אקונומטריקה א'

החוג לכלכלה

GOOL

בשביל התירגול

גול זה בול. בשבילך!

תוכן העניינים:

הקדמה: תזכורת של סטטיסטיקה ומתמטיקה

<u>הגדרות וסימונים</u>

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

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

<u>משתנה מקרי</u> - 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 <u>קבועים</u>- המקבלים ערך קבוע ומסומנים באות לועזית ללא אינדקס (כמו למשל a או b)

באקונומטריקה נעסוק בעיקר במשתנים מקריים.

לכל משתנה מקרי X_t יש <u>תוחלת</u> המייצגת את מרכז ההתפלגות.

התוחלת מסומנת μ_X או $E\left(X\right)$.

<u>השונות</u> מייצגת את מידת הפיזור של ההתפלגות.

השונות מסומנת σ_X^2 או $V\left(X\right)$.

<u>סטית התקן</u> היא השורש של השונות והיא מסומנת σ_X.

לשני משתנים מקריים X ו- Y יש <u>שונות משותפת</u> (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}

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

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

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**

אמידה

פרמטר- ערך המשתנה הנחקר המתאר את כל האוכלוסיה

סטטיסטי/אומד-ערך המשתנה הנחקר המתאר את המדגם

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

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:**

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

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

<u>הגדרות ופיתוחים</u>

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:**

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

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:**

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

<u>הערה חשובה:</u>

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

חוקי השונות המשותפת

1) **שונות משותפת בין משתנה לקבוע =0:**

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

2) שונות משותפת של משתנים המוכפלים בקבוע:

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

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

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

פרק ראשון: מבוא - מה זו אקונומטריקה?

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

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

נתחיל בדוגמא:
מתווך דירות בתל אביב רצה לבדוק איך משפיע גודלה של דירה על המחיר שבו היא נמכרת.
הוא הניח 2 הנחות מקדימות:
1) רק גודל הדירה משפיע על מחיר הדירה באופן שיטתי. כל שאר הדברים המשפיעים על מחיר הדירה הם אקראיים ולא ניתנים לחיזוי.
2) ההשפעה של גודל הדירה על מחיר הדירה היא 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

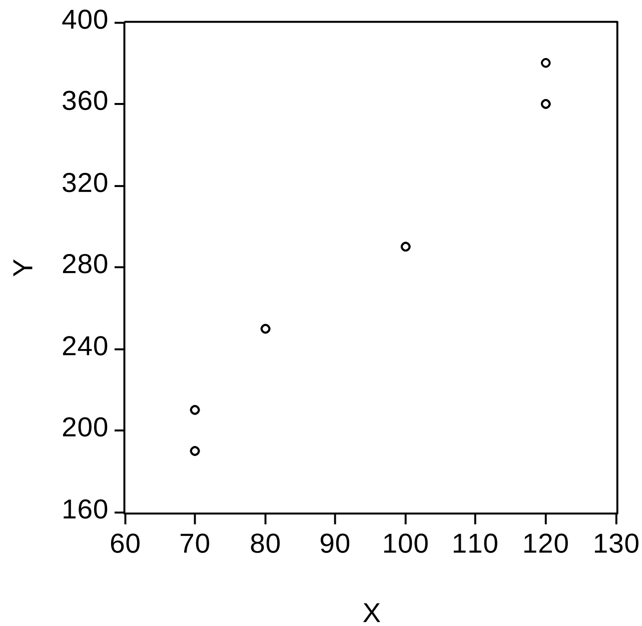
פרמטרים:
α **- חותך**
β **- שיפוע**

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.**

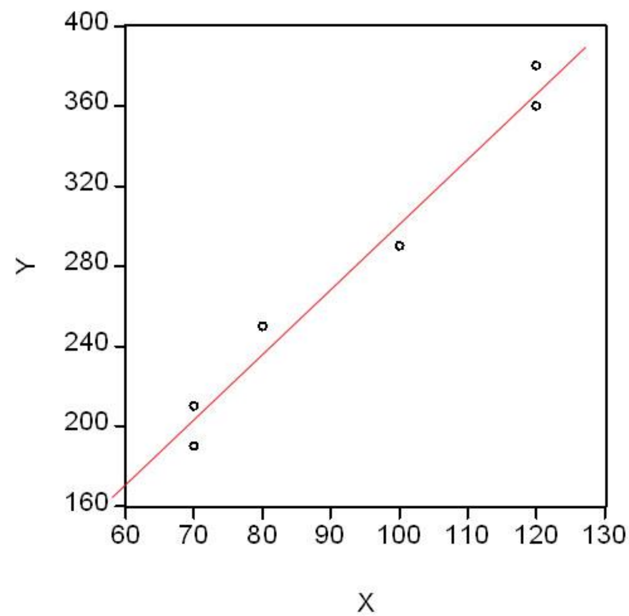
מספר הדירה	גודל הדירה במ״ר	מחיר הדירה באלפי דולרים
1	$X_1 = 70$	$Y_1 = 190$
2	$X_2 = 70$	$Y_2 = 210$
3	$X_3 = 80$	$Y_3 = 250$
4	$X_4 = 100$	$Y_4 = 290$
5	$X_5 = 120$	$Y_5 = 360$
6	$X_6 = 120$	$Y_6 = 380$

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-290,3 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' - $\tilde{\beta}$.

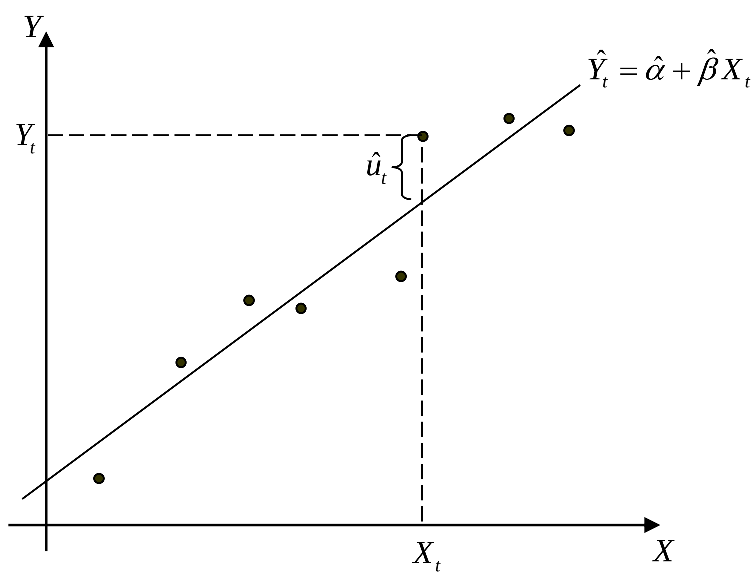
4) בעוד α ו- β הם קבועים, $\hat{\alpha}$ ו- $\hat{\beta}$ הם משתנים מקריים. מדוע? מפני שבכל מדגם מתקבלים $\hat{\alpha}$ ו- $\hat{\beta}$ אחרים.

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

פרק שני: 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 מתקבלית 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	$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)\left(Y_t - \bar{Y}\right)}{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)^2}$ $= \frac{\sum_{t=1}^{T}(X_t - \bar{X})Y_t}{\sum_{t=1}^{T}(X_t - \bar{X})^2}$ $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$	$\hat{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$	$\hat{\alpha} = \bar{Y}$
תוחלת האומדים	$E\left(\hat{\beta}\right) = \beta$ $E\left(\hat{\alpha}\right) = \alpha$	$E\left(\hat{\beta}\right) = \beta$	$E\left(\hat{\alpha}\right) = \alpha$

$V(\hat{\alpha})=\frac{\sigma_u^2}{T}$	$V(\hat{\beta})=\frac{\sigma_u^2}{\sum_{i=1}^{T}X_t^2}$	$V(\hat{\beta})=\frac{\sigma_u^2}{S_{XX}}$ $V(\hat{\alpha})=\sigma_u^2\left(\frac{1}{T}+\frac{\bar{X}^2}{S_{XX}}\right)$	**שונות האומדים**

****<u>הערה חשובה:</u> בתהליך הגזירה של פונקציית הריבועים הפחותים מתקבלות "המשוואות הנורמליות":**

<u>עבור המודל הקלאסי (עם חותך):</u>

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

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

<u>עבור מודל ללא חותך:</u>

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

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:**
$\text{cov}(X_t, u_t) = 0$ **(שכן המתאם בין X לבין U שווה ל-0 עבור כל t).**

5) הומוסקדסטיות: השונות של ההפרעה האקראית זהה לכל תצפית ותצפית:
$V(u_t) = \sigma_u^2$ **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** $\Leftarrow$

$\operatorname{cov}(X_t, u_t) = 0$

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

$V(u_t) = \sigma_u^2$ **לכל** t

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

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

<u>תכונות האומדים</u>

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

1) <u>לינאריות</u>

אר"פ ניתנים להצגה כקומבינציה לינארית 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 לינארי?</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. $$\hat{\beta}=\frac{\sum_{t=1}^{T}(X_t-\overline{X})Y_t}{\sum_{t=1}^{T}(X_t-\overline{X})^2}$$

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

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

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 $E\left(\hat{\beta}\right) = \beta$, אז $\hat{\beta}$ הוא אומד חסר הטיה ל- β.

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

המשך הדוגמא שלעיל:

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

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 $\Leftarrow$ מבלי לחשב את שונותו נדע לפי משפט גאוס-מרקוב שהיא גדולה יותר משל אומד הריבועים הפחותים.

אם האומד איננו ליניארי ו/או חסר הטיה $\Leftarrow$ לא ניתן להשתמש במשפט גאוס-מרקוב ואז היחס בין שונות האומד לשונות אומד הריבועים הפחותים המקביל איננו ידוע.

<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**

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

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

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 $\hat{\beta}$ הינו אומד חסר הטיה ל- β נכון / לא נכון

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

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

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

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

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

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

(יש לשים לב המודל ללא חותך)

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

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

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

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

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

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

(יש לשים לב המודל ללא חותך)

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

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 $\frac{\sum X_t Y_t}{\sum X_t^2}$ **?**

<u>שאלות נוספות מתוך מבחנים</u>

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 שלישי: 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 הלוגריתמים) 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.**

גמישות

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 לינארי.

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.

סיכום:

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% Y ישתנה 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**	$\frac{\beta}{X}$	$\frac{\beta}{Y}$

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

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:

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

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

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

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

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

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

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

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

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

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

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

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

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

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 רביעי: 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 (Xים)?

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

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	$s^2_{\hat{\alpha}}$	$\mathrm{cov}\left(\hat{\alpha},\hat{\beta}\right)$
X	$\mathrm{cov}\left(\hat{\alpha},\hat{\beta}\right)$	$s^2_{\hat{\beta}}$

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 $\hat{\gamma}$ **(** $S_{\hat{\gamma}}^2$ **):**

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

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

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

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 שה- $\alpha \neq 5\beta$

חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר (SALARY) לפי המודל: $\ln(\text{SALARY}_t) = \alpha + \beta \cdot \text{EXP}_t + u_t$ **. 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:

<u>עריכת תחזית וקריאת פלטים (של SPSS)</u>

אמידה נקודתית

אמידה נקודתית עבור X_0 מסוים (תחזית) מחושבת על פי קו הרגרסיה <u>במדגם</u>:

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

אמידת מרווח ל- $E(Y)$

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

<u>נוסחת הרב״ס:</u>

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

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

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

אמידת מרווח ל-Y

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

<u>נוסחת הרב"ס:</u>

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

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

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.

<u>*התחזית מדויקת יותר (שונות* התחזית קטנה יותר) כאשר:</u>

1. **n (גודל המדגם) גדול יותר**
2. **שונות המשתנה המסביר X גדולה יותר**
3. **X_0 קרוב יותר ל $\overline{X}$**
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			

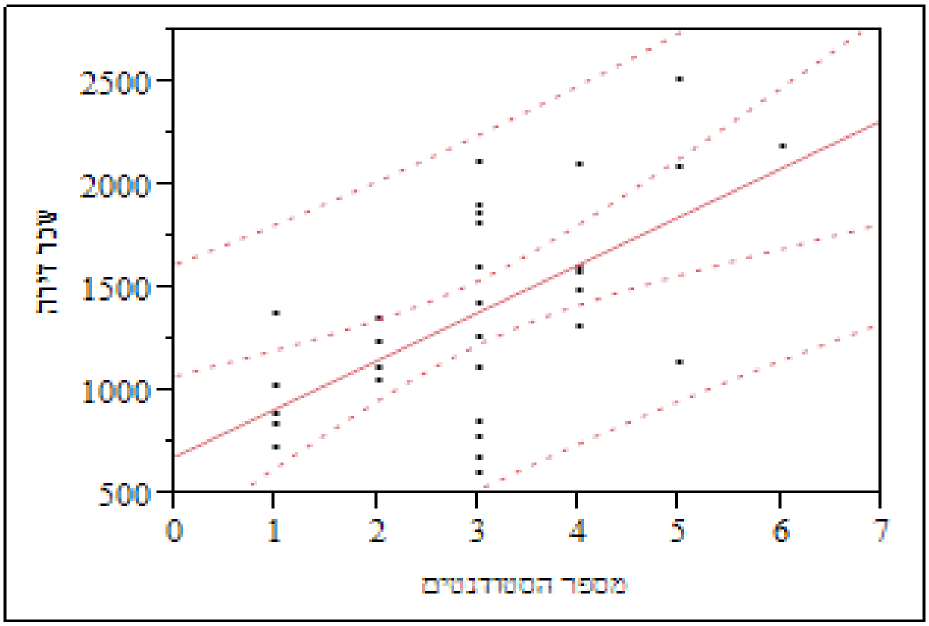
Descriptive Statistics

	Mean	Std. Deviation	N
שכר הדירה	1386.7667	509.46027	30

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

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
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 חמישי: רגרסיה רב משתנית

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 היטובסקי

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 $\left|t_{\hat{\beta}}\right| > 1$.

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

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

כאשר $1 < \left|t_{\hat{\beta}}\right| < 2$ 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:**

$$WALD_{stat} = \frac{\dfrac{ESS_R - ESS_U}{m}}{\dfrac{ESS_U}{T-k-1}}$$ **(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 **:**

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

<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?

<u>המשך השאלה</u>

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?

<u>מקרים פרטיים של מבחן WALD</u>

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 :

$$WALD_{stat} = F$$

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

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

למשל במודל:

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

רוצים לבדוק את מובהקות β_2

<u>השערות:</u>

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

ניתן לבדוק זאת גם במבחן WALD:

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

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

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.**

<u>תרגול מסכם</u>

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-6.0 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.**

סיכום: מבחנים לדוגמא

מבחן לדוגמא מס' 1

שאלה 1 (55 נקודות)

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-4.0. 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:

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

8) מהי השערה 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

9) באיזו רמת מובהקות נקבל את טענת החוקר?

10) 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.

11) מהי השערת האפס לבדיקת הטענה?

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

13) 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** $\tilde{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{S_{XX}}$

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**

מבחן לדוגמא מס' 2

שאלה 1 (60 נקודות)

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:

כאשר: $Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + \gamma_2 \cdot Z_2 + \nu$

$Z_0 =$ _______________

$Z_1 =$ _______________

$Z_2 =$ _______________

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: $\tilde{\beta} = \frac{S_{XY}}{\sum_{t=1}^{T} X_t^2}$**.**

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 $\hat{\beta}$ 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 $E\left(\hat{\beta}\right) < \beta$:

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 $\hat{\beta}$ 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 $\hat{\beta}$ 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

<u>מבחן לדוגמא מס' 3</u>

<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:

$$\ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + U_t \qquad \textbf{(3)}$$

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

מבחן לדוגמא מס' 4

שאלה מס׳ 1 (55 נקודות)

בנק מעוניין לאמוד את סך הפעילות בכרטיסי אשראי של לקוחותיו. לשם כך אסף נתונים על 35,971 מלקוחותיו ואמד את המשוואה הבאה:

$$CREDIT_t = \alpha + \beta \cdot SAVINGS_t + U_t \quad (1)$$

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:

$$CREDIT_t = \alpha + \beta_1 \cdot PIKADON1_t + \beta_2 \cdot PIKADON2_t + U_t \quad \textbf{(2)}$$

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

<u>שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)</u>

עני 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

<u>שאלה 3 (12 נקודות)</u>

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

מבחן לדוגמא מס' 5

שאלה מס' 1 (55 נקודות)

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 למשואה (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:__________

D_0 : _____________

כאשר: 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

<u>שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)</u>

עני על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל:

$Y = \alpha + \beta \cdot X + U$ **ומתקיימות כל ההנחות הקלאסיות).**

1. $R^2 < \overline{R}^2$ **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.

<u>שאלה 3 (17 נקודות)</u>

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: $\tilde{\beta} = \dfrac{\sum X_t Y_t}{\sum X_t^2}$

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
INTERCE					
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