{המודל הלא מוגבל (U) כולל את שני חלקי המדגם:}$$

$$CHOW_{stat} = \frac{\frac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\frac{ESS_f + ESS_m}{DF_f + DF_m}} = WALT_{stat}$$

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

1. אם דחינו את HO במבחן CHOW נתקשה לברר את מקור ההבדל שנמצא.

2. בהרצת שתי רגרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.

**?** חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים

בקטעי כביש בינעירוניים, בהינתן נפח התנועה.

החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

כבישים מהירים בלבד

(1)

כבישים לא - מהירים בלבד

(2)

שני סוגי הכביש (כל המדגם)

(3)

(4)

כאשר :  $=$  מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה

= נפח התנועה בקטע כביש t ליום באלפים

= משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר ,

ו- 0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

(1) בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.

(2) חשבו את הערכים המספריים עבור אומדני משוואה (4).

(3) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

הועלתה הטענה כי המקדם להשפעה של נפח התנועה בדרכים מהירות הינו כפול מזה שבדרכים לא - מהירות.

(4) מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4)?

(5) מהי הרגרסיה "תחת" למבחן WALT ?

## משוואה (1) - כבישים מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 344

Number of Observations Used 344

### Analysis of Variance

|        | Source          | DF  | Sum of Squares | Mean Square | F Value | Pr |
|--------|-----------------|-----|----------------|-------------|---------|----|
| > F    |                 |     |                |             |         |    |
| <.0001 | Model           | 1   | 4700.81174     | 4700.81174  | 89.12   |    |
|        | Error           | 342 | 18039          | 52.74684    |         |    |
|        | Corrected Total | 343 | 22740          |             |         |    |

|                |           |          |        |
|----------------|-----------|----------|--------|
| Root MSE       | 7.26270   | R-Square | 0.2067 |
| Dependent Mean | 5.10465   | Adj R-Sq | 0.2044 |
| Coeff Var      | 142.27617 |          |        |

### Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1  | 1.55289            | 0.54303        | 2.86    | 0.0045  |
| avgd      | 1  | 0.02098            | 0.00222        | 9.44    | <.0001  |

## משוואה (2) - כבישים לא מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 410

Number of Observations Used 410

### Analysis of Variance

| Source          | DF  | Sum of Squares | Mean Square | F Value | Pr > F |
|-----------------|-----|----------------|-------------|---------|--------|
| Model           | 1   | 971.99073      | 971.99073   | 145.83  | <.0001 |
| Error           | 408 | 2719.34830     | 6.66507     |         |        |
| Corrected Total | 409 | 3691.33902     |             |         |        |

|                |           |          |        |
|----------------|-----------|----------|--------|
| Root MSE       | 2.58168   | R-Square | 0.2633 |
| Dependent Mean | 1.38780   | Adj R-Sq | 0.2615 |
| Coeff Var      | 186.02612 |          |        |

### Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1  | 0.14978            | 0.16360        | 0.92    | 0.3605  |
| avgd      | 1  | 0.02877            | 0.00238        | 12.08   | <.0001  |

### משוואה (3) - שני סוגי הכביש (כל המדגם)

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

#### Analysis of Variance

|        | Source          | DF  | Sum of Squares | Mean Square | F Value | Pr |
|--------|-----------------|-----|----------------|-------------|---------|----|
| > F    |                 |     |                |             |         |    |
| <.0001 | Model           | 1   | 8052.00804     | 8052.00804  | 288.84  |    |
|        | Error           | 752 | 20964          | 27.87730    |         |    |
|        | Corrected Total | 753 | 29016          |             |         |    |
|        | Root MSE        |     | 5.27990        | R-Square    | 0.2775  |    |
|        | Dependent Mean  |     | 3.08355        | Adj R-Sq    | 0.2765  |    |
|        | Coeff Var       |     | 171.22758      |             |         |    |

#### Parameter Estimates

| Variable  | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|-----------|----|--------------------|----------------|---------|---------|
| Intercept | 1  | 0.73903            | 0.23665        | 3.12    | 0.0019  |
| avgd      | 1  | 0.02330            | 0.00137        | 17.00   | <.0001  |

**משוואה (4)**

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

## Analysis of Variance

|        | Source          | DF  | Sum of Squares | Mean Square | F Value | Pr |
|--------|-----------------|-----|----------------|-------------|---------|----|
| > F    | Model           | 3   | 8256.966       | 2752.322    | 99.44   |    |
| <.0001 | Error           | 750 | 20759          | 27.678      |         |    |
|        | Corrected Total | 753 | 29016          |             |         |    |

Root MSE 5.26102 R-Square 0.2846

Dependent Mean 3.08355 Adj R-Sq 0.2817

Coeff Var 170.61553

## Parameter Estimates

|  | Variable  | DF | Parameter Estimate | Standard Error | t Value | Pr >  t |
|--|-----------|----|--------------------|----------------|---------|---------|
|  | Intercept | 1  | 0.14978            | 0.33340        | 0.45    | 0.6534  |
|  | type      | 1  |                    |                |         | 0.0067  |
|  | avgd      | 1  |                    |                |         | <.0001  |

avgdtype 1  
0.1283

### סיכום ביניים:

| משתנה דמי לכול הפונקציה                                                                                                                                                                                                                                                                                   | משתנה דמי לשיפוע                                               | משתנה דמי לחותך                                       |                  |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------|-------------------------------------------------------|------------------|
| $Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$                                                                                                                                                                                                                                          | $Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$              | $Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$ | המודל            |
| קיים הבדל בין הקטגוריות<br>במשוואת הרגרסיה כולה (בחותרך<br>ובשיפוע).                                                                                                                                                                                                                                      | קיים הבדל בין הקטגוריות<br>בתוספת ל- $Y$ בגין $X$<br>(בשיפוע). | קיים הבדל בין הקטגוריות<br>ב- $Y$ ההתחלתי (בחותרך).   | ההשערה<br>במילים |
| מבחן WALT להפרש בין הפונקציות<br>(החותכים והשיפועים):<br>$H0: \alpha_1 = \beta_1 = 0$<br><br>**ניתן לבדוק את ההשערה בדבר הבדל<br>בין הפונקציות גם במבחן CHOW.<br><br>אם דוחים את $H0$ יש לברר את מקור<br>ההבדל באמצעות מבחני $t$ (אפשרי רק<br>ב-WALT):<br><br>$H0: \alpha_1 = 0$<br><br>$H0: \beta_1 = 0$ | מבחן $t$ להפרש השיפועים:<br><br>$H0: \beta_1 = 0$              | מבחן $t$ להפרש החותכים:<br><br>$H0: \alpha_1 = 0$     | בדיקת<br>ההשערה  |

**משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים**

כאשר המשתנה האיכותי כלל שני ערכים בלבד (למשל, מגדר: גבר, אישה) הסתפקנו במשתנה דמי אחד.

במקרים רבים המשתנה האיכותי כולל יותר משני ערכים/קטגוריות. במקרה כזה נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

**למשל**, את המשתנה האיכותי של עונות השנה הכולל 4 ערכים: אביב, קיץ, סתיו, חורף נייצג באמצעות 3 משתני דמי:

$D_1$  יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

$D_2$  יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

$D_3$  יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס.

נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

$$V_t = \text{מדד מחירי הירקות}$$

$$P_t = \text{מדד המחירים לצרכן}$$

**(1) משתני דמי לחותך**

**הטענה:** יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות

$$\text{המודל: } V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t$$

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב-  $\beta$ .

למחיר זה יתווסף  $\alpha_0$  בחורף,  $\alpha_0 + \alpha_1$  באביב,  $\alpha_0 + \alpha_2$  בקיץ ו-  $\alpha_0 + \alpha_3$  בסתיו.

ניתן לראות כי:  $\alpha_0$ : החותך בקטגוריה שהושמטה

$\alpha_0 + \alpha_i$ : החותך בקטגוריה i.

בדיקת השערות:

השערות:

$$H0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t \quad (U)$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (R)$$

\*\*שימו לב שהחותך במשוואה המוגבלת איננו  $\alpha_0$  שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את  $H0$  במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין החותכים על ידי מבחני t:

1. האם יש הבדל במחיר ההתחלתי של הירקות בין האביב לחורף:  $H0: \alpha_1 = 0$

2. האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:  $H0: \alpha_2 = 0$

3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:  $H0: \alpha_3 = 0$

? א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.

1. מהי השערת האפס לבדיקת הטענה?

2. פרטו שני מבחנים סטטיסטיים בעזרתם ניתן לבדוק את הטענה.

ב. הועלתה הטענה כי יש רק שתי עונות המשפיעות על מחיר הירקות ההתחלתי: קיץ+אביב, חורף+סתיו.

1. מהי השערת האפס לבדיקת הטענה?

2. מהו המבחן הסטטיסטי המתאים? פרטו.

**(2) משתני דמי לשיפוע**

**הטענה:** יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן

$$V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t \quad \text{המודל:}$$

המחיר ההתחלתי של הירקות שווה בין עונות השנה ( $\alpha$ ) אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב:

$$\beta_0 \text{ בחורף, } \beta_0 + \beta_1 \text{ באביב, } \beta_0 + \beta_2 \text{ בקיץ ו- } \beta_0 + \beta_3 \text{ בסתיו.}$$

ניתן לראות כי:  $\beta_0$  : השיפוע בקטגוריה שהושמטה

$$\beta_0 + \beta_i \text{ : השיפוע בקטגוריה } i.$$

**בדיקת השערות:**

**השערות:**

$$H0: \beta_1 = \beta_2 = \beta_3 = 0$$

$$H1: OTHERWISE$$

**המבחן הסטטיסטי : מבחן WALT:**

$$V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t \quad (U)$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (R)$$

**\*\*שימו לב** שהשיפוע במשוואה המוגבלת איננו  $\beta_0$  שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את  $H0$  במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני  $t$ .

**(3) משתני דמי לכל הפונקציה**

**הטענה:** יש הבדל בין עונות השנה בפונקצית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t \quad \text{המודל:}$$

**בדיקת השערות:**

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

**השערות:**

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$$

**המבחן הסטטיסטי: מבחן WALT:**

$$V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t \quad (U)$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (R)$$

**אם דוחים את  $H_0$ , יש לבדוק במבחן WALT האם ההבדל הוא בין החותכים או בין השיפועים:**

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \quad H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

**באם דוחים את  $H_0$  יש להמשיך לבדוק באמצעות מבחן t:**

$$H_0: \beta_j = 0 \quad H_0: \alpha_j = 0$$

**משתני דמי עבור שני משתנים איכותיים**

נתבונן בדוגמא שבה יש שני משתנים איכותיים המשפיעים על פונקצית השכר-

מגדר (אישה, גבר) וגזע (לבן, שחור).

נגדיר משתנה דמי G שיקבל 1 אם מדובר בגבר ו-0 אחרת (אישה).

נגדיר משתנה דמי R שיקבל 1 אם מדובר בלבן ו-0 אחרת (שחור).

נבדוק כיצד מגדר וגזע משפיעים על השכר ההתחלתי (החותך), כאשר השכר תלוי

גם בשנות לימוד ( $S_t$ ).

**(1) הבדל בחותך ללא אינטראקציה**

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t \quad \text{המודל:}$$

**במודל זה- אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.**

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים לא תלוי בגזע (זהה עבור שחורים ועבור לבנים) ולהיפך-ההבדל בשכר ההתחלתי בין לבנים לשחורים לא תלוי במגדר (זהה עבור נשים וגברים).

ניתן לבדוק השערות על כל אחד מהמשתנים ה"ת האיותיים בנפרד:

$$1. \text{ הבדל בשכר ההתחלתי בין גברים לנשים: } H_0: \alpha_1 = 0$$

$$2. \text{ הבדל בשכר ההתחלתי בין שחורים ללבנים: } H_0: \alpha_2 = 0$$

## (2) הבדל בחותך עם אינטראקציה

$$\text{המודל: } W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים תלוי בגזע (שונה אם מדובר בשחורים או בלבנים) ולהיפך.

במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

$$3. \text{ } H_0: \alpha_3 = 0$$

## דרך נוספת ליצירת מודל עם אינטראקציה:

הגדרת משתני דמי המייצגים שילוב בין המשתנים האיותיים גזע ומגדר באופן הבא:

$D_1$  יקבל 1 אם מדובר בגבר לבן ו-0 אחרת

$D_2$  יקבל 1 אם מדובר בגבר שחור ו-0 אחרת

$D_3$  יקבל 1 אם מדובר באשה לבנה ו-0 אחרת

הנשים השחורות מהוות כאן את קבוצת הייחוס.

$$\text{המודל: } W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה:

| הפרש                  | אישה                  | גבר                   |      |
|-----------------------|-----------------------|-----------------------|------|
| $\gamma_1 - \gamma_3$ | $\gamma_0 + \gamma_3$ | $\gamma_0 + \gamma_1$ | לבן  |
| $\gamma_2$            | $\gamma_0$            | $\gamma_0 + \gamma_2$ | שחור |
|                       | $\gamma_3$            | $\gamma_1 - \gamma_2$ | הפרש |

ההשערות לבדיקת קיום האינטראקציה:

$$HO: \gamma_1 - \gamma_2 = \gamma_3 \quad \text{או} \quad HO: \gamma_1 - \gamma_3 = \gamma_2$$

התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת:

$$WALD = t^2$$

$$PF = Pt$$

**? שאלה מס' 1**

חוקר בדק השפעות של השכלה, גזע (שחור, לבן) וניסיון (EXP) על לוג השכר  $\ln(Y)$  במדגם בן 306 תצפיות:

$$\ln(Y)_i = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 EXP_i + \beta_2 EXP_i^2 + u_i$$

$\ln(Y)$ -לוג השכר

EXP - שנות ניסיון

$D_1$  מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (ו-0 אחרת)

$D_2$  מקבל את הערך 1 עבור שחורים בעלי השכלה נמוכה (ו-0 אחרת)

$D_3$  מקבל את הערך 1 עבור לבנים בעלי השכלה גבוהה (ו-0 אחרת)

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

| Analysis of Variance |     |                |             |         |     |
|----------------------|-----|----------------|-------------|---------|-----|
| Source               | DF  | Sum of Squares | Mean Square | F Value | Pr  |
| Model                | 5   | -----          | -----       | -----   | --- |
| Error                | 300 | 140            | -----       |         |     |
| Corrected Total      | 305 | 210            |             |         |     |
| Root MSE             |     | -----          | R-Square    | -----   |     |
| Dependent Mean       |     | -----          | Adj R-Sq    | -----   |     |
| Coeff Var            |     | -----          |             |         |     |

| Parameter Estimates |    |                    |                |         |      |
|---------------------|----|--------------------|----------------|---------|------|
| Variable            | DF | Parameter Estimate | Standard Error | t Value | Pr > |
| Intercept           | 1  | -----              | -----          | 60.84   | 0.00 |
| D1                  | 1  | -----              | -----          | -3.20   | 0.00 |
| D2                  | 1  | -----              | -----          | -5.56   | 0.00 |
| D3                  | 1  | -----              | -----          | 7.23    | 0.00 |
| EXP                 | 1  | -----              | -----          | 8.11    | 0.00 |
| $EXP^2$             | 1  | -----              | -----          | -7.45   | 0.00 |

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

| הפרש                  | השכלה גבוהה           | השכלה נמוכה           |        |
|-----------------------|-----------------------|-----------------------|--------|
| $\alpha_3$            | $\alpha_0 + \alpha_3$ | $\alpha_0$            | לבנים  |
| $\alpha_1 - \alpha_2$ | $\alpha_0 + \alpha_1$ | $\alpha_0 + \alpha_2$ | שחורים |
|                       | $\alpha_1 - \alpha_3$ | $\alpha_2$            | הפרש   |

(א) לפי המשוואה הניסיון זהה עבור שחורים ולבנים: נכון/לא נכון/ לא ניתן לדעת

(ב) בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.

(ג) בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.

(ד) הי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?

(ה) לבדיקת ההשערה של הסעיף הקודם בוצע מבחן W.L.D.

הרגרסיה המוגבלת תחת השערת האפס הינה:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + \nu$$

מהם ה-Zים?

(ו) בדוק את ההשערה אם ידוע שבמודל המוגבל  $R^2 = 0.33$

(ז) החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

כאשר: S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים)

E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).

מה הקשר בין המקדמים של שני המודלים?

(ח) אם יאמוד החוקר את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

## שאלה מס' 2

חוקרת בדקה השפעות השכלה, מגדר וניסיון על הכנסה מעבודה לפי המשוואה הבאה:

$$\ln(MWAGE) = \alpha_0 + \alpha_1 \cdot S + \alpha_2 \cdot E + \alpha_3 \cdot (S \cdot E) + \beta_0 \cdot EXP + \beta_1 \cdot (EXP \cdot S) + \beta_2 \cdot (EXP \cdot E) + \beta_3 \cdot (EXP \cdot S \cdot E) + U$$

כאשר : S משתנה דמי : 1 עבור נשים, 0 גברים

E משתנה דמי : 1 עבור השכלה גבוהה (scl > 12), 0 השכלה

נמוכה

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

| הפרש                        | השכלה נמוכה (E=0)                                | השכלה גבוהה (E=1)                                                                          |             |
|-----------------------------|--------------------------------------------------|--------------------------------------------------------------------------------------------|-------------|
| חותך: $\alpha_2 + \alpha_3$ | $\alpha_0 + \alpha_1 + (\beta_0 + \beta_1)EXP_t$ | $\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + (\beta_0 + \beta_1 + \beta_2 + \beta_3)EXP_t$ | נשים (S=1)  |
| שיפוע: $\beta_2 + \beta_3$  |                                                  |                                                                                            |             |
| חותך: $\alpha_2$            | $\alpha_0 + \beta_0 EXP_t$                       | $\alpha_0 + \alpha_2 + (\beta_0 + \beta_2)EXP_t$                                           | גברים (S=0) |
| שיפוע: $\beta_2$            |                                                  |                                                                                            |             |
| הפרש                        | חותך: $\alpha_1$<br>שיפוע: $\beta_1$             | חותך: $\alpha_1 + \alpha_3$<br>שיפוע: $\beta_1 + \beta_3$                                  |             |

א. רשמו את הפונקציה לחישוב:

1. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.
2. תחזית לוג השכר ההתחלתי עבור נשים משכילות.
3. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

1. אין השפעה של מגדר והשכלה על השכר.
2. השפעת ההשכלה אינה תלויה במגדר.
3. אין השפעות השכלה אצל גברים.
4. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

## פרק 11 - הטרנסקדסטיות

הטרנסקדסטיות הוא מצב שבו מופרת הנחה מס' 5, הנחת ההומוסקדסטיות, הגורסת כי השונות של ה"קריזה" היא אותה שונות עבור כל תצפית ותצפית:  $V(u_t) = \sigma^2$  לכל  $t$ , כלומר התצפיות מפוזרות באופן אחיד סביב קו הרגרסיה.

במצב של הטרנסקדסטיות הקריזה של כל תצפית היא בעלת שונות אחרת:

$$V(u_t) = \sigma_t^2$$

למשל כאשר בודקים את הקשר שבין הכנסה ותצרוכת מגלים כי ברמות הכנסה נמוכות אין גמישות רבה בהוצאות ולכן התצפיות מרוכזות סביב קו הרגרסיה. לעומת זאת ברמות גבוהות של הכנסה התצפיות מפוזרות יותר ויש שונות גבוהה יותר בתצרוכת. השונות, אם כן, איננה אחידה סביב קו הרגרסיה והיא תלויה בתצפית - זהו מצב של הטרנסקדסטיות.

### **ההשלכות של הטרנסקדסטיות על אומדי OLS**

מבין התכונות של אר"פ (ליניאריות, חוסר הטיות, עקיבות ויעילות) היחידה שמופרת בהינתן הטרנסקדסטיות היא: **יעילות** של אומדי הרבועים הפחותים. זאת מכיוון שבכדי להוכיח יעילות של האומדים השתמשנו בהנחה מס' 5 של שונות קבועה. לכן הפרתה גוררת פגיעה ביעילות האומד.

### **מבחנים לזיהוי הטרנסקדסטיות**

החשד לקיומה של בעיית הטרנסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות. גרף השאריות גם מאפשר לנו לבחון את הצורה הפונקציונאלית של השונות, כלומר באיזה אופן השונות משתנה בין תצפית לתצפית.

קיימות שתי שיטות לזיהוי הטרנסקדסטיות: מבחן Goldfeld-Quandt (GQ) ומבחן White.

ההבדל בין שני המבחנים הוא בהחלטה על הצורה הפונקציונאלית של השונות. כלומר על האופן שבו השונות משתנה בין תצפית לתצפית, (על פי דיאגרמת הפיזור).

מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונות שונות בלבד. ואילו מבחן White מניח כי לכל תצפית ושונות שונה של טעויות.

### (1) מבחן GQ

ההנחה העומדת בבסיס מבחן זה היא כי קיימות שתי שונות של טעויות.

ביצוע המבחן:

- מחלקים את המדגם לשני חלקים:

1. החלק שבו אנו חושדים שיש שונות גבוהה יותר

2. החלק שבו אנו חושדים שיש שונות נמוכה יותר

מקובל להשמיט מס' תצפיות (בין 1/6 ל-1/3) במרכז המדגם.

- אומדים כל אחד מהחלקים בנפרד ומקבלים את ה-ESS של כל חלק.

- מחשבים את הסטטיסטי:  $F_{stat} = \frac{ESS_1 / T_1 - K - 1}{ESS_2 / T_2 - K - 1}$  (תמיד השונות הגבוהה חלקי הקטנה).

- סטטיסטי זה מתפלג  $F_{(\alpha; T_1 - K - 1, T_2 - K - 1)}$

- כלל ההכרעה: אם  $F_{stat} > F_C$  אז דוחים את  $H_0$ .

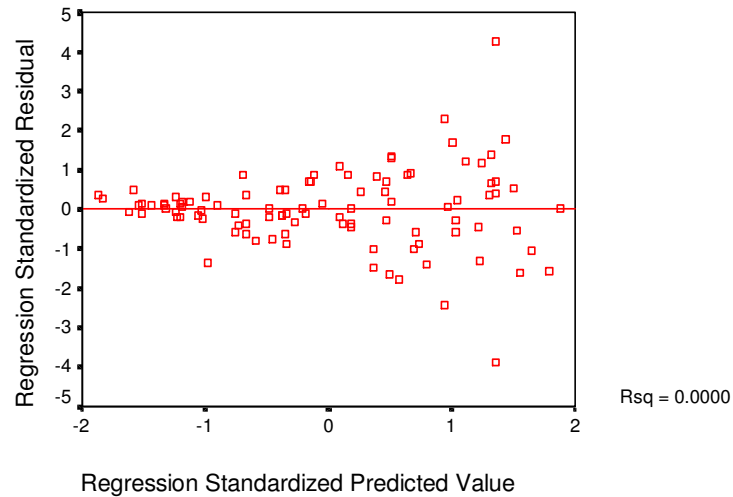
- ההשערות:  
 $H_0: \sigma_1^2 = \sigma_2^2$   
 $H_1: \sigma_1^2 > \sigma_2^2$

**לדוגמא:** נאמד הקשר שבין הכנסה לתצרוכת:  $Y_t = \alpha + \beta X_t + u_t$

גרף השאריות של הרגרסיה הנ"ל נתון להלן:

## Scatterplot

Dependent Variable: Y



מגרף זה אנו למדים כי ככל שעולים ברמת ההכנסה כך השונות בהוצאות הפרט עולה. כלומר השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה - זהו מצב של הטרוסקדסטיות.

בכדי לבצע מבחן GQ :

- התצפיות של המשתנה הכנסה סודרו מהגדול לקטן והמדגם חולק לשלוש קבוצות שוות.
- רגרסיה נפרדת הורצה על השליש הראשון ועל השליש האחרון.

התוצאות של אמידת הקשר בין הכנסה לתצרוכת מוצג בפלטים 1 ו-2 בהתאמה:

## משוואה (1)

The REG Procedure

Model: MODEL1

Dependent Variable: Y

Number of Observations Read 16

Number of Observations Used 16

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

## Analysis of Variance

|     | Source          | DF  | Sum of<br>Squares | Mean<br>Square | F Value | Pr   |
|-----|-----------------|-----|-------------------|----------------|---------|------|
| > F | Model           | --  | -----             | -----          | ----    | ---- |
| --  | Error           | 14  | 166452.9          | -----          |         |      |
|     | Corrected Total | --- | -----             |                |         |      |

## משוואה (2)

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read 16

Number of Observations Used 16

## Analysis of Variance

|     | Source          | DF  | Sum of<br>Squares | Mean<br>Square | F Value | Pr  |
|-----|-----------------|-----|-------------------|----------------|---------|-----|
| > F | Model           | --- | -----             | -----          | -----   | --- |
| --  | Error           | 14  | 2934638           | -----          |         |     |
|     | Corrected Total | --- | -----             |                |         |     |

השערות:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

סטטיסטי המבחן:

$$F_{stat} = \frac{ESS_1 / T_1 - K - 1}{ESS_2 / T_2 - K - 1} = \frac{2,934,638}{166,452.9} = 17.62$$

כלל הכרעה:

לכן יש סיבה מספקת לדחות את  $H_0$  ברמת  $F_{stat} = 17.62 > F(0.05;14,14) = 2.48$

מובהקות של 5%.

מסקנה:

יש עדות להטרוסקדסטיות בנתונים.

? על מנת לבחון את פונקציית הייצור בענף מסוים נאספו נתונים על 150 פירמות.

נסמן:

$Q$  תפוקה שנתית באלפי שקלים.

$L$  מספר עובדים

המודל הנאמד:  $\ln(Q) = \alpha + \beta \cdot \ln(L)$

החוקר חשש שההפרעה המקרית איננה הומוסקדסטית.

לשם כך הוא מיין את התצפיות בסדר עולה של מספר העובדים, השמיט  $1/3$  מהתצפיות האמצעיות והריץ שתי רגרסיות נפרדות עם מספר שווה של תצפיות:

ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא קיבל:

$$ESS=279.3 \quad R^2 = 0.403$$

ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא קיבל:

$$ESS=493.8 \quad R^2 = 0.238$$

האם יש עדות לקיום הטרוסקדסטיות בנתונים? (בצעו את המבחן המתאים: רשמו השערות, חשבו סטטיסטי מבחן, רשמו כלל הכרעה והגיעו למסקנה).

2) מבחן White

ההנחה העומדת בבסיס מבחן זה כי לכל תצפית ותצפית שונות שונה של טעויות.

הביטוי המתמטי של הנחה זו היא היותה של השונות פונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:

לפתרון מלא בסרטון וידאו היכנסו ל- [www.GooL.co.il](http://www.GooL.co.il)

כתבה ופתרה - קרן ברוסרד ©

$$\sigma_t^2 = f(x_j, x_j^2, x_j x_j)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \beta_1 x_k^2 + \dots + \beta_k x_k^2 + \gamma_{12} x_1 x_2 + \gamma_{13} x_1 x_3 \dots$$

האומד ל- $\sigma_t^2$  הוא  $\hat{u}_t^2$

המבחן הוא מבחן LM:

- אומדים את המודל המקורי ומקבלים את הסטיות מקו הרגרסיה  $\hat{u}_t$  (המכונה ב SAS - RES).

- אומדים את  $\hat{u}_t^2$  כפונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:  $\hat{u}_t^2 / x_j, x_j^2, x_j x_j$  זוהי רגרסיית העזר.

- נחשב את סטטיסטי LM:  $LM_{stat} = T_y \cdot R_y^2$

- אם  $LM_{stat} > \chi_m^2$  כאשר  $m =$  מס' המשתנים ברגרסיית העזר.

- השערות:

$$H_0: \alpha_j = \beta_j = \gamma_{jj} = 0$$

$$H_1: \text{OTHERWISE}$$

**נדגים** על אותו הקשר שבין הכנסה לתצרוכת:

$$Y_t = \alpha + \beta X_t + u_t$$

שאנו חושדים על פי גרף השאריות כי קיים בו מצב של הטרוסקדסטיות.

בכדי לבצע את מבחן WHITE:

- נחשב את השאריות של הרגרסיה:  $\hat{u}_t = Y_t - \hat{Y}_t$

- נעלה את השאריות בריבוע:  $\hat{u}_t^2$

- נאמוד את המשוואה:  $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + v_t$

תוצאות האמידה מוצגות להלן:

The REG Procedure

Model: MODEL1

Dependent Variable:  $RES^2$ 

Number of Observations Read 48

Number of Observations Used 48

## Analysis of Variance

|     | Source          | DF  | Sum of Squares | Mean Square | F Value  | Pr  |
|-----|-----------------|-----|----------------|-------------|----------|-----|
| > F | Model           | --- | -----          | -----       | -----    | --- |
| --  | Error           | --- | -----          | -----       |          |     |
|     | Corrected Total | --- | ----           |             |          |     |
|     | Root MSE        |     | -----          | R-Square    | 0.390763 |     |
|     | Dependent Mean  |     | -----          | Adj R-Sq    | -----    |     |

השערות:

$$H0: \alpha_1 = \beta_1 = 0$$

$$H1: OTHERWISE$$
סטטיסטי המבחן:

$$LM_{stat} = T_y \cdot R_y^2 = 48 \cdot 0.3907 = 18.75$$

כלל הכרעה:

$$LM_{stat} = 18.75 > \chi_{2(0.05)}^2 = 5.991$$

מובהקות של 5%.

מסקנה

יש עדות לקיום של הטרוסקדסטיות בנתונים.

**\*\*הערה חשובה:** אם חלק מערכי  $\hat{u}_t^2$  יצאו שליליים (והרי שונות לא יכולה להיות

שלילית) יש להוציא לוג ולאמוד את  $\ln \hat{u}_t^2$  כך התחזית תצא תמיד חיובית.

**?** חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות

והאפשרות של מכירת עיתונים.

נסמן:

Y\_SALES=מכירות חודשיות (ש"ח)

X1\_SQUARES=שטח החנות(מ"ר)

X2\_RENT=דמי שכירות(\$)

PAPERS=משתנה איכותי המקבל 1-אם החנות מוכרת גם עיתונים ו-0 אם לא.

החוקר חשד כי קיימת בעיה של הטרוסקדסטיות בנתונים.

החוקר ביצע מבחן לזיהוי הטרוסקדסטיות שתוצאותיו נתונות להלן:

Dependent Variable:

Number of Observations Read 20

Number of Observations Used 20

Analysis of Variance

|     | Source          | DF  | Sum of Squares | Mean Square | F Value  | Pr  |
|-----|-----------------|-----|----------------|-------------|----------|-----|
| > F | Model           | --- | -----          | -----       | -----    | --- |
| --  | Error           | --- | -----          | -----       |          |     |
|     | Corrected Total | --- | ----           |             |          |     |
|     | Root MSE        |     | -----          | R-Square    | 0.086942 |     |
|     | Dependent Mean  |     | -----          | Adj R-Sq    | -----    |     |

| Parameter Estimates |                   |    |                    |                |         |      |
|---------------------|-------------------|----|--------------------|----------------|---------|------|
|                     | Variable          | DF | Parameter Estimate | Standard Error | t Value | Pr > |
| t                   | Intercept         | 1  | -----              | -----          | -----   | ---  |
| --                  |                   |    |                    |                |         |      |
| ---                 | X1_SQUARE         | 1  | -----              | -----          | -----   | ---  |
| ---                 |                   |    |                    |                |         |      |
| ---                 | X1_SQUARE^2       | 1  | -----              | -----          | -----   | ---  |
| ---                 |                   |    |                    |                |         |      |
| ---                 | X1_SQUARE*X2_RENT | 1  | -----              | -----          | -----   | --   |
| ---                 |                   |    |                    |                |         |      |
| --                  | X2_RENT           | 1  | -----              | -----          | -----   | --   |
| ---                 |                   |    |                    |                |         |      |
| ---                 | X2_RENT^2         | 1  | -----              | -----          | -----   | --   |
| ---                 |                   |    |                    |                |         |      |
| ---                 | X2_RENT*PAPERS    | 1  | -----              | -----          | -----   | --   |
| ---                 |                   |    |                    |                |         |      |
| ---                 | PAPERS            | 1  | -----              | -----          | -----   | -    |

הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת : \_\_\_\_\_

על ידי מבחן : \_\_\_\_\_

המשתנה התלוי הינו: \_\_\_\_\_

המשתנים הב"ת: \_\_\_\_\_

ההשערות הינן: \_\_\_\_\_

גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): \_\_\_\_\_

המסקנה המתקבלת היא: \_\_\_\_\_

### פיתרון בעיית ההטרוסקדסטיות- ריבועים פחותים משוקללים (WLS)

נניח שאנו רוצים לאמוד את המודל :  $Y_t = \alpha + \beta X_t + u_t$  וידוע כי לכל קריזה שונות אחרת.

ההנחה שעומדת בבסיס שיטת ה-WLS היא כי השונות המשתנה כוללת בתוכה מרכיב קבוע ומרכיב משתנה :

$$\sigma_t^2 = Z_t \cdot \sigma^2$$

את המרכיב המשתנה בשונות ( $Z_t$ ) יש לנטרל.

לשם כך ניצור משתנה חדש-  $W_t$  שיהווה השורש ההופכי לאותו מרכיב משתנה :

$$W_t = \frac{1}{\sqrt{Z_t}}$$

נכפיל כל תצפית במשתנה החדש  $W_t$  וניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית :

$$\begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ Y_t W_t &= \alpha \cdot W_t + \beta (X_t W_t) + u_t W_t \end{aligned}$$

בצורתה המפורשת המשוואה החדשה נראית כך :

$$\frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$$

קיבלנו מודל חדש שבו :

$$\frac{u_t}{\sqrt{Z_t}} : \text{הקריזה היא}$$

$$\frac{Y_t}{\sqrt{Z_t}} : \text{המשתנה המוסבר}$$

$$\frac{1}{\sqrt{Z_t}} \text{ ו- } \frac{X_t}{\sqrt{Z_t}} : \text{המשתנים המסבירים הם}$$

במודל זה אין חותך.

הפרמטרים  $\alpha$  ו- $\beta$  הם של המשוואה המקורית (שימו לב כי ל- $\alpha$  אין משמעות של חותך).

למשוואה זו אין משמעות כלכלית אך היא מאפשרת לנו לאמוד את הפרמטרים  $\alpha$  ו- $\beta$  בשיטת OLS כך שנקבל אומדים  $\hat{\alpha}$  ו- $\hat{\beta}$  יעילים.

ניתן להוכיח כי הטרגנספורמציה הליניארית של השונות הפכה אותה לקבועה:

$$V\left(\frac{u_t}{\sqrt{Z_t}}\right) = \frac{1}{Z_t} V(u_t) = \frac{1}{Z_t} \cdot \sigma^2 \cdot Z_t = \sigma^2$$

הערה: אם ביצענו את מבחן WHITE והוצאנו  $\log$  לנתונים אז  $Z_t = e^{\ln \hat{u}_t^2}$  כך שיש

$$\text{לשקלל את הרגרסיה ב: } W_t = \frac{1}{\sqrt{e^{\ln \hat{u}_t^2}}}.$$

### ? שאלה מס' 1

$$Y_t = \alpha + \beta \cdot X_t + U_t \quad (1) \quad \text{נתון המודל :}$$

$$\text{ונתון כי : } \text{VAR}(U_t) = \frac{\sigma^2}{Z_t^2} \quad (Z_t \text{ משתנה ידוע}).$$

א. מהי הבעיה שנוצרת באמידת משוואה (1) ?

ב. מהן תכונות אומדי הריבועים הפחותים של משוואה (1) ?

כדי לפתור את הבעיה שנוצרה, נאמדה המשוואה הבאה:

$$Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + U_t \cdot W_t \quad (2)$$

ג. מהו  $W_t$  שבעזרתו ניתן לאמוד את  $\alpha$  ו- $\beta$  בצורה יעילה ?

ד. מהו האומד היעיל של  $\sigma^2$  ?

ה. האם ניתן להשוות בין המודלים על בסיס  $R^2$ ? אם לא, האם ניתן להחליט בכל זאת איזה מודל טוב יותר?

ו. חוו דעתכם על הטענות הבאות, ונמקו:

1. אם נתון כי  $Z_t = a + b \cdot \bar{X}$ , התשובות לסעיפים א' ו- ב' נשארות ללא שינוי.

2. המשוואה הנורמאלית:  $\sum \hat{\varepsilon}_t = 0$  (כאשר  $\varepsilon_t = U_t \cdot W_t$ ) היא אחת המשוואות הנורמאליות לאמידת משוואה (2).

### שאלה מס' 2

עני על שאלה מס' 1, כאשר נתון כי  $\text{VAR}(U_t) = \sigma^2 \cdot X_t^2$ .

### שאלה מס' 3

נתון המודל:  $Y_t = \alpha + \beta X_t + u_t$

וקיים מדגם של 100 תצפיות

כאשר נתון כי:

$$V(u_t) = \sigma_t^2 = \begin{cases} X_t \sigma^2 & \Leftrightarrow t \geq 50 \\ X_t^2 \sigma^2 & \Leftrightarrow t \leq 50 \end{cases}$$

(שאר ההנחות הקלאסיות מתקיימות)

א. במשוואה מס' 1 יש בעיה של: \_\_\_\_\_

ב. אמידת משוואה (1) תניב אומדים בלתי מוטים ועקיבים:

נכון/ לא נכון/ לא ניתן לדעת

ג. פיתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה:

$$Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t \quad (2)$$

$$W_t = \text{כאשר: } \underline{\hspace{2cm}}$$

$$V(u_t) = \sigma_t^2 = \begin{cases} 3\sigma^2 & \Leftrightarrow t \geq 50 \\ \sigma^2 & \Leftrightarrow t \leq 50 \end{cases} \quad \text{ד. ס נתון כי:}$$

כן/ לא/ לא ניתן

האם ישתנו תשובותיכם לסעיפים א ו-ב :

לדעת

## פרק 12 - מתאם סידרתי

**מתאם סידרתי עוסק במצב שבו מופרת ההנחה הקלאסית מס' 6 - אי תלות בין**

$$\text{הטעויות: } \text{cov}(u_t, u_s) = 0$$

**זהו מצב שבו קיימת תלות סטטיסטית בין הטעויות במודל:  $\text{cov}(u_t, u_s) \neq 0$ .**

**תלות כזו בין הטעויות קיימת בדרך כלל כאשר הנתונים הנאספים הם נתוני סדרות עיתיות ולא נתוני חתך בהם עסקנו עד כה.**

**נתוני חתך - מתייחסים לפרטים שונים בתקופת זמן נתונה.**

**נתוני סדרות עיתיות - מתייחסים לאותו הפרט לאורך זמנים שונים.**

**בנתוני החתך סביר שמאחר ומדובר בפרטים שונים - הטעויות בניבוי שלהם תהיינה בלתי תלויות. לעומת זאת בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר דווקא שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.**

**אם אנחנו מדברים, למשל, על פונקציות תצורות:**

$$Y_t = \alpha + \beta X_t + u_t$$

**אם מדובר בנתוני חתך - החלטות התצורות של פרט t לא אמורות להשפיע על החלטותיו של פרט s והנחה 6 תתקיים:**

$$\text{cov}(u_t, u_s) = 0$$

**אולם אם מדובר על נתוני סדרה עיתית: החלטותיו של פרט מסוים היום סביר שיושפעו מהחלטות שעשה בעבר או שישפיעו על החלטות שיעשה בעתיד:**

$$\text{cov}(u_t, u_{t-s}) \neq 0$$

**כאשר  $u_t$  - "קריזה" בזמן t.**

**$u_{t-s}$  - "קריזה" בזמן t-s.**

## השלכות על אומדי הריבועים הפחותים (OLS)

מבין התכונות של אר"פ (ליניאריות, חוסר הטיה, עקיבות ויעילות) היחידה שמופרת כאשר קיים מתאם סדרתי היא:

### תכונת היעילות.

משום שתכונת היעילות היא היחידה מבין תכונות אר"פ התלויה להוכחתה בקיומה של הנחת אי התלות בין הטעויות.

משום הפגיעה בתכונת היעילות, בדיקת ההשערות לא תהיה תקפה.

**\*\*שימו לב כי במידה וקיים מתאם סדרתי חיובי בין הטעויות ולמשתנים יש מגמת זמן (X עולה או יורד עם הזמן) אומד השונות (ESS) יהיה מוטה כלפי מטה ואז נקבל  $R^2$ , F ו-t מוטים כלפי מעלה.**

### מבנה המתאם הסדרתי

הגדרנו מתאם סדרתי כהפרה של הנחה מס' 6:

$$Y_t = \alpha + \beta X_t + u_t$$

$$\text{cov}(u_t, u_{t-s}) \neq 0$$

נשאלת השאלה כיצד נראית הפרה זו?

### מתאם סדרתי מסדר ראשון:

ההנחה היא כי יש מתאם בין קריזות במרחק אחד, כלומר  $u_t$  תלוי ישירות רק ב-

$$u_{t-1}$$

$$\text{cov}(u_t, u_{t-1}) \neq 0$$

את המתאם בין ה"קריזות" מסדר ראשון ניתן לנסח באופן הבא:

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

כך ש:

(1)  $\rho \neq 0$  (כי אם  $\rho = 0$  אין מתאם)

(2)  $1 < \rho < -1$  (כי אם חורג מ-1 ה"קריזה" הולכת וגדלה עם הזמן)

(3)  $\rho$  חיובי פירושו מתאם סדרתי חיובי ואילו  $\rho$  שלילי פירושו מתאם סדרתי שלילי (לא נפוץ).

(4)  $\varepsilon_t$  מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין (בניגוד ל- $u_t$ ) כך ש:

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma_\varepsilon^2$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t-s}) = 0$$

המודל יכול שתי משוואות-המשוואה העיקרית והגדרת המתאם הסדרתי (מסדר ראשון):

$$Y_t = \alpha + \beta X_t + u_t$$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

מלבד  $\alpha$  ו- $\beta$  נרצה לאמוד גם את  $\rho$ .

**מתאם סדרתי מסדר שני:**

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \varepsilon_t$$

הן  $u_{t-1}$  והן  $u_{t-2}$  משפיעים ישירות על  $u_t$ .

**מתאם סדרתי מסדר P:**

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \dots + \rho_P \cdot u_{t-P} + \varepsilon_t$$

$u_t$  מושפע מתקופות שונות בעבר.

**תכונות המתאם הסדרתי**

ניתן להוכיח כי מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן:

$$r_{(u_t, u_{t-s})} = \rho^s$$

מכיוון שכך המתאם של  $u_t$  הולך ופוחת עם הזמן:

$$\rho_{u_t, u_{t-1}} > \rho_{u_t, u_{t-2}}^2 > \rho_{u_t, u_{t-3}}^3 > \dots > \rho_{u_t, u_{t-s}}^s$$

בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t) = 0$$

$$V(u_t) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

$$COV(u_t, u_{t-s}) = \rho^s \sigma_u^2 = \rho^s \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

**?** נתון מתאם סדרתי מסדר ראשון:  $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

נתון כי  $\rho = 0.9$  וכי  $V(\varepsilon_t) = \sigma_\varepsilon^2 = 1$

מצאו את:

א. המתאם בין  $u_t$  ל- $u_{t-1}$

ב. המתאם בין  $u_t$  ל- $u_{t-4}$ . הסבר את ההבדל בין המתאמים (סעיף א' ו-ב').

ג. השונות  $\sigma_u^2$

ד. חזרו על סעיפים א' עד ג' עבור  $\rho = 0.4$ . הסבירו את ההבדל בין התוצאות.

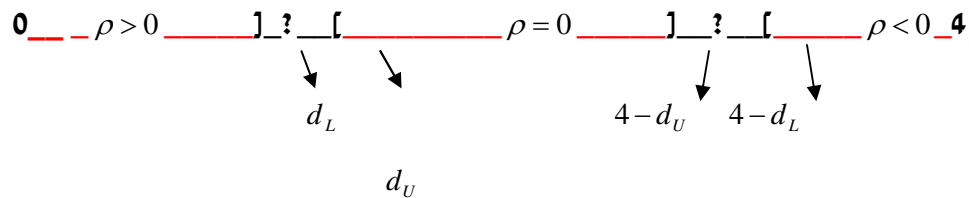
**מבחנים לזיהוי מתאם סדרתי**

קיימים שני מבחנים סטטיסטיים לזיהוי קיומו של מתאם סדרתי:

מבחן DW (דרבין ווטסון) ומבחן LM.

### 1) מבחן DW (דרבין ווטסון) לקיום מתאם סדרתי מסדר ראשון:

- נניח תחילה כי אין מתאם סדרתי ונאמוד את המשוואה הראשית בשיטת OLS.
- כחלק מתוצאות האמידה נקבל ציון DW (יכול לקבל ערכים בין 0 ל-4 בלבד).
- נתבונן בטבלת DW ולפי  $K = \text{מס' המשתנים ה'ב'ת במודל ו-T} = \text{מס' התצפיות במדגם}$  נשלוף שני ערכים:  $d_U$  ו-  $d_L$ .
- נחלק את הטווח שבין 0 ל-4 באופן הבא:



- נראה היכן נופל ציון ה-DW שהתקבל כחלק מתוצאות האמידה.
- ניתן לדעת אם יש מתאם ואיזה סוג של מתאם רק אם ציון ה-DW ייפול בחלקים המודגשים.

#### השערות:

$$H_0: \rho = 0$$

$$H_1: \rho > 0, \rho < 0$$

#### חישוב הסטטיסטי

$$DW_{stat} \cong 2 \cdot (1 - \hat{\rho})$$

אם אנו מקבלים ציון  $\hat{\rho}$  ניתן להציב בנוסחה ולקבל  $DW_{stat}$ .

**לדוגמא:** חוקר רצה לאמוד את מחיר סגירה של מניה כפונקציה של הזמן שעובר:

$$CLOSE_t = \alpha + \beta \cdot TIME_t + u_t$$

כאשר:

$$CLOSE_t = \text{מחיר סגירה של מניה ב-} \$ \text{ ביום } t.$$

$$TIME_t = \text{משתנה זמן שמקבל את הערכים: } 1, 2, 3, \dots$$

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

| Analysis of Variance |      |                |             |         |        |
|----------------------|------|----------------|-------------|---------|--------|
| Source               | DF   | Sum of Squares | Mean Square | F Value | Prob>F |
| Model                | 1    | -----          |             | 55.78   |        |
| Error                | 151  | -----          |             |         |        |
| C Total              | 152  | -----          |             |         |        |
| Root MSE             | ---- | R-square       | 0.181       |         |        |
| Dep Mean             | ---- | Adj R-sq       | -----       |         |        |
| C.V.                 | ---- |                |             |         |        |

| Parameter Estimates |    |                    |                |                       |         |
|---------------------|----|--------------------|----------------|-----------------------|---------|
| Variable            | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob> T |
| INTERCEP            | 1  | 1.3474             | 0.0148         | 91.047                | 0.0000  |
| TIME                | 1  | -0.00075           | 0.0001         | -7.468                | 0.0000  |
| Durbin-Watson D     |    | 0.150              |                |                       |         |

האם קיים מתאם סדרתי?

**נתבונן בטבלת DW כאשר נתון כי  $K=1$  ו- $T=153$  :**

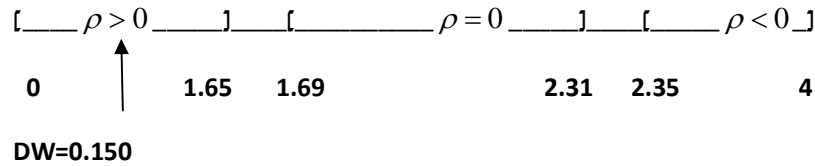
**TABLE 12** Cutoff Points for the Distribution of the Durbin-Watson Test Statistic

Let  $d_\alpha$  be the number such that  $P(d < d_\alpha) = \alpha$ , where the random variable  $d$  has the distribution of the Durbin-Watson statistic under the null hypothesis of no autocorrelation in the regression errors. For probabilities  $\alpha = .05$  and  $\alpha = .01$ , the tables show, for numbers of independent variables,  $K$ , values  $d_L$  and  $d_U$  such that  $d_L \leq d_\alpha \leq d_U$  for numbers  $n$  of observations.

| $\alpha = .05$ |       |       |       |       |       |       |       |       |       |       |
|----------------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $n$            | $K$   |       |       |       |       |       |       |       |       |       |
|                | 1     |       | 2     |       | 3     |       | 4     |       | 5     |       |
|                | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15             | 1.08  | 1.36  | 0.95  | 1.54  | 0.82  | 1.75  | 0.69  | 1.97  | 0.56  | 2.21  |
| 16             | 1.10  | 1.37  | 0.98  | 1.54  | 0.86  | 1.73  | 0.74  | 1.93  | 0.62  | 2.15  |
| 17             | 1.13  | 1.38  | 1.02  | 1.54  | 0.90  | 1.71  | 0.78  | 1.90  | 0.67  | 2.10  |
| 18             | 1.16  | 1.39  | 1.05  | 1.53  | 0.93  | 1.69  | 1.82  | 1.87  | 0.71  | 2.06  |
| 19             | 1.18  | 1.40  | 1.08  | 1.53  | 0.97  | 1.68  | 0.86  | 1.85  | 0.75  | 2.02  |
| 20             | 1.20  | 1.41  | 1.10  | 1.54  | 1.00  | 1.68  | 0.90  | 1.83  | 0.79  | 1.99  |
| 21             | 1.22  | 1.42  | 1.13  | 1.54  | 1.03  | 1.67  | 0.93  | 1.81  | 0.83  | 1.96  |
| 22             | 1.24  | 1.43  | 1.15  | 1.54  | 1.05  | 1.66  | 0.96  | 1.80  | 0.86  | 1.94  |
| 23             | 1.26  | 1.44  | 1.17  | 1.54  | 1.08  | 1.66  | 0.99  | 1.79  | 0.90  | 1.92  |
| 24             | 1.27  | 1.45  | 1.19  | 1.55  | 1.10  | 1.66  | 1.01  | 1.78  | 0.93  | 1.90  |
| 25             | 1.29  | 1.45  | 1.21  | 1.55  | 1.12  | 1.66  | 1.04  | 1.77  | 0.95  | 1.89  |
| 26             | 1.30  | 1.46  | 1.22  | 1.55  | 1.14  | 1.65  | 1.06  | 1.76  | 0.98  | 1.88  |
| 27             | 1.32  | 1.47  | 1.24  | 1.56  | 1.16  | 1.65  | 1.08  | 1.76  | 1.01  | 1.86  |
| 28             | 1.33  | 1.48  | 1.26  | 1.56  | 1.18  | 1.65  | 1.10  | 1.75  | 1.03  | 1.85  |
| 29             | 1.34  | 1.48  | 1.27  | 1.56  | 1.20  | 1.65  | 1.12  | 1.74  | 1.05  | 1.84  |
| 30             | 1.35  | 1.49  | 1.28  | 1.57  | 1.21  | 1.65  | 1.14  | 1.74  | 1.07  | 1.83  |
| 31             | 1.36  | 1.50  | 1.30  | 1.57  | 1.23  | 1.65  | 1.16  | 1.74  | 1.09  | 1.83  |
| 32             | 1.37  | 1.50  | 1.31  | 1.57  | 1.24  | 1.65  | 1.18  | 1.73  | 1.11  | 1.82  |
| 33             | 1.38  | 1.51  | 1.32  | 1.58  | 1.26  | 1.65  | 1.19  | 1.73  | 1.13  | 1.81  |
| 34             | 1.39  | 1.51  | 1.33  | 1.58  | 1.27  | 1.65  | 1.21  | 1.73  | 1.15  | 1.81  |
| 35             | 1.40  | 1.52  | 1.34  | 1.58  | 1.28  | 1.65  | 1.22  | 1.73  | 1.16  | 1.80  |
| 36             | 1.41  | 1.52  | 1.35  | 1.59  | 1.29  | 1.65  | 1.24  | 1.73  | 1.18  | 1.80  |
| 37             | 1.42  | 1.53  | 1.36  | 1.59  | 1.31  | 1.66  | 1.25  | 1.72  | 1.19  | 1.80  |
| 38             | 1.43  | 1.54  | 1.37  | 1.59  | 1.32  | 1.66  | 1.26  | 1.72  | 1.21  | 1.79  |
| 39             | 1.43  | 1.54  | 1.38  | 1.60  | 1.33  | 1.66  | 1.27  | 1.72  | 1.22  | 1.79  |
| 40             | 1.44  | 1.54  | 1.39  | 1.60  | 1.34  | 1.66  | 1.29  | 1.72  | 1.23  | 1.79  |
| 45             | 1.48  | 1.57  | 1.43  | 1.62  | 1.38  | 1.67  | 1.34  | 1.72  | 1.29  | 1.78  |
| 50             | 1.50  | 1.59  | 1.46  | 1.63  | 1.42  | 1.67  | 1.38  | 1.72  | 1.34  | 1.77  |
| 55             | 1.53  | 1.60  | 1.49  | 1.64  | 1.45  | 1.68  | 1.41  | 1.72  | 1.38  | 1.77  |
| 60             | 1.55  | 1.62  | 1.51  | 1.65  | 1.48  | 1.69  | 1.44  | 1.73  | 1.41  | 1.77  |
| 65             | 1.57  | 1.63  | 1.54  | 1.66  | 1.50  | 1.70  | 1.47  | 1.73  | 1.44  | 1.77  |
| 70             | 1.58  | 1.64  | 1.55  | 1.67  | 1.52  | 1.70  | 1.49  | 1.74  | 1.46  | 1.77  |
| 75             | 1.60  | 1.65  | 1.57  | 1.68  | 1.54  | 1.71  | 1.51  | 1.74  | 1.49  | 1.77  |
| 80             | 1.61  | 1.66  | 1.59  | 1.69  | 1.56  | 1.72  | 1.53  | 1.74  | 1.51  | 1.77  |
| 85             | 1.62  | 1.67  | 1.60  | 1.70  | 1.57  | 1.72  | 1.55  | 1.75  | 1.52  | 1.77  |
| 90             | 1.63  | 1.68  | 1.61  | 1.70  | 1.59  | 1.73  | 1.57  | 1.75  | 1.54  | 1.78  |
| 95             | 1.64  | 1.69  | 1.62  | 1.71  | 1.60  | 1.73  | 1.58  | 1.75  | 1.56  | 1.78  |
| 100            | 1.65  | 1.69  | 1.63  | 1.72  | 1.61  | 1.74  | 1.59  | 1.76  | 1.57  | 1.78  |

לפי הטבלה (עבור  $T=100$ ):  
 $d_L = 1.65$   
 $d_U = 1.69$

נציב בטווח:



מסקנה:

יש עדות לקיום מתאם סדרתי חיובי מסדר ראשון.

המודל הינו:  $Y_t = \alpha + \beta X_t + \varepsilon_t$   
 $u_t = \rho u_{t-1} + \varepsilon_t$

$u_t$  - הפרעה מקרית קלאסית

$T=100$  וידוע כי:

$$u_t = 0.9u_{t-1}$$

$$d_L = 1.57$$

$$d_U = 1.65$$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

למבחן DW יש שתי בעיות עיקריות:

- (1) מתאים רק למתאם סדרתי מסדר ראשון
- (2) יש אזהרים "מתים" בטווח בהם לא ניתן לדעת האם יש מתאם סדרתי.

בנוסף לכך על מספר תנאים להתקיים כדי שאפשר יהיה להשתמש במבחן DW:

1. הרגרסיה כוללת חותך

2. ה-Xים קבועים ולא משתנים

3. אין משתנים מסבירים שהם פיגור של המשתנה המוסבר

4. אין תצפיות חסרות באמצע

5. אם קיים מתאם סדרתי מסדר ראשון אז הוא מהצורה:  $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

## (2) מבחן LM

לעומת מבחן DW מבחן LM מתאים גם לבחון קיומו של מתאם סדרתי מסדרים גבוהים יותר מסדר ראשון.

$$Y_t = \alpha + \beta \cdot X_t + u_t$$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

ניתן לרשום את המודל הלא מוגבל:

$$Y_t = \alpha + \beta \cdot X_t + \rho \cdot u_{t-1} + \varepsilon_t \quad (U)$$

ניתן להתייחס לבחינת קיומו של מתאם סדרתי כהוספת משתנה מסביר:  $\hat{u}_{t-1}$

השלבים לביצוע המבחן:

- נאמוד את המודל המקורי ונחשב  $\hat{u}_t$  ו-  $\hat{u}_{t-1}$
- נאמוד את רגרסיית העזר:  $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}$
- נחשב סטטיסטי LM:  $LM_{stat} = T_y \cdot R_y^2$
- נדחה את  $H_0$  כאשר:  $LM_{stat} > \chi_m^2$  כאשר  $m =$  סדר המתאם הסדרתי.

אם נדחה את  $H_0$  נדע את סימנו של המתאם הסדרתי לפי המקדם של  $\hat{u}_{t-1}$

ברגרסיית העזר ששווה ל-  $\hat{\rho}$ .

שימו לב כי אם נרצה לבדוק מתאם סדרתי מסדרים גבוהים יותר:

ההשערות:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0$$

$H_1$ : אחרת

רגרסיית העזר:  $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-s}$

**לדוגמא:** עבור הדוגמא הקודמת- ניבוי מחיר סגירה של מניה כפונקציה של

$$CLOSE_t = \alpha + \beta \cdot TIME_t + u_t$$

נבחן את קיומו של מתאם סדרתי מסדר ראשון באמצעות מבחן LM.

נאמוד את רגרסיית העזר:

$$u_t = \gamma_0 + \gamma_1 \cdot TIME_t + \rho \cdot u_{t-1} + \varepsilon_t$$

תוצאות האמידה שהתקבלו:

Dependent variable: RES

| Analysis of Variance |       |                |             |         |        |
|----------------------|-------|----------------|-------------|---------|--------|
| Source               | DF    | Sum of Squares | Mean Square | F Value | Prob>F |
| Model                | 2     | -----          |             |         |        |
| Error                | 150   | -----          |             |         |        |
| C Total              | 152   | -----          |             |         |        |
| Root MSE             | ----- | R-square       | 0.855       |         |        |
| Dep Mean             | ----- | Adj R-sq       | -----       |         |        |
| C.V.                 | ----- |                |             |         |        |

| Parameter Estimates |    |                    |
|---------------------|----|--------------------|
| Variable            | DF | Parameter Estimate |
| INTERCEP            | 1  | -00096             |
| TIME                | 1  | 1.16331E-05        |
| RES1                | 1  | 0.927172           |

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסיה?

פיתרון בעיית המתאם הסדרתי – גרסיית הפרשים (שיטת קוקרן-אורקט)

ניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית שבה לא יהיה מתאם סדרתי ולכן ניתן יהיה לאמוד אותה בשיטת הריבועים הפחותים, האומדים יהיו יעילים וניתן יהיה לבצע בדיקת השערות.

$$\text{משוואה (1) : המודל בזמן } t: Y_t = \alpha + \beta X_t + u_t$$

$$\text{משוואה (2) : המודל בזמן } t-1 \text{ מוכפל ב- } \rho: \rho \cdot Y_{t-1} = \rho \cdot \alpha + \rho \cdot X_{t-1} + \rho \cdot u_t$$

החסרת משוואה (2) ממשוואה (1):

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

כדי לאמוד את הפרמטרים של רגרסיית הפרשים נגדיר:

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$\alpha^* = \alpha(1 - \rho)$$

$$\beta^* = \beta$$

$$X_t^* = X_t - \rho X_{t-1}$$

$$\varepsilon_t = u_t - \rho u_{t-1}$$

כך "נטרלנו" את המתאם הסדרתי:

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t - \rho \cdot u_{t-1} \quad \text{נזכור כי } \varepsilon_t \text{ מקיים את כל ההנחות הקלאסיות}$$

ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.

המשוואה "המתוקנת" אותה נאמוד:

$$Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$$

לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים:  $\hat{\alpha}, \hat{\beta}$

מאחר ש-  $\rho$  איננו ידוע יש צורך לאמוד אותו.

**?** סטודנט הניח כי במודל:  $Y_t = \alpha + \beta X_t + u_t$  קיים מתאם סדרתי מסדר ראשון

בשאריות כך שמתקיים:  $u_t = 0.7u_{t-1} + \varepsilon_t$  ולכן במקום לאמוד את המודל המקורי אמד את המודל:

$$Y_t - 0.7Y_{t-1} = \alpha(1 - 0.7) + \beta(X_t - 0.7X_{t-1}) + u_t$$

הסטודנט טען כי במודל החדש לא קיים מתאם סדרתי .

טענת הסטודנט: נכונה / לא נכונה / לא ניתן לדעת

### אמידת $\rho$ בשיטת קוקרן אורקוט

שיטת קוקרן אורקוט לאמידת  $\rho$  היא שיטה איטרטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות.

התהליך מתבצע באופן הבא:

אמידת המשוואה המקורית בה אנו מניחים כי קיים מתאם סדרתי.

חישוב  $\hat{u}_t$  ו-  $\hat{u}_{t-1}$  וקבלת האומד  $\hat{\rho}$  על ידי אמידת המשוואה:  $\hat{u}_t = \rho \hat{u}_{t-1} + \varepsilon_t$

הבעיה היא כי האומד איננו יעיל כיוון שהאומדים לפרמטרים במשוואה המקורית אינם יעילים (בשל קיומו של מתאם סדרתי). לכן נמשיך ונבצע את הפעולות הבאות:

הצבת האומד הזה בגרסיית ההפרשים וחילוץ  $\hat{\alpha}$  ו-  $\hat{\beta}$  חדשים. שוב נחשב  $\hat{u}_t$  ו-

$\hat{u}_{t-1}$  תוך שימוש באומדים החדשים ונקבל אומד  $\hat{\rho}$  חדש.

נחזור על התהליך הזה מספר פעמים נוספות עד שנגיע להתכנסות (עד שהאומד  $\hat{\rho}$  ישתווה לזה שקיבלנו בתחילה).

אין צורך לבצע תהליך זה ידנית משום שרוב תוכנות המחשב מבצעות הליך זה באופן מיידי ומדווחות על התוצאות הסופיות.

התהליך הממוחשב נקרא אוטו רגרסיה (AUTOREGRESION) מסדר ראשון, שני, שלישי וכו' (תלוי בסדר המתאם הסדרתי).

התיקון למתאם הסדרתי יתבצע על ידי הרצת רגרסיה עם משתנה  $AR(1)$  (אוטו רגרסיה מסדר ראשון),  $AR(1)$  ו- $AR(2)$  (אם מניחים קיום אוטו רגרסיה מסדר שני) וכו'.

אם משתנה  $AR$  מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

**לדוגמא:** נמשיך עם הדוגמא של ניבוי מחיר סגירה של מניה כפונקציה של הזמן:

$$CLOSE_t = \alpha + \beta \cdot TIME + u_t$$

נניח כי קיים מתאם סדרתי מסדר ראשון בנתונים.

תוצאות האמידה בשיטת אוטורגרסיה מסדר ראשון מוצגות להלן:

| Parameter Estimates |   |                    |                |                                  |
|---------------------|---|--------------------|----------------|----------------------------------|
| Variable DF         |   | Parameter Estimate | Standard Error | T for H0: Parameter=0<br>Prob> T |
| INTERCEP            | 1 | 1.333              |                |                                  |
| TIME                | 1 | -0.0006            |                |                                  |
| AR(1)               | 1 | 0.927              |                | 0.000                            |
| Durbin-Watson D     |   | 2.235              |                |                                  |

א. בדקו האם נפתרה בעיית המתאם הסדרתי.

ב. מהי המשוואה לאמידת מחיר הסגירה הצפוי ביום המסחר הבא?

**? שאלה מס' 1**

נאמד הקשר שבין הכנסה לתצרוכת לתקופה ינואר 1994 עד דצמבר 1997 ( $T=48$ ).

המודל הינו:  $C_t = \alpha + \beta Y_t + u_t$

בניסיון לבדוק האם מתקיים קשר מהסוג הבא:  $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$

נאמדה המשוואה הבאה:  $\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \gamma_2 \hat{u}_{t-2} + \gamma_3 \hat{u}_{t-3} + \gamma_4 Y_t + \omega_t$

Depended Variable: RES

תוצאות האמידה מוצגות להלן:

| Parameter Estimates |       |                    |                |                       |         |
|---------------------|-------|--------------------|----------------|-----------------------|---------|
| Variable            | DF    | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob> T |
| INTERCEP            | 1     | -54.709            | 710.85         | -0.076                | 0.939   |
| RESID1              | 1     | 0.705              | 0.152          | 4.631                 | 0.000   |
| RESID2              | 1     | -0.0066            | 0.188          | -0.035                | 0.972   |
| RESID3              | 1     | -0.337             | 0.167          | -2.012                | 0.051   |
| Y                   | 1     | 0.0027             | 0.032          | 0.085                 | 0.932   |
| Durbin-Watson D     | 1.954 |                    |                |                       |         |

נתון בנוסף כי  $R^2 = 0.479$

א. הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: \_\_\_\_\_

על ידי מבחן: \_\_\_\_\_

ההשערות הינן: \_\_\_\_\_

גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): \_\_\_\_\_

המסקנה המתקבלת היא: \_\_\_\_\_

בהנחה כי קיים מתאם סדרתי מסדר שלישי בנתונים נאמד מחדש הקשר שבין ההכנסה לתצרוכת בהתאם לשיטתם של קוקרן ואורקוט.

תוצאות האמידה מוצגות להלן:

Depended Variable:C

Parameter Estimates

Parameter Standard T for H0:

| Variable        | DF    | Estimate | Error   | Parameter=0 | Prob> T |
|-----------------|-------|----------|---------|-------------|---------|
| INTERCEP        | 1     | -2103.53 | 1128.38 | -1.864      | 0.069   |
| Y               | 1     | 0.71944  | 0.0510  | 14.086      | 0.000   |
| AR(1)           | 1     | 0.7070   | 0.1525  | 4.634       | 0.000   |
| AR(2)           | 1     | -0.0064  | 0.1889  | -0.034      | 0.972   |
| AR(3)           | 1     | -0.3282  | 0.1669  | -1.966      | 0.0562  |
| Durbin-Watson D | 1.954 |          |         |             |         |

ב. רשמו את המשוואה המתוקנת המשמשת לעריכת תחזיות.

## שאלה מס' 2

חוקר רצה לאמוד את עקומת הביקוש לטיסות לאירופה.

לרשותו נתונים שבועיים לאורך 3 שנים (52 שבועות).

נסמן:  $Y_t$  = מספר כרטיסי הטיסה לאירופה שנמכרו בשבוע t

$p_t$  = מחיר ממוצע ב-\$ של הכרטיסים שנמכרו בשבוע t

החוקר אמד את המודל:  $Y_t = e^\alpha \cdot P_t^{\beta_1} \cdot P_{t-1}^{\beta_2} \cdot e^{u_t}$

וקיבל לאחר הטרנספורמציה הלוגריתמית:  $R^2 = 0.81$ .

לבדיקת ההשערה כי קיים מתאם סדרתי בנתונים מסדר ראשון הוא חישב את

ערכי  $\hat{u}_t$  ולאחר מכן חישב את הרגרסיה:  $u_t = \gamma_0 + \gamma_1 \ln P_t + \gamma_2 \ln P_{t-1} + \gamma_3 u_{t-1} + v_t$

מקדם ההסבר המרובה ברגרסיה זו הוא 0.282

א. נסח את ההשערה ובחן אותה בר"מ של 0.05.

החוקר מניח שיש מתאם סדרתי מסדר ראשון.

לאחר תיקון Cochrane-Orcutt התקבל:

$$\ln Y_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$$

$$\hat{\rho} = 0.2$$

הניחו שהשבוע ובשבוע שעבר מחיר ממוצע של כרטיס היה \$500.

השבוע נמכרו 6,185 כרטיסים. בשבוע הבא צפוי מחיר של \$400.

ב. כמה כרטיסים יימכרו?

החוקר גם מנסה לקבוע האם בנתונים אלה קיים מתאם מסדר שני.

ג. רשמו את המשוואה הנוספת שעליו לאמוד.

במשוואה הנוספת התקבל מתאם מרובה השווה ל- 0.12.

ד. מהי המסקנה בר"מ של 0.05?

**פרק 13 - סיכום בעיית המתאם הסדרתי והטרוסקדסטיות**

| הטרוסקדסטיות                                              | מתאם סדרתי                                                                   |                           |
|-----------------------------------------------------------|------------------------------------------------------------------------------|---------------------------|
| למשל: $Y_t = \alpha + \beta X_t + u_t$                    |                                                                              | המשוואה העיקרית של המודל  |
| $V(u_t) = \sigma^2$                                       | $\text{cov}(u_t, u_{t-s}) = 0$                                               | ההנחה הקלאסית המופרת      |
| $V(u_t) = \sigma_t^2$                                     | $\text{cov}(u_t, u_{t-s}) \neq 0$                                            | המצב לאחר ההפרה           |
| $V(u_t) = W_t \sigma^2$                                   | $u_t = \rho u_{t-1} + \varepsilon_t$                                         | המשוואה המאפיינת את ההפרה |
| מתקבלים אומדים חסרי הטיה ועקיבים, אך תכונת היעילות נפגעת. |                                                                              | מה קורה אם אומדים ב-OLS   |
| מבחן GQ                                                   | מבחן DW                                                                      | זיהוי הבעיה               |
| מבחן White                                                | מבחן LM                                                                      |                           |
| שיטת WLS                                                  | שיטת קוקרן-אורקוט (רגרסיית ההפרשים)<br>הכנסת משתנה מוסבר בפיגור (מודל דינמי) | פתרון הבעיה               |