מבוא לאקונומטריקה

החוג לכלכלה

תוכן העניינים:

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 Y_t או X_t)

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_t יש תוחלת המייצגת את מרכז ההתפלגות.

התוחלת מסומנת μ_X או $E(X)$.

השונות מייצגת את מידת הפיזור של ההתפלגות.

השונות מסומנת σ_X^2 או $V(X)$.

סטית התקן היא השורש של השונות והיא מסומנת σ_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 cov (x,y)

כאשר:

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

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

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

X,Y בלתי תלויים $\Leftarrow$ X,Y בלתי מתואמים

<u>מקדם מתאם של פירסון</u> - מדד לכיוון ועוצמת הקשר הלניארי בין שני משתנים:

מקדם המתאם מסומן η_{xy}

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

$$-1\le\eta\le 1$$

כאשר:

$\eta=1$ **מתאם ליניארי חיובי מלא בין שני המשתנים**

$\eta=-1$ **מתאם ליניארי שלילי מלא בין שני המשתנים**

$\eta=0$ **לא קיים מתאם ליניארי בין שני המשתנים**

<u>אמידה</u>

<u>פרמטר</u>- ערך המשתנה הנחקר המתאר את כל האוכלוסיה

<u>סטטיסטי/אומד</u>-ערך המשתנה הנחקר המתאר את המדגם

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

נוסחאות וחוקים בסטטיסטיקה

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

חוקי הסיגמה

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

2) **סכום של קבוע:**

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

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

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

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:**

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

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

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

3) **מונה של השונות המשותפת:**

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

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 ≠ לכפל/חילוק התוחלות:**

$$E(X \cdot Y) \neq E(X) \cdot E(Y)$$
$$E(\frac{X}{Y}) \neq \frac{E(X)}{E(Y)}$$
$$E(X^2) \neq [E(X)]^2$$

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 השונויות

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

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

שונות של סכום/הפרש ≠ סכום/הפרש השונויות

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

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:**

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

2) **שונות משותפת של משתנים המוכפלים בקבוע:**

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

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

$$\operatorname{cov}(X,X)=V(X)$$
$$\operatorname{cov}(Y,Y)=V(Y)$$

פרק 1 - מבוא לקורס

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 לינארית.

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). α ו- β הם הפרמטרים של המודל. α נקרא חותך. β, או כל מקדם אחר של משתנה מסביר, נקרא שיפוע. u מכונה ההפרעה האקראית (לעיתים מכונה בבדיחות הדעת קריזה).

מודל: $$Y=\alpha+\beta X+u$$

הפרעה אקראית "קריזה"

משתנה מסביר

פרמטרים:
α - חותך
β - שיפוע

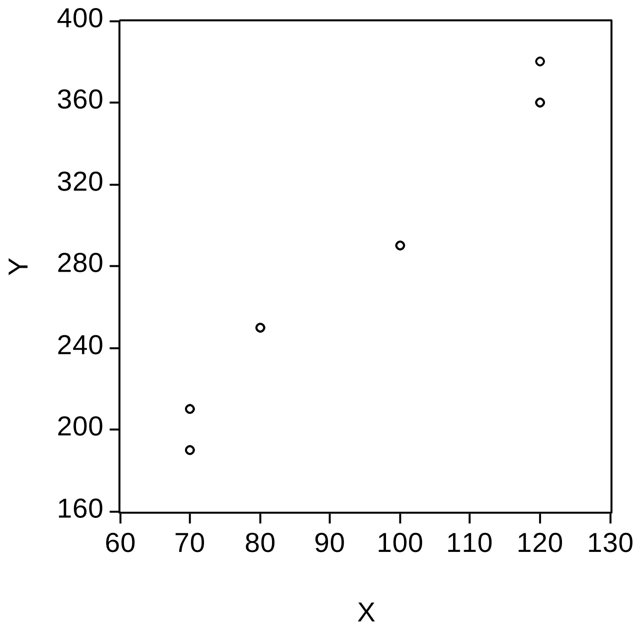
משתנה מוסבר

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.**

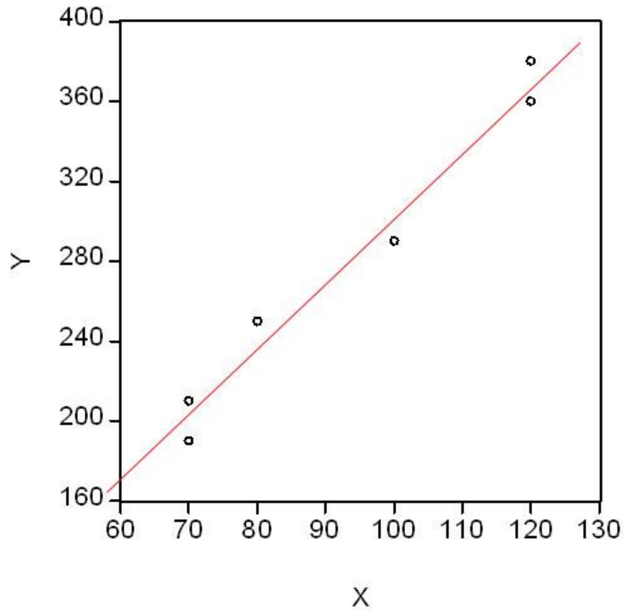
מספר הדירה	**גודל הדירה במ"ר**	**מחיר הדירה באלפי דולרים**
1	$X_1 = 70$	$Y_1 = 190$
2	$X_2 = 70$	$Y_2 = 210$
3	$X_3 = 80$	$Y_3 = 250$
4	$X_4 = 100$	$Y_4 = 290$
5	$X_5 = 120$	$Y_5 = 360$
6	$X_6 = 120$	$Y_6 = 380$

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 <u>בשיטת הריבועים הפחותים</u> (השיטה תתואר במלואה בהמשך):



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

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.

<u>לסיכום:</u>

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

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

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' - $\tilde{\beta}$.

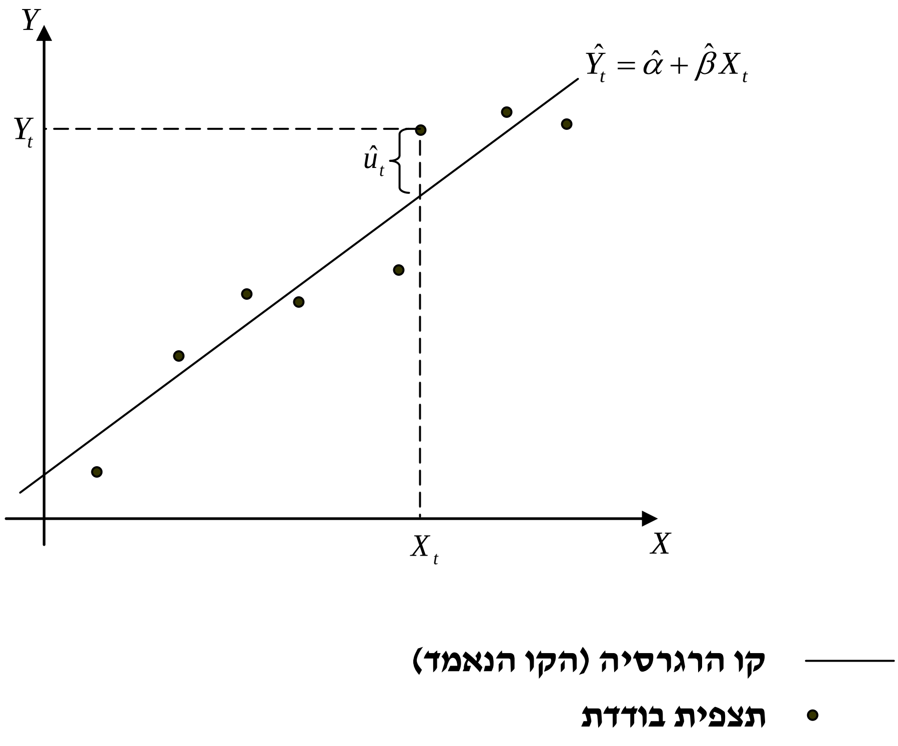
4) בעוד α ו- β הם קבועים, $\hat{\alpha}$ ו- $\hat{\beta}$ הם משתנים מקריים. מדוע? מפני שבכל מדגם מתקבלים $\hat{\alpha}$ ו- $\hat{\beta}$ אחרים.

5) את α ו- β 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 הוא $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$.**

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



קו הרגרסיה (הקו הנאמד)

תצפית בודדת

שלבי 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) וההנחות הקלאסיות

שיטת האמידה של α ושל β 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:

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

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

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

__הערה חשובה:__ בתהליך הגזירה של פונקציית הריבועים הפחותים מתקבלות "המשוואות הנורמליות":

__עבור המודל הקלאסי (עם חותך):__

בגזירה של α מתקבלת המשוואה הנורמאלית: $\sum \hat{u} = 0$

בגזירה של β מתקבלת המשוואה הנורמאלית: $\sum \hat{u} \cdot x = 0$

__עבור מודל ללא חותך:__

מתקבלת משוואה נורמאלית אחת מגזירת β בלבד: $\sum \hat{u} \cdot x = 0$

פיתרון המשוואות הנורמאליות נותן את נוסחאות האומדים- $\hat{\alpha}$ ו- $\hat{\beta}$ הנתונות בטבלה לעיל.

המשוואות הנורמאליות צריכות להתקיים על מנת שפונקציית הריבועים הפחותים תתקיים ($\sum \hat{u}_t^2 = \min$).

__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.

מוסבר=α+β מסביר+u

מקדם β: שיפוע הקו המתאר את הקשר בין המסביר למוסבר.

כדי שהקשר יהיה ליניארי שיפוע β 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) קיימים לפחות שני ערכי X ששונים 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) תוחלת ההפרעה האקראית היא אפס לכל תצפית: $E(u_t)=0$ **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-Xים 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 תצרוכת משפחה על סמך הכנסתה, כאשר נדגום משפחה ונשאל להכנסתה נצפה לקבל תשובה מסויימת (שההכנסה למשפחה לא תהיה אקראית) ולהניח כי זהו המשתנה המשפיע על התצרוכת ולא להיפךבמודל הניבוי הנוכחי בו אנו משתמשים.

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

5) הומוסקדסטיות: השונות של ההפרעה האקראית זהה לכל תצפית ותצפית: $V(u_t)=\sigma_u^2$ **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 התצרוכת של משפחות, אם דגמנו באופן אקראי את המשפחות, לא יהיה קשר בין הטעות בניבוי של תצרוכת משפחה מסוימת (u_t) לטעות בניבוי התצרוכת 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.

<u>לסיכום:</u>

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

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

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

4) X_t **אינם משתנים מקריים** $\Leftarrow$ **ניתן להוציא אותם מחוץ לתוחלת ולשונות** $\Leftarrow$ $\text{cov}(X_t,u_t)=0$

5) **הומוסקדסטיות: שונות ההפרעה האקראית קבועה לכל תצפית:** $V(u_t)=\sigma_u^2$ **לכל** t

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

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

<u>תכונות האומדים</u>

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

1) <u>ליניאריות</u>

אר"פ ניתנים להצגה כקומבינציה ליניארית של Y_t.

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

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

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

אומד זה ניתן להצגה בצורה הבאה:

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

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

לפיכך מדובר באומד ליניארי.

הוכחת ליניאריות עבור המודל הקלאסי:

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

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

<u>כלל אצבע-כיצד יודעים אם אומד הוא ליניארי?</u>

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).**

לסיכום: 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 האומדים הבאים הם אומדים ליניאריים?

א. $$\hat{\beta}=\frac{\sum_{t=1}^{T}(X_t-\bar{X})Y_t}{\sum_{t=1}^{T}(X_t-\bar{X})^2}$$

ב. $$\tilde{\beta}=\frac{\sum_{t=1}^{T}Y_t^3\sum_{t=1}^{T}X_t(Z_t+Y_t)}{\sum_{t=1}^{T}X_t^2}$$

2) <u>חוסר הטיה</u>

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

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

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

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

<u>כיצד יודעים אם אומד הוא חסר הטיה?</u>

בשלב הראשון יש לבצע עבודת הכנה –

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

**** יש לזכור כי:**

u_t / y_t **מהווים משתנים מקריים** $\Leftarrow$ **נשארים בתוך התוחלת, השונות וה-$\sum$.**

x_t **<u>איננו</u> משתנה מקרי (על פי הנחה מס' 4)** $\Leftarrow$ **יוצא מחוץ לתוחלת ולשונות אך נשאר בתוך ה-$\sum$**

α / β **קבועים** $\Leftarrow$ **יוצאים מחוץ לתוחלת, לשונות ול-$\sum$**

דוגמא:

עבור המודל $Y_t = \beta X_t + u_t$ **והאומד המתאים לו** $\hat{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$ **:**

$$\hat{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2} = \frac{\sum X_t(\beta X_t + u_t)}{\sum X_t^2} = \frac{\beta \sum X_t^2}{\sum X_t^2} + \frac{\sum X_t u_t}{\sum X_t^2} = \beta + \frac{\sum X_t u_t}{\sum X_t^2}$$

שלב מקדים זה יעשה לפני בדיקת חוסר הטייה, יעילות ועקיבות.

הוכחת חוסר הטיה – מפעילים תוחלת על האומד, ואם התוחלת שווה לפרמטר האמיתי אז האומד חסר הטיה.

בשפה מתמטית: אם $E(\hat{\beta}) = \beta$ **, אז** $\hat{\beta}$ **הוא אומד חסר הטיה ל-** β **.**

כדי שהדבר יתקיים הנחות (3) ו- (4) חייבות להתקיים.

המשך הדוגמא שלעיל:

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

מסקנה: האומד חסר הטיה!

כלל אצבע:

אם בעבודת ההכנה נשארים בסוף הפיתוח רק שני סוגי איברים:

1) הפרמטר האמיתי

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

אז האומד חסר הטיה.

למשל, בעבודת ההכנה שלעיל:

$$\hat{\beta} = \beta + \frac{\sum X_t u_t}{\sum X_t^2}$$

איבר המכיל את u_t   **הפרמטר האמיתי**

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) <u>יעילות</u>

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.

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

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

<u>משפט גאוס מרקוב:</u>

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

<u>כיצד מחשבים שונות של אומד?</u>

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

$$V\left(\hat{\beta}\right)=V\left(\frac{\sum X_t u_t}{\sum X_t^2}\right)=\frac{V\left(\sum X_t u_t\right)}{\left(\sum X_t^2\right)^2}=\frac{\sum V\left(X_t u_t\right)}{\left(\sum X_t^2\right)^2}=$$

$$=\frac{\sum X_t^2 V\left(u_t\right)}{\left(\sum X_t^2\right)^2}=\frac{\sum X_t^2\sigma_u^2}{\left(\sum X_t^2\right)^2}=\frac{\sigma_u^2\sum X_t^2}{\left(\sum X_t^2\right)^2}=\frac{\sigma_u^2}{\sum X_t^2}$$

2. במודל עם חותך

$$V(\hat{\beta})=V(\frac{\sum(X_t-\bar{X})u_t}{S_{XX}})=\frac{V\sum(X_t-\bar{X})u_t}{S_{XX}{}^2}=\frac{\sum V(X_t-\bar{X})u_t}{S_{XX}{}^2}=$$

$$\frac{\sum\left[(X_t-\bar{X})^2V(u_t)\right]}{S_{XX}{}^2}=\frac{\sum\left[(X_t-\bar{X})^2V(u_t)\right]}{S_{XX}{}^2}=\frac{\sigma^2 S_{XX}}{S_{XX}{}^2}=\frac{\sigma^2}{s_{xx}}$$

4) <u>עקיבות</u>

ככל שהמדגם יגדל כן יתקרב האומד לערך האמיתי של הפרמטר.

אם נגדיל את המדגם לאינסוף תצפיות ונחשב את האומד, הוא יהיה שווה לפרמטר האמיתי באוכלוסיה ($\underset{T \to \infty}{\hat{\theta} \to \vartheta}$)

<u>תנאי הכרחי לעקיבות:</u> 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.

****<u>הערה חשובה:</u> בכדי שאומד יהיה עקיב, הנחות 1-4 צריכות להתקיים.**

<u>סיכום: השלבים להוכחת התכונות</u>

1) הוכחת ליניאריות

2) הכנת האומד ⇐ להציב במקום Y_t את המודל האמיתי.

במודל עם חותך: $Y_t = \alpha + \beta X_t + u_t$

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 הטייה 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 התלוים בגודל המדגם: ככל שגודל המדגם גדול יותר כך שונות האומד קטנה והאומד יהיה יעיל יותר לפרמטר באוכ'.**

<u>תרגול ממבחנים</u>

<u>תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 25 נקודות)</u>

נתון המודל $Y_t = \alpha + \beta X_t + u_t$ **,** $T = 100$

כאשר מתקיימות כל ההנחות הקלאסיות.

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

א. האומד $\tilde{\beta}$ **הינו אומד חסר הטיה ל-** β **נכון / לא נכון**

ב. האומד $\tilde{\beta}$ **הינו אומד עקיב ל-** β **נכון / לא נכון**

ג. האומד $\tilde{\beta}$ **הינו אומד לינארי ל-** β **נכון / לא נכון**

ד. האומד $\tilde{\beta}$ **הינו אומד יעיל ל-** β **נכון / לא נכון**

ה. השונות האמיתית של $\tilde{\beta}$ **היא:**

<u>תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 14 נקודות)</u>

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

(יש לשים לב המודל ללא חותך)

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

א. האומד $\tilde{\beta}$ **הינו אומד מוטה ל-** β **: 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 $\tilde{\beta}$ **?**

<u>תרגיל המבוסס על שאלה ממבחן לדוגמא (בשווי של 16 נקודות)</u>

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

(יש לשים לב המודל ללא חותך)

נתון האומד: $\tilde{\beta} = \dfrac{\sum X_t Y_t}{\sum \left(X_t - \bar{X}\right)^2}$

א. מהי התוחלת של $\tilde{\beta}$?

ב. $E\left(\tilde{\beta}\right) < \beta$. **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 $\dfrac{\sum X_t Y_t}{\sum X_t^2}$?

שאלות נוספות מתוך מבחנים

בכל השאלות ההנחות הקלאסיות מתקיימות.

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

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

2. $\sum_{t=1}^{T}\left(X_t - \bar{X}\right)\bar{Y} \neq 0$ **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. $\sum_{t=1}^{T}\left(X_t-\bar{Y}\right)\hat{u}_t=0$

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

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 <u>קשר קוי</u> 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 <u>השינוי השולי</u> (השיפוע של הגרף): אם מגדילים את x ביחידה אחת, בכמה יחידות משתנה y.
במודל הליניארי- שינוי זה הוא קבוע ושווה ל- β**.**

בניגוד למודל הליניארי, שלושת המודלים האחרים (המודלים הלוגריתמיים) מתארים קשרים שאינם ליניאריים בין X ל-Y. במודלים אלו <u>השינוי השולי</u> (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:**

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

השינוי השולי איננו קבוע אלא תלוי ב-y: $\frac{\partial Y}{\partial X}=\beta\cdot Y$**.**

ככל ש-Y גדל כך השיפוע (β**) גדל.**

משמעות ה- β **במודל כזה היא <u>שיעור</u> השינוי השולי:** $\beta=\frac{\frac{\partial Y}{\partial X}}{Y}$

שיעור שינוי שולי אומר: אם מגדילים את X ביחידה, בכמה <u>%</u> 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 הבאה: $Y = e^{\alpha} \cdot X^{\beta} \cdot e^{u}$

השינוי השולי איננו קבוע אלא תלוי ב-x וב-y: $\frac{\partial Y}{\partial X} = \beta \cdot \frac{Y}{X}$.

משמעות ה- β **במודל כזה היא <u>הגמישות</u>:** $\beta = \frac{\partial Y}{\partial X} \cdot \frac{X}{Y} = \frac{\frac{\partial Y}{Y}}{\frac{\partial X}{X}}$

משמעות הגמישות היא שינוי שולי באחוזים: אם מגדילים את X ב-% אחד, בכמה % ישתנה Y.

במודל הלוגריתמי הכפול ה- β **מייצגת את הגמישות, כלומר אם נגדיל את X ב-% אחד, Y ישתנה ב-** β **%.**

במודלים אלו הגמישות היא קבועה למרות שהשינוי השולי איננו קבוע.

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

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

השינוי השולי איננו קבוע אלא תלוי ב-X. ככל ש-X עולה כך פוחת השינוי השולי: $\frac{\partial Y}{\partial X} = \frac{\beta}{X}$.

במודל זה אם X עולה ב-% אחד, Y עולה ב- β **:** $\beta = \frac{\partial Y}{\frac{\partial X}{X}}$

ל- β **אין משמעות כלכלית במודל זה.**

<u>גמישות</u>

בנוסף למשמעות ה- β בכל אחד מהמודלים, מושג נוסף שיש להכיר הוא מושג <u>הגמישות</u>.

כאמור, גמישות משמעה: שינוי שולי <u>באחוזים</u>. 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 ($\frac{\partial Y}{\partial X}$) ב- $\frac{X}{Y}$

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

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

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

4) **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.

<u>טרנספורמציות של המודלים הלא ליניאריים לקו ישר:</u>

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:

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

$$LN(X \cdot Y) = LN(X) + LN(Y)$$
$$LN(\frac{X}{Y}) = LN(X) - LN(Y)$$

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.

<u>סיכום:</u>

המודל	משמעות ה- β	השינוי השולי ($\frac{\partial Y}{\partial X}$) בכמה ישתנה Y אם נגדיל את X ביחידה?	הגמישות ($\frac{\partial Y}{\partial X}\cdot\frac{X}{Y}$) בכמה % ישתנה Y אם נגדיל את X ב-1%?
ליניארי $Y=\alpha+\beta X+u$	**השינוי השולי** אם נגדיל את X ביחידה Y ישתנה ב- β יחידות	β	$\frac{\beta X}{Y}$
חצי לוגריתמי $\ln Y=\alpha+\beta X+u$ ($Y=e^{\alpha+\beta x+u}$)	**שיעור השינוי השולי** אם נגדיל את X ביחידה Y ישתנה ב- $100\cdot\beta\%$	βY	βX
לוגריתמי כפול $\ln Y=\alpha+\beta\ln X+u$ ($Y=e^{\alpha}\cdot X^{\beta}\cdot e^{u}$)	**הגמישות** אם נגדיל את X ב-1% Yישתנה ב- β%	$\frac{\beta Y}{X}$	β
לוג ליניארי $Y=\alpha+\beta\ln X+u$ ($e^{y}=e^{\alpha}\cdot X^{\beta}\cdot e^{u}$)	**אין משמעות כלכלית** אם נגדיל את X ב-1% Y ישתנה ב- β	$\frac{\beta}{X}$	$\frac{\beta}{Y}$

<u>המשתנה שיש בו LN השינוי בו יהיה באחוזים</u>**

תרגול

על מנת לאמוד את התשואה להשכלה בישראל בשנים 1948-1990 נאמדו המודלים הבאים:

$$MWAGE_t = 139.547 + 118.628 \cdot SCL_t \qquad \textbf{(1)}$$

$$MWAGE_t = -1445.08 + 1239.60 \cdot LN(SCL)_t \qquad \textbf{(2)}$$

$$LN(MWAGE)_t = 5.244 + 0.778 \cdot LN(SCL)_t \qquad \textbf{(3)}$$

$$LN(MWAGE)_t = 6.292 + 0.070 \cdot SCL_t \qquad \textbf{(4)}$$

א. הסבירו את המשמעות של β בכל אחד מהמודלים

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

נתונים תוצאות האמידה של המודלים הבאים:

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

כאשר:

Q- הוצאות צריכה על מוצר מסוים על ידי פרט מסוים.

A- הוצאות צריכה על המוצר בהינתן רמת הכנסה ו/או שנות לימוד אפסיים.

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 - מבחני המובהקות וקריאת פלטים (של תוכנת SAS)

פלט ניתוח שונות (Analysis of Variance)

להלן פלט ניתוח שונות של SAS:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	k	RSS	$RSS/k = MSR$	$F = \frac{MSR}{MSE}$	PF
Error	$T-k-1$	ESS	$ESS/(T-k-1) = MSE$		
C Total	$T-1$	TSS			

Root MSE	$\sqrt{MSE} = s_u$	R-square	$R^2 = \frac{RSS}{TSS}$
Dep Mean	$\bar{Y}$	Adj R-sq	$\bar{R}^2 = 1 - \frac{ESS}{TSS} \cdot \frac{T-1}{T-k-1}$
C.V.	$\frac{S_u}{\bar{Y}} \cdot 100$		

פלט זה מתחלק לשני חלקים:

החלק הראשון (מעל לקו המקווקו) מתאר את מבחן F למובהקות מודל הרגרסיה.

החלק השני (מתחת לקו המקווקו) מתאר מדדים חשובים של מודל הרגרסיה.

מדדים חשובים של מודל הרגרסיה

1) מדד R^2 לטיב ההתאמה (R-square)

<u>עונה על השאלה:</u> איזה אחוז מהשונות של המשתנה התלוי (Y) מוסבר על ידי קו הרגרסיה, ההיגיון הכלכלי (X־ים)?

מדד לפרופורציית השונות המוסברת.

$$0 \le R^2 = \frac{RSS}{TSS} = 1 - \frac{ESS}{TSS} \le 1$$

- **R^2 נע בין 0 ל-1. ככל שקרוב יותר ל-1 ההתאמה טובה יותר.**
- **אר"פ מביא למקסימום את R^2**
- **בהוספת משתנים מסבירים נוספים למודל, R^2 יכול רק לעלות או "להתנפח" או לכל היותר להשאר ללא שינוי (כתוצאה מהירידה בשונות הלא מוסברת ESS).**

2) $\bar{R}^2$ - R^2 המתוקן (Adj R-sq)

<u>עונה על השאלה:</u> האם כדאי היה לי להוסיף משתנים ב"ת נוספים למודל?

משמש להשוואה בין מודלים בעלי מספר שונה של משתנים מסבירים:

$$\bar{R}^2 = R^2 \cdot \left(\frac{T-1}{T-k-1}\right)$$

- **המדד המתוקן לפרופורצית השונות המוסברת ($\bar{R}^2$) לוקח בחשבון את "ההפסד" בדרגות החופש כתוצאה מהוספת המשתנים למודל ולא רק את "הרווח" בירידת השונות הלא מוסברת (ה-ESS).**
- **לכן, בניגוד ל- R^2, ה- $\bar{R}^2$ יכול לרדת בהוספת משתנים למודל ולא רק לעלות.**

מבחן F למובהקות מודל הרגרסיה

<u>מבחן מובהקות העונה לשאלה:</u> האם מודל הרגרסיה שלנו לניבוי משתנה תלוי מסויים על ידי k משתנים ב"ת, מובהק באוכלוסיה?

<u>השערות:</u>

$$H_0: R^2 = 0$$
$$H_1: R^2 > 0$$

<u>סטטיסטי המבחן:</u>

$$F = \frac{MSR}{MSE} = \frac{\frac{RSS}{K}}{\frac{ESS}{T-K-1}} = \frac{\frac{R^2}{K}}{\frac{1-R^2}{T-K-1}}$$

- **את סטטיסטי F נחשב בעזרת לוח ניתוח השונות המוצג בחלק הראשון בפלט.**

<u>כלל החלטה ומסקנה:</u>

נדחה את H0 כאשר:

שימוש בטבלת F: $F>F(K,T-K-1;1-\alpha)$

שימוש בפלט: $PF<\alpha$

<u>מסקנה:</u> יש/אין עדות לכך שמודל הרגרסיה מובהק באוכלוסיה.

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

(במיליארדי $) שגובה מדינה במערב לפי המודל : $TAX_t = \alpha + \beta \cdot INCOME_t + u_t$.

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

Model: MODEL1

Dependent Variable: TAX

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2046.89694	2046.89694	8798.672	0.0001
Error	49	11.39922	0.23264		
C Total	50	2058.29615			

Root MSE	0.48232	R-square	0.9945
Dep Mean	5.4242	Adj R-sq	0.9943
C.V.	8.88711		

בדקו את ההשערה כי המודל מובהק ברמת מובהקות של 0.05.

פלט מקדמי הרגרסיה (Parameter Estimates)

להלן פלט מקדמי הרגרסיה של SAS:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	$\hat{\alpha}$	$s_{\hat{\alpha}}$	$\frac{\hat{\alpha}}{s_{\hat{\alpha}}}=t_{(\hat{\alpha}=0)}$	$Pt_{\hat{\alpha}}$
X	1	$\hat{\beta}$	$s_{\hat{\beta}}$	$\frac{\hat{\beta}}{s_{\hat{\beta}}}=t_{(\hat{\beta}=0)}$	$Pt_{\hat{\beta}}$

ניתוח הפלט

הפלט לעיל מתאר את מקדמי הרגרסיה: $\hat{\alpha}$ ו- $\hat{\beta}$ ומובהקותם.

- **שורה ראשונה מתייחסת ל- $\hat{\alpha}$ המכונה ה"חותך" של קו הרגרסיה (INTERCEP) ומובהקותו.**
- **השורות הבאות מתייחסות למקדם של המשתנים הבלתי תלויים , ל- $\hat{\beta}$ טות.**

  בדוגמא שלהלן קיים משתנה מסביר אחד בלבד (k=1).

  במודל של רגרסיה רבת משתנים יתווספו שורות נוספות כמספר המשתנים הבלתי תלויים במודל.
- **Parameter Estimate – מתאר את ערך המקדמים $\hat{\alpha}$ וה- $\hat{\beta}$ טות. כל שאר הנתונים מתייחסים למבחני המובהקות שלהם.**

מבחן t למובהקות ה-β

<u>מבחן מובהקות העונה לשאלה:</u>האם משתנה מסביר מסויים רלוונטי למודל (מובהק)?

<u>השערות:</u>

$$H_0 : \beta = 0$$
$$H_1 : \beta \neq 0$$

<u>סטטיסטי המבחן:</u>

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

<u>כלל הכרעה ומסקנה:</u>

נדחה את H0 אם:

שימוש בטבלת T: $|t_{\beta=0}| > t_{(T-K-1,1-\frac{\alpha}{2})}$

שימוש בפלט: $Pt_{\hat{\beta}} < \alpha$

<u>מסקנה:</u> יש/אין עדות לכך שהמשתנה הב"ת המסויים מובהק באוכ' (ולכן רלוונטי למודל).

<u>הערות:</u>

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

  <u>ההשערות:</u>

  $$H_0 : \beta = 0$$
  $$H_1 : \beta > / < 0$$

  <u>כלל ההכרעה:</u>

  שימוש בטבלת T:

  $$t_{\beta=0} > t_{(T-K-1,1-\alpha)}$$
  $$t_{\beta=0} < -t_{(T-K-1,1-\alpha)}$$

  שימוש בפלט:

  $$\frac{pt_{\hat{\beta}}}{2} < \alpha$$

• ניתן לבדוק בנוסף האם ה-β (השינוי השולי) שווה לערך מסויים באוכ׳.

<u>השערות לדוגמא:</u>

$$H_0 : \beta = 2$$
$$H_1 : \beta \neq 2$$

<u>סטטיסטי המבחן:</u>

$$t_{\beta=2} = \frac{\hat{\beta}-2}{S_{\hat{\beta}}}$$ **(זה לא נתון בפלט של SAS וצריך לחשב)**

<u>כלל הכרעה:</u>

ניתן להשתמש בטבלת T אך לא ניתן להשתמש ב- $Pt_{\hat{\beta}}$ **שבפלט.**

במקרה זה ניתן גם לחשב רווח בר סמך ל- β ולראות האם הוא מכיל את הערך המבוקש (את β_0) אם כן- נקבל את H0 ואם לא-נדחה אותה.

• **רב"ס ל-β:**

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

?בהמשך לדוגמא הקודמת- בדיקת השפעת ההכנסה על גודל המס, התקבלו גם התוצאות הבאות:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-0.086912	0.08953904	-0.971	0.3365
INCOME	1	0.152232	0.0016229	93.801	0.0001

א. אמדו את המודל : $TAX = \alpha + \beta \cdot INCOME + U$

מהי המשמעות הכלכלית של β ?

ב. האם המודל מובהק? בדקו על סמך הפלט הנ"ל ברמת מובהקות של 0.05.

ג. מהי רמת המובהקות הקטנה ביותר, עבורה עדיין תידחה השערת האפס מסעיף ב׳ ?

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס (שיפוע β חיובי) ברמת מובהקות של 0.01.

ה. בנו רווח-סמך ברמת סמך של 95% עבור β

ו. **בדקו את ההשערה שתוספת של מיליארד $ להכנסה תגדיל את המס ב - 0.2 מיליארד $, ברמת מובהקות של 0.05.**

- **שימו לב כי-**

במודל עם משתנה מסביר אחד בלבד קיימת זהות בין מבחן F למובהקות המודל לבין מבחן t למובהקות ה-β:

$$F_{(1,T-2;1-\alpha)} = t^2_{(T-2,1-\frac{\alpha}{2})}$$

$$F = t^2_{\hat{\beta}}$$

$$PF = Pt_{\hat{\beta}}$$

כלומר: כל החלטה המתקבלת במבחן אחד חייבת להיות זהה להחלטה המתקבלת במבחן השני.

בשאלה לדוגמא ניתן לראות כי:

$$F_{(1,49;0.95)} = 4 = t^2_{(49,0.975)} = 2^2$$

$$F = 8798.672 = t^2 = 93.801^2$$

$$PF = 0.0001 = Pt_{\hat{\beta}}$$

<u>פלט ה-Covariance of Estimates</u>

פלט שמתאר את השונות המשותפת (covariance) של האומדנים- $\hat{\alpha}$ ו- $\hat{\beta}$:

Covariance of Estimates		
COVB	**INTERCEP**	X
INTERCEP	$s^2_{\hat{\alpha}}$	$\text{cov}(\hat{\alpha},\hat{\beta})$
X	$\text{cov}(\hat{\alpha},\hat{\beta})$	$s^2_{\hat{\beta}}$

שימוש בטבלת ה-cov של SAS יעשה במקרה שבו בהשערת האפס מופיעים שני הפרמטרים- α ו- β

במקרה כזה יוצרים פרמטר חדש: γ (גמה) המהווה קומבינציה ליניארית של α ו- β

<u>השערות:</u>

$$H_0: \gamma = 0$$
$$H_1: \gamma \neq 0$$

<u>סטטיסטי המבחן:</u>

$$t = \frac{\hat{\gamma} - \gamma}{S_{\hat{\gamma}}}$$

מחשבים את $\hat{\gamma}$ על ידי הצבת האומדנים- $\hat{\alpha}$ ו- $\hat{\beta}$ (מתוך הפלט של SAS).

את השונות של $\hat{\gamma}$ מחשבים תוך שימוש בנוסחאות :

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

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

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

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.

דוגמא:

נתון פלט האמידה של המודל $Y_t = \alpha + \beta X_t + u_t$

שלצורך אמידתו נאספו 240 תצפיות:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.25	0.25	21	0.0000
X	1	0.96	0.12	8	0.0000

Covariance of Estimates

	INTERCEP	X
INTERCEP	0.0625	-0.003
X	-0.003	0.0144

יש לבדוק את ההשערה $H_0: \alpha = 5\beta$

ניצור פרמטר חדש γ השווה לקומבינציה ליניארית של α ו- β : $\gamma = \alpha - 5\beta$

<u>ההשערות:</u>

$$H_0: \gamma = 0$$
$$H_1: \gamma \neq 0$$

<u>סטטיסטי המבחן:</u>

$$t = \frac{\hat{\gamma}-\gamma}{S_{\hat{\gamma}}}$$

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

חישוב השונות של $\hat{\gamma}$ **(**$S^2_{\hat{\gamma}}$**):**

$$V(\hat{\gamma}) = V(\hat{\alpha}-5\hat{\beta}) = V(\hat{\alpha}) + V(5\hat{\beta}) - 2\,\mathrm{cov}(\hat{\alpha},5\hat{\beta}) =$$
$$V(\hat{\alpha}) + 5^2 V(\hat{\beta}) - 2\cdot 5\,\mathrm{cov}(\hat{\alpha},\hat{\beta}) =$$
$$0.0625 + 25\cdot 0.0144 - 10\cdot(-0.003) = 0.4525$$

סטית התקן היא: $S_{\hat{\gamma}} = \sqrt{0.4525} = 0.6726$

נחשב את הסטטיסטי: $t_{\hat{\gamma}} = \dfrac{0.45-0}{0.6726} = 0.66$

כלל הכרעה ומסקנה: $t_{\hat{\gamma}} = 0.66 < t_{(238,0.975)} = 1.96$ **, לכן אין סיבה מספקת לדחות את השערת האפס.**

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

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

המודל: $\ln(\mathrm{SALARY}_t) = \alpha + \beta\cdot \mathrm{EXP}_t + u_t$ **. הוא אסף 403 תצפיות, ואמד את הפרמטרים בתוכנת SAS. להלן חלקים מהפלט ויש להשלימו:**

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**
Model	**----**	**----**	**5.68015**	**----**	**----**
Error	**----**	**205.22539**	**----**		
C Total	**----**	**----**			

Root MSE	**----**	**R-square**	**----**
Dep Mean	**7.14247**	**Adj R-sq**	**0.0245**
C.V.	**10.01602**		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	---- (*)	----	----	----
EXP	1	-0.008740	----	----	0.0009

Covariance of Estimates

COVB	INTERCEP	EXP
INTERCEP	0.0047463101	----
EXP	-0.000154685	6.882844 E-6

(*) נתון נוסף: $\overline{\text{EXP}} = 22$

א. קיים קשר חיובי מובהק בין ותק ללוג השכר. **נכון / לא נכון**

ב. שיעור התשואה בשכר לשנת ותק הוא:

ג. תחזית לוג השכר עבור אדם בעל 10 שנות ותק היא:

עריכת תחזית וקריאת פלטים (של SPSS)

אמידה נקודתית

אמידה נקודתית עבור X_0 מסוים (תחזית) מחושבת על פי קו הרגרסיה במדגם:

$$\hat{Y}_0 = \hat{\alpha} + \hat{\beta} \cdot X_0$$

אמידת מרווח ל-$E(Y)$

כאשר אנו מתבקשים לאמוד את התחזית באוכלוסיה עבור X_0 מסוים, נחשב רווח בר סמך לערך ממוצע של Y באוכ' עבור X_0 מסוים ($E(Y)$) ברמת סמך $1-\alpha$.

נוסחת הרב"ס:

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

$$\hat{\sigma}_u = MSE = \frac{SSE}{n-2} \qquad \sum (X_i - \bar{X})^2 = S_{XX} = (n-1)S_X^2$$

רישום הרב"ס: $p(\_\_\_ \le E(Y) \le \_\_\_\_) = 1-\alpha$

אמידת מרווח ל-Y

כאשר אנו מתבקשים לאמוד ערך בודד של Y באוכלוסיה עבור X_0 מסוים, נחשב רווח בר סמך לערך בודד של Y באוכ' עבור X_0 מסוים (Y_0) ברמת סמך $1-\alpha$.

נוסחת הרב"ס:

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

רישום הרב"ס: $p(\_\_\_ \le Y \le \_\_\_\_) = 1-\alpha$

רב"ס לערך בודד יהיה רחב יותר מאשר רב"ס לערך ממוצע משום שטעות התקן בראשון גדולה מאשר באחרון.

רב"ס לערך ממוצע מנסה לאמוד את התחזית באוכ' עבור ערך מסוים של X ואילו רב"ס לערך בודד מנסה לאמוד את תחום ערכי Y באוכ' עבור ערך X מסוים.

התחזית מדויקת יותר ***(שונות*** התחזית קטנה יותר) כאשר:

1. n (גודל המדגם) גדול יותר
2. שונות המשתנה המסביר X גדולה יותר
3. X_0 קרוב יותר ל $\bar{X}$
4. האומד לשונות הטעויות- $\hat{\sigma}_u$, קטן יותר.

דוגמא:

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

Descriptive Statistics

	Mean	Std. Deviation	N
שכר הדירה	1386.7667	509.46027	30
מספר הסטודנטים	3.0000	1.31306	30

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602[a]	.362	.339	414.05503

a. Predictors: (Constant), number of students
b. Dependent Variable:rent

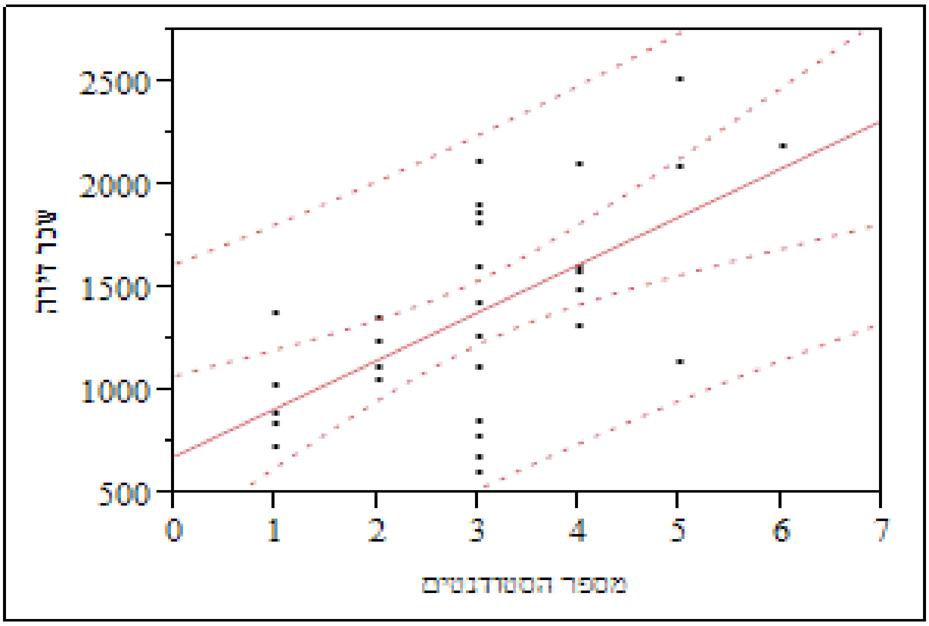
ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2726579.520	1	2726579.520	15.904	.000[a]
	Residual	4800363.847	28	171441.566		
	Total	7526943.367	29			

a. Predictors: (Constant), number of students
b. Dependent Variable: rent

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	T	Sig.
		B	Std. Error	Beta		
1	(Constant)	686.207	191.244		3.588	.001
	מספר הסטודנטים	233.520	58.556	.602	3.988	.000



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

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

3. **אמוד את שכר הדירה שישלם סטודנט יחיד החולק את הדירה עם שותף אחד בלבד, ברמת ביטחון של 95% .**

פרק 5 - 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-** α' **ו-** β' **ואת האומדים ב-** $\hat{\alpha}'$ **ו-** $\hat{\beta}'$**).**

השינויים מסוכמים בטבלה הבאה:

	$\hat{\beta}'$	$\hat{\alpha}'$	$S_{\hat{\beta}}$	$S\hat{\alpha}$
הוספת קבוע ל- X **:** $Y=\alpha'+\beta'(X+d)+v$	$\hat{\beta}'=\hat{\beta}$	$\hat{\alpha}'=\hat{\alpha}-\hat{\beta}d$	$s_{\hat{\beta}'}=s_{\hat{\beta}}$	$s_{\hat{\alpha}'}\neq s_{\hat{\alpha}}$
הוספת קבוע ל- Y **:** $Y+d=\alpha'+\beta'X+v$	$\hat{\beta}'=\hat{\beta}$	$\hat{\alpha}'=\hat{\alpha}+d$	$s_{\hat{\beta}'}=s_{\hat{\beta}}$	$s_{\hat{\alpha}'}=s_{\hat{\alpha}}$
הכפלת X **פי קבוע:** $Y=\alpha'+\beta'(dX)+v$	$\hat{\beta}'=\dfrac{\hat{\beta}}{d}$	$\hat{\alpha}'=\hat{\alpha}$	$s_{\hat{\beta}'}=\dfrac{s_{\hat{\beta}}}{d}$	$s_{\hat{\alpha}'}=s_{\hat{\alpha}}$
הכפלת Y **פי קבוע:** $dY=\alpha'+\beta'X+v$	$\hat{\beta}'=d\hat{\beta}$	$\hat{\alpha}'=d\hat{\alpha}$	$s_{\hat{\beta}'}=ds_{\hat{\beta}}$	$s_{\hat{\alpha}'}=ds_{\hat{\alpha}}$

מסקנות מהטבלה:

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

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

חוקר ביקש לאמוד את הקשר בין שכר בש"ח (MWAGE) לבין שנות לימוד (SCL) באמצעות 2 מודלים שונים.

להלן תוצאות האמידה :

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

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

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.**

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?

תזכורת של חוקי לוגים:

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

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

פרק 6 - רגרסיה מרובה

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.

המודל הקלאסי:

$$Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t$$

- **קבוע α יש אחד**
- **מספר ה-β טות כמספר המשתנים הב"ת במודל.**

מבחני המובהקות

מבחן F למובהקות המודל

השערות:

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0$$
$$H_1 : OTHERWISE$$

סטטיסטי המבחן F וכלל ההכרעה:

$$F = \frac{\frac{RSS}{k}}{\frac{ESS}{T-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{T-k-1}} > F(k, T-k-1; 1-\alpha)$$

מבחן t למובהקות ה-βטות

מבחן לבדיקת מובהקות β ספציפית.

השערות:

$$H_0 : \beta_i = 0$$
$$H_1 : \beta_i \neq 0$$

סטטיסטי המבחן t וכלל ההכרעה:

$$\left| t_{\hat{\beta}_i} \right| = \left| \frac{\hat{\beta}_i}{S_{\hat{\beta}_I}} \right| > t_{(T-k-1; 1-\frac{\alpha}{2})}$$

❓ נאמד המודל $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

Analysis of Variance

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

Root MSE	------------	R-square	-------------
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

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
X	1	-------------	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-------------	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

א. השלם את הנתונים החסרים בפלט.

ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05

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

השוואה בין מודלים- $\bar{R}^2$ וחוק חיטובסקי

לעיתים אנו מתבקשים להכריע בסוגיה האם כדאי לנו להוסיף למודל משתנה ב"ת מסויים. במקרה זה נשווה את פרופורציית השונות המוסברת המתוקנת $\bar{R}^2$ בין המודל ללא המשתנה המסביר לבין המודל עם המשתנה המסביר שהוספנו.

להזכירכם $\bar{R}^2$ בניגוד ל- R^2 לוקח בחשבון לא רק את הרווח שבהקטנת החלק הלא מוסבר בהוספת משתנים ב"ת למודל אלא גם את ההפסד שבירידה בדרגות החופש. לכן בניגוד ל- R^2 הוא יכול לקטון בהוספת משתנים ב"ת למודל.

לפי חוק חיטובסקי-בהוספת משתנה מסביר אחד בלבד למודל ה- $\bar{R}^2$ יעלה אך ורק אם $|t_{\hat{\beta}}|>1$.

כאשר $|t_{\hat{\beta}}|<1$ אז $\bar{R}^2$ ירד בהוספת המשתנה והוא גם לא יהיה רלוונטי למודל (מובהק).

כאשר $|t_{\hat{\beta}}|>2$ אז $\bar{R}^2$ יעלה והמשתנה שהוסף יהיה גם מובהק.

כאשר $1<|t_{\hat{\beta}}|<2$ אז ה- $\bar{R}^2$ יעלה אך יש לבדוק את רלוונטיות המשתנה שהוסף למודל על פי מבחן t.

❓ **במודל לניבוי ההכנסה על פי שנות לימוד וותק במקום העבודה, התקבל**

$\bar{R}^2=0.266$. הוסף המשתנה היקף המשרה. במבחן למובהקות המשתנה הנוסף התקבל: $t_{\hat{\beta}}=o.456$. האם ערך $\bar{R}^2$ יעלה/ירד/לא ישתנה בהוספת המשתנה הנוסף למודל?

<u>הערה חשובה:</u> ניתן להשתמש גם באומד המוטה- R^2 להשוואה בין מודלים אם מתקיימים שני התנאים הבאים:

1) **מספר המשתנים זהה**
2) **המשתנה המוסבר זהה**

מבחן WALD

אם רוצים לבדוק השערת אפס שיש בה מספר לא מוגבל של שוויונים משתמשים במבחן WALD (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 המודל, נסמן אותו ב- ESS_U.

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 המודל, נסמן אותו ב- ESS_R **.**

5) אם יודעים את מספר התצפיות, T**, מספר המשתנים המסבירים במודל הלא-מוגבל,** k**, ומספר השוויונים בהשערת האפס,** m**, אפשר לחשב את הסטטיסטי:**

$$WALD_{stat} = \frac{\frac{ESS_R - ESS_U}{m}}{\frac{ESS_U}{T-k-1}}$$ **(המודל המקורי עם חותך).**

m **- דרגות החופש של המונה**

$T-k-1$ **- דרגות החופש של המכנה**

אם במעבר ממודל U ל-R המשתנה המוסבר לא השתנה, ניתן לחשב את סטטיסטי המבחן WALD גם במונחים של R^2 **:**

$$WALD_{stat} = \frac{\frac{R_U^2 - R_R^2}{DF_R - DF_U}}{\frac{1-R_U^2}{DF_U}}$$

<u>כלל הכרעה לדחיית H0:</u>

$$WALD_{stat} > F_{(DF_R - DF_U, DF_U; 1-\alpha)}$$

אם דוחים את H0 המסקנה היא שהמודל המקורי (הלא-מוגבל) הוא הרלוונטי ולהיפך.

נאמד המודל $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ **והתקבלו התוצאות הבאות:**

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	646169.84	161542.46	51835.84	0.0000
Error	199	620.1683	3.1164236		
C Total	203	646790.01			

Root MSE	1.7653395	R-square	0.999041
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

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
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

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

להלן אמידת המודל המוגבל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	646166.01	323083.01		
Error	201	623.9983	3.104469		
C Total	203	646790.01			

Root MSE	1.7619504	R-square	0.999035
Dep Mean	173.6645	Adj R-sq	0.999026
C.V.	1.0145714		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
X	1	0.978491	0.036399	26.88240	0.0000
Z+3S	1	2.999995	0.003669	817.6080	0.0000
W	1	-5.043109	0.071218	-70.81249	0.0000

ג. חשב את הסטטיסטי של WALD.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

מקרים פרטיים של מבחן WALD

1) מבחן F למובהקות המודל הוא מקרה פרטי של WALD :

השערות:

$$H_0: \beta_1 = \beta_2 = ... = \beta_k = 0$$
$$H_1: OTHERWISE$$

אם נבצע זאת במבחן WALD :

המודל הלא מוגבל יהיה:

U: $$Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + u_t$$

המודל המוגבל יהיה:

R: $Y_t = \alpha + u_t$

במקרה זה :

$$WALD_{stat} = F$$

$$WALD_{stat} = \frac{\dfrac{R_U^2 - R_R^2}{DF_R - DF_U}}{\dfrac{1-R_U^2}{DF_U}} = \frac{\dfrac{R_U^2}{k}}{\dfrac{1-R_U^2}{T-k-1}} = F$$

2) מבחן t למובהקות ה-β הוא מקרה פרטי של מבחן WALD:

למשל במודל:

$$Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k + u_t$$

רוצים לבדוק את מובהקות β_2

השערות:

$$H_0 : \beta_2 = 0$$
$$H_1 : \beta_2 \neq 0$$

ניתן לבדוק זאת גם במבחן WALD:

U: $Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + ... + \beta_k X_k + u$

R: $Y_t = \alpha + \beta_1 X_1 + \beta_3 X_3 + ... + \beta_k X_k + u$

במקרה זה :

$$WALD_{stat} = t_{\hat{\beta}}^2$$

$$F_{(1,DF_U;1-\alpha)} = t^2_{(DF_U;1-\frac{\alpha}{2})}$$

$$PF_{WALD} = Pt_{\hat{\beta}}$$

3) כשיש מגבלה אחת ב-HO- ניתן לבצע גם מבחן t עם הגדרת סטטיסטי $\hat{\gamma}$.

במקרה זה:

$$WALD_{stat} = t_{\hat{\gamma}}^2$$

$$F_{(1,DF_U;1-\alpha)} = t^2_{(DF_{\hat{\gamma}};1-\frac{\alpha}{2})}$$

$$PF_{WALD} = Pt_{\hat{\gamma}}$$

לדוגמא:

על מנת לאמוד את פונקצית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה:

$$C_t = \alpha + \beta_1 \cdot W_t + \beta_2 \cdot P_t + u_t$$

להלן תוצאות האמידה של המשוואה הנ"ל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	---	---------	--------	-------	--------
Error	---	---------	52968		
C Total	---	---------			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-107.226	------	----------	-------
W	1	0.743	------		
P	1	0.561	-------		

Covariance of Estimates

COV	INTERCEP	W	P
INTERCEP	-----------	-----------	--------
W	-----------	0.0046	-0.0090
P	-----------	-0.0090	0.016

על מנת לבדוק את ההשערה שהנטיה השולית לצרוך מתוך ההכנסה זהה לנטיה השולית לצרוך מתוך ההון, נאמדה גם המשוואה הבאה:

$C_t = \alpha + \beta_1 \cdot Y_t + u_t$ **כאשר:** Y_t= **סה"כ ההכנסה של משק בית t**

התקבל: ESS=0.4566

בידקו את ההשערה בשתי דרכים.

<u>לסיכום</u>: מתי נשתמש במבחן-t ומתי במבחן-F ?

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

תרגול מסכם

חוקר אמד את התצרוכת של 500 משקי בית כפונקציה של הכנסה שלהן לפי המשוואה: $EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$.

$EXPENSE_t$ **- התצרוכת של משק הבית ה-t-י באלפי שקלים**

$INCOME_t$ **- ההכנסה של משק הבית ה-t-י באלפי שקלים**

ההפרעות האקראיות מקיימות את כל ההנחות הקלאסיות

התקבל הפלט הבא:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2013.105	2013.105	6495.745	0.0000
Error	498	154.3358	0.3099112		
C Total	499	2167.441			

Root MSE	0.556697	R-square	0.928794
Dep Mean	3.990208	Adj R-sq	0.928651
C.V.	13.95157		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	0.041995	0.054951	0.764236	0.4451
INCOME	1	0.713503	0.008853	80.59618	0.0000

1) **מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?**
2) **מהו אחוז השונות בתצרוכת המוסבר ע"י ההכנסה?**
3) **מהו אומדן לתצרוכת ההתחלתית של משק בית?**
4) **האם אומדן זה מובהק?**
5) **על עוזר מחקר הטיל החוקר לבדוק את ההשערה כי על כל 1000 ש"ח נוספים בהכנסה צורך הפרט 700 ש"ח, כנגד ההשערה כי הוא צורך יותר מ-700 ש"ח. נסח את השערת האפס ואת ההשערה האלטרנטיבית.**

6) מהו הסטטיסטי t לבדיקת ההשערה?

7) מהו הסטטיסטי WALD לבדיקת ההשערה?

8) התברר כי היתה טעות בנתונים, וכי יש להוסיף 1000 ש"ח לתצרוכת של כל משק בית:

א. ההוספה תגדיל את האומד ל- α : נכון / לא נכון / לא ניתן לדעת.

ב. בעקבות ההוספה האומד ל- α יהיה מובהק: נכון / לא נכון / לא ניתן לדעת.

ג. ההוספה תשנה את האומד ל- β : נכון / לא נכון / לא ניתן לדעת.

ד. ההוספה תשנה את R^2 : נכון / לא נכון / לא ניתן לדעת.

החוקר טען כי יש להוסיף לפונקצית התצרוכת גם את השפעת העושר. העושר של משק בית מורכב מתוכניות החסכון שלו ($SAVINGS$) ומניירות הערך שיש לו (NE). שתי סדרות הנתונים הן באלפי שקלים.

החוקר אמד את המשוואה:

$$EXPENSE_t = \alpha + \beta_1 \cdot INCOME_t + \beta_2 \cdot SAVINGS_t + \beta_3 \cdot NE_t + u_t$$

וקיבל כי סכום ריבועי הסטיות של הטעויות הוא 121.

9) מהי השערת האפס לבדיקת הטענה של החוקר (שהמודל החדש נכון ולא המקורי)?

10) מהו הסטטיסטי WALD לבדיקת ההשערה?

החוקר רצה לבדוק את ההשערה כי הנש"צ מתוך ההכנסה שווה ל-0.6 וכי השפעת ניירות הערך על התצרוכת היא פי 2 מהשפעת תוכניות החסכון.

11) מהי השערת האפס לבדיקה זו?

12) המודל המוגבל לבדיקת ההשערה יהיה מהצורה $Z_t = \gamma_0 + \gamma_1 W_t + v_t$.

בטא את Z_t , ו- W_t באמצעות המשתנים המקוריים.

פרק 7 - בעיות ספציפיקציה

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

ישנם שני סוגים של טעויות ספציפיקציה:

1) **הוספת משתנה בלתי תלוי שאיננו רלוונטי (איננו מובהק) למודל.**

2) **השמטת משתנה בלתי תלוי רלוונטי (מובהק) למודל.**

1) הוספת משתנה לא רלוונטי

נניח שהמודל האמיתי (הנכון) כולל שני משתנים בלתי תלויים, אבל אנחנו חושבים בטעות כי הוא כולל גם משתנה בלתי תלוי שלישי.

המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$

המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon$

אם נקבל את HO במבחן t למובהקות β_3 נסיק כי המשתנה איננו רלוונטי ונאמוד את המודל מחדש הפעם ללא המשתנה השלישי.

אולם, גם אם לא נוכל לאמוד מחדש, הימצאותו של משתנה שאיננו רלוונטי במודל הרגרסיה איננה פוגמת ברלוונטיות של המשתנים האחרים במודל ולא בתכונות החיוניות למבחני המובהקות שלהם.

כלומר, האומדים האחרים $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ יהיו חסרי הטיה, יעילים ועקיבים לפרמטרים באוכלוסיה $(\alpha, \beta_1, \beta_2)$ וגם השונויות הנאמדות תהיינה חסרות הטיה, כך שניתן יהיה לבצע על האומדים האחרים בדיקת השערות תקפה.

<u>**מסקנה:**</u>

טעות ספציפיקציה של הוספת משתנה שאיננו רלוונטי למודל איננה טעות שיוצרת בעיה.

****שימו לב כי לעיתים נרצה להשאיר משתנה שאיננו מובהק במודל. מדובר במקרים בהם המשתנה הנוסף מעלה את $AdjR^2$ למרות שאיננו מובהק, כיוון ש: $1 < t_{\hat{\beta}}$ (לפי חוק חיטובסקי).**

2) השמטת משתנה רלוונטי

נניח שהמודל האמיתי כולל שני משתנים בלתי תלויים אולם אנו אומדים בטעות מודל הכולל רק את אחד המשתנים הב"ת:

המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$

המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$

במקרה זה עלולה להיווצר בעיה בהשמטת משתנה x_2.

א. <u>בעיה באמידה של β_1 :</u>

הבעיה תיווצר במקרה ובו קיים קשר בין המשתנה הב"ת שהושמט לבין המשתנים הב"ת האחרים במודל.

כאשר <u>$r_{12} \neq 0$</u> , בהיעדר x_2 , $\hat{\beta}_1$ לא ישקף טוב את הפרמטר באוכלוסיה β_1 (יהיה אומד מוטה לפרמטר) מכיוון שהוא "יספח לעצמו" חלק מן ההשפעה שיש ל- x_2 על Y.

- **הדגמה באמצעות מעגלי ואן:**

****שימו לב- אם מקדם השיפוע של x_1 שונה במודל הנאמד לעומת במודל האמיתי ניתן להסיק מכך כי** $\text{cov}_{12} \neq 0$

ניתן להעריך את <u>כיוון ההטיה</u> (חיובית או שלילית) של האומד ל- $\hat{\beta}_1$ ביחס לפרמטר β_1 באוכלוסיה:

גודל ההטיה: $\beta_2 \cdot \frac{\operatorname{cov}(x_1, x_2)}{\operatorname{cov}(x_1, x_1)}$ **או** $\beta_2 \cdot \frac{S_{12}}{S_{11}}$

$$E(\hat{\beta}_1) = \beta_1 + \beta_2 \cdot \frac{S_{12}}{S_{11}}$$

אם $S_{12} = 0$ **האומד ל-** β_1 **יהיה חסר הטיה (למרות היעדרו של** x_2 **מהמודל).**

אם S_{12} **ו-** β_2 **הם שווי סימן-ההטיה חיובית-האומד ל-** β_1 **יהיה מוטה כלפי מעלה.**

אם S_{12} **ו-** β_2 **הם שוני סימן-ההטיה שלילית- האומד ל-** β_1 **יהיה מוטה כלפי מטה.**

ב. <u>בעיה באמידה של α:</u>

$$E(\hat{\alpha}) = \alpha + \beta_2 (\bar{x}_2 - \bar{x}_1 \frac{S_{12}}{S_{11}})$$

ניתן לראות כי גם כאשר $S_{12} = 0$ **האומד ל-** α **יהיה מוטה.**

רק כאשר $S_{12} = 0$ **וגם** $\bar{x}_2 = 0$ **האומד ל-** α **יהיה חסר הטיה.**

ג. <u>בעיה באמידת השונות של</u> β_1 **<u>ו-</u>** α **:**

אומד השונות יהיה מוטה תמיד כלפי מעלה כי הוא איננו לוקח בחשבון את החלק ש- x_2 **מסביר ב-Y:**

$S_{\hat{\beta}_1}^2 = \frac{ESS}{T-k-1}$ **מכיוון ש-ESS מוטה כלפי מעלה אז** $s_{\hat{\beta}_1}^2$ **מוטה כלפי מעלה.**

<u>מסקנה:</u>

בדיקת ההשערות על הפרמטרים של המודל הנאמד איננה תקפה.

<u>סיכום:</u>

בהיעדר x_2 , בדיקות ההשערות לפרמטרים של המודל הטעותי אינן תקפות::

	אומד ל- β_1	**אומד ל- α**	**אומד לשונות הפרמטרים**
$S_{12}=0$	**חסר הטיה**	**מוטה** **אלא אם-** $\bar{x}_2=0$	**מוטה (כלפי מעלה)**
$S_{12}\neq 0$	**מוטה** **<u>כיוון ההטיה:</u>** **חיובי:** S_{12} **ו-** β_2 **שווי סימן** **שלילי:** S_{12} **ו-** β_2 **מנוגדי סימן**	**מוטה**	

פרק 8 - 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 שמס' שנות הלימוד של אדם משפיעות על שכרו:

W_t=**השכר (המשתנה התלוי)**

S_t = **שנות לימוד (המשתנה הב"ת)**

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.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t$

החותך מייצג כאן את השכר ההתחלתי.

שכר ההתחלתי של אישה: α_0

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

הבדל בשכר בין נשים וגברים: α_1 **(הפרש בין החותכים)**

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש

החותכים: $H0: \alpha_1 = 0$

**** 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:**

$$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$$

(S.E) (134) (56) (24)

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.**

<u>פונקציית רגרסיה המכילה משתנים איכותיים בלבד</u>

המגדר הוא המשתנה היחיד במשוואה:

$$W_t = \alpha_0 + \alpha_1 D + u_t$$

החותך מייצג כאן את השכר ה<u>ממוצע</u> עבור כל קטגוריה.

שכר הממוצע של אישה: α_0

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

הבדל בשכר הממוצע בין נשים וגברים: α_1 **(הפרש בין החותכים)**

בדיקת השערות על משתנה הדמי: מבחן t: $H0:\alpha_1=0$ **(מבחן זהה 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:

$$W_t=5200+1120\cdot D$$

נתון: $S_{\hat{\alpha}_1}=63$

בדקו האם קיים הבדל מובהק בשכר ה<u>ממוצע</u> בין נשים וגברים?

(2) משתנה דמי לשיפוע

המגדר משפיע על התוספת לשכר בגין שנות הלימוד.

$$W_t=\alpha+\beta_0 S_t+\beta_1 DS_t+u_t$$

השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.

אצל אישה: התוספת לשכר בגין שנות לימוד- β_0

אצל גבר: התוספת לשכר בגין שנות לימוד- $\beta_0+\beta_1$

הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: β_1 **(הפרש השיפועים)**

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש השיפועים: $H0:\beta_1=0$

**** 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) (23) (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.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 D S_t + u_t$

השכר ההתחלתי של אישה: α_0

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

הבדל בשכר ההתחלתי בין המינים: α_1 **(הבדל בחותכים)**

אצל אישה-התוספת לשכר בגין שנות הלימוד: β_0

אצל גבר-התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$

הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: β_1 **(הבדל בשיפועים)**

<u>בדיקת השערות למשתני הדמי:</u>

באמצעות מבחן WALD יש לבדוק: $H0: \alpha_1 = \beta_1 = 0$

H1: לפחות אחד הפרמטרים שונה מ-0

אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים בנפרד:

$HO:\alpha_1 = 0$ ו- $H0:\beta_1 = 0$

<u>מבחן CHOW</u>

דרך נוספת לבדיקת ההבדל בין הקטגוריות, <u>בלא יצירת משתני דמי</u>:

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 .

נשים: $W_t = \alpha_f + \beta_f X_t + u_t$

גברים: $W_t = \alpha_m + \beta_m X_t + u_t$

השערות: $H_0 : \alpha_f = \alpha_m ; \beta_f = \beta_m$

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALD שהשתמשנו בו מקודם):

המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן לא יכלול את המדגם המאוחד כי אין צורך בשתי רגרסיות נפרדות:

$$W_t = \alpha + \beta X_t + u_t$$

המודל הלא מוגבל (U) כולל את שני חלקי המדגם:

$$ESS_U = ESS_f + ESS_m$$
$$DF_U = DF_f + DF_m$$

$$CHOW_{stat} = \frac{\dfrac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\dfrac{ESS_f + ESS_m}{DF_f + DF_m}} = WALD_{stat}$$

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

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 <u>הפונקציה</u> של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

(1) כבישים מהירים בלבד

(2) כבישים לא - מהירים בלבד

(3) שני סוגי הכביש (כל המדגם)

(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 WALD ?**

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
Model	1	4700.81174	4700.81174	89.12	<.0001
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
Model	1	8052.00804	8052.00804	288.84	<.0001
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 לחותך	**משתנה דמי לשיפוע**	**משתנה דמי לכל הפונקציה**
המודל	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$
השערה במילים	**קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).**	**קיים הבדל בין הקטגוריות בתוספת ל- Y בגין X (בשיפוע).**	**קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).**
בדיקת השערה	**מבחן t להפרש החותכים:** $H0: \alpha_1 = 0$	**מבחן t להפרש השיפועים:** $H0: \beta_1 = 0$	**מבחן WALD להפרש בין הפונקציות (החותכים והשיפועים):** $H0: \alpha_1 = \beta_1 = 0$ ****ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW.** **אם דוחים את HO יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALD):** $H0: \alpha_1 = 0$ $H0: \beta_1 = 0$

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:

$V_t=$ **מדד מחירי הירקות**

$p_t=$ **מדד המחירים לצרכן**

(1) <u>*משתני דמי לחותך*</u>

<u>**הטענה:**</u> **יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות**

<u>**המודל:**</u> $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- β.

למחיר זה יתווסף α_0 **בחורף,** $\alpha_0+\alpha_1$ **באביב,** $\alpha_0+\alpha_2$ **בקיץ ו-** $\alpha_0+\alpha_3$ **בסתיו.**

ניתן לראות כי: α_0**: החותך בקטגוריה שהושמטה**

$\alpha_0+\alpha_i$**: החותך בקטגוריה i.**

<u>**בדיקת השערות:**</u>

השערות:

$$H0:\alpha_1=\alpha_2=\alpha_3=0$$
$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

(U) $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$

(R) $V_t=\alpha\ \ +\beta\cdot P_t+u_t$

****שימו לב שהחותך במשוואה המוגבלת איננו α_0 שכן המשתנה המסביר של עונות השנה ירד.**

אם נדחה את HO 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) <u>משתני דמי לשיפוע</u>

<u>הטענה:</u> יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן

<u>המודל:</u> $V_t=\alpha+\beta_0P_t+\beta_1(D_{1i}P_t)+\beta_2(D_{2i}P_t)+\beta_3(D_{3i}P_t)+u_t$

המחיר ההתחלתי של הירקות שווה בין עונות השנה (α) אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב:

β_0 בחורף, $\beta_0+\beta_1$ באביב, $\beta_0+\beta_2$ בקיץ ו- $\beta_0+\beta_3$ בסתיו.

ניתן לראות כי: β_0 : השיפוע בקטגוריה שהושמטה

$\beta_0+\beta_i$: השיפוע בקטגוריה i.

<u>בדיקת השערות:</u>

השערות:

$$H0:\beta_1=\beta_2=\beta_3=0$$
$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

(U) $V_t=\alpha+\beta_0P_t+\beta_1(D_{1i}P_t)+\beta_2(D_{2i}P_t)+\beta_3(D_{3i}P_t)+u_t$

(R) $V_t=\alpha+\beta\cdot P_t+u_t$

****שימו לב שהשיפוע במשוואה המוגבלת איננו β_0 שכן המשתנה המסביר של עונות השנה ירד.**

אם נדחה את HO במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t.

(3) <u>משתני דמי לכל הפונקציה</u>

<u>הטענה:</u>יש הבדל בין עונות השנה בפונקצית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

<u>המודל:</u> $V_t=\alpha_0+\alpha_1D_{1i}+\alpha_2D_{2i}+\alpha_3D_{3i}+\beta_0P_t+\beta_1(D_{1i}P_t)+\beta_2(D_{2i}P_t)+\beta_3(D_{3i}P_t)+u_t$

<u>בדיקת השערות:</u>

השערות:

$$HO:\alpha_1=\alpha_2=\alpha_3=\beta_1=\beta_2=\beta_3=0$$

המבחן הסטטיסטי: מבחן WALD:

(U) $V_t=\alpha_0+\alpha_1 D_{1i}+\alpha_2 D_{2i}+\alpha_3 D_{3i}+\beta_0 P_t+\beta_1(D_{1i}P_t)+\beta_2(D_{2i}P_t)+\beta_3(D_{3i}P_t)+u_t$

(R) $V_t=\alpha+\beta\cdot P_t+u_t$

אם דוחים את HO, יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים או בין השיפועים:

$$HO:\beta_1=\beta_2=\beta_3=0 \quad HO:\alpha_1=\alpha_2=\alpha_3=0$$

באם דוחים את H0 יש להמשיך לבדוק באמצעות מבחני t:

$$HO:\beta_j=0 \qquad H0:\alpha_j=0$$

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) <u>הבדל בחותך ללא אינטראקציה</u>

<u>המודל:</u> $W_t=\alpha_0+\alpha_1 G+\alpha_2 R+\beta\cdot S_t+u_t$

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 הב"ת האיכותיים בנפרד:

1. **הבדל בשכר ההתחלתי בין גברים לנשים:** $HO:\alpha_1=0$

2. **הבדל בשכר ההתחלתי בין שחורים ללבנים:** $H0:\alpha_2=0$

(2) <u>הבדל בחותך עם אינטראקציה</u>

<u>המודל:</u> $W_t=\alpha_0+\alpha_1 G+\alpha_2 R+\alpha_3 G\cdot R+\beta\cdot S_t+u_t$

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. $HO:\alpha_3=0$

<u>דרך נוספת ליצירת מודל עם אינטראקציה:</u>

הגדרת משתני דמי המייצגים שילוב בין המשתנים האיכותיים 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.

<u>המודל:</u> $W_t=\gamma_0+\gamma_1 D_1+\gamma_2 D_2+\gamma_3 D_3+\delta\cdot S_t+u_t$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה:

	גבר	אישה	הפרש
לבן	$\gamma_0+\gamma_1$	$\gamma_0+\gamma_3$	$\gamma_1-\gamma_3$
שחור	$\gamma_0+\gamma_2$	γ_0	γ_2
הפרש	$\gamma_1-\gamma_2$	γ_3	

ההשערות לבדיקת קיום האינטראקציה:

$HO:\gamma_1-\gamma_3=\gamma_2$ **או** $HO:\gamma_1-\gamma_2=\gamma_3$

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:

$$\ln(Y)_t=\alpha_0+\alpha_1D_1+\alpha_2D_2+\alpha_3D_3+\beta_1EXP_t+\beta_2EXP_t^2+u_t$$

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 > F
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 > |t|
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:

	השכלה נמוכה	**השכלה גבוהה**	**הפרש**
לבנים	α_0	$\alpha_0+\alpha_3$	α_3
שחורים	$\alpha_0+\alpha_2$	$\alpha_0+\alpha_1$	$\alpha_1-\alpha_2$
הפרש	α_2	$\alpha_1-\alpha_3$	

(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 WALD.

הרגרסיה המוגבלת תחת השערת האפס הינה:

$Z_0=\gamma_0+\gamma_1 Z_1+\gamma_2 Z_2+\gamma_3 Z_3+\gamma_4 Z_4+v$ **מהם ה-Z-ים?**

(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').

<u>שאלה מס' 2</u>

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:

	השכלה גבוהה (E=1)	השכלה נמוכה (E=0)	הפרש
נשים (S=1)	$\alpha_0 + \alpha_1 + \alpha_2 + \alpha_3 + (\beta_0 + \beta_1 + \beta_2 + \beta_3)EXP_t$	$\alpha_0 + \alpha_1 + (\beta_0 + \beta_1)EXP_t$	חותך: $\alpha_2 + \alpha_3$ שיפוע: $\beta_2 + \beta_3$
גברים (S=0)	$\alpha_0 + \alpha_2 + (\beta_0 + \beta_2)EXP_t$	$\alpha_0 + \beta_0 EXP_t$	חותך: α_2 שיפוע: β_2
הפרש	חותך: $\alpha_1 + \alpha_3$ שיפוע: $\beta_1 + \beta_3$	חותך: α_1 שיפוע: β_1	

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 9 - משוואות סימולטניות

עוסקות בהפרה של הנחה 4 המדברת על כך שהמשתנים הב"ת במשוואת הרגרסיה אינם משתנים מקריים ולכן לא מתואמים עם הטעויות: $\text{cov}(x,u)=0$.

במילים אחרות ה-Xים במשוואה נחשבו <u>משתנים אקסוגניים</u>- משפיעים על Y אך לא מושפעים ממנו בחזרה.

זהו סוג של משתנים כלכליים הנשלט על ידי קובעי המדיניות או גורמים חיצוניים אחרים.

דוגמא למשתנה אקזוגני: שער חליפין קבוע (הנקבע באופן אקזוגני על ידי הבנק המרכזי) המשפיע על כמות הכסף במשק אך לא מושפע ממנה בחזרה.

לעומת זאת, קיים סוג נוסף של משתנים כלכליים ב"ת הקרויים <u>משתנים אנדוגניים</u>- משפיעים על Y אך גם מושפעים ממנו בחזרה.

מאחר ומשתנים אלו הם גם מסבירים וגם מוסברים, הם נחשבים כמשתנים מקריים, המתואמים עם הטעויות במודל: $\text{cov}(x,u)\neq 0$.

לדוגמא: התצרוכת הפרטית מושפעת בדר"כ מרמת ההכנסה, וזו מושפעת מן הביקושים השונים לתוצר, ולכן גם ההכנסה וגם התצרוכת נקבעים ביחד באופן אנדוגני במערכת.

משוואות המבנה(משוואות סימולטניות)

מערכת משוואות הכוללות משתנים מסבירים אנדוגניים ואקסוגניים.

בדר"כ מדובר בשתי משוואות אשר המשתנה המוסבר בראשונה הוא משתנה מסביר בשנייה והמשתנה המוסבר בשנייה הוא משתנה מסביר בראשונה.

משתנים המופיעים באחת המשוואות כמוסברים ובאחרת כמסבירים הם משתנים אנדוגניים. יתר המשתנים במשוואות הם אקסוגניים.

לדוגמא:

(1)

(2)

X ו-Y הם משתנים אנדוגניים הנקבעים סימולטנית במערכת (למשל, תצרוכת והכנסה) ואילו ה-Z-ים הם משתנים אקזוגניים (כמו למשל, שער הדולר ומגדר).

המטרה היא , כמו תמיד, לאמוד בצורה יעילה את הפרמטרים (אלפות ובטות) ולבצע בדיקת השערות.

הבעיה היא כי ברגע שקיים משתנה <u>מסביר</u> אנדוגני במערכת, אמידת כל אחת מהמשוואות המבניות בנפרד בשיטת OLS תניב אומדים <u>לא ליניאריים, מוטים, לא יעילים ולא עקיבים.</u>

השלכות על אר"פ

הנחה 4 שימשה אותנו להוכחת ליניאריות, חוסר הטיה ועקיבות.

לכן הפרתה של הנחה זו משמעה פגיעה בכל תכונות אר"פ. האומדים לא ליניאריים, מוטים לא עקיבים ולכן גם לא יעילים (לפי גאוס מרקוב). אומד השונות מוטה גם הוא ובדיקת ההשערות לא תקפה (ללא תלות בגודל המדגם).

נלמד 3 שיטות אמידה אחרות: IV, ILS ו- TSLS. אך ראשית יש לבטא את המשוואות המבניות בצורה אחרת הנקראת: "הצורה המצומצמת".

הצורה המצומצמת של מודל עם משוואות סימולטניות

משוואות הצורה המצומצמת הן פיתרון עבור המשתנים האנדוגניים במערכת: הגדרת המשתנים האנדוגניים כפונקציה של המשתנים האקזוגניים במערכת בלבד.

מספר המשוואות המצומצמות הוא כמספר המשתנים האנדוגניים במערכת (במקרה זה שניים).

בדוגמא שלנו:

$$Y_t = \mu_0 + \mu_1 \cdot Z_{1t} + \mu_2 \cdot Z_{2t} + \nu_t$$

$$X_t = \lambda_0 + \lambda_1 \cdot Z_{1t} + \lambda_2 Z_{2t} + \zeta_t$$

<u>תכונות המשוואות מהצורה המצומצמת</u>

- **מס' המשוואות הוא כמספר המשתנים האנדוגניים במערכת (X ו-Y).**
- **המשתנה המוסבר הוא אנדוגני וכל המסבירים אקסוגניים.**
- **המשתנים המסבירים הם זהים בכל המשוואות (ה-Zים).**
- **מכיוון שכל המשתנים המסבירים הם אקסוגניים ניתן לאמוד את הפרמטרים (ה- λ ות וה- μ ים) ב-OLS ולקבל אומדים ליניאריים, חסרי הטיה, יעילים ועקיבים עם יכולת לבצע בדיקת השערות.**

<u>אמידת הפרמטרים של משוואות המבנה באמצעות משוואות הצורה המצומצמת</u>

משוואות הצורה המצומצמת מאפשרות, כאמור, לאמוד את הפרמטרים (ה- λ ות וה- μ ים) בשיטת OLS אבל אנחנו מעוניינים למעשה לאמוד את הפרמטרים של המשוואות המקוריות- משוואות המבנה (ה- α ות וה- β ות).

מתוך הפרמטרים של הצורה המצומצמת נחלץ את הפרמטרים של משוואות המבנה: מתוך ה- λ ות נחלץ את ה- α ות ומתוך ה- μ ים נחלץ את ה- β ות.

בתהליך החילוץ של הפרמטרים המבניים ייתכנו 3 מצבים:

1) <u>אין זיהוי:</u> לא ניתן לחלץ את הפרמטרים המבניים (ה- α ות וה- β ות) מתוך הפרמטרים של הצורה המצומצמת (ה- λ ות וה- μ ים). זה קורה כשיש

פחות משוואות מנעלמים (כלומר פחות λ ות ו- μ ים מ- α ות ו- β ות) אז יש אינסוף פתרונות.

2) <u>זיהוי מדויק:</u> יש רק דרך אחת לחלץ את הפרמטרים המבניים מהפרמטרים של הצורה המצומצמת, זה קורה כשיש בדיוק אותו מספר משוואות ונעלמים (פיתרון יחיד).

3) <u>זיהוי יתר:</u> יש יותר מדרך אחת לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת. זה קורה כשיש יותר משוואות מנעלמים (יותר מפתרון אחד).

בכדי להקל על בעיית הזיהוי מומלץ לאמץ את הכלל הבא:

עבור כל אחת מהמשוואות המבניות יש לחשב :

1) $g-1$: מס' אנדוגניים במשוואה הספציפית פחות 1

ולהשוות עם:

2) $K-k$: מספר אקסוגניים סה"כ בשתי המשוואות כולל חותך (K) פחות מספר אקסוגניים במשוואה הספציפית כולל חותך (k).

אם 1= 2 זיהוי מדויק ; 1 > 2 זיהוי יתר; 1 >2 אין זיהוי

בדוגמא שלנו:

משוואה מבנית 1: $\begin{matrix} g-1=2-1=1 \\ K-k=3-3=0 \end{matrix}$ $\Leftarrow$ **המשוואה לא מזוהה**

משוואה מבנית 2: $\begin{matrix} g-1=2-1=1 \\ K-k=3-2=1 \end{matrix}$ $\Leftarrow$ **המשוואה מזוהה בדיוק**

שיטות לפיתרון משוואות סימולטניות

קיימות 3 שיטות לפיתרון משוואות סימולטניות: ILS, IV ו- TSLS.

1) <u>שיטת ריבועים פחותים עקיפה (ILS)</u>

א. יש להציג את מערכת משוואות המבנה בצורתה המצומצמת.

ב. יש לאמוד בשיטת OLS את הפרמטרים של המשוואות בצורה המצומצמת (ה- λ ות וה- μ ים).

ג. יש לחלץ מן הפרמטרים של המערכת המצומצמת את הפרמטרים של הצורה המבנית.

משום שתהליך החילוץ איננו ליניארי האומדים המבניים המתקבלים הם **מוטים אך עקיבים**.

כאשר הזיהוי מדויק האומדים יהיו גם אסימפטוטית **יעילים** (במדגמים גדולים).

כאשר הזיהוי הוא יתר: האומדים **לא יהיו יעילים**.

בדוגמא שלנו:

משוואה מס' 1 לא מזוהה ולכן איננה ניתנת לאמידה.

משוואה מס' 2 מזוהה בדיוק ולכן יתקבל בשיטת ILS אומד אחד בלבד שיהיה מוטה אך עקיב וגם יעיל במדגמים גדולים.

2) שיטת ריבועים פחותים בשני שלבים (2SL2)

לשיטה זו שני שלבים:

א. אמידת משוואות הצורה המצומצמת בשיטת OLS ושימוש בתוצאות האמידה כדי לחשב את המשתנים האנדוגניים (המסבירים).

ב. הצבת המשתנים האנדוגניים שהתקבלו במשוואות המבנה ואמדתן ב-OLS.

אם משוואות המבנה מזוהות בדיוק או ביתר- האומדים שיתקבלו יהיו אמנם מוטים אבל עקיבים ויעילים אסימפטוטית.

האומדים שיתקבלו יהיו זהים לאומדים שהתקבלו בשיטת הריבועים הפחותים העקיפה.

כאשר אין זיהוי: אין אקסוגניים ולכן אין משתנים מסבירים בצורה המצומצמת או שכל האקזוגניים בצורה המצומצמת כבר קיימים במשוואה המקורית ולכן החלפת x ב- $\hat{x}$ תיצור בעיה של מולטיקוליניאריות מלאה.

בהמשך לדוגמתנו:

שלב א:

$$\hat{Y}_t = \hat{\mu}_0 + \hat{\mu}_1 \cdot Z_{1t} + \hat{\mu}_2 \cdot Z_{2t\,t}$$

$$\hat{X}_t = \hat{\lambda}_0 + \hat{\lambda}_1 \cdot Z_{1t} + \hat{\lambda}_2 Z_{2t}$$

שלב ב:

(1) $$Y_t = \alpha_0 + \alpha_1 \cdot \hat{X}_t + \alpha_2 \cdot Z_{1t} + \alpha_3 \cdot Z_{2t} + \varepsilon_t$$

(2) $$X_t = \beta_0 + \beta_1 \cdot \hat{Y}_t + \beta_2 \cdot Z_{1t} + \omega_t$$

מכיוון שמשוואה (1) איננה מזוהה, החלפת X ב- $\hat{x}$ תיצור בעיה של מולטיקוליניאריות מלאה (שכן משתנה זה מכיל רק משתנים אקזוגניים שכבר קיימים במשוואה).

מכיוון שמשוואה (2) מזוהה בדיוק, האומד $\hat{y}$ יהיה זהה לאומד שהתקבל בשיטת הריבועים הפחותים העקיפה.

3) <u>שיטת משתני העזר (IV)</u>

משתנה עזר הוא משתנה <u>שיחליף את המשתנה המסביר האנדוגני </u>במשוואת המבנה ויעזור לאמוד את הקשר בינו לבין התלוי.

<u>משתנה העזר צריך להיות</u>:

1. **משתנה אקזוגני או פונקציה ליניארית של משתנים אקזוגניים:**

$$\operatorname{cov}(Z,u) = 0$$

2. **מתואם עם המשתנה האנדוגני אותו הוא מחליף:** $\operatorname{cov}(Z,X) \neq 0$.

ככל שהמתאם גבוה יותר, האומד שיתקבל באמצעותו יהיה טוב יותר.

לדוגמא:

(2)

כאשר: $\operatorname{cov}(Y_t, \omega_t) \neq 0$ **ואילו-** $\operatorname{cov}(Z_{1t}, \omega_t) = 0$

<u>הבעיה:</u> אומדני OLS שיתקבלו יהיו מוטים, לא עקיבים ולא יעילים.

<u>הפיתרון בשיטת IV:</u> אמידת ההשפעה של Y על X עם משתנה אקסוגני שלא קיים במערכת ($Z_{2א}$, למשל) שמתואם עם Y (אותו הוא מחליף) אך לא עם u:

(2) $$X_t = \beta_0 + \beta_1 \cdot Z_{2t} + \beta_2 \cdot Z_{1t} + \omega_t$$

$$\operatorname{cov}(Z_{2t}, Y_t) \neq 0$$
$$\operatorname{cov}(Z_{2t}, \omega_t) = 0$$

<u>תכונות האומדים שיתקבלו באמצעות משתני העזר:</u>

אם יש יותר ממשתנה עזר אחד המקיימים את התנאים הנ"ל, האומדים שיתקבלו יהיו כולם מוטים אך עקיבים (ניתן להשתמש בהם במדגמים גדולים).

משתנה העזר <u>היחיד</u> שיניב אומד יעיל יהיה בעל המתאם הגבוה ביותר עם המשתנה האנדוגני אותו הוא בא להחליף. משתנה עזר זה יהיה אומדן לאנדוגני שהתקבל מאמידת משוואת הצורה המצומצמת בשלב הראשון של 2SLS.

למעשה, שיטת האמידה בשני שלבים- 2SLS היא מקרה פרטי של שיטת משתנה העזר: האומדן לאנדוגני שמתקבל בשלב הראשון מהווה את משתנה העזר היחיד שיניב אומדים יעילים (מבין אפשרויות אחרות של משתני עזר שיניבו אומדים עקיבים בלבד).

במקרה שלנו יהיה זה : $\hat{Y}_t = \hat{\mu}_0 + \hat{\mu}_1 \cdot Z_{1t} + \hat{\mu}_2 \cdot Z_{2t}$ **משתנה זה יחליף את Y במשוואה מבנית מס' (2).**

$\hat{Y}_t$ **מהווה משתנה עזר העומד בכל התנאים הדרושים:**

אינו נמצא במשוואת המבנה ($\hat{Y}_t \neq Y_t$).

הוא פונקציה של משתנים אקזוגניים בלבד ולכן לא מתואם עם הטעויות במשוואת המבנה ($\text{cov}(\hat{Y}_t, u) = 0$). כלומר, הוא מהווה משתנה אקזוגני במשוואת המבנה בניגוד למשתנה אותו הוא מחליף- Y_t.

מקיים קשר עם Y_t ($\text{cov}(\hat{Y}_t, Y_t) \neq 0$). למעשה, זהו האומד שהקשר שלו עם Y_t יהיה החזק ביותר מבין משתני העזר האפשריים ולכן היחיד שיניב אומדים יעילים לפרמטרים.

<u>משתנה לא יוכל לשמש כמשתנה עזר:</u>

אם נוסחתו מכילה רק משתנים אקזוגניים המצויים במשוואת המבנה בה הוא משמש כמשתנה עזר, שכן אז תיווצר בעיית מולטיקוליניאריות מלאה.

בדוגמא שלנו: Z_{1t} לא יוכל לשמש כמשתנה עזר המחליף את Y במשוואה (2) כי הוא קיים כבר במשוואה.

במילים אחרות, נוסחת משתנה העזר צריכה להיות מורכבת <u>מלפחות משתנה אקזוגני אחד שלא מופיע במשוואה</u> כדי שהמשתנה יוכל לשמש כמשתנה עזר.

<u>משתני עזר שונים יכולים להניב את אותם האומדים לפרמטרים:</u>

נבדוק זאת בצורה הבאה: נמחק מהנוסחאות של משתני העזר את המשתנים האקסוגניים המופיעים במשוואה. אם נשארנו עם שני ביטויים שהם מכפלה אחד של השני, יתקבלו אותם האומדים.

למשל:

עבור משוואה (2): **משתני העזר** Z_{2t} **ו-** $Z_{1t}+3Z_{2t}$ **יניבו את אותם האומדים שכן אם נמחק ממשתנה העזר השני את** Z_{1t} **שכבר נמצא במשוואה (2) נישאר עם** $3Z_{2t}$ **שהוא מכפלה של** Z_{2t}.

נוכיח באמצעות המשוואות הנורמאליות של משוואה (2):

$$X_t=\beta_0+\beta_1\cdot Y_t+\beta_2\cdot Z_{1t}+\omega_t$$

$$\sum\hat{\omega}_t=0$$
$$\sum\hat{\omega}_t Y_t=0$$
$$\sum\hat{\omega}_t Z_{1t}=0$$

נציב את משתנה העזר : Z_{2t} **ואת משתנה העזר :** $Z_{1t}+3Z_{2t}$ **במשוואה הנורמאלית במקום Y האנדוגני:**

$$\sum\hat{\omega}\cdot Z_{2t}$$
$$\sum\hat{\omega}(Z_{1t}+3Z_{2t})=\sum\hat{\omega}Z_{1t}+\sum\hat{\omega}3Z_{2t}=3\sum\hat{\omega}Z_{2t}$$

ניתן לראות כי משתנה העזר השני הוא מכפלה של משתנה העזר הראשון ולכן יניב את אותם האומדים לפרמטר.

? נתונות המשוואות הבאות :

(1)

(2)

נתון כי: משתנים אנדוגניים ו- משתנים אקסוגניים.

חוו דעתכם על כל אחת מהטענות הבאות, והסבירו:

א. ניתן להשתמש ב- כמשתנה עזר לאמידת משוואה מס' (1).

ב. ניתן להשתמש ב- כמשתנה עזר לאמידת משוואה מס' (2).

ג. יתכנו מספר אומדים עקיבים שונים זה מזה ל- במשוואה מס' (2).

ד. שימוש ב- כמשתנה עזר לאמידת משוואה מס' (2) יניב אומדים עקיבים וגם יעילים.

ה. משתנה העזר $Z_{1t} + 3Z_{2t}$ יניב אומדים זהים לאלו שהתקבלו בסעיף ב'.
הוכיחו את תשובתכם באמצעות המשוואות הנורמאליות של משוואה (2).

ו. משתנה העזר $3Z_{1t} + 5Z_{2t}$ יניב אותם אומדים כמו משתנה העזר בסעיף ד'.
הוכיחו את תשובתכם באמצעות המשוואות הנורמאליות של משוואה (2).

סיכום תוצאות אמידה של משוואות סימולטניות

מס' האומדים שיתקבלו בשלושת השיטות ותכונותיהם תלויים בזיהוי של המשוואה:

אם המשוואה לא מזוהה: לא ניתן להשתמש באף אחת מהשיטות.

כאשר המשוואה מזוהה (בדיוק או ביתר):האומדים שיתקבלו בשלושת השיטות יהיו תמיד מוטים אך עקיבים.

תכונת היעילות ומס' האומדים האפשרי מסוכמים בטבלה הבאה:

	מזוהה בדיוק	מזוהה ביתר
שיטת ILS	אומד אחד לפרמטר יעיל	יתכן יותר מאומד אחד לפרמטר לא יעילים
שיטת 2SLS	אומדן אחד למשתנה האנדוגני יעיל	
שיטת IV	אינסוף משתני עזר אם משתנה העזר זהה לאומדן לאנדוגני המתקבל בשלב הראשון בשיטת- 2SLS הוא יהיה גם יעיל.	

כאשר הזיהוי מדויק יתקבל אותו אומד מוטה אך עקיב ויעיל בשלושת השיטות: ILS, 2SLS ו- IV (במידה ומשתנה העזר הוא $\hat{X}_t$ מהשלב הראשון של 2SLS).

משתנים בפיגור ומשוואות סימולטניות

אם X_t אקסוגני אז גם המשתנים בפיגור X_{t-p} בוודאות אקסוגניים.

אם Y_t אנדוגני אז מעמדם של המשתנים בפיגור תלוי בקיומו של מתאם סדרתי:

אם יש מתאם סדרתי ($\text{cov}(Y_{t-1},u_t)\neq 0$) אז Y_{t-1} אנדוגני.

אם אין מתאם סדרתי ($\text{cov}(Y_{t-1},u_t)=0$) אז Y_{t-1} אקסוגני.

למשל:

אם נוסיף למשוואה (2) את Y_{t-1} אז מכיוון ש Y_t הוא אנדוגני -מעמדו של Y_{t-1} כאנדוגני או אקסוגני יקבע על פי קיומו של מתאם סידרתי בנתונים (אם מצוין שכל ההנחות הקלאסיות מתקיימות, הווה אומר שגם הנחה מס' 6 מתקיימת ומשתנה Y_{t-1} יהיה אקסוגני).

אם נוסיף למשוואה (2) את Z_{1t-1} אז מכיוון ש: Z_{1t} הוא אקסוגני אז גם Z_{1t-1} יהיה אקסוגני.

תרגיל מסכם

נתונות המשוואות הבאות:

$$Y_t=\alpha_0+\alpha_1X_t+\alpha_2Z_{1t}+\alpha_3Z_{2t}+\alpha_4Z_{3t}+\alpha_5Z_{4t}+u_t \quad \textbf{(1)}$$

$$X_t=\beta_0+\beta_1Y_t+\beta_2Z_{1t}+\beta_3Z_{2t}+\beta_4Z_{5t}+v_t \quad \textbf{(2)}$$

נתון כי: $\text{cov}(Z_j,u_t)=0$ עבור $j=1,...,5$ (כלומר ה-Z-ים אקסוגניים).

א. אמידת כל אחת מהמשוואות תניב אומדים:

1. מוטים נכון/לא נכון/ לא ניתן לדעת

2. עקיבים נכון/ לא נכון/ לא ניתן לדעת

ב. משוואה (1) מזוהה בדיוק/מזוהה ביתר/ בלתי מזוהה

משוואה (2) מזוהה בדיוק/ מזוהה ביתר/ בלתי מזוהה

ג. חווה דעתך על הטענות הבאות:

1. **תוך שימוש בשיטת ILS ניתן לאמוד את משוואה (1) באופן עקיב וחד ערכי:**

נכון/לא נכון/ לא ניתן לדעת

2. **תוך שימוש בשיטת ILS ניתן לאמוד את משוואה (2) באופן עקיב וחד ערכי:**

נכון/לא נכון/ לא ניתן לדעת

ד. משוואות הצורה המצומצמת הן:

$$Y_t = \lambda_0 + \lambda_1 Z_{1t} + \lambda_2 Z_{2t} + \lambda_3 Z_{3t} + \lambda_4 Z_{4t} + \lambda_5 Z_{5t} + \varepsilon_{1t}$$
$$X_t = \mu_0 + \mu_{1t} + \mu_2 Z_{2t} + \mu_3 Z_{3t} + \mu_4 Z_{4t} + \mu_5 Z_{5t} + \varepsilon_{2t}$$

נכון/ לא נכון/ לא ניתן לדעת

ה. אמידת משוואות הצורה המצומצמת ב-OLS תניב אומדים חסרי הטיה, עקיבים ויעילים: **נכון/לא נכון/ אי אפשר לדעת**

ו. להלן רשימה של משתני עזר פוטנציאליים:

1. Z_5
2. $\frac{Z_1 + Z_5}{2}$
3. $2Z_1 + 3Z_2 + Z_3$
4. $Z_3 + Z_4$
5. $3Z_3 + 4Z_4$
6. $3Z_3 + 3Z_4$
7. Z_1

עבור כל משתנה רשום באיזה משוואה ניתן להשתמש בו אם בכלל.

ז. איזה מבין משתני העזר הבאים יניבו את אותם האומדים עבור אותה המשוואה (תתכן יותר מתשובה אחת נכונה):

1. 1 ו-2

2. 4 ו-6

3. 5 ו-6

4. 4 ו-5

ח. האם משתנה עזר 1 (Z_5) יניב אומדים יעילים?

ט. אם ידוע כי אין מתאם סדרתי, האם Y_{t-1}, X_{t-1} הם אנדוגניים או אקסוגניים?

י. האם הוספה של משתנה אקזוגני נוסף למשוואה 1 תשנה את הזיהוי של משוואה 2?

יא. האם הוספה של משתנה אקסוגני נוסף למשוואה 2 תשנה את הזיהוי של משוואה 1?

יב. הנח כי הוטלו המגבלות הבאות על הפרמטרים המבניים:

$\alpha_2 = \beta_2 = 0$. האם ניתן כעת לזהות את יתר הפרמטרים במודל?

שאלה מס' 1

כדי לבדוק האם יש קשר בין שכר המורים וההוצאה לתלמיד בבי"ס ציבוריים נאמד המודל הבא: $Pay_i = \beta_1 + \beta_2 Spend + u_i$

כאשר pay הוא שכר המורים ו-Spend מייצג את ההוצאה לתלמיד שנמדדו בדולרים לשנה:

```
. reg pay  spending

      Source |       SS       df       MS              Number of obs =      51
-------------+------------------------------           F(  1,    49) =  112.60
       Model |  608555015     1   608555015           Prob > F      =  0.0000
    Residual |  264825250    49  5404596.94           R-squared     =  0.6968
-------------+------------------------------           Adj R-squared =  0.6906
       Total |  873380265    50  17467605.3           Root MSE      =  2324.8

------------------------------------------------------------------------------
         pay |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    spending |   3.307585   .3117043    10.61   0.000     2.681192    3.933978
       _cons |   12129.37   1197.351    10.13   0.000     9723.204    14535.54
------------------------------------------------------------------------------
```

א. פרשו את הרגרסיה (רמז: משמעות האומדים לפרמטרים, האם התוצאות מובהקות?). האם יש הגיון כלכלי לתוצאות שהתקבלו?

ב. מצאו תחזית נקודתית לשכר בממוצע אם ההוצאה לתלמיד היא 5000$

ג. נבדקו מדדים תיאוריים לגבי המשתנה הב"ת:

```
. su  spending

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
    spending |        51    3696.608    1054.761       2297       8349
```

ד. הסבירו את משמעות ההנחות הקלאסיות במונחי המשתנים.

ה. האם ההוצאה הממוצעת לתלמיד מסבירה הרבה משונות שכר המורים?

ו. מהו הפער החזוי בשכר המורים בין שני בתי ספר שהפער בהוצאה לתלמיד ביניהם הוא 20 $?

ז. האם ניתן לומר על סמך התוצאות כי המודל מובהק? אם כן, באיזה רמת מובהקות?

ח. חוקר ביצע מבחן שמטרתו לבדוק האם המודל יוצא מראשית הצירים.

1. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן:
2. ערך t סטטיסטי לבדיקת המבחן:
3. מסקנת המבחן: דוחים/לא דוחים את H0 ברמת סמך של 95%.

ט. חוקר ביצע מבחן שמטרתו לבדוק כי עבור עליה בדולר אחד בהוצאה השנתית הממוצעת לתלמיד שכר המורים הממוצע עולה בלפחות 3 דולר.

1. השערת האפס וההשערה האלטרנטיבית לבדיקת המובחן :
2. ערך t סטטיסטי לבדיקת המבחן:
3. ערך t קריטי לבדיקת המבחן:
4. מסקנת המבחן: לא דוחים את H0 ברמת ביטחון של 95%.

י. חשבו רווח בר סמך לאמידת השיפוע ברמת סמך של 90%.

יא. חוקר רצה לאמוד את המשוואה בשקלים במקום בדולרים ולכן אמד את המשוואה הבאה: $p\hat{a}y1 = \alpha_0 + \alpha_1 spend1$

הניחו כי שער הדולר עומד על 3.5 ₪.

1. מהי הפקודה ב-stata ליצירת המשתנים החדשים?
2. מה יהיה האומדן ל- α_1 ? מה יהיה האומדן לסטיית התקן של α_1 ?

על אותם הנתונים נאמדה גם המשוואה הבאה:

(3) $Pay_i = \beta_0 + \beta_1 Spend + \beta_2 Spend2 + u_i$

כאשר: $spend2 = spend^2$

יב. כתוב את הפקודה בSTATA ליצירת המשתנה החדש.

יג. מתוצאות האמידה התקבל ש:

$$p\hat{a}y_i = 23128.44 + 1.223 spend_i - 0.0000635 spend2$$

(9.55) (10.55) (8.78)

הערכים בסוגריים הם ערכי t סטטיסטי.

השערת האפס לבדיקה האם יש השפעה ליניארית של ההוצאה על התשלום למורים:

T סטטיסטי:

מסקנת הבדיקה: דוחים/לא דוחים את HO

יד. מהי הפקודה ב- STATA לבדיקת המתאם בין שני המשתנים הב"ת?

טו. בהינתן תוצאות המודל כיצד נראה פרופיל השכר כפונקציה של ההוצאה על התלמיד (אין צורך בציור מדויק רק במגמה).

טז. מהי ההוצאה לתלמיד שאחריה שכר המורים מתחיל לרדת?

יז. התקבל מתאם של : 0.97 בין שני המשתנים הב"ת. לנוכח המתאם הגבוה החוקר טען כי האומדים במשוואה (3) הם מוטים.

הטענה נכונה/הטענה אינה נכונה.

יח. לאור תוצאות האמידה של משוואה (3) ניתן להסיק כי האומדים של משוואה (1) הינם מוטים.

הטענה נכונה/הטענה איננה נכונה

יט. הניחו כי יש הטרוסקדסטיות במשוואה (3). מה יהיו ההשלכות לגבי האומדים של המשוואה:

1. האומדים יהיו חסרי הטיה ויעילים
2. האומדים יהיו חסרי הטיה אך לא יעילים
3. האומדים יהיו מוטים אך יעילים
4. האומדים יהיו מוטים ולא יעילים

שאלה מס' 2

נתונים 2 המודלים הבאים:

1. $Y_t = \alpha + \beta X_t + u_t$

2. $Y_t = \beta X_t + u_t$

א. **בהנחה כי** R^2 **של המודל הראשון גבוה מזה של המודל השני, האם ניתן לומר שהמודל הראשון טוב יותר?**

הניחו כי מודל (1) הוא המודל האמיתי.

ב. **מהן תכונות האומד לשיפוע של מודל (2)? מוטה/חסר הטיה?**

ג. **האם ניתן לומר כי האומד הינו עקיב?**

ד. **חוו דעתכם על הטענה:"על סמך משפט גאוס מרקוב, ניתן להסיק כי אומד הריבועים הפחותים של משוואה (1) הינו יעיל יותר מאומד הריבועים הפחותים של משוואה (2)". הסבירו את תשובתכם.**

ה. **חשבו את שונות האומד.**

ו. **לאיזה משני המודלים יהיה אומד לשיפוע גבוה יותר? לראשון/לשני/לא ניתן לדעת.**

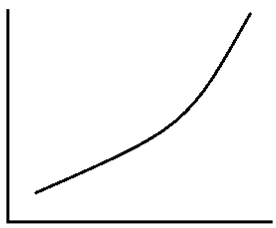
שאלה מס' 3

על מנת לאמוד את הקשר שבין השכלה(בשנים) להכנסה(באלפי שקלים) נאמדו שני המודלים הבאים:

```
. reg wage educ

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  253.20
       Model |  7888.51144     1  7888.51144           Prob > F      =  0.0000
    Residual |  31092.9858   998  31.1552964           R-squared     =  0.2024
-------------+------------------------------           Adj R-squared =  0.2016
       Total |  38981.4972   999  39.0205177           Root MSE      =  5.5817

------------------------------------------------------------------------------
        wage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   1.138517   .0715497    15.91   0.000      .998112    1.278922
       _cons |  -4.912181   .9667875    -5.08   0.000     -6.80935   -3.015011
------------------------------------------------------------------------------
```

```
. reg lwage educ

      Source |       SS       df       MS              Number of obs =    1000
-------------+------------------------------           F(  1,   998) =  272.72
       Model |  65.5213155     1  65.5213155           Prob > F      =  0.0000
    Residual |  239.767622   998  .240248118           R-squared     =  0.2146
-------------+------------------------------           Adj R-squared =  0.2138
       Total |  305.288937   999  .305594532           Root MSE      =  .49015

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        educ |   .1037608   .0062831    16.51   0.000     .0914313    .1160904
       _cons |   .7883743   .0848975     9.29   0.000     .6217761    .9549724
------------------------------------------------------------------------------
```

א. ידוע כי ניתן לתאר את הקשר בין שכר להשכלה על ידי הגרף הבא:



באיזה מודל כדאי לבחור? תנו 3 נימוקים לבחירתכם.

ב. מהי התשואה להשכלה על סמך המודל הנבחר?

ג. על פי מודל (1), מהי גמישות השכר ביחס להשכלה בנקודת הממוצעים: (7,12)?

ד. על בסיס נתוני המדגם חושב השכר הממוצע עבור התצפיות שלהן 16 שנות השכלה.

```
. su  wage if educ==16

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        wage |       186    13.30328    7.575015       2.54      60.19
```

הממוצע הוא 13.303

- **שימו לב: סימן ה- "=" משמש להשוואה בין משתנים (למשל gen price_new= price*1.2), לעומת זאת, סימן ה- "==" משמש להגדרת ערך משתנה (l educ if educ == 8).**

1. **מהי התחזית לשכר לעובד עם 16 שנות השכלה עבור כל אחד מהמודלים ?**
2. **איזה מהמודלים נותן תחזית נקודתית מדויקת יותר:**

המודל הראשון/המודל השני/ לא ניתן להשוות

ה. **איזה מודל מסביר חלק גדול יותר של השונות של המשתנה התלוי:**
המודל הראשון/המודל השני/לא ניתן להשוות

ו. **לפי האומדים בפלט (לא לפי מבחן סטטיסטי) התשואה להשכלה חיובית בהכרח:נכון/לא נכון**

ז. **רצו לבחון כיצד השכר מושפע גם מוותק העובד ולכן הכניסו את המשתנה: log(vetek). מה משמעות מקדם השיפוע של המשתנה החדש במודל הראשון ובמודל השני.**

ח. **בהתייחס למודל השני תוצאות האמידה מראות כי הן התשואה להשכלה והן גמישות הוותק אינם מובהקים. יחד עם זאת הרגרסיה עם שני המשתנים הב"ת יצאה מובהקת. כיצד ניתן להסביר זאת? האם מומלץ להשמיט את שני המשתנים ביחד מהמודל (לאמוד את השכר באמצעות משתנים אחרים)? האם מומלץ לאמוד את המודל ללא משתנה הוותק?**

שאלה מס' 4

על מנת לבחון את פונקצית הייצור של אורז נאמד המודל הבא:

$$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 \ln(LABOR) + \beta_4 \ln(FERT) + \varepsilon_i$$

כאשר:

PROD= כמויות אורז מדושן (נמדד בטונות)

AREA= גודל החלקות בהם האורז נשתל (נמדד בעשרות דונמים)

LABOR= סה"כ ימי עבודה של עובדים ובני משפחה (של החקלאי)

FERT= כמויות דשן בשימוש (נמדד בק"ג).

להלן תוצאות האמידה:

```
. reg lprod larea llabor lfert

      Source |       SS       df       MS              Number of obs =     352
-------------+------------------------------           F(  3,   348) =  646.51
       Model |  226.084875     3  75.361625           Prob > F      =  0.0000
    Residual |  40.5653554   348  .116567113           R-squared     =  0.8479
-------------+------------------------------           Adj R-squared =  0.8466
       Total |   266.65023   351  .759687266           Root MSE      =  .34142

------------------------------------------------------------------------------
       lprod |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       larea |   .3617359   .0639678     5.65   0.000     .2359237    .4875481
      llabor |   .4328479   .0668825     6.47   0.000      .301303    .5643928
       lfert |   .2095023   .0382654     5.47   0.000     .1342417    .2847628
       _cons |  -1.546786   .2556536    -6.05   0.000    -2.049607   -1.043966
------------------------------------------------------------------------------
```

א. בחנו את ההשערה כי גמישות הייצור ביחס לגודל החלקות (AREA) שווה לגמישות ביחס לימי העבודה (LABOR). השתמשו ברמת מובהקות של 5%, נסחו את ההשערה בצורה פורמאלית ודווחו את התוצאות תוך שימוש בפלט הבא:

```
. test  larea= llabor

 ( 1)  larea - llabor = 0

       F(  1,   348) =    0.34
            Prob > F =    0.5592
```

ב. כתבו את הפקודות ב-STATA לאמידת הרגרסיה המוגבלת מהסעיף הקודם.

ג. כעת, בחנו את ההשערה המורכבת מההשערה של סעיף א' + ההשערה כי פונקציית הייצור מקיימת תק"ל $(\beta_2+\beta_3+\beta_4=1)$, תוך שימוש בפלט הבא:

```
. gen l_pr_fe=log( prod/ fert)

. gen l_ar_la_fe=log( area* labor/fert^2)

. reg  l_pr_fe  l_ar_la_fe

      Source |       SS       df       MS              Number of obs =     352
-------------+------------------------------           F(  1,   350) =  439.63
       Model |  51.0075377     1  51.0075377           Prob > F      =  0.0000
    Residual |  40.6079092   350  .116022598           R-squared     =  0.5568
-------------+------------------------------           Adj R-squared =  0.5555
       Total |  91.6154469   351  .261012669           Root MSE      =  .34062

------------------------------------------------------------------------------
     l_pr_fe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  l_ar_la_fe |   .3940824    .018795    20.97   0.000     .3571172    .4310477
       _cons |  -1.402958   .0913195   -15.36   0.000    -1.582562   -1.223354
------------------------------------------------------------------------------

.
```

באיזה מבחן אתם משתמשים?

נסחו את ההשערה, את הסטטיסטי והקריטי.

מהי מסקנתכם?

שאלה מס' 5

חוקרים ביקשו לאמוד את פונקצית החיסכון המצרפי במשק הישראלי.

מפתח שמות משתנים:

GDS87 **– חסכון מקומי גולמי**

GDP87 **- תוצר מקומי גולמי**

GC87 **– הוצאות הממשלה**

כל הנתונים הינם במיליוני ש"ח במחירים קבועים של שנת 1987.

נאמדו 2 המודלים הבאים:

```
. reg   gds87   gdp87

      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  1,    24) =  106.04
       Model |  268713647     1   268713647           Prob > F      =  0.0000
    Residual |  60820138.1    24  2534172.42           R-squared     =  0.8154
-------------+------------------------------           Adj R-squared =  0.8077
       Total |  329533785    25  13181351.4           Root MSE      =  1591.9

------------------------------------------------------------------------------
       gds87 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gdp87 |   .1852456   .0179896    10.30   0.000     .1481169    .2223742
       _cons |  -2624.027    1018.32    -2.58   0.017    -4725.737   -522.3182
------------------------------------------------------------------------------

. reg   gds87 gdp87 gc87

      Source |       SS       df       MS              Number of obs =      26
-------------+------------------------------           F(  2,    23) =  157.31
       Model |  307085221     2   153542611           Prob > F      =  0.0000
    Residual |  22448563.7    23  976024.509           R-squared     =  0.9319
-------------+------------------------------           Adj R-squared =  0.9260
       Total |  329533785    25  13181351.4           Root MSE      =  987.94

------------------------------------------------------------------------------
       gds87 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       gdp87 |   .2955591   .0208369    14.18   0.000     .2524547    .3386635
        gc87 |  -.7784411   .1241513    -6.27   0.000    -1.035268   -.5216146
       _cons |   4217.674   1260.961     3.34   0.003     1609.177    6826.171
------------------------------------------------------------------------------
```

א. כתבו את המודל האקונומטרי שנאמד בכל אחת מהאמידות

ב. איזה מבין שני המודלים הקודמים הייתם מעדיפים? למה?

ג. מהו ההבדל בין המשמעות של האומדן למקדם המשתנה *GDP87* בשני המודלים?

ד. ביחס לאומד במשוואה 2 האומד ל- $\hat{\alpha}_2$ במשוואה 1 יהיה:

מוטה כלפי מעלה/מוטה כלפי מטה/ חסר הטיה/לא ניתן לדעת

ה. ביחס לאומד במשוואה 3 האומד ל- $\hat{\alpha}_2$ במשוואה 1 יהיה:

יעיל/לא יעיל/ לא ניתן לדעת

ו. **בהינתן תוצאות משוואה 1 הקורלציה בין שני המשתנים הב"ת היא: חיובית/שלילית/אפס/לא ניתן לדעת**

ז. **החוקר החליט להוסיף משתנה המודד את הצריכה הממשלתית במיליוני דולרים במקום בשקלים : $GC87 ואמד את המשוואה:**

$$\textbf{(3)}\ GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + \beta_4 \$GC87 + u_t$$

האומדים של משוואה 3 יהיו:

חסרי הטיה ויעילים/מוטים ולא יעילים/חסרי הטיה אך לא יעילים/לא מוגדרים.

ח. **החוקר החליט להוסיף מדד נוסף לתוצר המקומי. כתוצאה מהוספת המדד הנוסף משתנה התוצר המקומי הפך להיות לא מובהק. כיצד ניתן להסביר זאת:**

1. **מולטיקוליניאריות מלאה**
2. **מולטיקוליניאריות חלקית**
3. **הוספת משתנה לא רלוונטי**
4. **השמטת משתנה רלוונטי**

ט. **החוקר טען כי: "אם נוסיף משתנה נוסף לרגרסיה כלשהי אז האומדן ל σ^2 <u>לעולם לא יעלה</u>". נכון/לא נכון?**

י. **החוקר טוען כי "אם נוסיף משתנה נוסף לרגרסיה, אז האומדן ל- $\bar{R}^2$ יעלה בהכרח". נכון/לא נכון.**

שאלה מס' 6

מידע שנאסף לאחרונה על מכירות של 880 בתים בסטוקטון, קליפורניה, נמצא בקובץ *Stockton2*.

המשתנים הם

- **PRICE = מחיר בית בדולרים**
- **SQFT = גודל הבית (square feet)**
- **BEDS = מספר חדרי שינה.**
- **BATHS= מספר חדרי שירותים.**
- **AGE = גיל הבית.**
- **STORIES= הקומה של הבית.**
- **VACANT= משתנה דמי המקבל 1 אם הבית היה פנוי בזמן מכירתו ו-0 אחרת.**

להלן תוצאות אמידה המתבססת על המשתנים הנ"ל:

```
. gen price1000= price/1000

. gen lprice1000=ln( price1000)

. gen  sqft100= sqft/100

. reg  lprice1000 sqft100 age beds baths stories vacant

      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  6,   873) =  420.24
       Model |  92.5168833     6  15.4194805           Prob > F      =  0.0000
    Residual |  32.0322735   873  .03669218            R-squared     =  0.7428
-------------+------------------------------           Adj R-squared =  0.7410
       Total |  124.549157   879  .141694149           Root MSE      =  .19155

------------------------------------------------------------------------------
  lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sqft100 |   .0637441   .0020353    31.32   0.000     .0597495    .0677387
         age |  -.0024514    .000366    -6.70   0.000    -.0031698   -.0017331
        beds |  -.0848595   .0133702    -6.35   0.000    -.1111011    -.058618
       baths |   .0089069   .0181165     0.49   0.623     -.02665     .0444638
     stories |  -.0182728   .0219074    -0.83   0.404     -.06127     .0247245
      vacant |  -.0803092   .0132448    -6.06   0.000    -.1063045   -.0543138
       _cons |   3.994605    .037782   105.73   0.000     3.920451    4.068759
------------------------------------------------------------------------------
```

א. מהי משוואת הרגרסיה שנאמדה?

ב. פרשו את תוצאות האמידה. דונו בסימנים ובמשמעויות של כל אחד מהמשתנים.

ג. מה ההבדל במחיר הממוצע בין בית פנוי בזמן מכירתו לבין בית המאוכלס בזמן מכירתו?

ד. חוקר אחר הגדיר את משתנה VACANT= משתנה דמי המקבל 1 אם הבית היה מאוכלס בזמן מכירתו ו-0 אחרת. האם יש צורך לחשב מחדש את משוואת הרגרסיה? כן/לא.

ה. נניח כי רצו לתרגם את תוצאות המודל מ-sqrft למטרים מרובעים. יחידה אחת של sqrft שווה ל-0.093 מטר רבוע. מה יהיה הערך של β_1 במודל החדש? מה יהיה ערך t סטטיסטי לדחיית H0 של המשתנה החדש?

ו. מה הפער החזוי בין מחיר של בגודל 130 מטר בת 15 שנה לבין דירה בגודל 115 מטר בת 20 שנה, בהינתן שכל שאר המשתנים נותרים קבועים?

ז. מהי רמת הסמך הנמוכה ביותר עבורה ניתן לדחות את הטענה כי $\beta_5 = 0$?

ח. החוקר החליט לאמוד את המשוואה ללא משתנה VACANT כמשתנה מסביר . להלן תוצאות האמידה:

```
. reg  lprice1000 sqft100 age beds baths stories

      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  5,   874) =  477.40
       Model |  91.1678763     5  18.2335753           Prob > F      =  0.0000
    Residual |  33.3812805   874  .038193685           R-squared     =  0.7320
-------------+------------------------------           Adj R-squared =  0.7304
       Total |  124.549157   879  .141694149           Root MSE      =  .19543

------------------------------------------------------------------------------
  lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sqft100 |   .0654365   .0020569    31.81   0.000     .0613995    .0694734
         age |  -.0021886   .0003708    -5.90   0.000    -.0029163   -.0014608
        beds |  -.0823237   .0136344    -6.04   0.000    -.1090837   -.0555638
       baths |   .0062791   .0184781     0.34   0.734    -.0299876    .0425459
     stories |  -.0260982   .0223123    -1.17   0.242    -.0698901    .0176938
       _cons |   3.925237   .0367376   106.85   0.000     3.853132    3.997341
------------------------------------------------------------------------------
```

כיצד השמטת המשתנה השפיעה על משוואת הרגרסיה ועל מקדמיה?

ט. בנוסף אמדו על בסיס המדגם הנוכחי את שתי הרגרסיות הבאות:

```
. reg  lprice1000 sqft100 age beds baths stories if  vacant==0

      Source |       SS       df       MS              Number of obs =     415
-------------+------------------------------           F(  5,   409) =  233.63
       Model |  50.040644     5  10.0081288           Prob > F      =  0.0000
    Residual | 17.5202976   409  .042836913           R-squared     =  0.7407
-------------+------------------------------           Adj R-squared =  0.7375
       Total | 67.5609416   414  .16319068            Root MSE      =  .20697

------------------------------------------------------------------------------
  lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sqft100 |   .0684748   .0030057    22.78   0.000     .0625662    .0743833
         age |   -.001996   .0005387    -3.70   0.000     -.003055    -.000937
        beds |  -.0977677    .019958    -4.90   0.000    -.1370009   -.0585346
       baths |   .0193179   .0251737     0.77   0.443    -.030168     .0688038
     stories |  -.0654738    .033737    -1.94   0.053    -.1317933    .0008457
       _cons |   3.979725   .0554426    71.78   0.000     3.870737    4.088713
------------------------------------------------------------------------------

. reg  lprice1000 sqft100 age beds baths stories if  vacant==1

      Source |       SS       df       MS              Number of obs =     465
-------------+------------------------------           F(  5,   459) =  242.82
       Model | 37.2518359     5  7.45036717           Prob > F      =  0.0000
    Residual | 14.0830841   459  .030682101           R-squared     =  0.7257
-------------+------------------------------           Adj R-squared =  0.7227
       Total |   51.33492   464  .110635603           Root MSE      =  .17516

------------------------------------------------------------------------------
  lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     sqft100 |   .0593134   .0027696    21.42   0.000     .0538708    .0647561
         age |  -.0028851   .0004996    -5.78   0.000    -.0038668   -.0019034
        beds |  -.0678451   .0178018    -3.81   0.000    -.1028283   -.0328619
       baths |  -.0103401    .026461    -0.39   0.696    -.0623399    .0416596
     stories |   .0265265   .0284661     0.93   0.352    -.0294135    .0824665
       _cons |   3.924588   .0472086    83.13   0.000     3.831817     4.01736
------------------------------------------------------------------------------
```

השוו את תוצאות האמידה של שתי הרגרסיות הנ"ל

י. ערכו מבחן Chow לבדיקת שקילות (יציבות) המקדמים בשתי הרגרסיות מסעיף ז'.

שאלה מס' 7

חוקר מעוניין ללמוד על הקשר בין הכנסה של משפחה לבין מספר שנות הלימוד של הבעל, מספר שנות הלימוד של האישה והימצאות ילדים קטנים בבית. להלן מפתח שמות המשתנים:

FAMINC = **הכנסת המשפחה (דולרים בשנה)**

HEDU = **מספר שנות הלימוד של הבעל**

WEDU = **מספר שנות הלימוד של האישה**

$$KID6 = \begin{cases} 1 & \text{אם יש ילדים מתחת לגיל 6} \\ 0 & \text{אחרת} \end{cases},$$

הוא אמד שלוש רגרסיות וקיבל את התוצאות הבאות:

```
. */ Table 1

. reg  faminc Heduc Weduc k16

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  3,   424) =   30.43
       Model |  1.4725e+11     3  4.9082e+10           Prob > F      =  0.0000
    Residual |  6.8384e+11   424  1.6128e+09           R-squared     =  0.1772
-------------+------------------------------           Adj R-squared =  0.1714
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40160

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3211.526   796.7026     4.03   0.000     1645.547    4777.504
       Weduc |   4776.907   1061.164     4.50   0.000     2691.111    6862.704
         k16 |  -14310.92   xxxxxxxx    xxxxx   xxxx    xxxxxxxxxx   xxxxxxxxx
       _cons |  -7755.331   11162.93    -0.69   0.488    -29696.91    14186.25
------------------------------------------------------------------------------

. */ Table 2

. reg  faminc Heduc Weduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  2,   425) =   40.87
       Model |  1.3405e+11     2  6.7027e+10           Prob > F      =  0.0000
    Residual |  6.9703e+11   425  1.6401e+09           R-squared     =  0.1613
-------------+------------------------------           Adj R-squared =  0.1574
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40498

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3131.509    802.908     3.90   0.000     1553.344    4709.674
       Weduc |   4522.641   1066.327     4.24   0.000     2426.711    6618.572
       _cons |  -5533.631   11229.53    -0.49   0.622    -27605.97    16538.71
------------------------------------------------------------------------------

. */ Table 3

. reg  faminc Heduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  1,   426) =   61.30
       Model |  1.0455e+11     1  1.0455e+11           Prob > F      =  0.0000
    Residual |  7.2654e+11   426  1.7055e+09           R-squared     =  0.1258
-------------+------------------------------           Adj R-squared =  0.1237
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   41297

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   5155.484   658.4573     7.83   0.000     3861.254    6449.713
       _cons |   26191.27   8541.108     3.07   0.002     9403.308    42979.23
------------------------------------------------------------------------------
```

א. פרשו את משמעויות המקדמים ברגרסיה בטבלה הראשונה, התייחסו למובהקות המקדמים (אין צורך בבדיקה פורמאלית של מובהקות).

ב. התקבל כי ממוצע משתנה KID6 במדגם הוא 0.432. מה משמעות נתון זה?

ג. חשבו את סטית התקן הנאמדת של אומד הריבועים הפחותים ל β_{KL6} מהמודל הנאמד בטבלה הראשונה. הראו את החישובים שלכם ואת הנוסחאות עליהן אתם מתבססים.

רמז: השתמשו באינפורמציה שניתן לבדוק מובהקות המקדם הנ"ל ביותר מדרך אחת.

ד. הסבירו את הסיבה להבדל המשמעותי בין אומדני β_{HEDU} בשלוש הטבלאות? איך אפשר להסביר את העדר השוני (כמעט העדר שוני) באומדני β_{HEDU} ו β_{WEDU} בטבלה הראשונה והשנייה?

שאלה מס' 7

הורצו 2 רגרסיות על מדגם בן 400 תצפיות והתקבלו התוצאות הבאות :

$$1.\ \hat{CM}_i = 263.6416 - 0.0056\cdot PGNP_i - 2.2316\cdot FLR_i$$
$$(s.e)\quad (11.5932)\quad (0.0019)\qquad (0.2099)\qquad R^2 = 0.7077$$

$$2.\ \hat{CM}_i = 168.3067 - 0.0055\cdot PGNP_i - 1.768\cdot FLR_i + 12.8686\cdot TFR_i$$
$$(s.e)\quad (11.5932)\quad (0.0019)\qquad (0.2099)\qquad (?)\qquad R^2 = 0.7474$$

כאשר:

CM=Child Mortality מס' מקרי המוות של ילדים מתחת לגיל 5 לכל 1000 לידות חיים

PGNP=Per Capita GNP תוצר לנפש במחירים קבועים בדולרים

FLR=Female Literacy Rate אחוז נשים שיודעות לקרוא ולכתוב

TFR=Total Fertility Rate מס' הלידות הממוצע לאישה במדינה

הנתונים הם עבור 64 מדינות.

א. כיצד ניתן לפרש את המקדם למשתנה *TFR* ? אפריורית , האם תצפו לקשר חיובי/שלילי בין *CM* ל-*TFR* ? הסבירו.

ב. האם ערכי המקדמים של המשתנים *PGNP,FLR* מרגרסיה 1 שונים מאלו ברגרסיה 2 ? אם כן, מה יכולה להיות הסיבה/ות לשינוי זה ?

ג. חוו דעתכם על הטענה הבאה: "כיוון ש: FLR ו-TFR כל כך מתואמים אין לשים אותם באותה הרגרסיה".

ד. באיזה מודל תבחרו מבין השניים? באיזה מבחן סטטיסטי יש להשתמש כדי לענות על שאלה זו? הראו חישובכם. (רמז: הביעו את הסטטיסטי של מבחן F במונחי R^2)

ה. האם תוכלו לחשב את סטיית התקן הנאמדת של המקדם למשתנה *TFR* ? (רמז: היזכרו בקשר בין התפלגות *T* להתפלגות *F*)

ו. האם ניתן להשוות את מקדם ההסבר של שתי הרגרסיות? האם ניתן להשוות את מקדם ההסבר המתוקנן? אם כן, השוו ודווחו על התוצאות.

ז. ענו על השאלות התיאורטיות הבאות:

1) איזה מן הגורמים הבאים יכול לגרום לכך שאומדי OLS יהיו מוטים:

1. הטרוסקדסטיות
2. השמטת משתנה מסביר רלוונטי
3. מקדם מתאם גבוה מאוד בין שני משתנים מסבירים במודל

2) איזה מהגורמים הנ"ל יכול לגרום לכך שסטטיסטי t של OLS לא יהיה תקף?

3) התייחסו לטענה הבאה: "אם האומדים הינם עקיבים הם יהיו בהכרח גם חסרי הטיה". נכון/לא נכון

4) אם נתון ש u לא מתפלג נורמאלית אז אמידת המשוואה בשיטת OLS תניב אומדים שאינם עקיבים. נכון/לא נכון

שאלה מס' 8

חוקרת רצתה לבדוק עונתיות במחירי הירקות. לשם כך הגדירה את משתני הדמי הבאים:

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.**

D_4 **יקבל את הערך 1 אם מדובר בחורף ו-0 אחרת.**

כאשר:

$V_t=$ **מדד מחירי הירקות**

$p_t=$ **מדד המחירים לצרכן**

לשם כך אמדה את הרגרסיה הבאה על פני 30 שנה:

$$V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\alpha_4\cdot P_t+u_t$$

א. מדוע לא הכניסה החוקרת למשוואת הרגרסיה את משתנה D_4 ?

תוצאות האמידה שהתקבלו הן:

$$V_t=1379.11+99.18D_{1t}+2209.47D_{2t}-476.56D_{3t}+489.92\cdot P_t$$

$$R^2=0.0844$$

נאמדה בנוסף גם המשוואה הבאה:

$$V_t=\beta_0+\beta_1\cdot P_t+\varepsilon_t$$

תוצאות האמידה שהתקבלו:

$$V_t=11114.14+536.36\cdot P_t$$

$$R^2=0.0670$$

ב. על סמך התוצאות שהתקבלו, דרגו את העונות לפי רמת המחיר הבסיסית שלהן. הציגו לכל עונה את המיקום שלה ואת רמת המחיר הבסיסית כפי שבא לידי ביטוי במודל.

ג. בדקו את ההשערה כי עונתיות לא משפיע על מחיר הירקות

ד. כתבו את ההשערות הבאות:

1. מדד מחירי הירקות זהה בחורף ובאביב

2. מדד מחירי הירקות גבוה בקיץ מאשר בחורף ביותר מ-600

ה. האם יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן (בהנחה שהמחיר ההתחלתי של הירקות זהה בין עונות השנה)?

שאלה מספר 9

חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.
החוקר בדק האם <u>הפונקציה</u> של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את המשוואות הבאות:

$$\mathrm{NUM}_t = \gamma_3 + \delta_3 \cdot \mathrm{AVGD}_t + \varepsilon_{3t} \qquad \textbf{(1)}$$

$$\mathrm{NUM}_t = \alpha + \beta_1 \cdot \mathrm{TYPE}_t + \beta_2 \cdot \mathrm{AVGD}_t + \beta_3 \cdot (\mathrm{AVGD} \cdot \mathrm{TYPE})_t + \mathrm{U}_t \qquad \textbf{(2)}$$

כאשר : NUM_t = מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה

AVGD_t =נפח התנועה בקטע כביש t ליום באלפים

TYPE_t = משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר , ו- 0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מוצגות להלן:

$$NUM_t = 0.739 + 0.0233 \cdot AVGD_t \quad \textbf{(1)}$$

(2)

$$NUM_t = 0.14978 + 1.40311 \cdot TYPE_t + 0.002877 \cdot AVGD_t - 0.008 \cdot (AVGD \cdot TYPE)_t$$

$$ESS = 20963 \quad Pt_{\hat{\alpha}} = 0.0019; Pt_{\hat{\beta}} = 0.0001 \textbf{(1)}$$

$$Pt_{\hat{\alpha}} = 0.6534; Pt_{\hat{\beta}_1} = 0.0067; Pt_{\hat{\beta}_2} = 0.0001; Pt_{\hat{\beta}_3} = 0.1283 \qquad ESS = 20759 \quad \textbf{(2)}$$

1) בדקו את טענת החוקר.
2) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?
3) מהי השערת האפס לבדיקת הטענה?
4) הרגרסיה המוגבלת "תחת H_0" למבחן F (WALD) הינה:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$$

כאשר:

$Z_0 =$ ____________
$Z_1 =$ ____________
$Z_2 =$ ____________
$Z_4 =$ ____________

שאלה מס' 10

ברשותכם נתונים על מחירי ארוחת ביג מק במסעדות מקדונלדס ברחבי הארץ ב – 1/1/2008 וב- 1/1/2009. חלק מהחנויות מנוהלות ע"י הרשת והשאר מנוהלות ע"י זכיינים. אתם מעוניינים לבדוק את השפעתה של פסיקת בית משפט מ-7/2008 אשר מאפשרת לרשת לקבוע מחיר מקסימום עבור מוצרים הנמכרים בחנויות המנוהלות ע"י זכיינים.

לפניך הנתונים הבאים:

$Price=$**מחיר ארוחת ביג-מק**

$$D_{2009}=\begin{cases}0 & year=2008\\ 1 & year=2009\end{cases}$$

$$D_{Franchise}=\begin{cases}0 & \text{חנות רשת}\\ 1 & \text{חנות זכיין}\end{cases}$$

נתון המודל הבא המתאר את המחיר:

$$Price_i=\beta_1+\beta_2 D_{2009}+\beta_3 D_{Franchise}+\beta_4 D_{2009}*D_{Franchise}+\varepsilon_i$$

א. מה המקדם שמגלם את הפרש השכר הצפוי בין חנות רשת לחנות זכיין כשכל יתר המשתנים מוחזקים קבוע?

ב. נסחו את H_0 ו- H_1 לבדיקת ההשערות הבאות –

1. **מחיר ארוחת ביג מק ב-2008 היה זהה בחנויות המנוהלות ע"י הרשת ובחנויות המנוהלות ע"י זכיינים**
2. **פסיקת בית המשפט שינתה את פער המחירים בין חנויות המנוהלות ע"י הרשת לבין חנויות המנוהלות ע"י זכיינים בין השנים 2008 ו-2009.**

ג. הגדירו את משתנה האינטראקציה על ידי הגדרת 4 הקבוצות הבאות:

D1-רשת לפני 2009

D2-רשת אחרי 2009

D3-זכיין לפני 2009

D4-זכיין אחרי 2009

1. **מהי הפקודה הרלוונטית ב-STATA ליצירת D1?**
2. **כמה ממשתני הדמי יש להכניס לתוך הרגרסיה?**
3. **כתבו את משוואת הרגרסיה**

4. **נסחו שוב את ההשערות של סעיף א'**
5. **חוקר טוען כי התוצאות שהתקבלו בבדיקת ההשערות על ידי הגדרת משתני הדמי בשני האופנים יהיו זהות. נכון/לא נכון.**

שאלה מס' 11

היצע העבודה של נשים נשואות היה נושא מרכזי במחקר הכלכלי. לצרוך אמידת היצע זה נבחר המודל הבא:

$$HOURS_i = \beta_1 + \beta_2 WAGE_i + \beta_3 EDUC_i + \beta_4 AGE_i + \beta_5 KIDSL6 + \beta_6 KIDS618 + \beta_7 NWIFEINC_i + u_i$$

כאשר:

HOURS- היצע העבודה בשעות

WAGE- שכר לשעה

EDUC-מספר שנות הלימוד

AGE- גיל

KIDSL6-מספר הילדים בבית מתחת לגיל 6

KIDS618-מספר הילדים בגילאים 6-18

NWIFEINC-הכנסת משק הבית ממקורות שאינם מעבודתה של האישה

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

ב. הסבירו מדוע לא ניתן לאמוד את משוואת ההיצע הנ"ל בשיטת הריבועים הפחותים. איזה מתכונות אר"פ עלולות להיפגע?

ג. אם כתוצאה מאמידת המשוואה בשיטת OLS התקבל כי $\hat{\beta}_2 > 0$, האם ניתן להסיק כי $\text{cov}(wage_i, hours_i) > 0$ במדגם?

ד. החוקר טען שכדאי להשמיט את משתנה ה-wage בשל היותו מתואם עם הטעויות במודל. לטענתו, השמטתו תאפשר לאמוד את המשוואה ב- OLS ולקבל אומדים חסרי הטיה. חוו דעתכם על טענתו.

ה. הניחו כי אנחנו משתמשים בניסיון של האישה בשוק העבודה (EXPER) ובריבועו ($EXPER^2$) כמשתני עזר למשתנה WAGE. הסבירו מדוע משתני העזר הללו עונים על הדרישות שלנו ממשתני עזר וכיצד ניתן לבדוק זאת.

ו. האם ניתן לאמוד את משוואת ההיצע על ידי שימוש במשתני העזר הנ"ל? הסבירו.

ז. תארו את השלבים (לא בפקודות מחשב) שתבצעו כדי לקבל את האומדים בשיטת TSLS.

ח. כתבו את פקודת המחשב לביצוע TSLS ב-STATA:

פרק 11 - מבחן לדוגמא מס' 1

<u>שאלה 1 (55 נקודות)</u>

חוקר רצה לבדוק את השפעת התל"ג על ההשקעה במשק לפי המודל הבא:

$\ln I_t = \alpha + \beta \ln Y_t + u_t$, כאשר: I_t היא ההשקעה באלפי שקלים, Y_t הוא התוצר באלפי שקלים, וההפרעה האקראית, u_t, מקיימת את כל ההנחות הקלאסיות.

באמידה התקבל הפלט הבא:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	0.38523	0.38523	72.14	<.0001
Error	199	1.06266	0.00534		
C Total	200	1.44789			

Root MSE	0.073075	R-square	0.733936
Dep Mean	10.01722	Adj R-sq	0.732104
C.V.	0.729494		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|	95% conf. lim.
INTERCEPT	1	3.472013	0.85463	4.06259	0.0002	1.79 – 5.15
lnY	1	0.570042	0.06452	8.493526	0.0000	---- – ----

1) **מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?**
2) **אם נגדיל את התוצר ב-1% בכמה תגדל ההשקעה?**
3) **מהו רווח הסמך ל- α? מהו רווח הסמך ל- β?**
4) **הועלתה הטענה כי הגמישות שווה ל-0.4. מהן ההשערות לבדיקת הטענה?**
5) **מהי הרגרסיה המוגבלת למבחן WALD תחת H0?**
6) **מהו הסטטיסטי של WALD למבחן זה (אם ניתן לחישוב)?**
7) **אם ההשקעה נמדדת בשקלים במקום באלפי שקלים:**

א) המקדם של $\ln Y$ לא ישתנה. **נכון / לא נכון / לא ניתן לדעת**

ב) החותך לא ישתנה. נכון / לא נכון / לא ניתן לדעת

ג) הסטטיסטי t לבדיקת המובהקות של β לא ישתנה.

נכון / לא נכון / לא ניתן לדעת

ד) הסטטיסטי F לבדיקת מובהקות המודל לא ישתנה.

נכון / לא נכון / לא ניתן לדעת

ה) R^2 לא ישתנה. נכון / לא נכון / לא ניתן לדעת

החוקר טען כי גם גודל האוכלוסיה, P, משפיע על ההשקעה לפי המודל הבא:

$$\ln I_t = \alpha + \beta_1 \ln Y_t + \beta_2 \ln P_t + u_t$$

1) מהי השערת האפס לבדיקת הטענה?

התקבל הפלט הבא:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T
INTERCEPT	1	1.131853	1.43547	0.788489	0.4435
lnY	1	1.035467	0.25756	4.020294	0.0004
lnP	1	-1.77456	0.94657	-1.874727	0.0736

2) באיזו רמת מובהקות נקבל את טענת החוקר?

3) R^2 של המשוואה החדשה קטן מזה של המשוואה המקורית.

נכון / לא נכון / לא ניתן לדעת

במשוואה החדשה הועלתה הטענה כי סכום הגמישויות שווה ל-0.

4) מהי השערת האפס לבדיקת הטענה?

5) מהו הסטטיסטי t לבדיקת ההשערה? (נתון כי $\text{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -0.25$)

6) האם ניתן לדחות את השערת האפס?

<u>שאלה 2 (24 נקודות)</u>

1) ברגרסיה מרובה, כמו ברגרסיה חד משתנית, מבחן F למובהקות המודל שווה לריבוע של מבחן t למובהקות של β. **נכון / לא נכון / לא ניתן לדעת**

2) אם הערך 0 נמצא בתוך רווח הסמך ל- β, אזי β מובהקת.

נכון / לא נכון / לא ניתן לדעת

3) בהוספת משתנה לא רלוונטי למודל האומד המתוקן לפרופורצית השונות המוסברת ירד בהכרח.

נכון / לא נכון / לא ניתן לדעת

4) אומדי הריבועים הפחותים אינם חסרי הטיה אם ידוע שהשונות של u_t אינה קבועה (הפרה של הנחה קלאסית). **נכון / לא נכון / לא ניתן לדעת**

5) אם דוחים H0 ברמת מובהקות מסויימת, אזי דוחים H0 בכל רמות המובהקות הקטנות יותר. **נכון / לא נכון / לא ניתן לדעת**

6) אומד חסר הטיה הוא אינו בהכרח אומד עקיב

נכון / לא נכון / לא ניתן לדעת

<u>שאלה 3 (14 נקודות)</u>

נתון מודל ללא חותך $Y_t = \beta X_t + u_t$, ונתון האומד $\tilde{\beta} = \dfrac{\sum_{t=1}^{T} X_t Y_t}{S_{XX}}$

1) האומד $\tilde{\beta}$ הוא אר"פ. **נכון / לא נכון / לא ניתן לדעת**

2) האומד $\tilde{\beta}$ הוא אומד חסר הטיה. **נכון / לא נכון / לא ניתן לדעת**

3) האומד $\tilde{\beta}$ הוא אומד לינארי. **נכון / לא נכון / לא ניתן לדעת**

4) אר"פ יעיל יותר מ- $\tilde{\beta}$. **נכון / לא נכון / לא ניתן לדעת**

5) מהי השונות של $\tilde{\beta}$?

<u>שאלה 4 (7 נקודות)</u>

נתון מודל ללא חותך $Y_t = \beta X_t + u_t$ **, ונתון האומד** $\hat{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$

1) האומד $\hat{\beta}$ **הוא אר"פ.** **נכון / לא נכון / לא ניתן לדעת**

2) האומד $\hat{\beta}$ **הוא אומד חסר הטיה.** **נכון / לא נכון / לא ניתן לדעת**

3) האומד $\hat{\beta}$ **הוא אומד לינארי.** **נכון / לא נכון / לא ניתן לדעת**

4) מהי השונות של $\hat{\beta}$ **?**

5) האומד $\hat{\beta}$ **הוא אומד עקיב.** **נכון / לא נכון / לא ניתן לדעת**

פרק 12 - מבחן לדוגמא מס' 2

<u>שאלה 1 (60 נקודות)</u>

חוקר בדק את השפעת שעות העבודה בשבוע (HOURS) על השכר החודשי ברוטו בשקלים (SALARY) לפי המודל: $SALARY_t = \alpha + \beta \cdot HOURS_t + u_t$.

הסטיה המקרית מקיימת את כל ההנחות הקלאסיות.

השלם את הפלט הבא, אם ידוע כי $\bar{X} = 46.040873$, $S_{XX} = 35079$:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	---	---	---	---
Error	401	402271435	---		
C Total	---	449757359			

Root MSE	---	R-square	---
Dep Mean	1580	Adj R-sq	---
C.V.	---		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEPT	1	---	---	---	0.7476
HOURS	1	36.06745	---	---	0.0001

1) **מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?**
2) **מהו האומדן לשכר התחלתי?**

החוקר רצה לבדוק את הטענה כי אם יעבוד שעה אחת נוספת בשבוע, שכרו יגדל ב-40 ש"ח.

3) **מהן ההשערות לבדיקת הטענה?**
4) **מהו הסטטיסטי t למבחן?**
5) **מהו הסטטיסטי WALD למבחן?**

6) מהי התחזית לשכר של עובד העובד 55 שעות בשבוע?

7) החוקר טען כי יש לבדוק את הקשר בין השכר לשעות העבודה ע"י שימוש בנתונים שנתיים, כלומר, שכר שנתי (בהנחה שהשכר החודשי קבוע כל השנה) ושעות עבודה שנתיות (בהנחה ששעות העבודה קבועות בכל 52 השבועות בשנה).

שימוש בנתונים שנתיים:

א. ישנה את הסטטיסטי t לבדיקת המובהקות של α.

נכון / לא נכון / לא ניתן לדעת

ב. יכפיל את האומד של β ב-0.23. נכון / לא נכון / לא ניתן לדעת

ג. יכפיל את סטית התקן של $\hat{\beta}$ ב-0.23. נכון / לא נכון / לא ניתן לדעת

ד. ישנה את Pvalue לבדיקת מובהקות המודל. נכון / לא נכון / לא ניתן לדעת

החוקר טען כי יש להוסיף למשוואה גם את השפעת הגיל (AGE) ומספר שנות הלימוד (SCL). לשם כך הוא אמד את המשוואה הבאה:

$$SALARY_t = \alpha + \beta_1 \cdot HOURS_t + \beta_2 \cdot AGE_t + \beta_3 \cdot SCL_t + u_t$$

8) מהי השערת האפס לבדיקת הטענה?

9) מהו הנתון הנדרש כדי לחשב את הסטטיסטי של WALD לבדיקת טענת החוקר?

10) בפלט האמידה של המשוואה החדשה לא היה ברור אם ערכו של נתון זה הוא 315968434 או 515968434 (בשל בעיה במדפסת). מהו הסטטיסטי של WALD לבדיקת טענת החוקר?

11) מהם הנתונים הנדרשים לחישוב הסטטיסטי t?

החוקר רוצה לבדוק את הטענה כי השפעת ההשכלה על השכר גדולה פי 8 מהשפעת הגיל על השכר.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEPT	1	-1995.0275	331.7857	-6.013	0.0001
HOURS	1	36.408461	4.710021	7.730	0.0001
AGE	1	13.674254	3.816426	3.583	0.0004

SCL	1	109.93799	10.63745	10.335	0.0001

12) הנתונים בפלט אינם מספיקים לבדיקת ההשערה לפי מבחן t. מהו הנתון החסר? באיזה פלט של SAS ניתן למצוא אותו?

13) בהנחה שנתון זה הוא 8.3969, חשב את הסטטיסטי t לבדיקת הטענה. מהי מסקנתך לגבי נכונות הטענה?

14) אם תרצה לבדוק את הטענה לפי מבחן WALD, יהיה המודל המוגבל:

$Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + \gamma_2 \cdot Z_2 + v$ **כאשר:**

$Z_0 =$ ______________

$Z_1 =$ ______________

$Z_2 =$ ______________

15) אם יש מספיק נתונים, חשב את הסטטיסטי של WALD לבדיקת הטענה?

<u>שאלה 2 (30 נקודות)</u>

נתון המודל $Y_t = \alpha + \beta \cdot X_t + u_t$**. ידוע כי כל ההנחות הקלאסיות מתקיימות.**

נתון האומד: $\tilde{\beta} = \dfrac{S_{XY}}{\sum_{t=1}^{T} X_t^2}$.

1) אומד זה הוא הפתרון של המשוואות הנורמליות $\sum_{t=1}^{T} \hat{u}_t = 0$, $\sum_{t=1}^{T} \hat{u}_t X_t = 0$

נכון / לא נכון / לא ניתן לדעת

2) התוחלת של $\tilde{\beta}$ **היא:**

א. β

ב. $\dfrac{\beta \cdot \sum_{t=1}^{T} X_t}{S_{XX}}$

ג. $\dfrac{\beta \cdot \sum_{t=1}^{T} X_t^2}{S_{XX}}$

ד. $\dfrac{\beta \cdot S_{XX}}{\sum_{t=1}^{T} X_t^2}$

ה. כל התשובות אינן נכונות.

3) הטענה כי $E\left(\tilde{\beta}\right)<\beta$ **:**

א. תמיד נכונה

ב. אינה נכונה

ג. נכונה אם ורק אם $\bar{X}>0$

ד. נכונה אם ורק אם $\bar{X}\neq 0$

ה. כל התשובות אינן נכונות

4) אם $\bar{X}=0$ **אז השונות של** $\tilde{\beta}$ **היא:**

א. $\dfrac{\sigma^2}{\left(\sum_{t=1}^{T} X_t\right)^2}$

ב. $\dfrac{\sigma^2}{S_{XX}^2}$

ג. $\dfrac{\sigma^2\sum_{t=1}^{T} X_t^2}{S_{XX}}$

ד. $\dfrac{\sigma^2}{S_{XX}}$

ה. כל התשובות אינן נכונות

5) אם $\bar{X}=0$ **, אז** $\tilde{\beta}$ **הינו האומד הלינארי חסר ההטיה בעל השונות הקטנה ביותר.**

נכון / לא נכון / לא ניתן לדעת

<u>שאלה 3 (10 נקודות)</u>

נתון המודל $Y_t=\alpha+\beta\cdot X_t+u_t$**. ידוע כי כל ההנחות הקלאסיות מתקיימות.**

נתון כי $\tilde{\beta}$ **הוא אומד לינארי וחסר הטיה ל-** β **, אך איננו אומד עקיב ל-** β**.**

מאחר ש- $\tilde{\beta}$ **אינו אומד עקיב, לא נוכל להשתמש במשפט גאוס מרקוב ולקבוע כי** $\hat{\beta}=\dfrac{S_{XY}}{S_{XX}}$ **(אר"פ) הינו אומד יעיל יותר.**

נכון / לא נכון / לא ניתן לדעת

פרק 13 - מבחן לדוגמא מס' 3

<u>שאלה מס' 1 (50 נקודות)</u>

על מנת לאמוד את פונקצית הייצור נאספו נתונים על 150 פירמות בשנת 2007 ונאמדה המשוואה הבאה: (1) $\ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + U_t$ **(1)**

כאשר:

$\ln(Y)_t =$ **תפוקה שנתית באלפי ש"ח בלוגים**

$\ln(L)_t =$ **מספר העובדים בלוגים**

U_t=**הטעות המיקרית המקיימת את כל ההנחות הקלאסיות.**

משוואה מס' (1) נאמדה בפלט מס' 1

(3) א. סטטיסטי F לבדיקת מובהקות המודל:

1. **לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.**
2. **ניתן לחשבו וערכו הוא:__________________**

(3) ב. סטטיסטי t לבדיקת מובהקות המודל:

1. **לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה**
2. **לא ניתן לחשבו בעזרת הנתונים הקיימים.**
3. **ניתן לחשבו וערכו הוא:_________________**

הועלתה הטענה כי עליה ב-1% במס' העובדים תגדיל את התפוקה בפחות מ-1%

(3) ג. ההשערות לבדיקת הטענה

הן: ___________________________________:H0

___________________________________:H1

(4) ד. הסטטיסטי לבדיקת הטענה הינו:

1. **לא ניתן לחשבו בנתונים הקיימים.**
2. **5.5**
3. **-5.5**
4. **-15.5**
5. **15.5**

(3) ה. הסטטיסטי של WALD לבדיקת הטענה:

1. **לא ניתן לחשבו בעזרת הנתונים הקיימים.**
2. **ניתן לחשבו וערכו הוא:________________**

(3) ו. לאור התשובות לסעיפים הקודמים, אחוז התפוקה קטן ככל שאחוז מס' העובדים גדל: נכון/לא נכון/אי אפשר לדעת

החוקרת טענה כי יש משתנים נוספים המסבירים את תפוקת הפירמה ואמדה את המשוואה הבאה: $\ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + \beta_3 \cdot \ln(PY)_t + U_t$ **(2)**

כאשר:

$\ln(K)_t =$ **מלאי ההון של הפירמה באלפי ₪ בלוגים**

$\ln(PY)_t =$ **הוצאות למחקר ופיתוח באלפי ₪ בלוגים**

משוואה מס' (2) נאמדה בפלט מס' 2

(2) ז. ההשערות לבדיקת הטענה הינן :

H_0:____________________________

H_1:____________________________

(2) ח. הסטטיסטי של WALD לבדיקת הטענה הינו:

1. **לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים**
2. **ניתן לחישוב וערכו:___________________________________.**

(2) ט. הסטטיסטי של t לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

2. לא ניתן לחשב סטטיסטי t לטענה מסוג זה

3. ניתן לחישוב וערכו:__________________________________.

החוקרת טענה כי השפעת הוצאות למחקר ופיתוח אינה מובהקת ולכן יש לאמוד את המשוואה הבאה:

$$(3) \quad \ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + U_t$$

משוואה מס' (3) נאמדה בפלט מס' 3

(3) י. ההשערות לבדיקת הטענה הינן :

H_0:______________________________

H_1:______________________________

(2) יא. הסטטיסטי של WALD לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

2. ניתן לחישוב וערכו:__________________________________.

הועלתה הטענה כי גמישות התפוקה ביחס להון גדולה פי 2 מגמישות התפוקה ביחס לעבודה.

בדקי את הטענה במשוואה (3)

(4) יב. השערת האפס לבדיקת הטענה היא :

H_0:______________________________

(2) יג. הסטטיסטי t לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

2. ניתן לחישוב וערכו:________________________.

(4) יד. הרגרסיה המוגבלת כאשר H_0 נכונה ("תחת H_0") למבחן WALD

הינה : $Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + V$

כאשר: $Z_1 =$ ______________________ $Z_0 =$ ______________________

(4) טו. הסטטיסטי של WALD לבדיקת הטענה (חשבי ישירות) :

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

2. ניתן לחישוב וערכו: ________________________

(3) טז. נטען כי אם נמדוד את המשתנים הב"ת במודל בדולרים במקום בשקלים , האומדים ל-β ול-α יישארו ללא שינוי (הנח כי שער הדולר הוא 3.5 ₪):

נכון/לא נכון/אי אפשר לדעת

(3) יז. נטען שאם נוריד את משתנה PY מהמודל ה- $\bar{R}^2$ יעלה:

נכון/לא נכון/אי אפשר לדעת

<u>שאלה מס' 2 (33 נקודות- כל סעיף 3 נקודות)</u>

עני על כל השאלות הבאות. כל שאלה בפני עצמה. בכל השאלות מונח המודל: $Y = \alpha + \beta X + U$ **(ומתקיימות כל ההנחות הקלאסיות)**

1. **במודל לוגריתמי כפול β מייצגת את שיעור השינוי השולי:**

   נכון/לא נכון/ אי אפשר לדעת

2. **במודל ללא חותך מתקיימת המשוואה הנורמאלית:** $\sum \hat{u}_t x_t = 0$ **בלבד:**

   נכון/לא נכון/אי אפשר לדעת

3. **כאשר מוסיפים משתנה ב"ת למודל, עליה ב-** $\bar{R}^2$ **מעידה על כך שהמשתנה שהוסף מובהק באוכלוסיה:**

   נכון/לא נכון/אי אפשר לדעת

4. **אם הנחה מס' 3)** $E(\hat{u})=0$ **לכל t) איננה מתקיימת, האומדים של המודל לא יהיו יעילים:**
נכון/לא נכון/ אי אפשר לדעת

5. **ככל ש-** S_{XX} **גדול יותר, קל יותר לדחות את HO למובהקות ה-β:**
נכון/לא נכון/אי אפשר לדעת

6. $R^2 > \overline{R}^2$ **מתקיים תמיד:**
נכון/לא נכון/ אי אפשר לדעת

7. **מבחן F למובהקות המודל מהווה מקרה פרטי של מבחן t למובהקות ה-β :**
נכון/לא נכון/ אי אפשר לדעת

8. **ככל שגודל המדגם גדל כך האומד יהיה יעיל יותר לפרמטר באוכלוסיה:**
נכון/לא נכון אי אפשר לדעת

9. **ה-PVALUE גדל ביחס הפוך לרמת המובהקות של המבחן (ה-α):**
נכון/לא נכון/אי אפשר לדעת

10. **אם דחינו את HO במבחן t למובהקות ה-β כאשר האומד חיובי, נדחה אותה בהכרח גם ביחס להשערה כי מקדם השיפוע חיובי באוכלוסיה:**
נכון/לא נכון/ אי אפשר לדעת

11. **אם ידוע כי הקשר בין X ל-Y מובהק באוכלוסיה, הדבר מעיד בהכרח על מובהקות המודל:** **נכון/לא נכון/ אי אפשר לדעת**

<u>שאלה מס' 3 (14 נק'):</u>

נתון המודל: $Y_t = \beta X_t + U_t$

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

(5)א. $\tilde{\beta}$ הינו אומד חסר הטייה ל-β: **נכון/לא נכון/אי אפשר לדעת**

(3)ב. שונותו של האומד: ______________________________

(2)ג. על סמך משפט גאוס מרקוב ניתן להסיק כי אר"פ הינו אומד יעיל יותר מ- $\tilde{\beta}$: **נכון/לא נכון/אי אפשר לדעת**

(2)ד. המשוואות הנורמאליות : $\sum \hat{u}_t = 0$ ו- $\sum \hat{u}_t x_t = 0$ הינן המשוואות לאמידת הפרמטרים של המודל בשיטת הריבועים הפחותים:
נכון/לא נכון/אי אפשר לדעת

(2)ה. אם נתון ש: $\bar{X} = 0$ אזי $\tilde{\beta}$ הינו אומד הריבועים הפחותים:
נכון/לא נכון/אי אפשר לדעת

פלט מס '1 - משוואה מס' (1)

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	1	8.54211			0.0001
Error	35969	40.42584			
C Total	35970	48.96795			

Root MSE	0.52264	R-square	0.1744
Dep Mean	5.54003	Adj R-sq	0.1689
C. V.	9.43380		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	4.389949	0.21003743	20.901	0.0001
lnL	1	0.257487	0.04767276		0.0001

פלט מס '2 – משוואה מס' (2)

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	3	15.63370	5.21123	22.825	0.0001
Error	146	33.33425	0.22832		
C Total	149	48.96795			

Root MSE	0.47783	R-square	0.3193
Dep Mean	5.54003	Adj R-sq	0.3053
C. V.	8.62496		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	0.542062	1.66317350	0.326	0.7450
lnL	1	0.267771	0.08146608	3.287	0.0013
lnK	1	0.405694	0.09700769	4.182	0.0001
lnPY	1	0.406149	0.30781185	1.319	0.1891

פלט מס '3 – משוואה מס' (3)

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	15.23620	7.61810	33.199	0.0001
Error	147	33.73175	0.22947		
C Total	149	48.96795			

Root MSE	0.47903	R-square	0.3111
Dep Mean	5.54003	Adj R-sq	0.3018
C. V.	8.64667		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	2.681787	0.37024512	7.243	0.0001
lnL	1	0.177813	0.04470595	3.977	0.0001
lnK	1	0.465154	0.08612163	5.401	0.0001

Covariance of Estimates

COVB	INTERCEP	lnL	lnK
INTERCEP	0.1370814505	-0.003289697	-0.02723683
lnL	-0.003289697	0.0019986217	-0.001270417
lnK	-0.02723683	-0.001270417	0.0074169359

פרק 14 - מבחן לדוגמא מס' 4

<u>שאלה מס' 1 (55 נקודות)</u>

בנק מעוניין לאמוד את סך הפעילות בכרטיסי אשראי של לקוחותיו. לשם כך אסף נתונים על 35,971 מלקוחותיו ואמד את המשוואה הבאה:

$CREDIT_t = \alpha + \beta \cdot SAVINGS_t + U_t$ **(1)**

כאשר:

$CREDIT_t =$ **סך הפעילות בכרטיסי אשראי בש"ח**

$SAVINGS_t =$ **סך הפעילות בחשבונות חיסכון בש"ח**

U_t **-סטיה מקרית המקיימת את כל ההנחות הקלאסיות.**

משוואה (1) נתונה בפלט מס' 1.

(3) א. סטטיסטי F לבדיקת מובהקות המודל הינו:

1. לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: __________________________

(3) ב. PVALUE של סטטיסטי t לבדיקת מובהקות ה-β :

1. לא ניתן לחשבו בעזרת הנתונים הקיימים.

2. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.

3. ניתן לחשבו וערכו:___________________________

הבנק טען שאם יגדילו לקוחותיו את הפעילות בחשבונות חיסכון שלהם אפילו בשקל אחד, הפעילות בכרטיסי אשראי תגדל ביותר מ 40 אגורות.

(3) ג. ההשערות לבדיקת הטענה הינן:

H0:____________________________________

HI :____________________________________

(3) ד. הסטטיסטי לבדיקת טענת הבנק הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.
2. הסטטיסטי לבדיקת הטענה צריך להיות שלילי.
3. 19.67
4. 5.797

(3) ה. הסטטיסטי של WALD לבדיקת טענת הבנק :

1. לא ניתן לחשבו בעזרת הנתונים הקיימים.
2. ניתן לחשבו וערכו:______________________

(3) ו. ברמת ביטחון של 95% מהו טווח הגידול בפעילות בכרטיסי אשראי, על כל שקל נוסף בפעילות בחשבונות חיסכון?

---

(3)ז. ברמת ביטחון 95% מהו האומד לתוחלת פעילות בכרטיסי אשראי עבור סך פעילות בחשבונות חיסכון של 50,000 ₪?

(5)ח. אם פעילות כרטיסי האשראי של כל לקוח תגדל ב- 1000 ש"ח:

1. האומד של α ישתנה: נכון/לא נכון/ אי אפשר לדעת

2. האומד של β ירד: נכון/לא נכון/ אי אפשר לדעת

3. סטטיסטי F לבדיקת מובהקות המודל לא ישתנה:

נכון/לא נכון/ אי אפשר לדעת

נטען שסה"כ פעילות הלקוח בחשבונות חיסכון איננו המשתנה המשפיע על הפעילות בכרטיסי האשראי, אלא הרכב החסכונות.

לשם כך נאמדה המשוואה הבאה:

$$CREDIT_t = \alpha + \beta_1 \cdot PIKADON1_t + \beta_2 \cdot PIKADON2_t + U_t \quad (2)$$

כאשר:

$PIKADON1_t$ = סה"כ הפקדה לפקדונות יומיים בש"ח.

$PIKADON2_t =$ **סה"כ הפקדה לפקדונות חודשיים בש"ח.**

משוואה (2) נאמדה בפלט מס' 2.

(3) ט. השערת האפס לבדיקת הטענה

הינה:________________________________HO:

(4) י. הסטטיסטי של WALD לבדיקת הטענה:

1. **לא ניתן לחשבו בעזרת הנתונים הקיימים.**
2. **ניתן לחשבו וערכו:____________________**

(3) יא. הסטטיסטי של t לבדיקת הטענה:

1. **לא ניתן לחשבו בעזרת הנתונים הקיימים.**
2. **לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.**
3. **ניתן לחישוב וערכו:______________________**

נטען שהגדלת הפעילות בחשבונות חיסכון של הלקוח על ידי העברה לפקדונות חודשיים משפיעה על הפעילות בכרטיסי אשראי פי 10 מאשר הגדלת הפעילות בחשבונות חיסכון על ידי העברה לפקדונות יומיים.

(3) יב. השערת האפס לבדיקת הטענה

הינה:___________________________HO:

(4) יג. הסטטיסטי t לבדיקת הטענה הינו:

1. **לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.**
2. **ניתן לחשבו וערכו הוא:________________**

(2) יד. PVALUE של סטטיסטי t מהסעיף הקודם:

1. **לא ניתן לחשב את הסטטיסטי t בעזרת הנתונים הקיימים.**
2. **ניתן לחשבו וערכו הוא:______________________**

(4) טו. הרגרסיה המוגבלת כאשר HO נכונה למבחן WALD הינה:

$$D_0 = \gamma_0 + \gamma_1 \cdot D_1 + \gamma_2 \cdot D_2 + \nu$$

כאשר:

D_0 : ____________

D_1 : ____________

D_2 : ____________

(4) טז. על פי משוואה מס' 2, כל שקל שיועבר לפיקדון הראשון יוסיף כ-0.07552
לסה"כ הפעילות בכרטיסי אשראי: נכון/לא נכון/ אי אפשר לדעת

שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)

ענה על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל $Y = \alpha + \beta \cdot X + U$ ומתקיימות כל ההנחות הקלאסיות).

1. **אם המודל מובהק אזי שיפוע הרגרסיה מובהק בהכרח:**
נכון/לא נכון/אי אפשר לדעת

2. **הגמישות במודל חצי לוגרתמי היא קבועה: נכון/לא נכון/אי אפשר לדעת**

3. **אם X_2 מהווה קומבינציה ליניארית של X_1 לא ניתן לאמוד את הרגרסיה המרובה:**
$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + U$: **נכון/לא נכון/ אי אפשר לדעת**

4. $\bar{R}^2 > R^2$ **רק בתנאי שהמודל מובהק: נכון/לא נכון/ אי אפשר לדעת**

5. **ליניאריות וחוסר הטיה של האומדים מהווים תנאי הכרחי לעקיבותם:**
נכון/לא נכון/ אי אפשר לדעת

6. **נתון כי רווח הסמך לאמידת β ברמת סמך של 95% הוא: [-2,-5]. מכך ניתן להסיק כי שיפוע הרגרסיה מובהק ברמת מובהקות של 5%:**

נכון/לא נכון/ אי אפשר לדעת

7. ככל שפיזור U_t גדול יותר כך קשה יותר לדחות את HO למובהקות המודל:

נכון/לא נכון/ אי אפשר לדעת

8. מודלים לא ליניאריים מתארים קשרים שאינם ליניאריים בין המשתנה המסביר למוסבר:

נכון/לא נכון/ אי אפשר לדעת

9. אם הנחה 5 (שונות קבועה) לא מתקיימת, אומדי הריבועים הפחותים אינם חסרי הטיה:

נכון/לא נכון/ אי אפשר לדעת

10. אם דחינו את HO לבדיקת הטענה כי שיפוע הרגרסיה הוא שלילי בוודאי שמודל הרגרסיה הוא מובהק:

נכון/לא נכון/ אי אפשר לדעת

שאלה 3 (12 נקודות)

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

נתון האומד: $\tilde{\beta} = \frac{\sum Y_t}{S_{XX}}$

(2)א. $E(\tilde{\beta}) =$ ____________________

(2)ב. על סמך משפט גאוס מרקוב אומד זה יעיל פחות מאומד הריבועים הפחותים:

נכון/ לא נכון/ אי אפשר לדעת

(2) ג. אומד $\tilde{\beta}$ מוגדר רק כאשר $S_X^2 \neq 0$:

נכון/ לא נכון/ אי אפשר לדעת

(3) ד. חשבו את השונות של $\tilde{\beta}$ עבור מודל שבו $\alpha \neq 0$

(3) ה. שונות האומד (שחושבה בסעיף הקודם) הינה גדולה משונות המודל הנתון:

נכון/לא נכון/ אי אפשר לדעת

פלט מס '1 - משוואה מס' (1)

Dependent Variable: credit

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	---	-----	------	------	<0.0001
Error	---	-----	------		
C Total	---	-----			

Root MSE	43859	R-square	0.0106
Dep Mean	7433.60809	Adj R-sq	0.0106
C. V.	589.99662		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0:	Prob>lTl	95% Confidence	
INTERCEP	1	11151.91516	394.35144	2.92	0.0035	378.97	1924.8
savings	1	0.56719	0.02884	19.67		0.51	0.623

פלט מס '2 – משוואה מס' (2)

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	1.00791E12	5.003955E11	261.10	0.0001
Error	35968	6.893195E13	1916479937		
C Total	35970	6.993274E13			

Root MSE	43778	R-square	0.0143
Dep Mean	7433.68809	Adj R-sq	0.0143
C. V.	588.90847		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	1259.36230	379.00751	3.32	0.0009
Pikadon1	1	0.07552	0.05539	1.36	0.1728
Pikadon2	1	0.72350	0.03199	22.62	0.0001

Covariance of Estimates

COVB	INTERCEP	Pikadon1	Pikadon2
INTERCEP	143646.69097	-8.178835194	-9.154578973
Pikadon1	-8.176835154	0.0030678685	0.0003564263
Pikadon2	-9.15457897	0.0003564263	0.0010231462

פרק 15 - מבחן לדוגמא מס' 5

<u>שאלה מס' 1 (55 נקודות)</u>

על מנת לאמוד את הקשר בין רמת המחירים במשק (P) לכמות הכסף (M), נאספו נתונים חודשיים בשנים 86-94 (סה"כ 105 תצפיות) ונאמדה המשוואה הבאה:

(1) $M_t = e^{\alpha} + p^{\beta} + e^{u}$

כאשר:

m= כמות הכסף במשק לחודש (מזומנים+עו"ש)

p= מדד המחירים לצרכן במשק

U_t**-סטיה מקרית המקיימת את כל ההנחות הקלאסיות.**

משוואה מס' (1) נאמדה בפלט מס' 1.

(2) א. כתבי את המשוואה בצורה ליניארית בעזרת הטרנספורמציה המתאימה.

(2)ב. האומדן למשוואה (1) הינו:____

(2) ג. המשמעות הכלכלית של β היא:______________

(4) ד. גבולות רווח-סמך ברמת סמך של 95% עבור β הינם:

גבול תחתון:____________________

גבול עליון:____________________

(3) ה. ערך t לחישוב מובהקות ה- β הינו:

1. לא ניתן לחשב ערך זה בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: ____________________

(2) ו. אם נגדיל את מדד המחירים לצרכן ביחידה אחת, כמות הכסף במשק תגדל ב:

1. 71.7233

2. 1.69267

3. 169.267

4. 1.69267%

5. אף תשובה איננה נכונה

הועלתה הטענה שתוספת של אחוז אחד במדד המחירים לצרכן תגדיל את כמות הכסף במשק ביותר מאחוז אחד.

(3) ז. ההשערות לבדיקת הטענה: ________________

(3)ח. סטטיסטי t לבדיקת הטענה הינו:

1. לא ניתן לחשבו באמצעות הנתונים הקיימים.
2. ניתן לחשבו וערכו הוא: ______________________________

(3) ט. על פי התשובות לסעיפים הקודמים ניתן להסיק כי ערכו של סטטיסטי F לבדיקת מובהקות המודל הינו:

1. לא ניתן לחשב את ערכו של סטטיסטי F על סמך סטטיסטי t
2. 861.4225
3. 5144.23
4. 71.7233

(3) י. אם נוציא שורש ריבועי למדד המחירים לצרכן במשק:

1. האומד של α ישתנה: נכון/לא נכון/ אי אפשר לדעת
2. האומד של β יעלה: נכון/לא נכון/ אי אפשר לדעת
3. סטטיסטי F לבדיקת מובהקות המודל לא ישתנה: נכון/לא נכון/ אי אפשר לדעת

הועלתה הטענה כי יש צורך להוסיף למשוואה גם את הפעילות הכלכלית במשק (Y) כמשתנה מסביר, ולכן יש לאמוד את המשוואה הבאה:

$$LN(M)_t = \alpha + \beta_1 \cdot LN(P)_t + \beta_2 \cdot LN(Y)_t + U_t \qquad (2)$$

משוואה (2) נתונה בפלט מס' 2.

(2) יא. סטטיסטי t לבדיקת הטענה הינו:

1. לא ניתן לחשבו בעזרת הנתונים הקיימים.
2. ניתן לחשבו וערכו
הוא:_______________________________

(2) יב. על פי התשובה לסעיף הקודם, ניתן להסיק את ערכו של סטטיסטי F למובהקות המודל . נכון/ לא נכון/ אי אפשר לדעת

(2) יג. על פי התשובה לסעיף יא' ניתן להסיק את ערכו של סטטיסטי WALD לבדיקת הטענה. נכון/לא נכון/ אי אפשר לדעת

הועלתה הטענה כי הגמישות ביחס למחיר גבוהה פי 10 מהגמישות ביחס לפעילות הכלכלית במשק.

(3) יד. סטטיסטי WALD לבדיקת הטענה הינו:
1. לא ניתן לחשבו בעזרת הנתונים הקימים.
2. ניתן לחשבו וערכו הוא:_________________________________

(3) טו. הרגרסיה המוגבלת כאשר HO נכונה למבחן WALD הינה:__________

כאשר: D_0 : ____________
D_1 : ____________

(4) טז. א. איזה מבין המודלים המוצעים במשוואות 1 ו-2 עדיף?
משוואה 1/משוואה2/אין הבדל בין המודלים
ב. אם משתנה רמת המחירים במשק היה מובהק במשוואה מס' 1, הוא יהיה מובהק בהכרח גם במשוואה מס' 2 :
נכון/לא נכון/אי אפשר לדעת

שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)

ענה על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל: $Y = \alpha + \beta \cdot X + U$ ומתקיימות כל ההנחות הקלאסיות).

1. $\overline{R}^2 < R^2$ מתקיים תמיד: נכון/לא נכון/ אי אפשר לדעת.

2. אם דוחים H_0 במבחן חד צדדי ברמת מובהקות α, אזי בהכרח גם נדחה H_0 במבחן הדו צדדי באותה רמת מובהקות: נכון/לא נכון/ אי אפשר לדעת.

3. אם ערך האומד ל-β גבוה, השערת האפס למובהקות השיפוע תידחה בוודאות: נכון/לא נכון/אי אפשר לדעת.

4. הוספת משתנה מסביר למשוואת הרגרסיה עשויה להקטין את R^2:
נכון/לא נכון/אי אפשר לדעת.

5. אם דוחים H_0 במבחן דו צדדי ברמת מובהקות α, אזי בהכרח גם נדחה H_0 במבחן החד צדדי באותה רמת מובהקות: נכון/לא נכון/ אי אפשר לדעת.

6. אם רווח בר סמך לשיפוע כולל את הערך אפס, ניתן לומר כי השערת האפס למובהקות השיפוע מתקבלת בהכרח: נכון/לא נכון/אי אפשר לדעת.

7. האומדים היעילים ביותר לפרמטרים באוכלוסיה יהיו בהכרח אומדי הריבועים הפחותים: נכון/לא נכון/ אי אפשר לדעת.

8. בהוספת משתנה מסביר מובהק למודל, ערך $\overline{R}^2$ יעלה בהכרח.
נכון/לא נכון/אי אפשר לדעת.

9. מבחן WALD הוא מקרה פרטי של מבחן F למובהקות המודל:
נכון/לא נכון/ אי אפשר לדעת.

10.שיטת הריבועים הפחותים מביאה למקסימום את $\overline{R}^2$:
נכון/לא נכון/אי אפשר לדעת.

שאלה 3 (17 נקודות)

נתון המודל : $Y_t = \alpha + \beta X_t + U_t$

נתון כי אר"פ למודל זה הינו: $\hat{\beta} = S_{xy} / S_{xx}$

א. הוכיחי כי $\hat{\beta}$ אומד ליניארי וחסר הטיה של β.

ב. חשבי את $VAR(\hat{\beta})$.

ג. נתון האומד : $\tilde{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$

הוכיחי כי $\tilde{\beta}$ אומד ליניארי אך איננו חסר הטיה ל- β.

ד. מהם התנאים בהם מתקיים : $E(\tilde{\beta}) = \beta$?

פלט מס' 1 - משוואה מס' (1)

Dependent Variable: lnm

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	1				<0.0001
Error	103				
C Total	104	44.91976			

Root MSE	0.09251	R-square	0.9804
Dep Mean	8.53854	Adj R-sq	0.9802
C. V.	1.08344		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP					
P	1	1.49372	0.09862	15.15	<.0001
lnp	1	1.69267	0.02360		<.0001

פלט מס' 2 – משוואה מס' (2)

Dependent Variable: lnm

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	44.05069	22.02535	2585.05	<0.0001
Error	102	0.86907	0.00852		
C Total	104	44.91976			

Root MSE	0.09231	R-square	0.9807
Dep Mean	8.53854	Adj R-sq	0.9803
C. V.	1.08104		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	0.78242	0.59739	1.31	0.1932
lnp	1	1.63491	0.05332	30.66	<.0001
lny	1	0.20001	0.16568	-----	0.2302

Covariance of Estimates

COVB	INTERCEP	lnp	lny
INTERCEP	0.35687	0.025884	-0.09762
lnp	0.02588	0.002843	-0.00792
lny	-0.09762	-0.00792	0.02745