

מבוא לאקונומטריקה

החוג לכלכלה

תוכן העניינים:

2.....	הקדמה: תזכורת של סטטיסטיקה ומתמטיקה
7.....	מה זו אקונומטריקה?
13.....	אומדי הריבועים הפחותים (אר"פ) וההנחות הקלאסיות
26.....	מודלים לא ליניאריים
28.....	מבחני המובהקות וקריאת פלטים (של תוכנת SAS)
41.....	שינוי יחידות מדידה
43.....	רגרסיה רב משתנית
53.....	מבחנים לדוגמא
82.....	מבחני ספציפיקציה- מבחן LM (כופלי לגרנג')
88.....	טעויות ספציפיקציה
92.....	מולטיקוליניאריות
101.....	משתני דמי
123.....	הפרה של ההנחות הקלאסיות
124.....	הטרוסקדסטיות
137.....	מתאם סידרתי
154.....	שאלות חזרה למבחן מבוססות על פלטי STATA

הקדמה: תזכורת של סטטיסטיקה ומתמטיקה

הגדרות וסימונים

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

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

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

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (כמו למשל X_t או Y_t)

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

באקונומטריקה נעסוק בעיקר במשתנים מקריים.

לכל משתנה מקרי X_t יש תוחלת המייצגת את מרכז ההתפלגות.

התוחלת מסומנת μ_x או $E(X)$.

השונות מייצגת את מידת הפיזור של ההתפלגות.

השונות מסומנת σ_x^2 או $V(X)$.

סטית התקן היא השורש של השונות והיא מסומנת σ_x .

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

להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר

ביניהם (יחס ישר או יחס הפוך).

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

כאשר:

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

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

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

$$X, Y \text{ בלתי תלויים} \Leftrightarrow X, Y \text{ בלתי מתואמים}$$

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

מקדם המתאם מסומן η_{xy}

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

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

כאשר:

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

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

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

אמידה

פרמטר- ערך המשתנה הנחקר המתאר את כל האוכלוסיה

סטטיסטי/אומד- ערך המשתנה הנחקר המתאר את המדגם

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

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

נוסחאות וחוקים בסטיסטיקה

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

חוקי הסיגמה

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

(2) סכום של קבוע:

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

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

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

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

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

(5) יש לשים לב כי:

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

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

הגדרות ופיתוחים

(1) סכום הסטיות מהממוצע = 0:

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

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

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

(3) מונה של השונות המשותפת:

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

חוקי התוחלת

(1) תוחלת של קבוע=קבוע:

$$E(a) = a$$

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

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

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

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

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

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

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

(4) השפעת טרנספורמציה ליניארית על התוחלת:

$$E\left(a/\frac{1}{a} X \pm b\right) = a/\frac{1}{a} \cdot E(X) \pm b$$

חוקי השונות

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

שונות של סכום/הפרש = סכום/הפרש השונות

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

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

(2) עבור X ו-Y תלויים/מתואמים מתקיים:

שונות של סכום/הפרש \neq סכום/הפרש השונות

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

(3) שונות של קבוע=0:

$$V(a) = 0$$

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

(4) השפעת טרנספורמציה ליניארית על השונות:

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

הערה חשובה:

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

חוקי השונות המשותפת

(1) שונות משותפת בין משתנה לקבוע = 0:

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

(2) שונות משותפת של משתנים המוכפלים בקבוע:

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

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

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

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

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 רגרסה.

נתחיל בדוגמא:

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

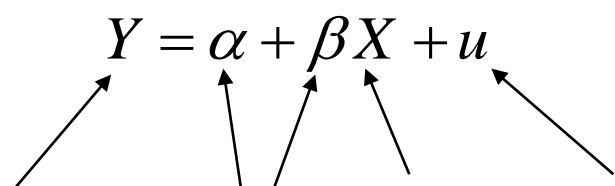
הוא הניח 2 הנחות מקדימות:

- רק גודל הדירה משפיע על מחיר הדירה באופן שיטתי. כל שאר הדברים המשפיעים על מחיר הדירה הם אקראיים ולא ניתנים לחיזוי.
- 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 הפרעה האקראית (לעיתים מכונה בבדיחות הדעת קריזה).

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

מודל:



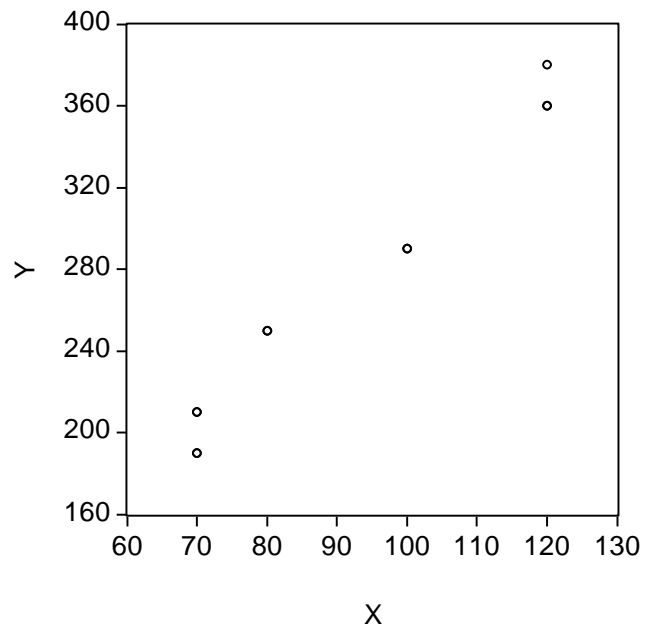
משתנה	פרמטרים:	משתנה	הפרעה אקראית
מוסבר	α - חותך	מסביר	"קריזה"
	β - שיפוע		

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

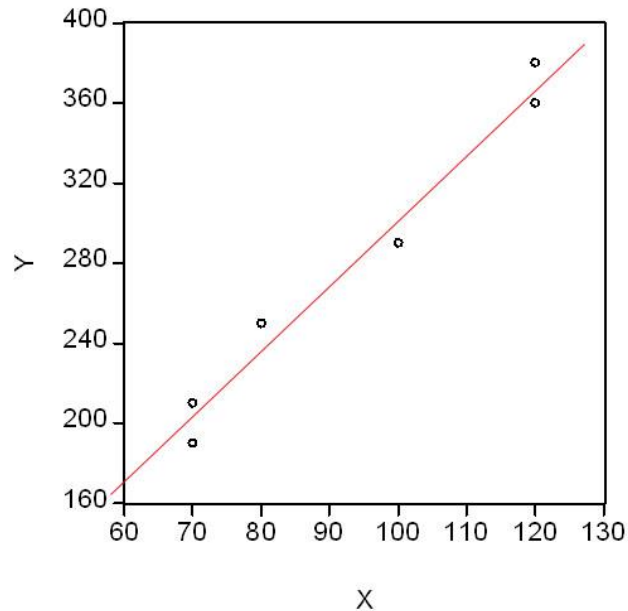
זהו המדגם של המתווך. במדגם יש 6 תצפיות. נוהגים להציג את המודל כאשר לכל משתנה נוסף אינדקס $Y_t = \alpha + \beta X_t + u_t$. האינדקס מייצג את מספר התצפית.

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 בשיטת הריבועים הפחותים (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 הפרעה 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.

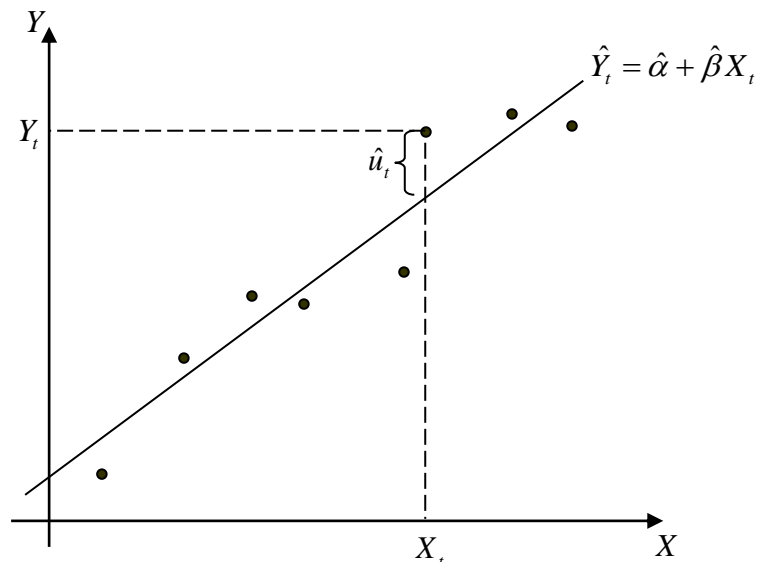
5) את α ו- β אי אפשר לדעת, ולכן אי אפשר לדעת מהו הקו האמיתי, וכן אי אפשר לדעת את u_t .

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

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

* 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

שלב התהליך האקונומטרי

הגדרת מודל

כל הגורמים המשפיעים באופן שיטתי חייבים להופיע במודל. כל השפעות 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.

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:

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

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

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

****הערה חשובה:** בתהליך הגזירה של פונקציית הריבועים הפחותים מתקבלות
 "המשוואות הנורמליות":

עבור המודל הקלאסי (עם חותך):

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

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

עבור מודל ללא חותך:

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

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

המשוואות הנורמליות 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.

$$u + \text{מסביר} = \beta + \alpha = \text{מוסבר}$$

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:

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

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

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

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

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

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

$$(4) \text{ ה-} X \text{ אינם משתנים מקריים.}$$

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:

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

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

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

הפיזור סביב קו הרגרסיה הוא אחיד.

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

"הקריזות" של תצפיות שונות אינן תלויות אחת בשניה.

הדבר תלוי בדגימה האקראית של התצפיות.

למשל, אם אנו בוחנים השפעה של ההכנסה על 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.

לסיכום:

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

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

(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

$$\text{ולשוונות} \Leftarrow \text{cov}(X_t, u_t) = 0$$

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

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

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

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

תכונות האומדים

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

1) לינאריות

אר"פ ניתנים להצגה כקומבינציה לינארית של Y_t .

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

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

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

אומד זה ניתן להצגה בצורה הבאה:

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

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

לפיכך מדובר באומד לינארי.

הוכחת לינאריות עבור המודל הקלאסי:

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

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

כלל אצבע-כיצד יודעים אם אומד הוא לינארי?

הלכה למעשה יש לבדוק האם מתקיימים 3 התנאים הבאים:

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

בחזקה או בשורש).

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

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

כזכור, x_t איננו משתנה מקרי).

לסיכום: אם בנוסחה של האומדן 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 שהאומדן לינארי.

? האם האומדנים הבאים הם אומדנים לינאריים?

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

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

2) תוסר השיה

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

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

זהו מושג תאורטי (ולא קונקרטי) שאומר כי ממוצע כל האומדנים ($\hat{\theta}$) של אינסוף המדגמים האפשריים בגודל מסוים שווה לפרמטר (θ). עבור מדגם מקרי אחד האומדן איננו שווה לפרמטר ($\hat{\theta} \neq \theta$) אבל על פני אינסוף המדגמים האפשריים, ממוצע האומדנים ($E(\hat{\theta})$) צריך להיות שווה לפרמטר (θ) כדי שהאומדן יהיה אח"ה.

כיצד יודעים אם אומדן הוא חסר הטיה?

בשלב הראשון יש לבצע עבודת הכנה –

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

**** יש לזכור כי:**

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

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

נשאר בתוך ה- \sum

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

דוגמא:

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

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

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

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

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

המשך הדוגמא שלעיל:

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

מסקנה: האומד חסר הטייה!

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:

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

הפרמטר האמיתי איבר המכיל את u_t

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

יעילות (3)

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

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

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

משפט גאוס מרקוב:

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

כלומר:

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

כיצד מחשבים שונות של אומד?

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

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

1. במודל ללא חותך

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

2. במודל עם חותך

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

4) עקיבות

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

$$\left(\begin{array}{l} \hat{\theta} \rightarrow \theta \\ T \rightarrow \infty \end{array} \right)$$

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

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

סיכום: השלבים להוכחת התכונות

1) הוכחת ליניאריות

2) הכנת האומד \Leftarrow להציב במקום Y_t את המודל האמיתי.

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

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

3) פיתוח האלגברה

4) חישוב תוחלת, שונות, עקיבות.

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

תרגול ממבחנים

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

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

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

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

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

נתון המודל $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות.
(יש לשים לב המודל ללא חותך)

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

- א. האומד $\tilde{\beta}$ הינו אומד מוטה ל- 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)

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

א. מהי התוחלת של $\tilde{\beta}$?

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

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

ד. מהי השונות האמיתית של האומדן $\frac{\sum X_t Y_t}{\sum X_t^2}$?

? 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 הטיות, אם נתון שהשונות של u_t אינה קבועה. נכון / לא נכון / אי אפשר לדעת

7. אומד חסר הטיות הוא אינו בהכרח גם אומד עקיב. נכון / לא נכון / אי אפשר לדעת

מודלים לא ליניאריים

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

בניגוד למודל הליניארי, שלושת המודלים האחרים מתארים קשרים שאינם

ליניאריים בין X ל-Y.

אולם אם נתייחס למשתנה המסביר או המוסבר בתוספת הלוג, ניתן יהיה לתאר

את הקשר בין המשתנה המסביר למשתנה המוסבר באופן ליניארי.

? 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 : (1600.01,12.311) עבור כל אחד מהמודלים.

טרנספורמציות לקו ישר:

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

? 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 $X=6$

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)

עונה על השאלה: איזה אחוז מהשונות של המשתנה התלוי (\bar{Y}) מוסבר על ידי קו

הרגרסיה, ההיגיון הכלכלי (X)?

מדד לפרופורציית השונות המוסברת.

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

- 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)).

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

עונה על השאלה: 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

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

מסויים על ידי k משתנים ב"ת, מובהק באוכלוסיה?

השערות:

$$H_0 : R^2 = 0$$

$$H_1 : R^2 > 0$$

סטטיסטי המבחן:

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

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

בפלט.

כלל החלטה ומסקנה:

נדחה את H0 כאשר:

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

שימוש בפלט: $PF < \alpha$

מסקנה: 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

מבחן מובהקות העונה לשאלה: האם משתנה מסביר מסויים רלוונטי למודל

(מובהק)?

השערות:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

סטטיסטי המבחן:

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

כלל הכרעה ומסקנה:

נדחה את H_0 אם:

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

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

מסקנה: יש/אין עדות לכך שהמשתנה ה- β המסויים מובהק באוכ' (ולכן רלוונטי

למודל).

הערות:

• 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'?

ההשערות:

$$H_0 : \beta = 0$$

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

כלל ההכרעה:

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'.

השערות לדוגמא:

$$H_0 : \beta = 2$$

$$H_1 : \beta \neq 2$$

סטטיסטי המבחן:

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

כלל הכרעה:

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 H_0 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

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

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:

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

$$F = t_{\hat{\beta}}^2$$

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

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

בשאלה לדוגמא ניתן לראות כי:

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

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

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

פלט ה-Covariance of Estimates

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

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

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

השערות:

$$H_0 : \gamma = 0$$

$$H_1 : \gamma \neq 0$$

סטטיסטי המבחן:

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

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:

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

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

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

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.

דוגמא:

נתון פלט האמידה של המודל $Y_i = \alpha + \beta X_i + u_i$

שלצורך אמידתו נאספו 240 תצפיות:

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0:	
		Estimate	Error	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 : $\gamma = \alpha - 5\beta$

ההשערות:

$$H_0 : \gamma = 0$$

$$H_1 : \gamma \neq 0$$

סטטיסטי המבחן:

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

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

חישוב השונות של $\hat{\gamma}$ ($S_{\hat{\gamma}}^2$):

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

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

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

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.
ב. שיעור התשואה בשכר לשנת ותק הוא:
ג. תחזית לוג השכר עבור אדם בעל 10 שנות ותק היא:
- נכון / לא נכון

עריכת תחזית וקריאת פלטים (של SPSS)

אמידה נקודתית

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

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

אמידת מרווח ל- $E(Y)$

כאשר אנו מתבקשים לאמוד את התחזית באוכלוסיה עבור X_0 מסוים, נחשב רווח

בר סמך לערך ממוצע של Y באוכ' עבור X_0 מסוים ($E(Y)$) ברמת סמך $1-\alpha$.

נוסחת הרב"ס:

$$\hat{Y} \pm t_{n-2, 1-\frac{\alpha}{2}} \hat{\sigma}_u \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$$
$$\hat{\sigma}_u = MSE = \frac{SSE}{n-2} \quad \sum (X_i - \bar{X})^2 = S_{xx} = (n-1)S_x^2$$

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

אמידת מרווח ל- Y

כאשר אנו מתבקשים לאמוד ערך בודד של Y באוכלוסיה עבור X_0 מסוים, נחשב

רווח בר סמך לערך בודד של Y באוכ' עבור X_0 מסוים (Y_0) ברמת סמך $1-\alpha$.

נוסחת הרב"ס:

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

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

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 מסוים.

התחזית מדויקת יותר (שונות התחזית קטנה יותר) כאשר:

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

דוגמא:

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

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

Model Summary ^b				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602 ^a	.362	.339	414.05503

a. Predictors: (Constant), number of students

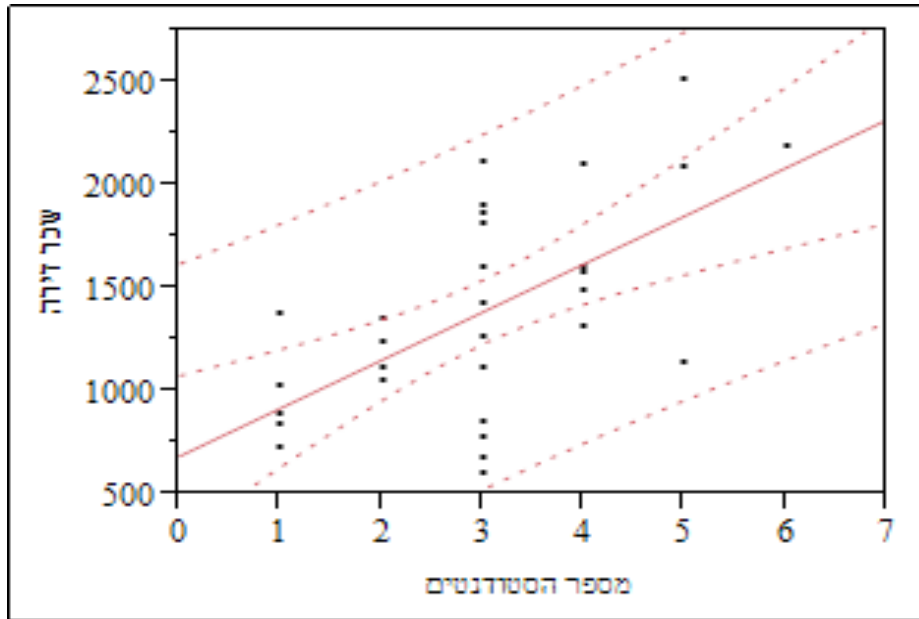
b. Dependent Variable: rent

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2726579.520	1	2726579.520	15.904	.000 ^a
	Residual	4800363.847	28	171441.566		
	Total	7526943.367	29			

a. Predictors: (Constant), number of students

b. Dependent Variable: rent

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



1. חשב אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.
2. אמוד את שכר הדירה הממוצע שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.
3. אמוד את שכר הדירה שישלם סטודנט יחיד החולק את הדירה עם שותף אחד בלבד, ברמת ביטחון של 95%.

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: $Y = \alpha' + \beta'(dX) + v$
$s_{\hat{\alpha}'} = ds_{\hat{\alpha}}$	$s_{\hat{\beta}'} = ds_{\hat{\beta}}$	$\hat{\alpha}' = d\hat{\alpha}$	$\hat{\beta}' = d\hat{\beta}$	הכפלת Y פי קבוע: $dY = \alpha' + \beta'X + v$

מסקנות מהטבלה:

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

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

? בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:

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

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

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

? בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:

החוקר טען כי הקשר בין השכר לותק אינו לינארי אלא ריבועי, ולכן אמד את

$$\ln(\text{SALARY}) = \alpha + \beta(\text{EXP}^2) + u$$

המודל הבא: כיצד ישתנו האומדים וסטיות התקן שלהם?

רגרסיה רב משתנית

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.

המודל הקלאסי:

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

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

מבחני המובהקות

מבחן F למובהקות המודל

השערות:

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

$$H_1 : \text{OTHERWISE}$$

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- β טות

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 בבלט.
 ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05.
 ג. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.

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

לעיתים אנו מתבקשים להכריע בסוגיה האם כדאי לנו להוסיף למודל משתנה ב"ת מסויים. במקרה זה נשווה את פרופורציית השונות המוסברת המתוקנת \bar{R}^2 בין המודל ללא המשתנה המסביר לבין המודל עם המשתנה המסביר שהוספנו. להזכירכם \bar{R}^2 בניגוד ל- R^2 לוקח בחשבון לא רק את הרווח שבהקטנת החלק הלא מוסבר בהוספת משתנים ב"ת למודל אלא גם את ההפסד שבירידה בדרגות החופש. לכן בניגוד ל- R^2 הוא יכול לקטון בהוספת משתנים ב"ת למודל.

לפי חוק חיטובסקי-בהוספת משתנה מסביר אחד בלבד למודל ה- \bar{R}^2 יעלה אך ורק

$$\text{אם } |t_{\hat{\beta}}| > 1$$

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

(מובהק).

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

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

למודל על פי מבחן t.

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

$\bar{R}^2 = 0.266$. הוסף המשתנה היקף המשרה. במבחן למובהקות המשתנה הנוסף התקבל: $t_{\hat{\beta}} = 0.456$. האם ערך \bar{R}^2 יעלה/ירד/לא ישתנה בהוספת המשתנה הנוסף למודל?

****הערה חשובה:** 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 השויונים בהשערת האפס, m , אפשר לחשב את הסטטיסטי:

$$WALD_{stat} = \frac{ESS_R - ESS_U}{\frac{m}{\frac{ESS_U}{T-k-1}}}$$

(המודל המקורי עם חותך).

m - דרגות החופש של המונה
 $T - k - 1$ - דרגות החופש של המכנה

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

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

כלל הכרעה לדחיית H_0 :

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

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

? נאמד המודל $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ 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	Standard	T for H0:	
		Estimate	Error	Parameter=0	Prob> T
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

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

א. מהי השערת האפס?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

המשך השאלה

להלן אמידת המודל המוגבל:

Analysis of Variance

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

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

Parameter Estimates

Variable	DF	Parameter	Standard	T for H0:	
		Estimate	Error	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 W.ALD.

ד. כמה דרגות חופש יש במונה וכמה במכנה?

ה. האם דוחים או מקבלים את השערת האפס?

מקרים פרטיים של מבחן WALD

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

השערות:

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

$$H_1 : \text{OTHERWISE}$$

אם נבצע זאת במבחן WALD :

המודל הלא מוגבל יהיה:

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

המודל המוגבל יהיה:

$$\text{R: } Y_t = \alpha + u_t$$

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:

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

רוצים לבדוק את מובהקות β_2

השערות:

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0$$

ניתן לבדוק זאת גם במבחן WALD :

$$\text{U: } Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u$$

$$\text{R: } Y_t = \alpha + \beta_1 X_1 + \beta_3 X_3 + \dots + \beta_k X_k + u$$

במקרה זה :

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

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

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

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

במקרה זה :

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

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

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

לדוגמא:

על מנת לאמוד את פונקציית 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:

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

התקבל: $ESS=0.4566$

בידקו את ההשערה בשתי דרכים.

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

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

תרגול מסכם

חוקר אמד את התצרוכת של 500 משקי בית כפונקציה של הכנסה שלהן לפי

$$\text{המשוואה: } EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$$

$EXPENSE_t$ - 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 השערה כי על כל 1000 ש"ח

נוספים בהכנסה צורך הפרט 700 ש"ח, כנגד השערה כי הוא צורך יותר

מ-700 ש"ח. נסח את השערת האפס ואת השערה האלטרנטיבית.

(6) מהו הסטטיסטי t לבדיקת השערה?

7) מהו הסטטיסטי WALS 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 לפונקציית 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 WALS לבדיקת ההשערה?

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

11) מהי השערת האפס לבדיקה זו?

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

בטא את Z_t , ו- W_t באמצעות המשתנים המקוריים.

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-0.4. מהן השערות לבדיקת הטענה?
- (5) מהי הרגרסיה המוגבלת למבחן WALT תחת H_0 ?
- (6) מהו הסטטיסטי של WALT 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 t לבדיקת המובהקות של α 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	Standard	T for H0:	
		Estimate	Error	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 השערה? (נתון כי $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.25$)

13) האם ניתן לדחות את השערת האפס?

שאלה 2 (24 נקודות)

- 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 H_0 ברמת מובהקות מסויימת, אזי דוחים H_0 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

שאלה 3 (14 נקודות)

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 $\tilde{\beta}$?

שאלה 4 (7 נקודות)

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

נתון מודל ללא חותך $Y_t = \beta X_t + u_t$, ונתון האומד

- (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 $\hat{\beta}$? נכון / לא נכון / לא ניתן לדעת
- (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: $SALARY_t = \alpha + \beta \cdot HOURS_t + u_t$.

הסטייה 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 WALS 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 WALT 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

WALT לבדיקת טענת החוקר?

11) מהם הנתונים הנדרשים לחישוב הסטטיסטי t ?

החוקר רוצה לבדוק את הטענה כי השפעת ההשכלה על השכר גדולה פי 8 מהשפעת הגיל על השכר.

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEPT	1	-1995.0275	331.7857	-6.013	0.0001
HOURS	1	36.408461	4.710021	7.730	0.0001
AGE	1	13.674254	3.816426	3.583	0.0004
SCL	1	109.93799	10.63745	10.335	0.0001

12) הנתונים בפלט אינם מספיקים לבדיקת ההשערה לפי מבחן t. מהו הנתון

החסר? באיזה פלט של SAS ניתן למצוא אותו?

13) בהנחה שנתון זה הוא 8.3969, חשב את הסטטיסטי t לבדיקת הטענה. מהי

מסקנתך לגבי נכונות הטענה?

14) אם תרצה לבדוק את הטענה לפי מבחן WALD, יהיה המודל המוגבל:

$$Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + \gamma_2 \cdot Z_2 + v$$

$$Z_0 = \underline{\hspace{2cm}}$$

$$Z_1 = \underline{\hspace{2cm}}$$

$$Z_2 = \underline{\hspace{2cm}}$$

15) אם יש מספיק נתונים, חשב את הסטטיסטי של WALD לבדיקת הטענה?

שאלה 2 (30 נקודות)

נתון המודל $Y_t = \alpha + \beta \cdot X_t + u_t$. ידוע כי כל ההנחות הקלאסיות מתקיימות.

$$\text{נתון האומדן: } \tilde{\beta} = \frac{S_{XY}}{\sum_{t=1}^T X_t^2}$$

1) אומדן זה הוא הפתרון של המשוואות הנורמליות $\sum_{t=1}^T \hat{u}_t X_t = 0$, $\sum_{t=1}^T \hat{u}_t = 0$

נכון / לא נכון / לא ניתן לדעת

2) התוחלת של $\tilde{\beta}$ היא:

א. β

ב. $\frac{\beta \cdot \sum_{t=1}^T X_t}{S_{XX}}$

ג. $\frac{\beta \cdot \sum_{t=1}^T X_t^2}{S_{XX}}$

ד. $\frac{\beta \cdot S_{XX}}{\sum_{t=1}^T X_t^2}$

ה. כל התשובות אינן נכונות.

3) הטענה כי $E(\tilde{\beta}) < \beta$:

א. תמיד נכונה

ב. אינה נכונה

ג. נכונה אם ורק אם $\bar{X} > 0$

ד. נכונה אם ורק אם $\bar{X} \neq 0$

ה. כל התשובות אינן נכונות

4) אם $\bar{X} = 0$ אז השונות של $\tilde{\beta}$ היא:

א.
$$\frac{\sigma^2}{\left(\sum_{t=1}^T X_t\right)^2}$$

ב.
$$\frac{\sigma^2}{S_{XX}}$$

ג.
$$\frac{\sigma^2 \sum_{t=1}^T X_t^2}{S_{XX}}$$

ד.
$$\frac{\sigma^2}{S_{XX}}$$

ה. כל התשובות אינן נכונות

5) אם $\bar{X} = 0$, אז $\tilde{\beta}$ הינו האומד הלינארי חסר ההטיה בעל השונות הקטנה ביותר.

נכון / לא נכון / לא ניתן לדעת

שאלה 3 (10 נקודות)

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

נכון / לא נכון / לא ניתן לדעת. $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$ (אר"פ) הינו אומד יעיל יותר.

מבחן לדוגמא מס' 3

שאלה מס' 1 (50 נקודות)

על מנת לאמוד את פונקציית הייצור נאספו נתונים על 150 פירמות בשנת 2007 ונאמדה

$$\ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + U_t \quad (1) \quad \text{המשוואה הבאה: (1)}$$

כאשר:

$$\ln(Y)_t = \text{תפוקה שנתית באלפי ש"ח בלוגים}$$

$$\ln(L)_t = \text{מספר העובדים בלוגים}$$

$$U_t = \text{הטעות המיקרית המקיימת את כל ההנחות הקלאסיות.}$$

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

H0: _____ הן:

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 WALT 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

$$(2) \ln(Y)_t = \alpha + \beta_1 \cdot \ln(L)_t + \beta_2 \cdot \ln(K)_t + \beta_3 \cdot \ln(PY)_t + U_t$$

כאשר:

$$\ln(K)_t = \text{מלאי ההון של הפירמה באלפי ש"ח בלוגים}$$

$$\ln(PY)_t = \text{הוצאות למחקר ופיתוח באלפי ש"ח בלוגים}$$

משוואה מס' (2) נאמדה בפלט מס' 2

(2) ז. ההשערות לבדיקת הטענה הינן:

$$H_0: \underline{\hspace{10cm}}$$

$$H_1: \underline{\hspace{10cm}}$$

(2) ח. הסטטיסטי של WALT לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

2. ניתן לחישוב וערכו: _____

(2) ט. הסטטיסטי של t לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים
2. לא ניתן לחשב סטטיסטי t לטענה מסוג זה
3. ניתן לחישוב וערכו: _____.

החוקרת טענה כי השפעת הוצאות למחקר ופיתוח אינה מובהקת ולכן יש לאמוד את המשוואה הבאה:

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

משוואה מס' (3) נאמדה בפלט מס' 3

(3) י. ההשערות לבדיקת הטענה הינו:

$$H_0: \underline{\hspace{10cm}}$$

$$H_1: \underline{\hspace{10cm}}$$

(2) יא. הסטטיסטי של WALT 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:

$$H_0: \underline{\hspace{10cm}}$$

(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 WALT

$$Z_0 = \gamma_0 + \gamma_1 \cdot Z_1 + V \quad \text{הינה:}$$

$$Z_0 = \underline{\hspace{4cm}} \quad Z_1 = \underline{\hspace{4cm}}: \text{כאשר}$$

(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 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

שאלה מס' 2 (33 נקודות- כל סעיף 3 נקודות)

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 באוכלוסייה:

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 H_0 למובהקות ה- β :
נכון/לא נכון/אי אפשר לדעת
6. $R^2 > \bar{R}^2$ 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 H_0 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

שאלה מס' 3 (14 נק'):

נתון המודל: $Y_t = \beta X_t + U_t$

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

(5)א. $\tilde{\beta}$ הינו אומדן חסר הטייה ל- β : נכון/לא נכון/אי אפשר לדעת

(3)ב. שונותו של האומדן: _____

(2)ג. על סמך משפט גאוס מרקוב ניתן להסיק כי אר"פ הינו אומדן 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 אומדן 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> T
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> T
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

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> T
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:

$$CREDIT_t = \text{סך הפעילות בכרטיסי אשראי בש"ח}$$

$$SAVINGS_t = \text{סך הפעילות בחשבונות חיסכון בש"ח}$$

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 בכרטיסי אשראי שלהם אפילו

בשקל אחד, הפעילות בחשבונות חיסכון 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 WALT 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?

תשובה:

לפי פלט Parameter Estimates:

$$p(0.51 \leq \beta \leq 0.623) = 0.95$$

(3) ז. ברמת ביטחון 95% מהו האומד לתוחלת פעילות בכרטיסי אשראי עבור סך

פעילות בחשבונות חיסכון של 50,000 ₪?

(5) ח. אם פעילות כרטיסי האשראי של כל לקוח תגדל ב- 1000 ש"ח:

1. האומד של α ישתנה: נכון/לא נכון/ אי אפשר לדעת

2. האומד של β ירד: נכון/לא נכון/ אי אפשר לדעת

3. סטטיסטי F לבדיקת מובהקות המודל לא ישתנה:

נכון/לא נכון/ אי אפשר לדעת

נטען שסה"כ פעילות הלקוח בחשבונות חיסכון איננו המשתנה המשפיע על הפעילות בכרטיסי האשראי, אלא הרכב החסכונות. לשם כך נאמדה המשוואה הבאה:

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

כאשר:

$PIKADON1_t$ = סה"כ הפקדה לפקדונות יומיים בש"ח.

$PIKADON2_t$ = סה"כ הפקדה לפקדונות חודשיים בש"ח.

משוואה (2) נאמדה בפלט מס' 2.

(3) ח. השערת האפס לבדיקת הטענה

HO: _____ הינה:

(4) ט. הסטטיסטי של WALT לבדיקת הטענה:

1. לא ניתן לחשבו בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו: _____

(3) י. 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) יא. השערת האפס לבדיקת הטענה

HO: _____ הינה:

(4) יב. הסטטיסטי t לבדיקת הטענה הינו:

1. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: _____

(2) יג. PVALUE של סטטיסטי t מהסעיף הקודם:

1. לא ניתן לחשב את הסטטיסטי t בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: _____

(4) יד. הרגרסיה המוגבלת כאשר HO נכונה למבחן WALT הינה:

$$D_0 = \gamma_0 + \gamma_1 \cdot D_1 + \gamma_2 \cdot D_2 + v$$

D_0 : _____

D_1 : _____ כאשר:

D_2 : _____

(4) טו. על פי משוואה מס' 2, כל שקל שיועבר לפיקדון הראשון יוסיף כ-0.07552

נח לסה"כ הפעילות בכרטיסי אשראי: נכון/לא נכון/ אי אפשר לדעת

שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)

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

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

1. אם המודל מובהק אזי שיפוע הרגרסיה מובהק בהכרח:

נכון/לא נכון/אי אפשר לדעת

2. הגמישות במודל חצי לוגריתמי היא קבועה: נכון/לא נכון/אי אפשר לדעת

3. אם X_2 מהווה קומבינציה ליניארית של X_1 לא ניתן לאמוד את הרגרסיה

המרובה:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + U \text{ : נכון/לא נכון/ אי אפשר לדעת}$$

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_i גדול יותר כך קשה יותר לדחות את H_0 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 (שוונות קבועה) לא מתקיימת, אומדי הריבועים הפחותים אינם חסרי הטיה:
נכון/לא נכון/ אי אפשר לדעת

10. אם דחינו את H_0 לבדיקת הטענה כי שיפוע הרגרסיה הוא שלילי בוודאי שמודל הרגרסיה הוא מובהק:
נכון/לא נכון/ אי אפשר לדעת

שאלה 3 (12 נקודות)

נתון המודל: $Y_i = \beta \cdot X_i + U_i$, כאשר כל ההנחות הקלאסיות מתקיימות.

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

א. $E(\tilde{\beta}) =$ _____.

ב. על סמך משפט גאוס מרקוב אומד זה יעיל פחות מאומד הריבועים הפחותים:
נכון/ לא נכון/ אי אפשר לדעת

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

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

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> T	95% Confidence	
INTERCEP							
P	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> T
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 (P) במשק לכמות הכסף (M), נאספו נתונים חודשיים בשנים 86-94 (סה"כ 105 תצפיות) ונאמדה המשוואה הבאה:

$$M_t = e^\alpha + p^\beta + e^u \quad (1)$$

כאשר:

m = כמות הכסף בחודש (מזומנים+עו"ש)

p = מדד המחירים לצרכן במשק

U_t - סטייה מקרית המקיימת את כל ההנחות הקלאסיות.

משוואה מס' (1) נאמדה בפלט מס' 1.

(2) א. כתבי את המשוואה בצורה ליניארית בעזרת הטרנספורמציה המתאימה.

(2) ב. האומדן למשוואה (1) הינו: _____

(2) ג. המשמעות הכלכלית של β

היא: _____

(4) ד. גבולות רווח-סמך ברמת סמך של 95% עבור β הינם:

גבול תחתון: _____

גבול עליון: _____

(3) ה. ערך t לחישוב מובהקות ה- β הינו:

1. לא ניתן לחשב ערך זה בעזרת הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: _____

(2) ו. אם נגדיל את מדד המחירים לצרכן ביחידה אחת, כמות הכסף במשק

תגדל ב:

1. 71.7233 .

2. 1.69267

3. 169.267

4. 1.69267%

5. 169.267%

הועלתה הטענה שתוספת של אחוז אחד במדד המחירים לצרכן תגדיל את כמות הכסף במשק ביותר מאחוז אחד.

(3) ז. השערת האפס לבדיקת

הטענה: H_0 : _____

(3) ח. סטטיסטי t לבדיקת הטענה הינו:

1. לא ניתן לחשבו באמצעות הנתונים הקיימים.

2. ניתן לחשבו וערכו הוא: _____

(3) ט. על פי התשובות לסעיף הקודם 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 מדד המחירים לצרכן במשק יוכפל פי שניים:

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:

$$\text{LN}(M)_t = \alpha + \beta_1 \cdot \text{LN}(P)_t + \beta_2 \cdot \text{LN}(Y)_t + U_t \quad (2)$$

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 י. 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 הקיימים.

2. ניתן לחשבו וערכו הוא: _____

(3) טו. הרגרסיה המוגבלת כאשר HO נכונה למבחן WALD הינה:

D_0 : _____

כאשר: D_1 : _____

(2) טז. איזה מבין המודלים המוצעים במשוואות 1 ו-2 עדיף?

משוואה 1/משוואה 2/אין הבדל בין המודלים

(2)ז. אם משתנה רמת המחירים במשק היה מובהק במשוואה מס' 1, הוא

יהיה מובהק בהכרח גם במשוואה מס' 2 : נכון/לא נכון/אי אפשר לדעת

שאלה מס' 2 (30 נקודות- כל שאלה 3 נקודות)

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

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

1. $\bar{R}^2 < R^2$ כאשר $k > 0$: 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 \bar{R}^2 יעלה בהכרח.
 נכון/לא נכון/אי אפשר לדעת.

9. מבחן WALS הוא מקרה פרטי של מבחן F למובהקות המודל:
 נכון/לא נכון/אי אפשר לדעת.

10. שיטת הריבועים הפחותים מביאה למקסימום את \bar{R}^2 :
 נכון/לא נכון/אי אפשר לדעת.

שאלה 3 (17 נקודות)

נתון המודל: $Y_t = \alpha + \beta X_t + U_t$

נתון כי אר"פ למודל זה הינו: $\hat{\beta} = S_{xy} / S_{xx}$

א. הוכיחי כי $\hat{\beta}$ אומד ליניארי וחסר הטיה של β .

ב. חשבי את $VAR(\hat{\beta})$.

ג. נתון האומד: $\tilde{\beta} = \frac{\sum X_t Y_t}{\sum X_t^2}$

הוכיחי כי $\tilde{\beta}$ אומד ליניארי אך איננו חסר הטיה ל- β .

ד. מהם התנאים בהם מתקיים: $E(\tilde{\beta}) = \beta$?

פלט מס' 1 - משוואה מס' (1)

Dependent Variable: lnm

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	1				<0.0001
Error	103				
C Total	104	44.91976			
Root MSE	0.09251		R-square	0.9804	
Dep Mean	8.53854		Adj R-sq	0.9802	
C. V.	1.08344				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
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> T
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		lnp	lny
INTERCEP	0.35687	0.025884	-0.09762
lnp	0.02588	0.002843	-0.00792
lny	-0.09762	-0.00792	0.02745

מבחני ספציפיקציה- מבחן LM (כופלי לגרנג')

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

לדוגמא:

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

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

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

השערות:

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

$$H_1: \text{OTHERWISE}$$

המודל המוגבל (RESTRICTED) הינו:

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

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

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

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

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

הרציונאל של המבחן הוא:

אחת מן ההנחות הקלאסיות של מודל הרגרסיה היא כי המשתנים ה"ב"ת אינם

$$\text{מתואמים עם טעות האמידה: } \text{cov}(X_t, u_t) = 0$$

המשתנים המצויים בתוך המודל (השניים הראשונים), לא מתואמים עם טעות האמידה ($R_{\hat{u}_{12}}^2 = 0$).

לגבי שני המשתנים שיש לנו ספק לגביהם:

אם הם יהיו מתואמים עם טעות האמידה ($R_{\hat{u}_{34}}^2 \neq 0$) – זו אינדיקציה ש"שכחנו" להוסיף למודל וכי הם רלוונטיים (דוחים H_0).

אם הם אינם מתואמים עם טעות האמידה ($R_{\hat{u}_{34}}^2 = 0$), המודל מושלם כמו שהוא (מקבלים את H_0).

• הסבר באמצעות מעגלי וואן:

חישוב הסטטיסטי:

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

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

כלל הכרעה של H_0 :

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

m- מס' ההגבלות ב- H_0 .

UNRESTRICTED

Dependent variable: Y

Analysis of Variance

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

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

Parameter Estimates

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

RESTRICTED

Dependent variable: Y

Analysis of Variance

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

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

Parameter Estimates

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

רגרסיית עזר

Dependent variable :RES

Analysis of Variance

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

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

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	5.9608892	0.776604	7.675584	0.0000
X1	1	1.2077723	0.978845	1.233875	0.8455
X2	1	0.4840697	0.886754	0.545889	0.9976
X3	1	-5.029101	0.073149	-68.75146	0.0000
X4	1	8.974106	0.029075	308.6485	0.0000

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

(2) איזה מן המשתנים הנוספים רלוונטי למודל?

****שימו לב כי:**

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

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

הקשר בין משוואות ה-U, ה-R והעזר:

א. $U=R+עזר$

ב. $ESS_U = ESS_y$

ג. $ESS_R = TSS_y$

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

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

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

טעויות ספציפיקציה

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

ישנם שני סוגים של טעויות ספציפיקציה:

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

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

1) הוספת משתנה לא רלוונטי

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

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

המודל האמיתי:

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

המודל הנאמד (הטעותי):

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

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

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

מסקנה:

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

**שימו לב כי לעיתים נרצה להשאיר משתנה שאיננו מובהק במודל. מדובר במקרים בהם המשתנה הנוסף מעלה את $AdjR^2$ למרות שאיננו מובהק, כיוון ש:

$$1 < t_{\hat{\beta}} \text{ (לפי חוק חיטובסקי).}$$

(2) השמטת משתנה רלוונטי

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

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

$$Y_t = \alpha + \beta_1 x_{1t} + \varepsilon \quad \text{(הטעות):}$$

במקרה זה עלולה להיווצר בעיה בהשמטת משתנה x_2 .

א. בעיה באמידה של β_1 :

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

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

• הדגמה באמצעות מעגלי ואן:

**שימו לב- אם מקדם השיפוע של x_1 שונה במודל הנאמד לעומת במודל האמיתי

ניתן להסיק מכך כי $\text{cov}_{12} \neq 0$

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

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

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

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

אם $S_{12} > 0$ ו- β_2 הם שווי סימן-ההטיה חיובית-האומדן ל- β_1 יהיה מוטה כלפי מעלה.

אם $S_{12} < 0$ ו- β_2 הם שוני סימן-ההטיה שלילית-האומדן ל- β_1 יהיה מוטה כלפי מטה.

ב. בעיה באמידה של α :

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

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

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

ג. בעיה באמידת השונות של β_1 ו- α :

אומדן השונות יהיה מוטה תמיד כלפי מעלה כי הוא איננו לוקח בחשבון את החלק

ש- x_2 מסביר ב- Y :

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

מסקנה:

בדיקת ההשערות על הפרמטרים של המודל הנאמד איננה תקפה.

סיכום:

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

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

מולטיקוליניאריות

מולטיקוליניאריות היא תופעה סטטיסטית בעייתית המתייחסת למתאם בין המשתנים המסבירים במודל.

נבחין בין מולטיקוליניאריות מלאה לחלקית.

(1) מולטיקוליניאריות מלאה

מתאם מלא בין המשתנים המסבירים במודל.

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

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

**שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל

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

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 ה"ב"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:

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

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

$$\sum \hat{u}_t x_{1t} = 0 : \hat{\beta}_1$$

$$\sum \hat{u}_t x_{2t} = 0 : \hat{\beta}_2$$

$$\sum \hat{u}_t (bx_2) = 0$$

מכיוון ש: $x_1 = bx_2$ (שער הדולר = b) אז גזירת $\hat{\beta}_1$: $b \sum \hat{u}_t x_2 = 0$ והמשוואה
 $0 = 0$

מתבטלת.

פיתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

(2) מולטיקוליניאריות חלקית

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

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

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

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

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

זיהוי מולטיקוליניאריות חלקית:

(1) כאשר קיימת סתירה בין התוצאה במבחן F למובהקות המודל (המודל

מובהק) לבין מבחני t למובהקות השיפועים (אף אחד מן השיפועים איננו מובהק).

הסתירה נוצרת כתוצאה מהגדלת השונות של כל אחד מהשיפועים בשל המתאם הגבוה בין ה"ב", באופן שלא מאפשר לדחות את השערת האפס למובהקות

$$s_{\hat{\beta}_1}^2 = \frac{MSE}{SSX_1(1-r_{12})}$$

השיפועים:

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

(2) רגישות לספציפיקציה - הורדת משתנה ב"ת שאיננו מובהק תהפוך

משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקוליניאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים ה"ב"ת האחרים.

(3) סימנים הפוכים - כאשר השיפועים של המשתנים ה"ב"ת מקבלים סימנים

הפוכים מכיוון ההשפעה שלהם על המשתנה התלוי. אם למשל, x_1 משפיע חיובית על Y ואילו x_2 משפיע שלילית על Y אבל הם יופיעו במשוואת הרגרסיה עם סימנים הפוכים ($\hat{\beta}_1$ שלילית ואילו $\hat{\beta}_2$ חיובית), יש לחשוד שקיימת בעיה של מולטי קולינאריות במודל (מתאם גבוה מאוד בין המשתנים ה"ב"ת ש"מבלבל" את הסימנים שלהם).

השלכות של מולטיקוליניאריות חלקית:

מולטיקוליניאריות חלקית איננה פוגעת בתכונות של אר"פ (הם נותרים ליניאריים, חסרי הטיה, יעילים ועקיבים) ולא באומד השונות של האומדים (שנותר חסר הטיה) כך שבדיקת השערות תוך שימוש באומדים הללו תהיה תקפה (זאת בניגוד למולטיקוליניאריות מלאה).

במובן הזה, בעיה של מולטיקוליניאריות חלקית דומה לבעיה של הוספת משתנה ב"ת שאיננו רלוונטי.

פיתרונות למולטיקוליניאריות חלקית:

(1) ברוב המקרים נשקול להוריד את אחד המשתנים.

יחד עם זאת, כאשר המובהקות של המשתנים היא גבולית $1 < t_{\hat{\beta}} < 2$, יתכן ונותיר את שניהם בתוך המודל כיוון שבסך הכול יש עליה ב- $AdjR^2$ (לפי חוק חיטובסקי).

2) ניתן לעיתים לאחד את שני המשתנים למשתנה אחד. למשל, בדוגמא של ניבוי ציוני ה-B.A. על ידי ציוני הפסיכומטרי והבגרות, יש לשקול להגדיר משתנה חדש שיהווה שקלול ציוני הפסיכומטרי והבגרות ולהשתמש בו לניבוי ה-B.A.

שלב בדיקת ההשערות:

- 1) מבצעים מבחן F לבדיקת מובהקות המודל.
- 2) במידה והמודל מובהק, מבצעים מבחן t למובהקות כל אחד מהשיפועים.
- 3) ביצוע מבחן WALT לבדיקת כל השיפועים שלא יצאו מובהקים:
 - א. אם מקבלים את H_0 : אין סתירה עם מבחני t- אין בעיה של מולטיקוליניאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.
 - ב. אם דוחים את H_0 : יש סתירה עם מבחני t- קיימת בעיה של מולטיקוליניאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALT בלעדיו, עד שמזהים את המשתנה/משתנים שיש להוריד מהמודל.

סיכום ותרגול של טעויות ספציפיקציה ומולטי קולינאריות

פתרון	השלכות	זיהוי	הגדרה	הבעיה

<p>הורדת* המשתנה</p>	<p>ניתן לבצע בדיקת השערות</p> <p>אר"פ ($\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$) חסרי הטיה</p> <p>אומדי השונות ($S_{\hat{\alpha}}^2, S_{\hat{\beta}_1}^2, S_{\hat{\beta}_2}^2$) חסרי הטיה</p>	<p>קבלת H0 במבחן t</p> <p>למובהקות β_2</p>	<p>המודל האמיתי:</p> $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$ <p>המודל הנאמד (הטעותי):</p> $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$	<p>הוספת משתנה</p> <p>לא רלוונטי</p>												
<p>הוספת המשתנה</p>	<p>לא ניתן לבצע בדיקת השערות</p> <table border="1" data-bbox="119 504 710 1041"> <tr> <td data-bbox="119 504 263 593">אומד לשונות הפרמטרים</td> <td data-bbox="263 504 367 593">אומד ל-α</td> <td data-bbox="367 504 598 593">אומד ל-β_1</td> <td data-bbox="598 504 710 593">בהיעדר x_2:</td> </tr> <tr> <td data-bbox="119 593 263 795">מוטה חיובית</td> <td data-bbox="263 593 367 795">מוטה אלא-אם- $\bar{x}_2 = 0$</td> <td data-bbox="367 593 598 795">חסר הטיה</td> <td data-bbox="598 593 710 795">$S_{12} = 0$</td> </tr> <tr> <td data-bbox="119 795 263 1041">מוטה חיובית</td> <td data-bbox="263 795 367 1041">מוטה</td> <td data-bbox="367 795 598 1041">מוטה חיובית: S_{12} ו-β_2 שווי סימן מוטה שלילית: S_{12} ו-β_2 מנוגדי סימן</td> <td data-bbox="598 795 710 1041">$S_{12} \neq 0$</td> </tr> </table>	אומד לשונות הפרמטרים	אומד ל- α	אומד ל- β_1	בהיעדר x_2 :	מוטה חיובית	מוטה אלא-אם- $\bar{x}_2 = 0$	חסר הטיה	$S_{12} = 0$	מוטה חיובית	מוטה	מוטה חיובית: S_{12} ו- β_2 שווי סימן מוטה שלילית: S_{12} ו- β_2 מנוגדי סימן	$S_{12} \neq 0$	<p>דחית H0 במבחן t</p> <p>למובהקות β_2</p>	<p>המודל האמיתי:</p> $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ <p>המודל הנאמד (הטעותי):</p> $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$	<p>השמטת משתנה</p> <p>רלוונטי</p>
אומד לשונות הפרמטרים	אומד ל- α	אומד ל- β_1	בהיעדר x_2 :													
מוטה חיובית	מוטה אלא-אם- $\bar{x}_2 = 0$	חסר הטיה	$S_{12} = 0$													
מוטה חיובית	מוטה	מוטה חיובית: S_{12} ו- β_2 שווי סימן מוטה שלילית: S_{12} ו- β_2 מנוגדי סימן	$S_{12} \neq 0$													
<p>הורדת אחד המשתנים</p>	<p>לא ניתן לבצע בדיקת השערות</p> <p>אר"פ ($\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2$) בלתי מוגדרים.</p>	<p>אם: $x_1 = a + bx_2$</p> <p>אז: $r_{12} = 1$</p>	<p>מתאם מלא בין המשתנים המסבירים במודל</p> $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ <p>כאשר $r_{12} = \pm 1$</p>	<p>מולטיקוליניאריות מלאה</p>												
<p>הורדת** אחד המשתנים או איחודם</p>	<p>ניתן לבצע בדיקת השערות</p> <p>אין פגיעה בתכונות אר"פ ושונותם</p>	<p>א.סתירה בין מבחן F ל-t</p> <p>ב. רגישות לספציפיקציה</p> <p>ג. סימנים הפוכים</p>	<p>מתאם חזק בין המשתנים המסבירים במודל</p> $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ <p>כאשר $0.7 < r_{12} < 1$</p>	<p>מולטיקוליניאריות חלקית</p>												

* במידה והמובהקות גבולית ($1 < t_{\hat{\beta}_2} < 2$) נשקול להשאיר משתנה לא רלוונטי כי מעלה את $AdjR^2$ (חוק חיטובסקי).

** במידה ומובהקותם גבולית ($1 < t_{\hat{\beta}} < 2$) נשקול להשאיר את שניהם בשל העלייה ב- $AdjR^2$ (חוק חיטובסקי).

תרגול

שאלה מס' 1

להלן מודל של שכר W_t , כפונקציה של שנות לימוד S_t :

$$W_t = \alpha + \beta \cdot S_t + u_t \quad (1)$$

להלן מודל של שכר W_t , כפונקציה של שנות לימוד S_t ושל גיל A_t :

$$W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + v_t \quad (2)$$

כל האומדים חיוביים ומובהקים וקיים קשר שלילי בין גיל להשכלה.

א. $\hat{\beta}_1$ במשוואה (1) הוא:

1. אומד חסר הטיה

2. אומד מוטה שלילית

3. אומד מוטה חיובית

4. אומד מוטה, אך לא ניתן לדעת את כיוון ההטיה.

ב. ניתן להשתמש במבחן t לבדיקת מובהקות השיפוע במשוואה (1). נכון/לא נכון/ לא ניתן לדעת

ג. בנוסף למשתנים במשוואה השנייה, החליט החוקר להוסיף גם את משתנה הוותק, EXP_t . 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' 3:

$$W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + \beta_3 \cdot EXP_t + w_t \quad (3)$$

חווה דעתך על המשוואה השלישית.

שאלה מס' 2

נתונות ארבע משוואות הרגרסיה הבאות (כאשר הסטיות במודל האמיתי מקיימות את הנחות הרגרסיה הקלאסיות):

$$\sum \hat{V}_t^2 = \sum (X_{2t} - \bar{X}_{2t})^2$$
 כאשר התקבל : (1)

(2)

(3)

$$Y_t = \alpha + \beta_1 \cdot X_{1t} + \Sigma_t \quad (4)$$

(6.3)

(המספרים בסוגריים הם ערכי t של אומדני המקדמים).

לגבי הטענות הבאות, קבעו לגבי כל טענה אם היא נכונה או לא, והסבירו :

א. האומד של β_1 במשוואה (2) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

ב. האומד של β_1 במשוואה (3) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

ג. האומד של β_1 במשוואה (4) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

ד. האומדן $\hat{\beta}_1$ במשוואה (4) זהה ל $\hat{\beta}_1$ במשוואה (2).

ה. השונות התיאורטית של האומדן $\hat{\beta}_1$ במשוואה (4) זהה לשונות התיאורטית של $\hat{\beta}_1$ במשוואה (2), אך אומדני השונות שונים.

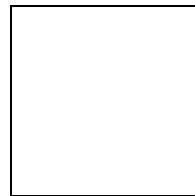
ו. האומד ל α במשוואה (4) הינו חסר הטיה.

ז. האומד ל α במשוואה (3) הינו חסר הטיה.

ח. R^2 של משוואה (2) גדול מ- R^2 של משוואה (3).

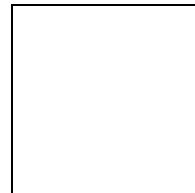
ט. \bar{R}^2 של משוואה (2) גדול מ- \bar{R}^2 של משוואה (3).

שאלה מס' 3



נתון המודל :

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



$$R^2 = 0.92$$

א. בהנחה כי מתקיים:

(0.5)

(0.3)

הערכים בסוגריים הם ערכי t למובהקות הבטות.

יש טעות במודל כי המודל מובהק והמקדמים לא: נכון/לא נכון/ לא ניתן לדעת

$$X_{1t} - 2X_{2t} = 1$$

ב. בהנחה כי מתקיים: נכון/לא נכון/ לא ניתן לדעת

ג. בהנחה כי מתקיים: $x_{1t} = x_{2t}^2$.

לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/ לא ניתן לדעת.

ד. הוכיחו תשובותיכם לסעיפים א ו- ב.

ה. 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.

3. בהנחה שהמודל יצא מובהק אולם הבטות אינן מובהקות וערכי t

למובהקות הבטות הן כדלקמן: $t_{\hat{\beta}_1} = 1.31$, מה יהיה הפיתרון הטוב $t_{\hat{\beta}_2} = 1.45$

ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף 2)?

א. להוריד את x_1

ב. להוריד את x_2

ג. להוריד את שני המשתנים.

ד. להותיר את שני המשתנים.

משתני דמי

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

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

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

$$W_t = \text{השכר (המשתנה התלוי)}$$

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

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

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

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

נשאלת השאלה כיצד נכניס אותו לתוך משוואת הרגרסיה?

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

משתנה כזה נקרא משתנה דמי (dummy variable).

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

ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

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

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

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

(1) משתנה דמי לחותך

המין משפיע על השכר ההתחלתי בלבד.

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

החותך מייצג כאן את השכר ההתחלתי.

שכר ההתחלתי של אישה: α_0

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

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

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש

החותכים: $H_0: \alpha_1 = 0$

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

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

הבאות:

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

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

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

- א. מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?
- ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?
- ג. האם הבדל זה מובהק באוכלוסיה?
- ד. בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ-500 ₪ מזה של נשים.
- ה. בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.

פונקצית רגרסיה המכילה משתנים איכותיים בלבד

המגדר הוא המשתנה היחיד במשוואה:

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

החותך מייצג כאן את השכר הממוצע עבור כל קטגוריה.

שכר הממוצע של אישה: α_0

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

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

בדיקת השערות על משתנה הדמי: מבחן t : $H_0: \alpha_1 = 0$ (מבחן זה למבחן t להבדל בין ממוצעים).

? על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים. תוצאות האמידה:

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

$$S_{\hat{\alpha}_1} = 63 \text{ נתון:}$$

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

(2) משתנה דמי לשיפוע

המגדר משפיע על התוספת לשכר בגין שנות הלימוד.

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

השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.

אצל אישה: התוספת לשכר בגין שנות לימוד - β_0

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

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

בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש

$$H_0: \beta_1 = 0$$

** החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.

? על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים

לנשים בתוספת לשכר בגין שנות הלימוד. תוצאות האמידה נתונות להלן:

$$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$

$$(68) \quad (23) \quad (25)$$

בדוק את ההשערה.

(3) משתנה דמי לכל הפונקציה

המין משפיע גם על החותך וגם על השיפוע. הווה אומר, גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.

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

α_0 : השכר ההתחלתי של אישה:

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

הבדל בשכר ההתחלתי בין המינים: α_1 (הבדל בחותכים)

אצל אישה- התוספת לשכר בגין שנות הלימוד: β_0

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

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

בדיקת השערות למשתני הדמי:

$$H_0: \alpha_1 = \beta_1 = 0$$

באמצעות מבחן WALD יש לבדוק:

לפחות אחד הפרמטרים שונה מ-0: H_1

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

$$H_0: \beta_1 = 0 \quad \text{ו-} \quad H_0: \alpha_1 = 0$$

מבחן CHOW

דרך נוספת לבדיקת ההבדל בין הקטגוריות, בלא יצירת משתני דמי:

חלוקת המדגם לפי הקטגוריות של המשתנה האיכותי. מדגם של גברים (T_m) ושל נשים (T_f).

עבור כל קבוצה לאמוד משוואות רגרסיה לניבוי שכר על ידי שנות לימוד.

$$W_t = \alpha_f + \beta_f X_t + u_t \quad \text{נשים}$$

$$W_t = \alpha_m + \beta_m X_t + u_t \quad \text{גברים}$$

$$H_0 : \alpha_f = \alpha_m ; \beta_f = \beta_m \quad \text{השערות}$$

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

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

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

$$ESS_U = ESS_f + ESS_m \quad \text{המודל הלא מוגבל (U) כולל את שני חלקי המדגם}$$

$$DF_U = DF_f + DF_m$$

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

למרות התוצאות הזרות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. אם דחינו את HO במבחן CHOW נתקשה לברר את מקור ההבדל שנמצא.

2. בהרצת שתי רגרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.

? חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים

בקטעי כביש בינעירוניים, בהינתן נפח התנועה.

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

כבישים מהירים בלבד

(1)

כבישים לא - מהירים בלבד

(2)

שני סוגי הכביש (כל המדגם)

(3)

(4)

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

= נפח התנועה בקטע כביש t ליום באלפים

= משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר ,

ו-0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

1) בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.

2) חשבו את הערכים המספריים עבור אומדני משוואה (4).

3) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

הועלתה הטענה כי המקדם להשפעה של נפח התנועה בדרכים מהירות הינו כפול מזה שבדרכים לא-מהירות.

4) מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?

5) מהי הרגרסיה "תחת"  " למבחן WALS ?

משוואה (1) - כבישים מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 344

Number of Observations Used 344

Analysis of Variance

	Source	DF	Sum of Squares	Mean Square	F Value	Pr
> F	Model	1	4700.81174	4700.81174	89.12	
<.0001	Error	342	18039	52.74684		

Corrected Total 343 22740

Root MSE 7.26270 R-Square 0.2067
 Dependent Mean 5.10465 Adj R-Sq 0.2044
 Coeff Var 142.27617

Parameter Estimates

Variable	DF	Parameter	Standard	t Value	Pr > t
		Estimate	Error		
Intercept	1	1.55289	0.54303	2.86	0.0045
avgd	1	0.02098	0.00222	9.44	<.0001

משוואה (2) - כבישים לא מהירים בלבד

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 410

Number of Observations Used 410

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	971.99073	971.99073	145.83	<.0001
Error	408	2719.34830	6.66507		
Corrected Total	409	3691.33902			

Root MSE 2.58168 R-Square 0.2633
 Dependent Mean 1.38780 Adj R-Sq 0.2615
 Coeff Var 186.02612

Parameter Estimates

Variable	DF	Parameter	Standard	t Value	Pr > t
		Estimate	Error		
Intercept	1	0.14978	0.16360	0.92	0.3605
avgd	1	0.02877	0.00238	12.08	<.0001

משוואה (3) - שני סוגי הכביש (כל המדגם)

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr
Model	1	8052.00804	8052.00804	288.84	<.0001
Error	752	20964	27.87730		
Corrected Total	753	29016			

Root MSE 5.27990 R-Square 0.2775

Dependent Mean 3.08355 Adj R-Sq 0.2765

Coeff Var 171.22758

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	0.73903	0.23665	3.12	0.0019
avgd	1	0.02330	0.00137	17.00	<.0001

(4) משוואה

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 754

Number of Observations Used 754

Analysis of Variance

	Source	DF	Sum of Squares	Mean Square	F Value	Pr
> F	Model	3	8256.966	2752.322	99.44	
<.0001	Error	750	20759	27.678		
	Corrected Total	753	29016			

Root MSE	5.26102	R-Square	0.2846
Dependent Mean	3.08355	Adj R-Sq	0.2817
Coeff Var	170.61553		

Parameter Estimates

	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >
t	Intercept	1	0.14978	0.33340	0.45	
0.6534						
	type	1				
0.0067						
	avgd	1				
<.0001						
	avgdtype	1				
0.1283						

סיכום ביניים:

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

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים

כאשר המשתנה האיכותי כלל שני ערכים בלבד (למשל, מגדר: גבר, אישה)
הסתפקנו במשתנה דמי אחד.

במקרים רבים המשתנה האיכותי כולל יותר משני ערכים/קטגוריות. במקרה כזה
נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

למשל, את המשתנה האיכותי של עונות השנה הכולל 4 ערכים: אביב, קיץ, סתיו,
חורף נייצג באמצעות 3 משתני דמי:

D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת
הייחוס.

נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

$$V_t = \text{מדד מחירי הירקות}$$

$$p_t = \text{מדד המחירים לצרכן}$$

(1) משתני דמי לחותך

הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t \quad \text{המודל:}$$

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

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

ניתן לראות כי: α_0 : החותך בקטגוריה שהושמטה

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

בדיקת השערות:

השערות:

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

$H1: OTHERWISE$

המבחן הסטטיסטי : מבחן WALT :

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

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

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

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

1. האם יש הבדל במחיר ההתחלתי של הירקות בין האביב לחורף: $H0: \alpha_1 = 0$

2. האם יש הבדל במחיר ההתחלתי של הקיץ לחורף: $H0: \alpha_2 = 0$

3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף: $H0: \alpha_3 = 0$

? א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.

1. מהי השערת האפס לבדיקת הטענה?

2. פרטו שני מבחנים סטטיסטיים בעזרתם ניתן לבדוק את הטענה.

ב. הועלתה הטענה כי יש רק שתי עונות המשפיעות על מחיר הירקות ההתחלתי: קיץ+אביב, חורף+סתיו.

1. מהי השערת האפס לבדיקת הטענה?

2. מהו המבחן הסטטיסטי המתאים? פרטו.

משתני דמי לשיפוע (2)

הטענה: יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן

$$V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t \quad \text{המודל:}$$

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

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

ניתן לראות כי: β_0 : השיפוע בקטגוריה שהושמטה

$\beta_0 + \beta_i$: השיפוע בקטגוריה i.

בדיקת השערות:

השערות:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי: מבחן WALT:

$$V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t \quad (\text{U})$$

$$V_t = \alpha + \beta \cdot P_t + u_t \quad (\text{R})$$

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

אם נדחה את H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t.

(3) משתני דמי לכל הפונקציה

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

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

בדיקת השערות:

השערות:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$$

המבחן הסטטיסטי: מבחן WALS:

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

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

אם דוחים את H_0 , יש לבדוק במבחן WALS האם ההבדל הוא בין החותכים או בין השיפועים:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0 \quad H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

באם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחן t:

$$H_0: \beta_j = 0 \quad H_0: \alpha_j = 0$$

משתני דמי עבור שני משתנים איכותיים

נתבונן בדוגמא שבה יש שני משתנים איכותיים המשפיעים על פונקציית השכר-

מגדר (אישה, גבר) וגזע (לבן, שחור).

נגדיר משתנה דמי G שיקבל 1 אם מדובר בגבר ו-0 אחרת (אישה).

נגדיר משתנה דמי R שיקבל 1 אם מדובר בלבן ו-0 אחרת (שחור).

נבדוק כיצד מגדר וגזע משפיעים על השכר ההתחלתי (החותך), כאשר השכר תלוי גם בשנות לימוד (S_t).

(1) הבדל בחותך ללא אינטראקציה

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t \quad \text{המודל:}$$

במודל זה- אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים לא תלוי בגזע (זהה עבור שחורים ועבור לבנים) ולהיפך- ההבדל בשכר ההתחלתי בין לבנים לשחורים לא תלוי במגדר (זהה עבור נשים וגברים).

ניתן לבדוק השערות על כל אחד מהמשתנים הב'ת האיכותיים בנפרד:

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

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

(2) הבדל בחותך עם אינטראקציה

$$\text{המודל: } W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים תלוי בגזע (שונה אם מדובר בשחורים או בלבנים) ולהיפך.

במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

3. $HO: \alpha_3 = 0$

דרך נוספת ליצירת מודל עם אינטראקציה:

הגדרת משתני דמי המייצגים שילוב בין המשתנים האיכותיים גזע ומגדר באופן הבא:

D_1 יקבל 1 אם מדובר בגבר לבן ו-0 אחרת

D_2 יקבל 1 אם מדובר בגבר שחור ו-0 אחרת

D_3 יקבל 1 אם מדובר באשה לבנה ו-0 אחרת

הנשים השחורות מהוות כאן את קבוצת הייחוס.

$$\text{המודל: } W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה:

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

ההשערות לבדיקת קיום האינטראקציה:

$$HO: \gamma_1 - \gamma_2 = \gamma_3 \quad \text{או} \quad HO: \gamma_1 - \gamma_3 = \gamma_2$$

התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת:

$$WALD = t^2$$

$$PF = Pt$$

? שאלה מס' 1

חוקר בדק השפעות של השכלה, גזע (שחור, לבן) וניסיון (EXP) על לוג השכר ($\ln(Y)$) במדגם בן 306 תצפיות:

$$\ln(Y)_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 EXP_t + \beta_2 EXP_t^2 + u_t$$

$\ln(Y)$ -לוג השכר

EXP - שנות ניסיון

D_1 מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (0-אחרת)

D_2 מקבל את הערך 1 עבור שחורים בעלי השכלה נמוכה (0-אחרת)

D_3 מקבל את הערך 1 עבור לבנים בעלי השכלה גבוהה (0-אחרת)

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

Analysis of Variance						
	Source	DF	Sum of Squares	Mean Square	F Value	Pr
> F	Model	5	-----	-----	-----	---
---	Error	300	140	-----		
	Corrected Total	305	210			
	Root MSE			-----	R-Square	-----
	Dependent Mean			-----	Adj R-Sq	-----
	Coeff Var			-----		

Parameter Estimates						
	Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >
t	Intercept	1	-----	-----	60.84	0.00
	D1	1	-----	-----	-3.20	0.00
	D2	1	-----	-----	-5.56	0.00
	D3	1	-----	-----	7.23	0.00
	EXP	1	-----	-----	8.11	0.00
	EXP ²	1	-----	-----	-7.45	0.00

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

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

(א) לפי המשוואה הניסיון זהה עבור שחורים ולבנים: נכון/לא נכון/ לא ניתן לדעת

(ב) בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.

(ג) בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.

(ד) הי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?

(ה) לבדיקת ההשערה של הסעיף הקודם בוצע מבחן W.L.D.

הרגרסיה המוגבלת תחת השערת האפס הינה:

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

(ו) בדוק את ההשערה אם ידוע שבמודל המוגבל $R^2 = 0.33$

(ז) החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

כאשר: S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים)

E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).

מה הקשר בין המקדמים של שני המודלים?

(ח) אם יאמוד החוקר את המשוואה:

$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$

האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו', ו-ז').

שאלה מס' 2

חוקרת בדקה השפעות השכלה, מגדר וניסיון על הכנסה מעבודה לפי המשוואה הבאה:

$$\ln(MWAGE) = \alpha_0 + \alpha_1 \cdot S + \alpha_2 \cdot E + \alpha_3 \cdot (S \cdot E) + \beta_0 \cdot EXP + \beta_1 \cdot (EXP \cdot S) + \beta_2 \cdot (EXP \cdot E) + \beta_3 \cdot (EXP \cdot S \cdot E) + U$$

כאשר S : משתנה דמי = 1 עבור נשים, = 0 גברים

E משתנה דמי = 1 עבור השכלה גבוהה (scl > 12), = 0 השכלה

נמוכה

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

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

א. רשמו את הפונקציה לחישוב:

1. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.
 2. תחזית לוג השכר ההתחלתי עבור נשים משכילות.
 3. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?
- ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:
1. אין השפעה של מגדר והשכלה על השכר.
 2. השפעת ההשכלה אינה תלויה במגדר.
 3. אין השפעות השכלה אצל גברים.
 4. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

הפרה של ההנחות הקלאסיות

ארבעת הנושאים האחרונים שנלמד עוסקים במצב של הפרת אחת ההנחות הקלאסיות הדרושות לאמידת הפרמטרים בשיטת OLS.

הנושא הראשון - הטרוסקדסטיות, עוסק בהפרת הנחה מס' 5 המדברת על שונות קבועה ויחידה לאורך קו הרגרסיה: $V(u_t) = \sigma^2$.

הנושא השני - מתאם סידרתי, עוסק בהפרת הנחה מס' 6 המדברת על אי תלות בין הטעויות: $\text{cov}(u_t, u_s) = 0$

הנושא השלישי - מודלים דינמיים, והרביעי - משוואות סימולטניות, עוסקים בהפרת הנחה מס' 4 המדברת על אי תלות בין המשתנים ה"ת" לטעויות: $\text{cov}(x, u) = 0$

בכל אחד מן הנושאים נלמד:

- מהן ההשלכות של הפרת ההנחות הללו על אומדי הריבועים הפחותים.
- מהם המבחנים הסטטיסטיים המשמשים לזיהוי קיומה של הפרה.
- כיצד נתקן את משוואת הרגרסיה כך שניתן יהיה לאמוד את הפרמטרים בשיטת OLS.

הטרוסקדסטיות

הטרוסקדסטיות הוא מצב שבו מופרת הנחה מס' 5, הנחת ההומוסקדסטיות, הגורסת כי השונות של ה"קריזה" היא אותה שונות עבור כל תצפית ותצפית: $V(u_t) = \sigma^2$ לכל t , כלומר התצפיות מפוזרות באופן אחיד סביב קו הרגרסיה.

במצב של הטרוסקדסטיות הקריזה של כל תצפית היא בעלת שונות אחרת:

$$V(u_t) = \sigma_t^2$$

למשל כאשר בודקים את הקשר שבין הכנסה ותצרוכת מגלים כי ברמות הכנסה נמוכות אין גמישות רבה בהוצאות ולכן התצפיות מרוכזות סביב קו הרגרסיה. לעומת זאת ברמות גבוהות של הכנסה התצפיות מפוזרות יותר ויש שונות גבוהה יותר בתצרוכת. השונות, אם כן, איננה אחידה סביב קו הרגרסיה והיא תלויה בתצפית - זהו מצב של הטרוסקדסטיות.

ההשלכות של הטרוסקדסטיות על אומדי OLS

מבין התכונות של אר"פ (ליניאריות, חוסר הטיות, עקיבות ויעילות) היחידה שמופרת בהינתן הטרוסקדסטיות היא: **יעילות** של אומדי הרבועים הפחותים.

זאת מכיוון שבכדי להוכיח יעילות של האומדים השתמשנו בהנחה מס' 5 של שונות קבועה. לכן הפרתה גוררת פגיעה ביעילות האומד.

מבחנים לזיהוי הטרוסקדסטיות

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות. גרף השאריות גם מאפשר לנו לבחון את הצורה הפונקציונאלית של השונות, כלומר באיזה אופן השונות משתנה בין תצפית לתצפית.

קיימות שתי שיטות לזיהוי הטרוסקדסטיות: מבחן GQ (Goldfeld-Quandt) ומבחן White.

ההבדל בין שני המבחנים הוא בהחלטה על הצורה הפונקציונאלית של השונות. כלומר על האופן שבו השונות משתנה בין תצפית לתצפית, (על פי דיאגרמת הפיזור).

מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונות שונות בלבד. ואילו מבחן White מניח כי לכל תצפית ותצפית שונות שונה של טעויות.

(1) מבחן GQ

ההנחה העומדת בבסיס מבחן זה היא כי קיימות שתי שונות שונות של טעויות. ביצוע המבחן:

- מחלקים את המדגם לשני חלקים:
 1. החלק שבו אנו חושדים שיש שונות גבוהה יותר
 2. החלק שבו אנו חושדים שיש שונות נמוכה יותר
 מקובל להשמיט מס' תצפיות (בין 1/6 ל-1/3) במרכז המדגם.
- אומדים כל אחד מהחלקים בנפרד ומקבלים את ה-ESS של כל חלק.
- מחשבים את הסטטיסטי: $F_{stat} = \frac{ESS_1 / T_1 - K - 1}{ESS_2 / T_2 - K - 1}$ (תמיד השונות הגבוהה חלקי הקטנה).

- סטטיסטי זה מתפלג $F_{(\alpha; T_1 - K - 1, T_2 - K - 1)}$

- כלל ההכרעה: אם $F_{stat} > F_C$ אז דוחים את H_0 .

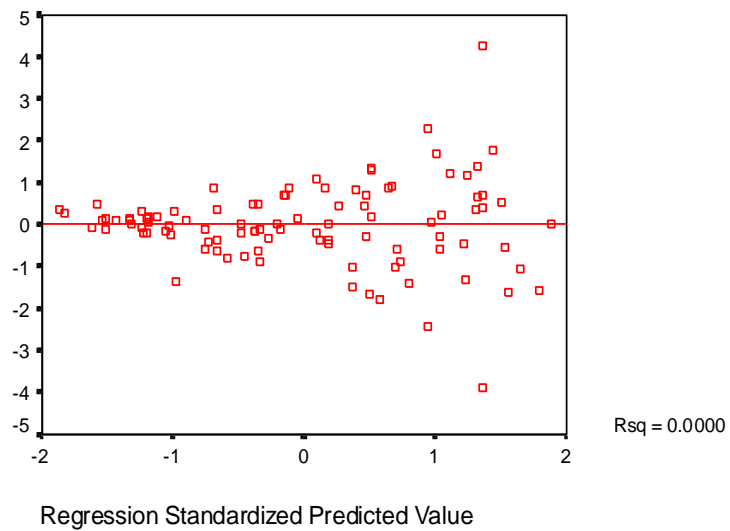
- ההשערות: $H_0: \sigma_1^2 = \sigma_2^2$
 $H_1: \sigma_1^2 > \sigma_2^2$

לדוגמא: נאמד הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$

גרף השאריות של הרגרסיה הנ"ל נתון להלן:

Scatterplot

Dependent Variable: Y



מגרף זה אנו למדים כי ככל שעולים ברמת ההכנסה כך השונות בהוצאות הפרט עולה. כלומר השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה - זהו מצב של הטרוסקדסטיות.

בכדי לבצע מבחן GQ :

- התצפיות של המשתנה הכנסה סודרו מהגדול לקטן והמדגם חולק לשלוש קבוצות שוות.
- רגרסיה נפרדת הורצה על השליש הראשון ועל השליש האחרון.

התוצאות של אמידת הקשר בין הכנסה לתצרוכת מוצג בפלטים 1 ו-2 בהתאמה:

משוואה (1)

The REG Procedure

Model: MODEL1

Dependent Variable: Y

Number of Observations Read 16
 Number of Observations Used 16

Analysis of Variance

	Source	DF	Sum of Squares	Mean Square	F Value	Pr
> F	Model	--	-----	-----	----	----
--	Error	14	166452.9	-----		
	Corrected Total	---	-----			

משוואה (2)

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read 16
 Number of Observations Used 16

Analysis of Variance

	Source	DF	Sum of Squares	Mean Square	F Value	Pr
> F	Model	---	-----	-----	-----	---
--	Error	14	2934638	-----		
	Corrected Total	---	-----			

השערות:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 > \sigma_2^2$$

סטטיסטי המבחן:

$$F_{stat} = \frac{ESS_1 / T_1 - K - 1}{ESS_2 / T_2 - K - 1} = \frac{2,934,638}{166,452.9} = 17.62$$

כלל הכרעה:

מובנהקות של 5%. $F_{stat} = 17.62 > F(0.05; 14, 14) = 2.48$ לכן יש סיבה מספקת לדחות את H_0 ברמת

מסקנה:

יש עדות להטרוסקדסטיות בנתונים.

? על מנת לבחון את פונקציית הייצור בענף מסוים נאספו נתונים על 150 פירמות.

נסמן:

Q תפוקה שנתית באלפי שקלים.

L מספר עובדים

$$\ln(Q) = \alpha + \beta \cdot \ln(L) \quad \text{המודל הנאמד:}$$

החוקר חשש שההפרעה המקרית איננה הומוסקדסטית.

לשם כך הוא מיין את התצפיות בסדר עולה של מספר העובדים, השמיט 1/3 מהתצפיות האמצעיות והריץ שתי רגרסיות נפרדות עם מספר שווה של תצפיות:

ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא קיבל:

$$ESS=279.3 \quad R^2 = 0.403$$

ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא קיבל:

$$ESS=493.8$$

$$R^2 = 0.238$$

האם יש עדות לקיום הטרוסקדסטיות בנתונים? (בצעו את המבחן המתאים: רשמו השערות, חשבו סטטיסטי מבחן, רשמו כלל הכרעה והגיעו למסקנה).

White (2) מבחן

ההנחה העומדת בבסיס מבחן זה כי לכל תצפית ותצפית שונות שונה של טעויות.

הביטוי המתמטי של הנחה זו היא היותה של השונות פונקציה ליניארית של כל

המשתנים המסבירים, ריבועיהם והאיברים הצולבים:

$$\sigma_t^2 = f(x_j, x_j^2, x_j x_j)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_1 + \dots + \alpha_k x_k + \beta_1 x_k^2 + \dots + \beta_k x_k^2 + \gamma_{12} x_1 x_2 + \gamma_{13} x_1 x_3 \dots$$

האומד ל- σ_t^2 הוא \hat{u}_t^2

המבחן הוא מבחן LM:

- אומדים את המודל המקורי ומקבלים את הסטיות מקו הרגרסיה \hat{u}_t (המכונה ב SAS-RES).

- אומדים את \hat{u}_t^2 כפונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים: $\hat{u}_t^2 / x_j, x_j^2, x_j x_j$ זוהי רגרסיית העזר.

- נחשב את סטטיסטי LM: $LM_{stat} = T_y \cdot R_y^2$

- אם $LM_{stat} > \chi_m^2$ כאשר $m =$ מס' המשתנים ברגרסיית העזר.

- השערות:

$$H_0: \alpha_j = \beta_j = \gamma_{jj} = 0$$

$$H_1: OTHERWISE$$

נדגים על אותו הקשר שבין הכנסה לתצרוכת:

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

שאנו חושדים על פי גרף השאריות כי קיים בו מצב של הטרוסקדסטיות.

בכדי לבצע את מבחן WHITE :

• נחשב את השאריות של הרגרסיה: $\hat{u}_t = Y_t - \hat{Y}_t$,

• נעלה את השאריות בריבוע: \hat{u}_t^2

• נאמוד את המשוואה: $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + v_t$

תוצאות האמידה מוצגות להלן:

The REG Procedure

Model: MODEL1

Dependent Variable: RES^2

Number of Observations Read 48

Number of Observations Used 48

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr
Model	---	-----	-----	-----	---
Error	---	-----	-----		
Corrected Total	---	----			

Root MSE ----- R-Square 0.390763

Dependent Mean ----- Adj R-Sq -----

השערות:

$H_0: \alpha_1 = \beta_1 = 0$

$H_1: OTHERWISE$

סטטיסטי המבחן:

$$LM_{stat} = T_y \cdot R_y^2 = 48 \cdot 0.3907 = 18.75$$

כלל הכרעה:

$LM_{stat} = 18.75 > \chi_{2(0.05)}^2 = 5.991$ לכן יש סיבה מספקת לדחות את H_0 ברמת

מובהקות של 5%.

מסקנה

יש עדות לקיום של הטרוסקדסטיות בנתונים.

**הערה חשובה: אם חלק מערכי \hat{u}_t^2 יצאו שליליים (והרי שונות לא יכולה להיות

שלילית) יש להוציא לוג ולאמוד את $\ln \hat{u}_t^2$ כך התחזית תצא תמיד חיובית.

? חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות

והאפשרות של מכירת עיתונים.

נסמן:

Y_SALES = מכירות חודשיות (ש"ח)

X1_SQUARES = שטח החנות (מ"ר)

X2_RENT = דמי שכירות (\$)

PAPERS = משתנה איכותי המקבל 1-אם החנות מוכרת גם עיתונים ו-0 אם לא.

החוקר חשד כי קיימת בעיה של הטרוסקדסטיות בנתונים.

החוקר ביצע מבחן לזיהוי הטרוסקדסטיות שתוצאותיו נתונות להלן:

Dependent Variable:

Number of Observations Read 20
 Number of Observations Used 20

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr
Model	---	-----	-----	-----	---
Error	---	-----	-----		
Corrected Total	---	----			
Root MSE		-----	R-Square	0.086942	
Dependent Mean		-----	Adj R-Sq	-----	

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >
Intercept	1	-----	----	-----	---
X1_SQUARE	1	-----	-----	-----	---
X1_SQUARE^2	1	-----	-----	-----	---
X1_SQUARE*X2_RENT	1	-----	-----	-----	--
X2_RENT	1	-----	-----	-----	--
X2_RENT^2	1	-----	-----	-----	--
X2_RENT*PAPERS	1	-----	-----	-----	--
PAPERS	1	-----	-----	-----	-

הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: _____

על ידי מבחן: _____

המשתנה התלוי הינו: _____

המשתנים הב"ת: _____

ההשערות הינן: _____

גודל הסטטיסטי למבחן הינו (רשמו תוצאה

מספרית): _____

המסקנה המתקבלת היא: _____

פיתרון בעיית ההטרוסקדסטיות- ריבועים פחותים משוקללים (WLS)

נניח שאנו רוצים לאמוד את המודל: $Y_t = \alpha + \beta X_t + u_t$ וידוע כי לכל קריזה שונות אחרת.

ההנחה שעומדת בבסיס שיטת ה-WLS היא כי השונות המשתנה כוללת בתוכה מרכיב קבוע ומרכיב משתנה:

$$\sigma_t^2 = Z_t \cdot \sigma^2$$

את המרכיב המשתנה בשונות (Z_t) יש לנטרל.

לשם כך ניצור משתנה חדש- W_t שיהווה השורש ההופכי לאותו מרכיב משתנה:

$$W_t = \frac{1}{\sqrt{Z_t}}$$

נכפיל כל תצפית במשתנה החדש W_t וניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית:

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

$$Y_t W_t = \alpha \cdot W_t + \beta (X_t W_t) + u_t W_t$$

בצורתה המפורשת המשוואה החדשה נראית כך:

$$\frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$$

קיבלנו מודל חדש שבו:

$$\frac{u_t}{\sqrt{Z_t}} : \text{ הקריזה היא}$$

$$\frac{Y_t}{\sqrt{Z_t}} : \text{ המשתנה המוסבר}$$

$$\frac{1}{\sqrt{Z_t}} \text{ ו-} \frac{X_t}{\sqrt{Z_t}} : \text{ המשתנים המסבירים הם}$$

במודל זה אין חותך.

הפרמטרים β ו- α הם של המשוואה המקורית (שימו לב כי ל- α אין משמעות של חותך).

למשוואה זו אין משמעות כלכלית אך היא מאפשרת לנו לאמוד את הפרמטרים α ו- β בשיטת OLS כך שנקבל אומדים $\hat{\alpha}$ ו- $\hat{\beta}$ יעילים.

ניתן להוכיח כי הטרגנספורמציה הליניארית של השונות הפכה אותה לקבועה:

$$V\left(\frac{u_t}{\sqrt{Z_t}}\right) = \frac{1}{Z_t} V(u_t) = \frac{1}{Z_t} \cdot \sigma^2 \cdot Z_t = \sigma^2$$

הערה: אם ביצענו את מבחן WHITE והוצאנו \log לנתונים אז $Z_t = e^{\ln \hat{u}_t^2}$ כך שיש

$$\text{לשקלל את הרגרסיה ב: } W_t = \frac{1}{\sqrt{e^{\ln \hat{u}_t^2}}}$$

שאלה מס' 1 ?

$$Y_t = \alpha + \beta \cdot X_t + U_t \quad (1) \quad \text{נתון המודל :}$$

$$\text{ונתון כי :} \quad \text{VAR}(U_t) = \frac{\sigma^2}{Z_t^2} \quad (Z_t \text{ משתנה ידוע}).$$

א. מהי הבעיה שנוצרת באמידת משוואה (1) ?

ב. מהן תכונות אומדי הריבועים הפחותים של משוואה (1) ?

כדי לפתור את הבעיה שנוצרה, נאמדה המשוואה הבאה :

$$Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + U_t \cdot W_t \quad (2)$$

ג. מהו W_t שבעזרתו ניתן לאמוד את α ו- β בצורה יעילה ?

ד. מהו האומד היעיל של σ^2 ?

ה. האם ניתן להשוות בין המודלים על בסיס R^2 ? אם לא, האם ניתן להחליט בכל זאת איזה מודל טוב יותר ?

ו. חוו דעתכם על הטענות הבאות, ונמקו :

1. אם נתון כי $Z_t = a + b \cdot \bar{X}$, התשובות לסעיפים א' ו-ב' נשארות ללא שינוי.

2. המשוואה הנורמאלית: $\sum \hat{\varepsilon}_t = 0$ (כאשר $\varepsilon_t = U_t \cdot W_t$) היא אחת המשוואות הנורמאליות לאמידת משוואה (2).

שאלה מס' 2

עני על שאלה מס' 1, כאשר נתון כי $\text{VAR}(U_t) = \sigma^2 \cdot X_t^2$.

שאלה מס' 3

נתון המודל: $Y_t = \alpha + \beta X_t + u_t$

וקיים מדגם של 100 תצפיות

כאשר נתון כי:

$$V(u_t) = \sigma_t^2 = \begin{cases} X_t \sigma^2 & \Leftrightarrow t \geq 50 \\ X_t^2 \sigma^2 & \Leftrightarrow t \leq 50 \end{cases}$$

(שאר ההנחות הקלאסיות מתקיימות)

א. במשוואה מס' 1 יש בעיה של: _____

ב. אמידת משוואה (1) תניב אומדים בלתי מוטים ועקיבים:

נכון/ לא נכון/ לא ניתן לדעת

ג. פיתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה
הבאה:

$$Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t \quad (2)$$

כאשר: $W_t =$ _____

$$V(u_t) = \sigma_t^2 = \begin{cases} 3\sigma^2 & \Leftrightarrow t \geq 50 \\ \sigma^2 & \Leftrightarrow t \leq 50 \end{cases} \quad \text{ד. ם נתון כי:}$$

כן/ לא/ לא ניתן

האם ישתנו תשובותיכם לסעיפים א ו-ב :
לדעת

מתאם סידרתי

מתאם סידרתי עוסק במצב שבו מופרת ההנחה הקלאסית מס' 6 - אי תלות בין

$$\text{cov}(u_t, u_s) = 0$$

הטעויות: זהו מצב שבו קיימת תלות סטטיסטית בין הטעויות במודל: $\text{cov}(u_t, u_s) \neq 0$.

תלות כזו בין הטעויות קיימת בדרך כלל כאשר הנתונים הנאספים הם נתוני סדרות עיתיות ולא נתוני חתך בהם עסקנו עד כה.

נתוני חתך - מתייחסים לפרטים שונים בתקופת זמן נתונה.

נתוני סדרות עיתיות - מתייחסים לאותו הפרט לאורך זמנים שונים.

בנתוני החתך סביר שמאחר ומדובר בפרטים שונים - הטעויות בניבוי שלהם תהיינה בלתי תלויות. לעומת זאת בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר דווקא שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.

אם אנחנו מדברים, למשל, על פונקציות תצרוכת:

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

אם מדובר בנתוני חתך - החלטות התצרוכת של פרט t לא אמורות להשפיע על החלטותיו של פרט s והנחה 6 תקייה:

$$\text{cov}(u_t, u_s) = 0$$

אולם אם מדובר על נתוני סדרה עיתית: החלטותיו של פרט מסוים היום סביר שיושפעו מהחלטות שעשה בעבר או שישפיעו על החלטות שיעשה בעתיד:

$$\text{cov}(u_t, u_{t-s}) \neq 0$$

כאשר: u_t - "קריזה" בזמן t.

u_{t-s} - "קריזה" בזמן t-s.

השלכות על אומדי הריבועים הפחותים (OLS)

מבין התכונות של אר"פ (ליניאריות, חוסר הטיות, עקיבות ויעילות) היחידה שמופרת כאשר קיים מתאם סידרתי היא:

תכונת היעילות.

משום שתכונת היעילות היא היחידה מבין תכונות אר"פ התלויה להוכחתה בקיומה של הנחת אי התלות בין הטעויות.

משום הפגיעה בתכונת היעילות, בדיקת ההשערות לא תהיה תקפה.

****שימו לב כי במידה וקיים מתאם סידרתי חיובי בין הטעויות ולמשתנים יש מגמת זמן (X עולה או יורד עם הזמן) אומד השונות (ESS) יהיה מוטה כלפי מטה ואז נקבל F, R^2 ו-t מוטים כלפי מעלה.**

מבנה המתאם הסדרתי

הגדרנו מתאם סדרתי כהפרה של הנחה מס' 6:

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

$$\text{cov}(u_t, u_{t-s}) \neq 0$$

נשאלת השאלה כיצד נראית הפרה זו?

מתאם סדרתי מסדר ראשון:

ההנחה היא כי יש מתאם בין קריזות במרחק אחד, כלומר u_t תלוי ישירות רק ב-

$$u_{t-1}$$

$$\text{cov}(u_t, u_{t-1}) \neq 0$$

את המתאם בין ה"קריזות" מסדר ראשון ניתן לנסח באופן הבא:

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

כך ש:

$$(1) \rho \neq 0 \text{ (כי אם } \rho = 0 \text{ אין מתאם)}$$

$$(2) 1 < \rho < -1 \text{ (כי אם חורג מ-1 ה"קריזה" הולכת וגדלה עם הזמן)}$$

$$(3) \rho \text{ חיובי פירושו מתאם סדרתי חיובי ואילו } \rho \text{ שלילי פירושו מתאם סדרתי שלילי (לא נפוץ).}$$

$$(4) \varepsilon_t \text{ מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין (בניגוד ל- } u_t \text{) כך ש:}$$

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma_\varepsilon^2$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t-s}) = 0$$

המודל יכול שתי משוואות-המשוואה העיקרית והגדרת המתאם הסדרתי (מסדר ראשון):

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

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

מלבד α ו- β נרצה לאמוד גם את ρ .

מתאם סדרתי מסדר שני:

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \varepsilon_t$$

הן u_{t-1} והן u_{t-2} משפיעים ישירות על u_t .

מתאם סדרתי מסדר P:

$$u_t = \rho_1 \cdot u_{t-1} + \rho_2 \cdot u_{t-2} + \dots + \rho_P \cdot u_{t-P} + \varepsilon_t$$

u_t מושפע מתקופות שונות בעבר.

תכונות המתאם הסדרתי

ניתן להוכיח כי מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן:

$$r_{(u_t, u_{t-s})} = \rho^s$$

מכיוון שכך המתאם של u_t הולך ופוחת עם הזמן:

$$\rho_{u_t, u_{t-1}} > \rho_{u_t, u_{t-2}}^2 > \rho_{u_t, u_{t-3}}^3 > \dots > \rho_{u_t, u_{t-s}}^s$$

בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t) = 0$$

$$V(u_t) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

$$COV(u_t, u_{t-s}) = \rho^s \sigma_u^2 = \rho^s \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

? נתון מתאם סדרתי מסדר ראשון: $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

$$V(\varepsilon_t) = \sigma_\varepsilon^2 = 1 \text{ וכי } \rho = 0.9$$

מצאו את:

א. המתאם בין u_t ל- u_{t-1}

ב. המתאם בין u_t ל- u_{t-4} . הסבר את ההבדל בין המתאמים (סעיף א' ו-ב').

ג. השונות σ_u^2

ד. חזרו על סעיפים א' עד ג' עבור $\rho = 0.4$. הסבירו את ההבדל בין התוצאות.

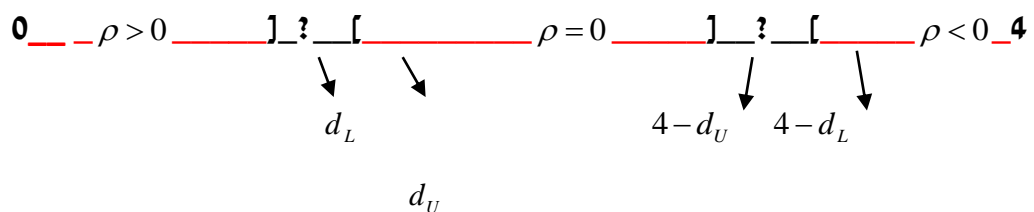
מבחנים לזיהוי מתאם סדרתי

קיימים שני מבחנים סטטיסטיים לזיהוי קיומו של מתאם סדרתי:

מבחן DW (דרבין ווטסון) ומבחן LM.

1) מבחן DW (דרבין ווטסון) לקיום מתאם סדרתי מסדר ראשון:

- נניח תחילה כי אין מתאם סדרתי ונאמוד את המשוואה הראשית בשיטת OLS.
- כחלק מתוצאות האמידה נקבל ציון DW (יכול לקבל ערכים בין 0 ל-4 בלבד).
- נתבונן בטבלת DW ולפי $K = \text{מס' המשתנים ה"ת במודל}$ ו- $T = \text{מס' התצפיות במדגם}$ נשלוף שני ערכים: d_U ו- d_L .
- נחלק את הטווח שבין 0 ל-4 באופן הבא:



- נראה היכן נופל ציון ה-DW שהתקבל כחלק מתוצאות האמידה.
- ניתן לדעת אם יש מתאם ואיזה סוג של מתאם רק אם ציון ה-DW ייפול בחלקים המודגשים.

השערות:

$$H_0: \rho = 0$$

$$H_1: \rho > 0, \rho < 0$$

חישוב הסטטיסטי

$$DW_{stat} \cong 2 \cdot (1 - \hat{\rho})$$

אם אנו מקבלים ציון $\hat{\rho}$ ניתן להציב בנוסחה ולקבל DW_{stat} .

לדוגמא: חוקר רצה לאמוד את מחיר סגירה של מניה כפונקציה של הזמן שעובר:

$$CLOSE_t = \alpha + \beta \cdot TIME_t + u_t$$

כאשר:

$$CLOSE_t = \text{מחיר סגירה של מניה ב-} \$ \text{ ביום } t.$$

$$TIME_t = \text{משתנה זמן שמקבל את הערכים: } 1, 2, 3, \dots$$

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	-----		55.78	
Error	151	-----			
C Total	152	-----			
Root MSE	----	R-square	0.181		
Dep Mean	----	Adj R-sq	-----		
C.V.	----				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	1.3474	0.0148	91.047	0.0000
TIME	1	-0.00075	0.0001	-7.468	0.0000
Durbin-Watson D		0.150			

האם קיים מתאם סדרתי?

נתבונן בטבלת DW כאשר נתון כי $K=1$ ו- $T=153$:

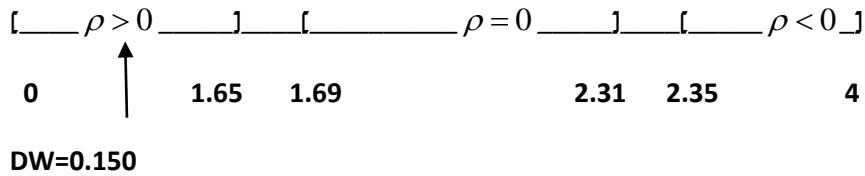
TABLE 12 Cutoff Points for the Distribution of the Durbin-Watson Test Statistic

Let d_α be the number such that $P(d < d_\alpha) = \alpha$, where the random variable d has the distribution of the Durbin-Watson statistic under the null hypothesis of no autocorrelation in the regression errors. For probabilities $\alpha = .05$ and $\alpha = .01$, the tables show, for numbers of independent variables, K , values d_L and d_U such that $d_L \leq d_\alpha \leq d_U$ for numbers n of observations.

$\alpha = .05$										
n	K									
	1		2		3		4		5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	1.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

$d_L = 1.65$
 $d_U = 1.69$: לפי הטבלה (עבור $T=100$)

נציב בטווח:



מסקנה:

יש עדות לקיום מתאם סדרתי חיובי מסדר ראשון.

$Y_t = \alpha + \beta X_t + \varepsilon_t$? המודל הינו:
 $u_t = \rho u_{t-1} + \varepsilon_t$

u_t - הפרעה מקרית קלאסית

$T=100$ וידוע כי:

$u_t = 0.9u_{t-1}$
 $d_L = 1.57$
 $d_U = 1.65$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

למבחן DW יש שתי בעיות עיקריות:

(1) מתאים רק למתאם סדרתי מסדר ראשון

(2) יש אזורים "מתים" בטווח בהם לא ניתן לדעת האם יש מתאם סדרתי.

בנוסף לכך על מספר תנאים להתקיים כדי שאפשר יהיה להשתמש במבחן DW:

1. הרגרסיה כוללת חותך

2. ה-Xים קבועים ולא משתנים

3. אין משתנים מסבירים שהם פיגור של המשתנה המוסבר

4. אין תצפיות חסרות באמצע

5. אם קיים מתאם סדרתי מסדר ראשון אז הוא מהצורה: $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

2) מבחן LM

לעומת מבחן DW מבחן LM מתאים גם לבחון קיומו של מתאם סדרתי מסדרים גבוהים יותר מסדר ראשון.

$$Y_t = \alpha + \beta \cdot X_t + u_t$$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

ניתן לרשום את המודל הלא מוגבל:

$$Y_t = \alpha + \beta \cdot X_t + \rho \cdot u_{t-1} + \varepsilon_t \quad (\text{U})$$

ניתן להתייחס לבחינת קיומו של מתאם סדרתי כהוספת משתנה מסביר: \hat{u}_{t-1}

השלבים לביצוע המבחן:

- נאמוד את המודל המקורי ונחשב \hat{u}_t ו- \hat{u}_{t-1}
- נאמוד את רגרסיית העזר: $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}$
- נחשב סטטיסטי LM: $LM_{stat} = T_y \cdot R_y^2$
- נדחה את H0 כאשר: $LM_{stat} > \chi_m^2$ כאשר $m =$ סדר המתאם הסדרתי.

אם נדחה את H0 נדע את סימנו של המתאם הסדרתי לפי המקדם של \hat{u}_{t-1}

ברגרסיית העזר ששווה ל- $\hat{\rho}$.

שימו לב כי אם נרצה לבדוק מתאם סדרתי מסדרים גבוהים יותר:

ההשערות:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_s = 0$$

H_1 : אחרת

רגרסיית העזר: $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-s}$

לדוגמא: עבור הדוגמא הקודמת- ניבוי מחיר סגירה של מניה כפונקציה של

$$CLOSE_t = \alpha + \beta \cdot TIME + u_t \quad \text{הזמן:}$$

נבחן את קיומו של מתאם סדרתי מסדר ראשון באמצעות מבחן LM.

נאמוד את רגרסיית העזר:

$$u_t = \gamma_0 + \gamma_1 \cdot TIME_t + \rho \cdot u_{t-1} + \varepsilon_t$$

תוצאות האמידה שהתקבלו:

Dependent variable: RES

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	-----			
Error	150	-----			
C Total	152	-----			
Root MSE	-----	R-square	0.855		
Dep Mean	----	Adj R-sq	-----		
C.V.	-----				

Parameter Estimates		
Variable	DF	Parameter Estimate
INTERCEP	1	-00096
TIME	1	1.16331E-05
RES1	1	0.927172

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסיה?

פיתרון בעיית המתאם הסדרתי – רגרסיית הפרשים (שיטת קוקרן-אורקט)
ניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית שבה לא יהיה מתאם סדרתי ולכן ניתן יהיה לאמוד אותה בשיטת הריבועים הפחותים, האומדים יהיו יעילים וניתן יהיה לבצע בדיקת השערות.

$$\text{משוואה (1): המודל בזמן } t: Y_t = \alpha + \beta X_t + u_t$$

$$\text{משוואה (2): המודל בזמן } t-1 \text{ מוכפל ב- } \rho: \rho \cdot Y_{t-1} = \rho \cdot \alpha + \rho \cdot X_{t-1} + \rho \cdot u_{t-1}$$

החסרת משוואה (2) ממשוואה (1):

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

כדי לאמוד את הפרמטרים של רגרסיית הפרשים נגדיר:

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$\alpha^* = \alpha(1 - \rho)$$

$$\beta^* = \beta$$

$$X_t^* = X_t - \rho X_{t-1}$$

$$\varepsilon_t = u_t - \rho u_{t-1}$$

כך "נטרלנו" את המתאם הסדרתי:

$$\varepsilon_t = u_t - \rho \cdot u_{t-1} \quad u_t = \rho \cdot u_{t-1} + \varepsilon_t \Leftarrow$$

ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.

המשוואה "המתוקנת" אותה נאמוד:

$$Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$$

לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים: $\hat{\alpha}, \hat{\beta}$

מאחר ש- ρ איננו ידוע יש צורך לאמוד אותו.

? סטודנט הניח כי במודל: $Y_t = \alpha + \beta X_t + u_t$ קיים מתאם סדרתי מסדר ראשון בשאריות כך שמתקיים: $u_t = 0.7u_{t-1} + \varepsilon_t$ ולכן במקום לאמוד את המודל המקורי אמד את המודל:

$$Y_t - 0.7Y_{t-1} = \alpha(1-0.7) + \beta(X_t - 0.7X_{t-1}) + u_t$$

הסטודנט טען כי במודל החדש לא קיים מתאם סדרתי.

טענת הסטודנט: נכונה / לא נכונה / לא ניתן לדעת

אמידת ρ בשיטת קוקרן אורקוט

שיטת קוקרן אורקוט לאמידת ρ היא שיטה איטרטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות. התהליך מתבצע באופן הבא:

אמידת המשוואה המקורית בה אנו מניחים כי קיים מתאם סדרתי.

חישוב $\hat{u}_t - \rho\hat{u}_{t-1}$ וקבלת האומד $\hat{\rho}$ על ידי אמידת המשוואה: $\hat{u}_t = \rho\hat{u}_{t-1} + \varepsilon_t$

הבעיה היא כי האומד איננו יעיל כיוון שהאומדים לפרמטרים במשוואה המקורית אינם יעילים (בשל קיומו של מתאם סדרתי). לכן נמשיך ונבצע את הפעולות הבאות:

הצבת האומד הזה ברגרסיית הפרשים וחילוץ $\hat{\alpha}$ ו- $\hat{\beta}$ חדשים. שוב נחשב $\hat{u}_t - \rho\hat{u}_{t-1}$ תוך שימוש באומדים החדשים ונקבל אומד $\hat{\rho}$ חדש.

נחזור על התהליך הזה מספר פעמים נוספות עד שנגיע להתכנסות (עד שהאומד $\hat{\rho}$ ישתווה לזה שקיבלנו בתחילה).

אין צורך לבצע תהליך זה ידנית משום שרוב תוכנות המחשב מבצעות הליך זה באופן מיידי ומדווחות על התוצאות הסופיות.

התהליך הממוחשב נקרא אוטו רגרסיה (AUTOREGRESION) מסדר ראשון, שני, שלישי וכו' (תלוי בסדר המתאם הסדרתי).

התיקון למתאם הסדרתי יתבצע על ידי הרצת רגרסיה עם משתנה AR(1) (אוטו רגרסיה מסדר ראשון), AR(1) ו-AR(2) (אם מניחים קיום אוטו רגרסיה מסדר שני) וכו'.

אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

לדוגמא: נמשיך עם הדוגמא של ניבוי מחיר סגירה של מניה כפונקציה של הזמן:

$$CLOSE_t = \alpha + \beta \cdot TIME + u_t$$

נניח כי קיים מתאם סדרתי מסדר ראשון בנתונים.

תוצאות האמידה בשיטת אוטורגרסיה מסדר ראשון מוצגות להלן:

Parameter Estimates				
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0 Prob> T
INTERCEP	1	1.333		
TIME	1	-0.0006		
AR(1)	1	0.927		0.000
Durbin-Watson D		2.235		

א. בדקו האם נפתרה בעיית המתאם הסדרתי.

ב. מהי המשוואה לאמידת מחיר הסגירה הצפוי ביום המסחר הבא?

? שאלה מס' 1

נאמד הקשר שבין הכנסה לתצרוכת לתקופה ינואר 1994 עד דצמבר 1997 (T=48).

$$C_t = \alpha + \beta Y_t + u_t \text{ : המודל הינו}$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t \text{ : בניסיון לבדוק האם מתקיים קשר מהסוג הבא}$$

$$\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \gamma_2 \hat{u}_{t-2} + \gamma_3 \hat{u}_{t-3} + \gamma_4 Y_t + \omega_t \text{ : נאמדה המשוואה הבאה}$$

Depended Variable: RES

תוצאות האמידה מוצגות להלן:

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	-54.709	710.85	-0.076	0.939
RESID1	1	0.705	0.152	4.631	0.000
RESID2	1	-0.0066	0.188	-0.035	0.972
RESID3	1	-0.337	0.167	-2.012	0.051
Y	1	0.0027	0.032	0.085	0.932
Durbin-Watson D	1.954				

$$R^2 = 0.479 \text{ נתון בנוסף כי}$$

א. הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: _____

על ידי מבחן: _____

ההשערות הינן: _____

גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): _____

המסקנה המתקבלת היא: _____

בהנחה כי קיים מתאם סדרתי מסדר שלישי בנתונים נאמד מחדש הקשר שבין ההכנסה לתצרוכת בהתאם לשיטתם של קוקרן ואורקוט.

תוצאות האמידה מוצגות להלן:

Depended Variable:C

Parameter Estimates		
Parameter	Standard	T for H0:

Variable	DF	Estimate	Error	Parameter=0	Prob> T
INTERCEP	1	-2103.53	1128.38	-1.864	0.069
Y	1	0.71944	0.0510	14.086	0.000
AR(1)	1	0.7070	0.1525	4.634	0.000
AR(2)	1	-0.0064	0.1889	-0.034	0.972
AR(3)	1	-0.3282	0.1669	-1.966	0.0562
Durbin-Watson D		1.954			

ב. רשמו את המשוואה המתוקנת המשמשת לעריכת תחזיות.

שאלה מס' 2

חוקר רצה לאמוד את עקומת הביקוש לטיסות לאירופה.

לרשותו נתונים שבועיים לאורך 3 שנים (52 שבועות).

נסמן: Y_t = מספר כרטיסי הטיסה לאירופה שנמכרו בשבוע t

p_t = מחיר ממוצע ב-\$ של הכרטיסים שנמכרו בשבוע t

החוקר אמד את המודל: $Y_t = e^\alpha \cdot P_t^{\beta_1} \cdot P_{t-1}^{\beta_2} \cdot e^{u_t}$

וקיבל לאחר הטרנספורמציה הלוגריתמית: $R^2 = 0.81$.

לבדיקת ההשערה כי קיים מתאם סדרתי בנתונים מסדר ראשון הוא חישב את

ערכי \hat{u}_t ולאחר מכן חישב את הרגרסיה: $u_t = \gamma_0 + \gamma_1 \ln P_t + \gamma_2 \ln P_{t-1} + \gamma_3 u_{t-1} + v_t$

מקדם ההסבר המרובה ברגרסיה זו הוא 0.282

א. נסח את ההשערה ובחן אותה בר"מ של 0.05.

החוקר מניח שיש מתאם סדרתי מסדר ראשון.

לאחר תיקון Cochrane-Orcutt התקבל:

$$\ln Y_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$$

$$\hat{\rho} = 0.2$$

הניחו שהשבוע ובשבוע שעבר מחיר ממוצע של כרטיס היה \$500.

השבוע נמכרו 6,185 כרטיסים. בשבוע הבא צפוי מחיר של \$400.

ב. כמה כרטיסים יימכרו?

החוקר גם מנסה לקבוע האם בנתונים אלה קיים מתאם מסדר שני.

ג. רשמו את המשוואה הנוספת שעליו לאמוד.

במשוואה הנוספת התקבל מתאם מרובה השווה ל- 0.12.

ד. מהי המסקנה בר"מ של 0.05?

סיכום בעיית המתאם הסדרתי והטרוסקדסטיות

הטרוסקדסטיות	מתאם סדרתי	
למשל: $Y_t = \alpha + \beta X_t + u_t$		המשוואה העיקרית של המודל
$V(u_t) = \sigma^2$	$\text{cov}(u_t, u_{t-s}) = 0$	ההנחה הקלאסית המופרת
$V(u_t) = \sigma_t^2$	$\text{cov}(u_t, u_{t-s}) \neq 0$	המצב לאחר ההפרה
$V(u_t) = W_t \sigma^2$	$u_t = \rho u_{t-1} + \varepsilon_t$	המשוואה המאפיינת את ההפרה
מתקבלים אומדים חסרי הטיה ועקיבים, אך תכונת היעילות נפגעת.		מה קורה אם אומדים ב-OLS
מבחן GQ	מבחן DW	זיהוי הבעיה
מבחן White	מבחן LM	
שיטת WLS	שיטת קוקרן-אורקוט (רגרסיית הפרשים) הכנסת משתנה מוסבר בפיגור (מודל דינמי)	פתרון הבעיה

שאלות חזרה למבחן מבוססות על פלטי STATA

שאלה מס' 1

כדי לבדוק האם יש קשר בין שכר המורים וההוצאה לתלמיד בני"ס ציבוריים

$$Pay_i = \beta_1 + \beta_2 Spend + u_i$$

נאמד המודל הבא:

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

```
. reg pay spending
```

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

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

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

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

ג. נבדקו מדדים תיאוריים לגבי המשתנה הב"ת:

```
. su spending
```

variable	Obs	Mean	Std. Dev.	Min	Max
spending	51	3696.608	1054.761	2297	8349

ד. הסבירו את משמעות ההנחות הקלאסיות במונחי המשתנים.

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

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

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

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

1. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן:

2. ערך t סטטיסטי לבדיקת המבחן:

3. מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת סמך של 95%.

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

1. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן:

2. ערך t סטטיסטי לבדיקת המבחן:

3. ערך t קריטי לבדיקת המבחן:

4. מסקנת המבחן: לא דוחים את H_0 ברמת ביטחון של 95%.

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

יא. חוקר רצה לאמוד את המשוואה בשקלים במקום בדולרים ולכן אמד את

$$\hat{pay}_1 = \alpha_0 + \alpha_1 \text{spend}_1$$

הניחו כי שער הדולר עומד על 3.5 ₪.

1. מהי הפקודה ב-stata ליצירת המשתנים החדשים?

2. מה יהיה האומדן ל- α_1 ? מה יהיה האומדן לסטיית התקן של α_1 ?

על אותם הנתונים נאמדה גם המשוואה הבאה:

$$(3) \text{Pay}_i = \beta_0 + \beta_1 \text{Spend} + \beta_2 \text{Spend}^2 + u_i$$

כאשר: $\text{spend}^2 = \text{spend}^2$

יב. כתוב את הפקודה ב-STATA ליצירת המשתנה החדש.

יג. מתוצאות האמידה התקבל ש:

$$\hat{pay}_i = 23128.44 + 1.223 \text{spend}_i - 0.0000635 \text{spend}_i^2$$

$$(9.55) \quad (10.55) \quad (8.78)$$

הערכים בסוגריים הם ערכי t סטטיסטי.

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

T סטטיסטי:

מסקנת הבדיקה: דוחים/לא דוחים את H_0

יז. מהי הפקודה ב- STATA לבדיקת המתאם בין שני המשתנים הב"ת?
טו. בהינתן תוצאות המודל כיצד נראה פרופיל השכר כפונקציה של ההוצאה על
התלמיד (אין צורך בציור מדויק רק במגמה).

טז. מהי ההוצאה לתלמיד שאחריה שכר המורים מתחיל לרדת?

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

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

הטענה נכונה/הטענה איננה נכונה

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

1. האומדים יהיו חסרי הטיה ויעילים

2. האומדים יהיו חסרי הטיה אך לא יעילים

3. האומדים יהיו מוטים אך יעילים

4. האומדים יהיו מוטים ולא יעילים

שאלה מס' 2

נתונים 2 המודלים הבאים:

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

$$Y_t = \beta X_t + u_t \quad .2$$

א. בהנחה כי R^2 של המודל הראשון גבוה מזה של המודל השני, האם ניתן לומר שהמודל הראשון טוב יותר? הניחו כי מודל (1) הוא המודל האמיתי.

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

ג. האם ניתן לומר כי האומדן הינו עקיב?

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

ה. חשבו את שונות האומדן.

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

שאלה מס' 3

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

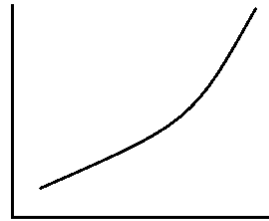
Source	SS	df	MS			
Model	7888.51144	1	7888.51144	Number of obs =	1000	
Residual	31092.9858	998	31.1552964	F(1, 998) =	253.20	
Total	38981.4972	999	39.0205177	Prob > F =	0.0000	
				R-squared =	0.2024	
				Adj R-squared =	0.2016	
				Root MSE =	5.5817	

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

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

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

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



באיזה מודל כדאי לבחור? תנו 3 נימוקים לבחירתכם.

ב. מהי התשואה להשכלה על סמך המודל הנבחר?

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

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

```
. su wage if educ==16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	186	13.30328	7.575015	2.54	60.19

הממוצע הוא 13.303

• שימו לב: סימן ה- "=" משמש להשוואה בין משתנים (למשל **gen**

price_new= price*1.2), לעומת זאת, סימן ה- "==" משמש להגדרת ערך משתנה (**l educ if educ == 8**).

1. מהי התחזית לשכר לעובד עם 16 שנות השכלה עבור כל אחד מהמודלים?

2. איזה מהמודלים נותן תחזית נקודתית מדויקת יותר:

המודל הראשון/המודל השני/ לא ניתן להשוות

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

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

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

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

שאלה מס' 4

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

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

כאשר:

PROD = כמויות אורז מדושן (נמדד בטונות)

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

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

FERT = כמויות דשן בשימוש (נמדד בק"ג).

להלן תוצאות האמידה:

. reg lprod larea llabor lfert

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

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

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

. test larea= llabor

(1) larea - llabor = 0

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

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

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

```

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

```

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

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

באיזה מבחן אתם משתמשים?
נסחו את ההשערה, את הסטטיסטי והקריטי.
מהי מסקנתכם?

שאלה מס' 5

חוקרים ביקשו לאמוד את פונקציית החיסכון המצרפי במשק הישראלי.

מפתח שמות משתנים:

GDS87 – חסכון מקומי גולמי

GDP87 - תוצר מקומי גולמי

GC87 – הוצאות הממשלה

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

נאמדו 2 המודלים הבאים:

. reg gds87 gdp87

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

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

. reg gds87 gdp87 gc87

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

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

א. כתבו את המודל האקונומטרי שנאמד בכל אחת מהאמידות

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

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

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

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

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

יעיל/לא יעיל/ לא ניתן לדעת

ו. בהינתן תוצאות משוואה 1 הקורלציה בין שני המשתנים ה"ת היא:

חיובית/שלילית/אפס/לא ניתן לדעת

ז. החוקר החליט להוסיף משתנה המודד את הצריכה הממשלתית במיליוני

דולרים במקום בשקלים : $GC87$ ואמד את המשוואה:

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

האומדים של משוואה 3 יהיו:

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

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

1. מולטיקוליניאריות מלאה

2. מולטיקוליניאריות חלקית

3. הוספת משתנה לא רלוונטי

4. השמטת משתנה רלוונטי

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

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

שאלה מס' 6

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

המשתנים הם

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

להלן תוצאות אמידה המתבססת על המשתנים הנ"ל:

```

.gen price1000= price/1000
.gen lprice1000=ln( price1000)
.gen sqft100= sqft/100
.reg lprice1000 sqft100 age beds baths stories vacant

```

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

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

- א. מהי משוואת הרגרסיה שנאמדה?
- ב. פרשו את תוצאות האמידה. דונו בסימנים ובמשמעויות של כל אחד מהמשתנים.
- ג. מה ההבדל במחיר הממוצע בין בית פנוי בזמן מכירתו לבין בית המאוכלס בזמן מכירתו?
- ד. חוקר אחר הגדיר את משתנה VACANT = משתנה דמי המקבל 1 אם הבית היה מאוכלס בזמן מכירתו ו-0 אחרת. האם יש צורך לחשב מחדש את משוואת הרגרסיה? כן/לא.
- ה. נניח כי רצו לתרגם את תוצאות המודל מ-sqrft למטרים מרובעים. יחידה אחת של sqrft שווה ל-0.093 מטר רבוע. מה יהיה הערך של β_1 במודל החדש? מה יהיה ערך t סטטיסטי לדחיית H_0 של המשתנה החדש?
- ו. מה הפער החזוי בין מחיר של בגודל 130 מטר בת 15 שנה לבין דירה בגודל 115 מטר בת 20 שנה, בהינתן שכל שאר המשתנים נותרים קבועים?
- ז. מהי רמת הסמך הנמוכה ביותר עבורה ניתן לדחות את הטענה כי $\beta_5 = 0$?

ח. החוקר החליט לאמוד את המשוואה ללא משתנה VACANT כמשתנה מסביר .

להלן תוצאות האמידה:

```
. reg lprice1000 sqft100 age beds baths stories
```

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

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

כיצד השמטת המשתנה השפיעה על משוואת הרגרסיה ועל מקדמיה ?

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

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

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

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


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

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

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

השוו את תוצאות האמידה של שתי הרגרסיות הנ"ל

י. ערכו מבחן Chow לבדיקת שקילות (יציבות) המקדמים בשתי הרגרסיות מהסעיף הקודם.

שאלה מס' 7

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

$FAMINC$ = הכנסת המשפחה (דולרים בשנה)

$HEDU$ = מספר שנות הלימוד של הבעל

$WEDU$ = מספר שנות הלימוד של האישה

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

הוא אמד שלוש רגרסיות וקיבל את התוצאות הבאות:

. */ Table 1

. reg faminc Heduc Weduc k16

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

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

. */ Table 2

. reg faminc Heduc Weduc

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

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

. */ Table 3

. reg faminc Heduc

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

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

- א. פרשו את משמעויות המקדמים ברגרסיה בטבלה הראשונה, התייחסו למובהקות המקדמים (אין צורך בבדיקה פורמאלית של מובהקות).
- ב. התקבל כי ממוצע משתנה KL6 במדגם הוא 0.432. מה משמעות נתון זה?
- ג. חשבו את סטית התקן הנאמדת של אומד הריבועים הפחותים ל β_{KL6} מהמודל הנאמד בטבלה הראשונה. הראו את החישובים שלכם ואת הנוסחאות עליהן אתם מתבססים.

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

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

שאלה מס' 8

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

- $$\hat{CM}_i = 263.6416 - 0.0056 \cdot PGNP_i - 2.2316 \cdot FLR_i$$

(s.e) (11.5932) (0.0019) (0.2099) $R^2 = 0.7077$
- $$\hat{CM}_i = 168.3067 - 0.0055 \cdot PGNP_i - 1.768 \cdot FLR_i + 12.8686 \cdot TFR_i$$

(s.e) (11.5932) (0.0019) (0.2099) (?) $R^2 = 0.7474$

כאשר:

$CM = \text{Child Mortality}$ מס' מקרי המוות של ילדים מתחת לגיל 5 לכל 1000 לידות חיים

$PGNP = \text{Per Capita GNP}$ תוצר לנפש במחירים קבועים בדולרים
 $FLR = \text{Female Literacy Rate}$ אחוז נשים שיודעות לקרוא ולכתוב
 $TFR = \text{Total Fertility Rate}$ מס' הלידות הממוצע לאישה במדינה הנתונים הם עבור 64 מדינות.

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

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

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

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

מבחן F במונחי R^2)

ה. האם תוכלו לחשב את סטיית התקן הנאמדת של המקדם למשתנה TFR ?

(רמז: היזכרו בקשר בין התפלגות T להתפלגות F)

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

2. ענו על השאלות התיאורטיות הבאות:

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

1. הטרוסקדסטיות

2. השמטת משתנה מסביר רלוונטי

3. מקדם מתאם גבוה מאוד בין שני משתנים מסבירים במודל

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

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

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

שאלה מס' 9

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

D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

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

כאשר:

$V_t =$ מדד מחירי הירקות

$p_t =$ מדד המחירים לצרכן

לשם כך אמדה את הרגרסיה הבאה על פני 30 שנה:

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

א. מדוע לא הכניסה החוקרת למשוואת הרגרסיה את משתנה D_4 ?
תוצאות האמידה שהתקבלו הן:

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

$$R^2 = 0.0844$$

נאמדה בנוסף גם המשוואה הבאה:

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

תוצאות האמידה שהתקבלו:

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

$$R^2 = 0.0670$$

- ב. על סמך התוצאות שהתקבלו, דרגו את העונות לפי רמת המחיר הבסיסית שלהן. הציגו לכל עונה את המיקום שלה ואת רמת המחיר הבסיסית כפי שבא לידי ביטוי במודל.
- ג. בדקו את ההשערה כי עונתיות לא משפיע על מחיר הירקות.
- ד. כתבו את ההשערות הבאות:
1. מדד מחירי הירקות זהה בחורף ובאביב
 2. מדד מחירי הירקות גבוה בקיץ מאשר בחורף ביותר מ-600
- ה. האם יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן (בהנחה שהמחיר ההתחלתי של הירקות זהה בין עונות השנה)?

שאלה מספר 10

חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.

החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את המשוואות הבאות:

$$NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t} \quad (1)$$

$$NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t \quad (2)$$

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

$AVGD_t$ = נפח התנועה בקטע כביש t ליום באלפים

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

כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מוצגות להלן:

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

(2)

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

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

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

(1) בדקו את טענת החוקר.

(2) מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה

עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?

(3) מהי השערת האפס לבדיקת הטענה?

(4) הרגרסיה המוגבלת "תחת" H_0 למבחן F (WALD) הינה:

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

כאשר:

$$Z_0 = \underline{\hspace{2cm}}$$

$$Z_1 = \underline{\hspace{2cm}}$$

$$Z_2 = \underline{\hspace{2cm}}$$

$$Z_4 = \underline{\hspace{2cm}}$$

שאלה מס' 11

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

לפניך הנתונים הבאים:

Price = מחיר ארוחת ביג-מק

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

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

נתון המודל הבא המתאר את המחיר:

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

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

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

1. מחיר ארוחת ביג מק ב-2008 היה זהה בחנויות המנוהלות ע"י הרשת ובחנויות המנוהלות ע"י זכיינים

2. פסיקת בית המשפט שינתה את פער המחירים בין חנויות המנוהלות ע"י הרשת לבין חנויות המנוהלות ע"י זכיינים בין השנים 2008 ו-2009.

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

D1-רשת לפני 2009

D2-רשת אחרי 2009

D3-זכיין לפני 2009

D4-זכיין אחרי 2009

1. מהי הפקודה הרלוונטיות ב-STATA ליצירת D1?
2. כמה ממשתני הדמי יש להכניס לתוך הרגרסיה?
3. כתבו את משוואת הרגרסיה
4. נסחו שוב את ההשערות של סעיף א'
5. חוקר טוען כי התוצאות שהתקבלו בבדיקת ההשערות על ידי הגדרת משתני הדמי בשני האופנים יהיו זהות. נכון/לא נכון.

