מבוא לאקונומטריקה א'

החוג לכלכלה

מכללת אריאל

נושאי הקורס:

גוּל זה בּוּל. בִּשְׁבִילְךָ!

<u>הקדמה: תזכורת של סטטיסטיקה ומתמטיקה</u>

<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 $E(X)$.

<u>השונות</u> מייצגת את מידת הפיזור של ההתפלגות.
השונות מסומנת σ_X^2 או $V(X)$.
<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 $\Rightarrow$ X,Y בלתי מתואמים

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

מקדם המתאם מסומן η_{xy}

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

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

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

$\eta_{XY}=\dfrac{\operatorname{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$	$r_{XY}=\dfrac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$

נוסחאות וחוקים בסטטיסטיקה

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

חוקי הסיגמה

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

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:**

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

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

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:**

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

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:**

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

<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:**

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

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:**

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

<u>פרק ראשון: מה זו אקונומטריקה?</u>

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 <u>רגרסיה</u>.

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

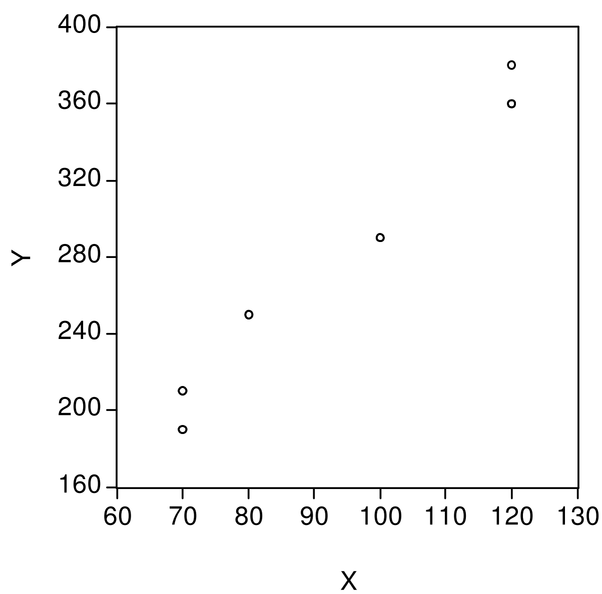
הפרעה אקראית	**משתנה**	**פרמטרים:**	**משתנה**
"קריזה"	**מסביר**	α **- חותך**	**מוסבר**
		β **- שיפוע**	

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

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

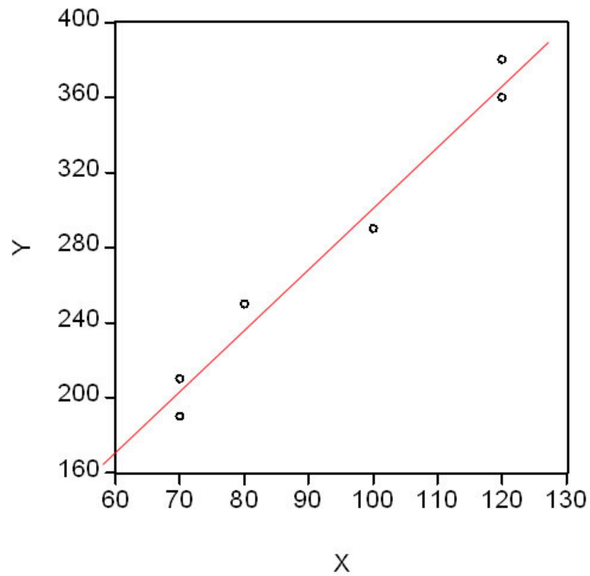
מספר הדירה	גודל הדירה במ"ר	מחיר הדירה באלפי דולרים
1	$X_1 = 70$	$Y_1 = 190$
2	$X_2 = 70$	$Y_2 = 210$
3	$X_3 = 80$	$Y_3 = 250$
4	$X_4 = 100$	$Y_4 = 290$
5	$X_5 = 120$	$Y_5 = 360$
6	$X_6 = 120$	$Y_6 = 380$

נציג את 6 התצפיות בגרף:



מהו הקו הישר המתאר את הקשר בין שני המשתנים בצורה הטובה ביותר? (הקו הוא ישר בגלל שהמתווך הניח לינאריות של המודל).

מסתבר שקו הרגרסיה הטוב ביותר הוא קו שחושב <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.

לסיכום:

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' - $\tilde{\beta}$.

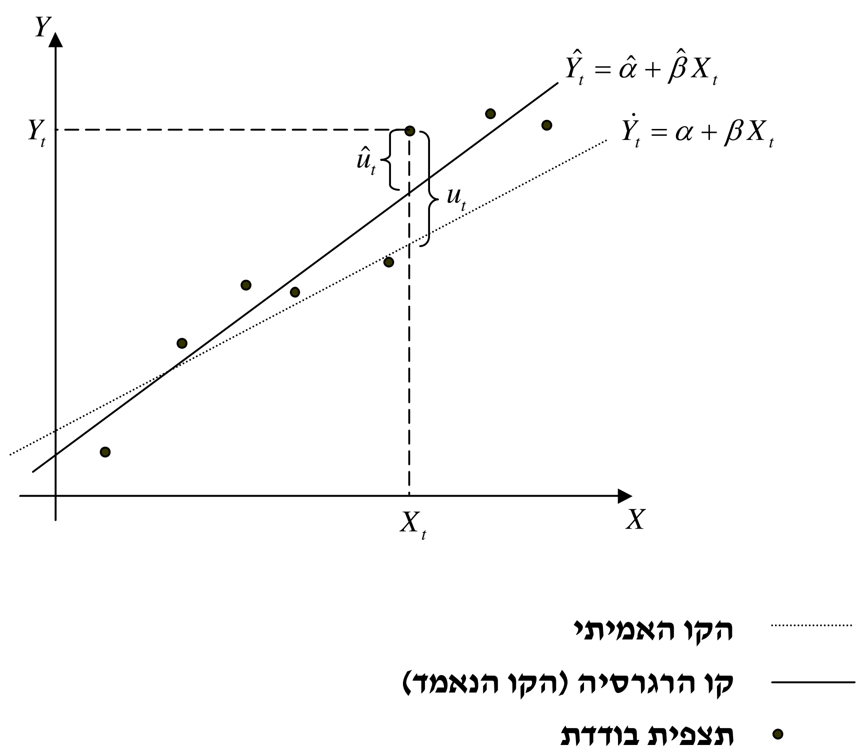
4) בעוד α ו- β הם קבועים, $\hat{\alpha}$ ו- $\hat{\beta}$ הם משתנים מקריים. מדוע? מפני שבכל מדגם מתקבלים $\hat{\alpha}$ ו- $\hat{\beta}$ אחרים.

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 $\hat{u}_t = Y_t - \hat{Y}_t$.**



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

<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: $\sum \hat{u}_t = 0$

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

<u>עבור מודל ללא חותך:</u>

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

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

<u>תירגול:</u>

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

א. נסחו את בעית ה-OLS.

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

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

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

ה. בהנחה והיינו בוחרים אומד אחר ל- β **שאינו אומד הריבועים הפחותים, מה היה יחס הביטויים:** $\sum e_i$ **ו-** $\sum e_i^2$ **של אומד זה ביחס לאומד הריבועים הפחותים?**

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

א. נסחו את בעית ה-OLS.

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

ג. מצאו נוסחה לקבלת $\hat{\beta}$.

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

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

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

תצפיות מ-5 סטודנטים:

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

איזה מבין המודלים הבאים נאמד?

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

ב. $\hat{score}_i = \hat{\beta} \cdot IQ_i$

ג. $\hat{score}_i = \hat{\alpha}$

ד. $\hat{score}_i = \bar{y}$

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

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

לשם כך נדגמו 10 מעסיקים באופן מקרי ונתקבלו התוצאות הבאות:

$$\bar{x} = 35$$

$$\bar{y} = 5.8$$

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

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

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

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

ההנחות הקלאסיות של מודל הרגרסיה:

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

1) קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.

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 להיפך במודל הניבוי הנוכחי בו אנו משתמשים.

**** שימו לב כי מהנחה זו משתמע גם כי המתאם בין הטעויות לערכי X שווה ל-0:**
$\operatorname{cov}(X_t, u_t) = 0$ **(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: $\operatorname{cov}(u_t, u_s) = 0$ **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: $S_{XX} = \sum_{t=1}^{T} (X_t - \bar{X})^2 \neq 0$

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:** $u_t \approx N$

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

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

ב. התייחסו לכל אחת מהטענות הבאות וקבעו האם היא:
הנחה קלאסית/משוואה נורמאלית (או תוצאה הנובעת ממשוואה נורמאלית)/אף אחד מהשניים:

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

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

3. $E(u_i) = 0$

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

5. $\bar{e} = 0$

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

7. $\sum u_i = 0$

8. $V(u_i) = \sigma_i$

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

תכונות האומדים

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

1) לינאריות

אר"פ ניתנים להצגה כקומבינציה לינארית של Y_t.

במילים אחרות, כדי ש- $\hat{\beta}$ למשל, תהיה אומד 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 האומדים של 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).**

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

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

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

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

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

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

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

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

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

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

מי מהאומדים הוא ליניארי ומהן המשקולות?

2) **חוסר הטיה**

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 ($\vartheta \neq \hat{\vartheta}$) 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 $\hat{\beta} = \dfrac{\sum X_t Y_t}{\sum X_t^2}$:

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

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:**

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

משפט גאוס מרקוב:

יעילות היא תמיד מושג השוואתי. לכן בכדי לדעת האם השונות של האומד היא המינימאלית האפשרית נשתמש במשפט גאוס מרקוב.

לפי משפט גאוס-מרקוב אר"פ הם בעלי השונות הנמוכה ביותר בקבוצה שלהם (קבוצת האומדים הלינאריים חסרי ההטיה), והם נקראים B.L.U.E. (Best Linear Unbiased Estimation).

כלומר:

אם האומד שלנו הוא ליניארי וחסר 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

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

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>סיכום: השלבים להוכחת התכונות</u>

1) הוכחת ליניאריות

2) הכנת האומד $\Leftarrow$ 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 ליניארי ל-** β **נכון / לא נכון**

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

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

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

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

(יש לשים לב המודל ללא חותך)

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 $\tilde{\beta}$?

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

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

? שאלות נוספות מתוך מבחנים

בכל השאלות ההנחות הקלאסיות מתקיימות.

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

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

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

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

נכון / לא נכון / אי אפשר לדעת

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

א. הוא בהכרח שלילי

ב. הוא בהכרח חיובי

ג. הוא בהכרח שווה לאפס

ד. לא ניתן לקבוע את סימנו על סמך הנתונים הקיימים

5. סמן את הטענה הנכונה בהכרח:

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

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

ג. $\sum_{t=1}^{T} X_t u_t = 0$

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

6. אומדי הריבועים הפחותים אינם חסרי הטיה, אם נתון שהשונות של u_t **אינה קבועה.** **נכון / לא נכון / אי אפשר לדעת**

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

נכון / לא נכון / אי אפשר לדעת

פרק רביעי: מודלים לא ליניאריים

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

נשאלת השאלה-מתי מודל מוגדר כליניארי?

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

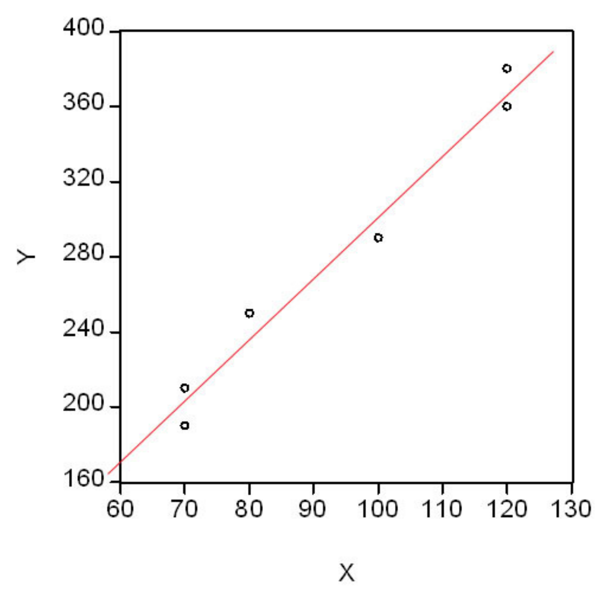
למשל המודל הליניארי הקלאסי: $Y = \alpha + \beta X + u$ **מתאר קשר קוי בין x ל-y .**

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

נגזרת זו מתארת את השינוי השולי (השיפוע של הגרף): אם מגדילים את x ביחידה אחת, בכמה יחידות משתנה y.

במודל הליניארי- שינוי זה הוא קבוע ושווה ל- β **.**

גרף המתאר את הקשר יראה כך:



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:

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

גרף המתאר את הקשר יראה כך:

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

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

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

גרף המתאר את הקשר יראה כך:

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

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

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:

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

גרף המתאר את הקשר יראה כך:

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

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 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**

תירגול

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

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

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

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

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

א. הסבירו את המשמעות של β בכל אחד מהמודלים

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

נתונים תוצאות האמידה של המודלים הבאים:

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

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

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

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

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

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

❓נתונים המודלים הבאים עבור התוצר במשק:

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

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

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

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

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

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

כאשר:

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

A- הוצאות צריכה על המוצר בהינתן רמת הכנסה אפסית.

K- הכנסת הפרט.

L- שנות לימוד.

א. מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?

ב. מי מבין המודלים שלא ניתנים לאמידה בשיטת OLS ניתן להביא למודל ליניארי בפרמטרים ועל כן לאמוד את הפרמטרים שלו?

ג. עבור כל אחד מהמודלים קבעו מיהו המשתנה המוסבר ומיהו המסביר במשוואת הרגרסיה הליניארית.

ד. עקומת אנג'ל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנג'ל?

? נתון המודל הבא:

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

א. האם ניתן לאמוד את המודל בשיטת OLS?

ב. מה המשוואה שצריך לאמוד על מנת לקבל את הפרמטרים למודל זה (כלומר כיצד הופכים את המודל לליניארי בפרמטרים)?

ג. נאמד המודל הבא :

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

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

מהם האומדנים עבור $A,\ \beta_1$?

פרק חמישי: רגרסיה מרובה ומולטיקוליניאריות

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

מודל רגרסיה מרובה נראה כך:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_J X_{Ji} + U_i$$

כאשר:

Y_i =**משתנה תלוי**

$X_{1i} \ldots X_{Ji}$ = **משתנים ב"ת**

U_i =**טעות מקרית המקיימת את כל ההנחות הקלאסיות.**

מקדמי מודל הרגרסיה המרובה:

α = **חותך אחד שמשמעותו: הציון המנובא כאשר כל המשתנים הב"ת=0.**

$\beta_1 \ldots \beta_J$ = **מקדמי השיפוע. מס' הבטות = למספר המשתנים הב"ת במודל.**

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

אמידת מודל הרגרסיה המרובה:

ברגרסיה מרובה, כמו ברגרסיה פשוטה , שיטת האמידה הטובה ביותר היא שיטת הריבועים הפחותים.
כלומר, נרצה להביא את סכום הטעויות בניבוי למינימום:

$$Min\sum e_i^2 = Min\sum (Y_i - \hat{\alpha} - \hat{\beta}_1 X_{1i} - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i} - \ldots - \hat{\beta}_J X_{Ji})^2$$

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

הפיתרון יהיה על פי נגזרות חלקיות לפי כל אומד:

$$\frac{\partial}{\partial\hat{\alpha}}=-2\sum_{1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji})=0$$

$$\frac{\partial}{\partial\hat{\beta}_1}=-2\sum_{1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji})X_{1i}=0$$

$$\frac{\partial}{\partial\hat{\beta}_2}=-2\sum_{1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji})X_{2i}=0$$

$$\frac{\partial}{\partial\hat{\beta}_3}=-2\sum_{1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji})X_{3i}=0$$

.

.

.

$$\frac{\partial}{\partial\hat{\beta}_J}=-2\sum_{1}^{n}(Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji})X_{Ji}=0$$

אם נשתמש בזהות: $e_i=Y_i-\hat{\alpha}-\hat{\beta}_1X_{1i}-\hat{\beta}_2X_{2i}-\hat{\beta}_3X_{3i}-...-\hat{\beta}_JX_{Ji}$ **נקבל את המשוואות הנורמאליות:**

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

$\sum_{i=1}^{n}e_iX_{1i}=0$ **בגלל שיש את** X_{1i}

$\sum_{i=1}^{n}e_iX_{2i}=0$ **בגלל שיש את** X_{2i}

$\sum_{i=1}^{n}e_iX_{3i}=0$ **בגלל שיש את** X_{3i}

.

.

.

$\sum_{i=1}^{n}e_iX_{Ji}=0$ **בגלל שיש את** X_{Ji}

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

דוגמא: מקרה פרטי, מודל עם שני משתנים מסבירים:

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

הנוסחאות הנורמאליות הן:

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

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

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

מפיתרון מערכת המשוואות נקבל את הנוסחאות הבאות לחישוב האומדים:

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

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

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

<u>הערה:</u> ניתן לראות כי אם לא קיים מתאם בין המשתנים הב"ת $r_{12} = 0$, שיפועי הרגרסיה המרובה זהים לשיפועי הרגרסיה הפשוטה:

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

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

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

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

א. מהי בעיית ה-OLS שעליו לפתור?

ב. מצאו את תנאי סדר ראשון של הבעיה.

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

לכן אסף מדגם בין ארבע תצפיות חודשיות.

להלן טבלה מסכמת:

חודש	X2 השקעות זרים בישראל (במיליוני דולרים)	X1 שער הריבית	Y דולר	טעות (e_i)
אוגוסט	100	3	3.2	-5
ספטמבר	95	3.5	3.6	6
אוקטובר	90	3.5	3.8	0
נובמבר	100	3	3.5	-2

מהו המודל אשר אותו אמד הכלכלן?

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

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

נתון האומד:

$$\tilde{\beta}_1 = \frac{\sum X_{1i}((Y_i - 8X_{2i} - 2) - \overline{(y - 8X_2 - 2)})}{\sum X_{1i}^2}$$

א. חשבו את תוחלת האומד

ב. חשבו את שונות האומד

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

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

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

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

אומדים את β_1 **באופן הבא:**

$$b_1 = \frac{\sum (x_{1i} - \bar{x})(y_i - 8x_{2i} - \overline{(y - 8x_2)})}{\sum (x_{1i} - \bar{x})^2}$$

א. האם האומד חסר הטיה?

ב. מהי שונות האומד?

<u>מולטיקוליניאריות</u>

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

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

מולטיקוליניאריות מלאה

מתאם מלא בין המשתנים המסבירים במודל.
הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני:

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

שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל $x_1 = x_2^2$**), אז בהכרח** $r_{12} \neq 1$.

<u>הערה:</u> מולטיקוליניאריות מלאה יכולה להיווצר גם כאשר קבוצה של משתנים מסבירים מהווה קומבינציה ליניארית מלאה של אחד המשתנים המסבירים:

$x_1 + x_2 = a + bx_3$

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

לדוגמא:
נניח שאנו רוצים לאמוד את הביקוש לדירות בתל אביב כפונקציה של מחירן בשקלים (x_1 **) ובדולרים (** x_2 **):**

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

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

מדוע זה בעייתי?

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

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

בדוגמא שלנו:

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

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

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

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

מכיוון ש: $x_1 = bx_2$ **(b = שער הדולר) אז גזירת** $\hat{\beta}_1$ **:** $\sum e_i(bx_{2i}) = 0$, $b\sum e_i x_{2i} = 0$, $0 = 0$ **והמשוואה מתבטלת.**

<u>פיתרון</u>: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקוליניאריות חלקית

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

$$x_1 = a + bx_2 + u_i$$
$$x_1 + x_2 = a + bx_3 + u_i$$

לדוגמא:

נניח שאנו רוצים לאמוד את הציונים בתואר ראשון ע"י ציוני הפסיכומטרי (x_1) וציוני הבגרות (x_2):

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

מכיוון שיש מתאם גבוה בין ציוני הפסיכומטרי וציוני הבגרות לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני ה-.B.A.

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

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

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

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

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

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

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

ג. הוכיחו תשובותיכם לסעיפים ב ו- ג.

ד. בהנחה כי מתקיים: $r_{12} = 0.98$

1. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים:

נכון/לא נכון/ לא ניתן לדעת.

2. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

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

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

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

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

כלכלן אמד את המודל הבא:

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

האם קיימת בעיה של מולטיקוליניאריות במודל?

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

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

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

להלן משוואה מס' 2:

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

חווה דעתך על המשוואה השניה.

פרק שישי: מבחן t

עד מעתה למדנו לאמוד את מקדמי הרגרסיה (ה- α וה-β) על סמך מדגם מקרי יחיד. כעת נלמד כיצד לבדוק האם המקדמים שאמדנו מובהקים באוכלוסיה.

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

מבחן למובהקות ה- β (מקדם השיפוע)

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

השערות:

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

סטטיסטי המבחן:

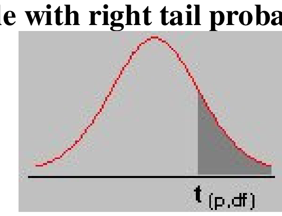
$$t_{\beta=0} = \frac{\hat{\beta} - \beta_0}{S_{\hat{\beta}}}$$

כלל הכרעה ומסקנה:

נדחה את H0 אם:

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

t table with right tail probabilities



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

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

inf	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905

<u>מסקנה:</u> יש/אין עדות לכך שהמשתנה הב"ת המסוים 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:

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

- **ניתן לבדוק בנוסף האם ה-β (השינוי השולי) שווה לערך מסוים באוכ'.**

  <u>השערות לדוגמא:</u>

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

  <u>סטטיסטי המבחן:</u>

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

  <u>כלל הכרעה:</u>

  על סמך טבלת T

- **רב"ס ל-β:**

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

ניתן לבדוק השערות באמצעות הרב"ס. צריך לבדוק האם הרב"ס מכיל את הערך המבוקש (את β_0) אם כן- נקבל את H0 ואם לא-נדחה אותה.

מבחן למובהקות ה- α (החותך)

עונה על השאלה- האם קו הרגרסיה יוצא מראשית הצירים (החותך של קו הרגרסיה=0):

השערות:

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

סטטיסטי המבחן:

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

כלל הכרעה ומסקנה:

נדחה את H0 אם:

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

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

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

  ההשערות:

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

  כלל ההכרעה:

  שימוש בטבלת T:

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

- **ניתן לבדוק בנוסף האם ה-α (החותך) שווה לערך מסוים באוכ'.**

  השערות לדוגמא:

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

  סטטיסטי המבחן:

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

<u>כלל הכרעה:</u>

על סמך טבלת T :

נדחה את H0 אם:

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

- **רב"ס ל-α:**

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

ניתן לבדוק השערות באמצעות הרב"ס. צריך לבדוק האם הרב"ס מכיל את הערך המבוקש (את α_0) אם כן- נקבל את H0 ואם לא-נדחה אותה.

? חוקר רצה לבחון את השפעת ההכנסה (INCOME) על גובה המס (TAX) (במיליארדי $) שגובה מדינה במערב לפי המודל : $TAX_i = \alpha + \beta \cdot INCOME_i + u_i$ **.**
לשם כך אסף נתונים מ-51 מדינות . להלן התוצאות:

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

(0.08953) (0.01622)

סטיות התקן של האומדים נתונות בסוגריים.

א. מהי המשמעות הכלכלית של β ושל α?

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

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

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס ברמת מובהקות של5% וברמת מובהקות של 1%

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

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

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

המודל: $\ln\left(\text{SALARY}_t\right)=\alpha+\beta\cdot\text{EXP}_t+u_t$ **. הוא אסף 403 תצפיות, ואמד את הפרמטרים. להלן תוצאות האמידה:**

$$\ln(SALARY)_i=7.334-0.0087\cdot EXP_i$$

(0.068) (0.0026)

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

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

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

נאמד המודל $Y_i=\alpha+\beta_1X_i+\beta_2Z_i+\beta_3W_i+\beta_4S_i+u_i$ **והתקבלו התוצאות הבאות:**

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

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

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

ב. בנו רווח בר סמך להשפעת X על Y

<u>תחזית</u>

המטרה של קו הרגרסיה הוא ביצוע תחזיות. תחזית נקודתית מחושבת על פי קו הרגרסיה שאמדנו. נציב במקום ה-Xים ערכים נתונים ונקבל למה שווה ה-Y המנובא.

לדוגמא:

נתונה משוואת הרגרסיה הבאה:

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

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

מה יהיה סה"כ הוצאות משק הבית אם גיל הילד הראשון הוא 2 שנים, של השני 4.5 שנים, השלישי הוא בן 5 ואילו הרביעי בן 8?

<u>תשובה:</u>

נציב במשוואת הרגרסיה:

$$\hat{y}_i=13+8\cdot2+7\cdot4.5+2\cdot5+9\cdot8=142.5$$

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

<u>ברגרסיה פשוטה</u>, כאשר המודל הינו: $Y_i = \alpha + \beta X_i + u_i$ אנו יכולים גם לאמוד את התחזית באוכלוסיה עבור ערך מסוים של X, באמצעות רווח בר סמך:

- **רווח בר סמך לתחזית עבור X_f מסוים:**

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

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

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

<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_f קרוב יותר ל $\bar{X}$**
4. **האומד לשונות הטעויות- S_u, קטן יותר.**

לדוגמא:

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

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

נתון בנוסף כי:

$$S_x^2 = 1.313^2$$
$$S_u^2 = 414.055^2$$
$$\bar{x} = 3$$

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

<u>תשובה:</u>

יש לחשב תחזית נקודתית עבור X=2:

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

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

<u>תשובה:</u>

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

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

חישובי עזר:

$$\sum (X_i - \overline{X})^2 = S_{XX} = (n-1)S_X^2 = (30-1) \cdot 1.31306^2 = 49.99$$
$$t_{28,0.025} = 2.048$$

$$1153.247 \pm 2.048 \cdot 414.055 \sqrt{1 + \frac{1}{30} + \frac{(2-3)^2}{49.99}}$$
$$1153.247 \pm 870.305$$
$$p(282.94 \le Y_{X=2} \le 2023.55) = 0.95$$

<u>מבחן t מורכב (בחינת קשרים ליניאריים בין הפרמטרים)</u>

לעיתים אנחנו מתבקשים לבדוק השערות העוסקות בקשרים בין הפרמטרים. כמו למשל: $H_0 : \alpha = 5\beta$ **או** $H0 : \beta_1 = 2 \cdot \beta_2$.

במקרים אלו נרשום את השערות האפס כך: $H0 : \alpha - 5\beta = 0$ **ו-** $H0 : \beta_1 - 2 \cdot \beta_2 = 0$

ונחשב את סטטיסטי המבחן t :

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

כאשר את טעות התקן של המבחן מחשבים תוך שימוש בנוסחאות :

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

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

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

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.

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

דוגמא:

נתון המודל $Y_t=\alpha+\beta X_t+u_t$

שלצורך אמידתו נאספו 240 תצפיות ונתקבל ש:

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

(0.25) (0.12)

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

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

<u>תשובה:</u>

$$H_0:\alpha-5\beta=0$$
$$H_1:\alpha-5\beta\neq0$$

<u>סטטיסטי המבחן:</u>

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

חישובי עזר:

נציב את האומדים: $\hat{\alpha}-5\hat{\beta}=5.25-5\cdot0.96=0.45$

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

$$V(\hat{\alpha}-5\hat{\beta})=V(\hat{\alpha})+V(5\hat{\beta})-2\operatorname{cov}(\hat{\alpha},5\hat{\beta})=$$
$$V(\hat{\alpha})+5^2V(\hat{\beta})-2\cdot5\operatorname{cov}(\hat{\alpha},\hat{\beta})=$$
$$0.0625+25\cdot0.0144-10\cdot(-0.003)=0.4525$$

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

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

מסקנה: אין עדות לכך שה- $\alpha\neq5\beta$

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

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

להלן תוצאות האמידה של המשוואה הנ"ל:

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

(0.0678) (0.4)

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

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

<u>תשובה:</u>

יש לבדוק את ההשערה כי: $\beta_1 = \beta_2$

השערות:

$$H_0 : \beta_1 - \beta_2 = 0$$
$$H_1 : \beta_1 - \beta_2 \neq 0$$

חישוב סטטיסטי המבחן:

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

חישובי עזר:

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

$$S^2_{\hat{\beta}_1 - \hat{\beta}_2} = S^2_{\hat{\beta}_1} + S^2_{\hat{\beta}_2} - 2 \cdot S^2{}_{\hat{\beta}_1} \cdot S^2{}_{\hat{\beta}_2} =$$
$$= 0.0046 + 0.016 - 2 \cdot -0.009 = 0.038$$
$$S_{\hat{\beta}_1 - \hat{\beta}_2} = \sqrt{0.038} = 0.194$$

כלל הכרעה:

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

מסקנה: אין עדות לכך שהנטיות השוליות לצריכה הן שונות.

<u>תירגול מסכם</u>

? כלכלן בנה עבור מכבי ת"א מודל החוזה את השכר שיש לשלם לשחקן כדורסל לחוזה של שנה:

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

כאשר:

Y:שכר השחקן באלפי $

X1:מס' נקודות שקולע השחקן בממוצע למשחק

X2:מס' האסיסטים שיש לשחקן בממוצע למשחק

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

הכלכלן דגם 34 משחקים וקיבל את התוצאות הבאות:

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

(2.2) (3) (4.4) (-5)

****הערכים שבסוגריים הם ערכי t.**

התקבל בנוסף כי:

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

א. תנו פירוש למקדמי הרגרסיה.

ב. איזה מהמשתנים הב"ת רלוונטי למודל?

ג. בנו רב"ס למשתנים המובהקים.

ד. מייקל ג'ורדן הצטרף למכבי והוא דורש 2 מיליון $ לעונה.
ידוע כי מייקל קולע 45 נקודות בממוצע למשחק, מוסר 15 אסיסטים בממוצע למשחק ויושב 5 דקות בממוצע על הספסל. כמה צריך לשלם לו?

ה. לטענת שימעון מזרחי מס' הנקודות הממוצע שקולע שחקן למשחק צריך להשפיע פי 4 ממספר האסיסטים הממוצע שלו. האם הוא צודק?

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

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

K =הכנסה חודשית באלפי שקלים

Q =צריכה שנתית באלפי שקלים

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

$\hat{\beta} = -2$, $\hat{\alpha} = 4$, $t_{\hat{\beta}} = -7$, $t_{\hat{\alpha}} = 3$ $\text{cov}(\hat{\alpha}, \hat{\beta}) = -0.05$, $S_K = 1.5, S_Q = 0.05$ **נקודת הממוצעים הינה: (6.7,0.4)**

א. הכלכלן ביקש לבדוק את ההשערה כי הגמישות במודל יחידתית ושווה ל:1-

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

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

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

<u>פרק שביעי: R^2 -מדד לטיב הרגרסיה ומבחן F</u>

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

<u>מדד R^2 לטיב הרגרסיה</u>

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

מדד לפרופורציית השונות המוסברת:

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

ניתן לכתוב את הנוסחה גם כך:

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

נוסחה זו מתבססת על הנוסחה לפירוק השונות של קו הרגרסיה:

$$TSS = RSS + ESS$$

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

<u>תכונות R^2 :</u>

- **מדד R^2 (פרופורצית השונות המוסברת) נע בין 0 ל-1:**

$$0 \le R^2 \le 1$$

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

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

כאשר $R^2 = 0$ הכל טעות ואין שום הסבר במודל (זה קורה כשאין שום משתנה מסביר במודל ויש רק חותך: $y_i = \alpha + u_i$).

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

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

$$\overline{R}^2 = 1-\left(1-R^2\right)\frac{n-1}{n-k}$$

K=מס' הפרמטרים במודל (כולל החותך)

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

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

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

<u>זהויות שכדאי לדעת לגבי R^2 :</u>

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

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

במודלים : $\begin{aligned} y_i &= \alpha_1 + \beta_1 x_i + u_i \\ x_i &= \alpha_2 + \beta_2 y_i + \varepsilon_i \end{aligned}$ **מתקיים:**

1. **הם בעלי אותו** R^2
2. $R^2 = \beta_1 \cdot \beta_2$

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

.1 $y_i = \alpha + \beta x_{1_i} + u_i \qquad R^2 = 0.15$

.2 $y_i = \alpha + u_i$

.3 $y_i = \beta x_{1_i} + u_i$

.4 $y_i = \alpha + \beta\ x_{1_i} + \beta_2 x_{2_i} + \varepsilon_i$

.5 $y_i = \alpha + \beta_2 x_{2_i} + u_i \qquad R^2 = 0.20$

? על סמך מדגם של 100 תצפיות נאמדו המודלים הבאים:

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

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

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

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

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

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

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

בהתיחס לטענות החוקרים ניתן לומר:

1. **רק הטענה של חוקר א נכונה**

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

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

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

נכון/לא נכון/ לא ניתן לדעת

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

נכון/לא נכון/ לא ניתן לדעת

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

נכון/לא נכון/לא ניתן לדעת

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

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

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

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

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

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

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

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

כאשר y_i הינו סה"כ הוצאות משק בית i, x_{ji} הינו גילו של הילד j, ונתון כי $w_i = 2x_{3i} + x_{1i} - x_{2i}$.

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

? נתונות שתי המשוואות הבאות:

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

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

1. 0.09

2. 0.69

3. 0.3

4. 0.72

5. אף תשובה לא נכונה.

מבחן F

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

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

אז קיים הקשר בין המבחנים:

$F = t^2$

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

ביצוע מבחן F (מבחן WALD)

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

איך עושים זאת?

1) **אומדים את המודל המקורי. מודל זה נקרא המודל החופשי או המודל הלא-מוגבל, ובאנגלית Unrestricted. בתהליך האמידה של המודל הלא-מוגבל מקבלים את סכום ריבועי הסטיות של הטעויות, נסמן אותו ב- $\sum e_{i\ UR}^2$.**
2) **מגדירים את כל השוויונים של השערת האפס.**
3) **מציבים את השוויונים של השערת האפס במודל המקורי. באופן הזה הופכים אותו למודל כפוי (כופים עליו את השערת האפס), או מודל מוגבל, או באנגלית Restricted.**
4) **אומדים את המודל המוגבל. בתהליך האמידה של המודל המוגבל מקבלים את סכום ריבועי הסטיות של הטעויות, נסמן אותו ב- $\sum e_{i\ R}^2$.**

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

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

(כש- m מספר המגבלות)

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

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

<u>כלל הכרעה לדחיית H0:</u>

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

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

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

וכי $R^2 = 0.99$

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

א. מהי השערת האפס ?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

<u>המשך השאלה</u>

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

ג. חשב את הסטטיסטי של WALD.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

במדגם של 82 תצפיות התקבל:

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

א. בחנו את ההשערה כי :

$$H0: \beta_2 = 0$$
$$H1: \beta_2 \neq 0$$

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

ב. חשבו את $S_{\hat{\beta}_2}$

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

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

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

על מנת לבדוק את ההשערה שהנטיה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטיה השולית לצרוך מתוך ההון נאמדה גם המשוואה הבאה:

$C_i = \alpha + \beta_1 \cdot Y_i + u_i$ **כאשר:** Y_i=**סה"כ ההכנסה של משק בית t** $(W_i + P_i)$**.**

התקבל: $\sum e^2 = 54156$

א. בדקו את ההשערה

ב. חשבו את סטטיסטי t לבדיקת ההשערה

<u>מבחן F למובהקות המודל</u>

מבחן מובהקות העונה לשאלה: האם מודל הרגרסיה שלנו לניבוי משתנה תלוי מסויים על ידי המשתנים ב"ת, מובהק באוכלוסיה?

השערות:

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

המודל הלא מוגבל יהיה:

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

המודל המוגבל יהיה:

R: $Y_t = \alpha + u_t$

מאחר ו- $R_R^2 = 0$ ו- $m = k-1$:

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

<u>הערה:</u> בדיקת מובהקות המודל ברגרסיה מרובה ניתנת לביצוע רק על ידי מבחן F מאחר ויש לי יותר ממגבלה אחת בהשערת האפס.

לעומת זאת בדיקת מובהקות המודל ברגרסיה חד משתנית ניתנת לביצוע גם על ידי מבחן t שכן יש לי רק מגבלה אחת בהשערת האפס: $F = t^2$

נתון המודל :

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

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

האם המודל מובהק?

<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>מולטיקולינאריות חלקית-סתירה בין מבחן F למבחני t</u>

לעיתים נוצר מצב שבו המודל הוא מובהק אולם אף אחד מהשיפועים לא יוצא מובהק.   מצב זה מעיד על בעיה של מולטיקולינאריות חלקית במודל (מתאמים גבוהים בין המשתנים הב"ת).

כיוון שמולטיקולינאריות מגדילה את טעות התקן של האומדים היא מקטינה את סטטיסטי t ולכן גורמת לכך שנקבל את H0.

תירגול מסכם

?נאמדו חמשת המודלים הבאים על 70 תצפיות:

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

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

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

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

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

המשתנה המוסבר הוא הכנסה מעבודה (I)

והמשתנים המסבירים שבחנו הם מספר שנות הלימוד (scl) , מספר שעות עבודה (workh) וותק בעבודה (exp)

הערה: הניחו כי ערך F הקריטי הוא 4.

א. האם לשעות עבודה (workh) ישנה השפעה מובהקת על ההכנסה במשוואה 2?

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

ג. האם רגרסיה 2 מובהקת? (בחנו האם יש הסבר במודל 2), כיצד זה מסתדר עם תשובתכם ל-א ו-ב.

ד. האם השפעת הוותק יכול להיות 0.15?

ה. כלכלן נוסף הציע להריץ את המודל :

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

איזו השערה ניתן לבחון באמצעות מודל זה? כמה דרגות חופש יש לסטטיסטי שנקבל? בחנו אותה.

?על סמך מדגם של 40 תצפיות נאמדו המשוואות הבאות:

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

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

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

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

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

ב. בדקו את ההשערה כי רגרסיה 2 מובהקת

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

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

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

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

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

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

חשבו את האומדן לסטית התקן של המקדם X3 ברגרסיה 3

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

א. מהי המשוואה לאמידת המקדמים של המודל?

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

ג. מהן דרגות החופש במונה ובמכנה?

ד. רשמו את הנוסחה לחישוב סטטיסטי המבחן

? המודל הבא מתאר את פונקצית הייצור של מוצר P:

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

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

מהו המודל המוגבל המתאים לבדיקת ההשערה כי פונקציית הייצור מקיימת תק"ל (תשואה קבועה לגודל)?

פרק שמיני: 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:

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

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 :

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

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

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

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

תזכורת של חוקי לוגים:

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

בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:

החוקר טען כי יש לבדוק את הקשר בין שכר לותק ע"י שימוש בשכר נטו (NET) ולא בשכר ברוטו (SALARY). (קיים שיעור מס קבוע של 20%). המודל הוא:

$\ln\left(\mathrm{NET}_t\right) = \alpha' + \beta' \cdot \mathrm{EXP}_t + v_t$.

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

מבחנים לדוגמא

מבחן לדוגמא מס' 1

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

1) הנח כי הקשר באוכ' בין X ל-Y נתון על ידי המשוואה הבאה: $Y_t = \beta \cdot X_t + U_t$, כאשר כל ההנחות הקלאסיות מתקיימות.

נתון האומד: $\tilde{\beta} = \dfrac{\sum Y_t}{S_{XX}}$

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

ג. אומד זה יעיל פחות מאומד הריבועים הפחותים:
נכון/ לא נכון/ אי אפשר לדעת

ד. האם אומד זה הוא blue?

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

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

ז. שונות האומד (שחושבה בסעיף הקודם) הינה גדולה משונות המודל הנתון: **נכון/לא נכון/ אי אפשר לדעת**

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

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

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

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

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

א. ההשערה שניתן לבדוק באמצעות המשוואות הנתונות הינה:

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

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

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

4. לא ניתן לדעת.

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

1. 56

2. 57

3. 112

4. 74.66

3) כלכלן הציע את המודלים הבאים:

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

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

האם ניתן לאמוד את המודלים בשיטת OLS?

א. אין בעיה לאמוד את שני המודלים

ב. לא ניתן לאמוד את המודל הראשון בלבד

ג. לא ניתן לאמוד את המודל השני בלבד

ד. לא ניתן לאמוד את שני המודלים

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

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

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

מכאן ש:

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

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

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

ד. כל התשובות שגויות

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

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

(2) (1)

הערכים שבסוגריים הם סטיות התקן של המקדמים.

א. בדוק האם המודל מובהק

ב. בדוק האם מקדמי השיפוע מובהקים

ג. מה תוכל להסיק מסעיפים א ו- ב?

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

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

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

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

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

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

ו- y_i **הכנסת משק בית i**

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

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

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

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

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

א. חשב את נוסחת אומד הריבועים הפחותים ל- α **על ידי פיתרון בעיית המינימיזציה של סכום ריבועי הסטיות.**

ב. חשב את נוסחת שונותו של האומד.

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

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

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

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

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

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

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

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

ב. בדוק את ההשערות שלמשתנים X2 ו-X3 ביחד אין השפעה על Y במודל (1).

ג. בדוק בהסתמך על מודל (2) האם המשתנה X2 מובהק ברגרסיה (1).

ד. ברגרסיה (1) נתונים כעת אומדי הטעויות הסטנדרטיות (סטיות התקן) של מקדמי X2 ו-X3 0.5 ו-2.5 בהתאמה. בדוק עבור כל אחד מהמקדמים הנ"ל האם מובהק ומה אפשר ללמוד מרגרסיה (1).

ה. איזו השערה ניתן לבדוק תוך שימוש במשוואות (6) ו-(3)?