אקונומטריקה למדעי הנתונים

תוכן העניינים

תוכן העניינים

אקונומטריקה למדעי הנתונים

פרק 1 - מבוא לקורס

תוכן העניינים

מבוא לקורס:

רקע:

הגדרות וסימונים:

משתנה אמפירי – תוצאותיו ידועות מראש (למשל: רמת הכנסה, גיל, מסי שנות לימוד במדגם מסוים).
משתנה מקרי – תוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קובייה או בהטלת מטבע). באקונומטריקה נעסוק בעיקר במשתנים מקריים.

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (למשל: Y_t או X_t).
קבוע – מקבל ערך אחד בלבד (מסומן באות לועזית ללא אינדקס – למשל a או b).
לכל משתנה מקרי X_t יש תוחלת המייצגת את מרכז ההתפלגות (μ_X או $E(X)$).
השונות – מייצגת את מידת הפיזור של ההתפלגות (σ_X^2 או $V(X)$).
סטית התקן – היא השורש של השונות (σ_X).
שונות משותפת (covariance) – מדד להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם ($\mathrm{Cov}(X,Y)$):
$\mathrm{Cov}(X,Y)=0 \Leftrightarrow X,Y$ בלתי מתואמים.
$\mathrm{Cov}(X,Y)>0 \Leftrightarrow$ מתאם חיובי בין המשתנים.
$\mathrm{Cov}(X,Y)<0 \Leftrightarrow$ מתאם שלילי בין המשתנים.
X,Y בלתי תלויים $\Leftarrow$ X,Y בלתי מתואמים.
מקדם המתאם של פירסון – מדד לכיוון ולעוצמת הקשר הליניארי בין שני משתנים: $\eta_{xy}=\dfrac{\mathrm{cov}(x,y)}{\sigma_x\cdot\sigma_y}$.
$-1\le\eta\le1$
$\eta=1$ מתאם ליניארי חיובי מלא בין שני המשתנים.
$\eta=-1$ מתאם ליניארי שלילי מלא בין שני המשתנים.
$\eta=0$ לא קיים מתאם ליניארי בין שני המשתנים.

אמידה:

פרמטר – ערך המשתנה הנחקר המתאר את כל האוכלוסייה.
סטטיסטי/אומד – ערך המשתנה הנחקר המתאר את המדגם.

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

נוסחאות וחוקים בסטטיסטיקה:

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

<u>חוקי הסיגמה</u>:

1. $\sum_{t=1}^{T}X_t=X_1+X_2+X_3+\ldots+X_T$.
2. סכום של קבוע: $\sum_{t=1}^{T}a=Ta$.
3. סכום של קבוע כפול משתנה = לקבוע כפול הסכום: $\sum_{t=1}^{T}aX_i=a\sum_{t=1}^{T}X_t$.
4. סכום של סכום/הפרש = לסכום/הפרש הסכומים: $\sum_{t=1}^{T}\left(X_t\pm Y\right)_t=\sum_{t=1}^{T}X_t\pm\sum_{t=1}^{T}Y_t$.
5. יש לשים לב כי: $\sum_{t=1}^{T}X_t^2\neq\left(\sum_{t=1}^{T}X_t\right)^2$

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

הגדרות ופיתוחים:

1. סכום הסטיות מהממוצע = 0: $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)=0$.
2. סכום הסטיות הריבועיות מהממוצע (מונה השונות):
$$S_{XX}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}X_t^2-T\bar{X}^2=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)X_t.$$
3. מונה של השונות המשותפת:
$$S_{XY}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t=\sum_{t=1}^{T}\left(Y_t-\bar{Y}\right)X_t.$$

חוקי התוחלת:

1. תוחלת של קבוע = קבוע: $E(a)=a$.
2. תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:
$$E(X\pm Y)=E(X)\pm E(Y)$$
$$E\left(\sum(X_i)\right)=\sum E(X_i).$$
3. תוחלת של כפל/חילוק ≠ לכפל/חילוק התוחלות:
$$E(X\cdot Y)\neq E(X)\cdot E(Y)$$
$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}.$$
$$E\left(X^2\right)\neq\left[E(X)\right]^2$$
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 השונויות:
$$V(X\pm Y)=V(X)+V(Y)$$
$$V\sum(X_i)=\sum V(X_i).$$
2. עבור X ו- Y תלויים/מתואמים מתקיים:
שונות של סכום/הפרש ≠ סכום השונויות:
$$V(X\pm Y)=V(X)+V(Y)\pm 2\cdot Cov(X,Y).$$

3. שונות של קבוע = 0: $V(a)=0$, $V(a\pm x)=V(X)$.

4. השפעת טרנספורמציה ליניארית על השונות: $V(aX+b)=a^2V(X)$.

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

<u>חוקי השונות המשותפת</u>:

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

המודל האקונומטרי:

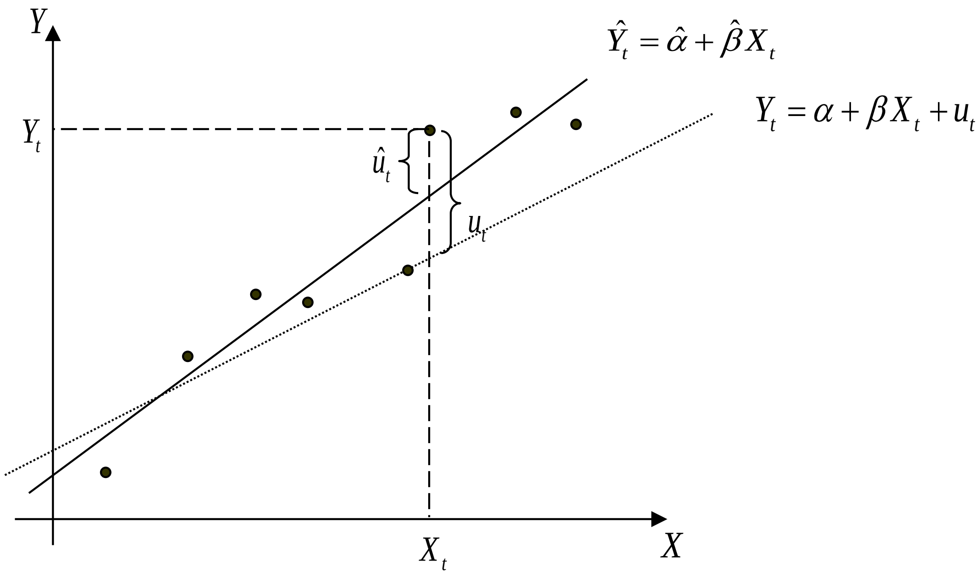
$$Y=\alpha+\beta X+u$$

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

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 ו- $\hat{\beta}$ הוא האומד ל- β.
3. אומדי ריבועים פחותים (אר"פ) הם אומדים שחושבו בשיטת 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 כיוון שבכל מדגם מתקבלים $\hat{\alpha}$ ו-$\hat{\beta}$ אחרים.
5. את α ו-β ו-u_t לא ניתן לדעת (אלא רק לאמוד מנתוני המדגם) – הקו האמיתי באוכ׳ לא ידוע.
6. אפשר לדעת את $\hat{u}_t$, שהיא הסטיה מקו הרגרסיה במדגם:

- עבור X_t, הערך הצפוי של המשתנה המוסבר $(\hat{Y}_t)$ המתקבל לפי הרגרסיה הוא: $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$.
- הסטיה של התצפית (Y_t) מהערך הצפוי לפי הרגרסיה $(\hat{Y}_t)$ היא: $\hat{u}_t = Y_t - \hat{Y}_t$.



— קו הרגרסיה הנאמד (במדגם)
..... קו הרגרסיה האמיתי באוכלוסייה (
• תצפית בודדת

שאלות:

1) הבא נוכיח את הזהויות הבאות:

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

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

ג. $\sum\left(X_t-\bar{X}\right)=0$.

ד. $\sum\frac{\left(X_t-\bar{X}\right)^2}{\sum\left(X_t-\bar{X}\right)X_t}=1$.

ה. $\sum\left(\hat{\alpha}+\hat{\beta}x_i\right)\left(x_i-\bar{x}\right)=\hat{\beta}\left(x_i-\bar{x}\right)^2$.

ו. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}$.

ז. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t$.

2) בטא באמצעות: $\operatorname{var}(y),\operatorname{var}(x),\operatorname{cov}(x,y)$ והקבועים a ו- b את הביטויים הבאים:

א. $Var(ax)$.

ב. $Var(x+y)$.

ג. $Var(ax+b)$.

ד. $Cov(x,ay)$.

ה. $Cov(x+a,y+b)$.

ו. מקדם המתאם בין x ל- y.

תשובות סופיות:

1) הוכחה.

2) א. $a^2\operatorname{var}(x)$. ב. $\operatorname{var}(x)+\operatorname{var}(y)+2\operatorname{cov}(x,y)$. ג. $a^2\operatorname{var}(x)$. ד. $a\operatorname{cov}(x,y)$.

ה. $\operatorname{cov}(x,y)$. ו. $r=\frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\cdot\sqrt{\operatorname{var}(y)}}$.

אקונומטריקה למדעי הנתונים

פרק 2 - אומדי הריבועים הפחותים

תוכן העניינים

אומדי הריבועים הפחותים:

רקע:

Ordinary Least Squares (OLS) – שיטת האמידה של α ושל β לקבלת אומדים $\hat{\alpha}$ ו-$\hat{\beta}$ שיביאו למינימום את סכום ריבועי טעויות האמידה:

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

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

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

"המשוואות הנורמליות" מתקבלות בתהליך הגזירה של פונקציית הריבועים הפחותים וחייבות להתקיים על מנת שהפונקציה תתקיים $\left(\sum \hat{u}_t^2 = \min\right)$:

עבור המודל הקלאסי (עם חותך):

בגזירה של α : $\sum \hat{u}_t = 0$.

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

עבור מודל ללא חותך:

בגזירת β בלבד: $\sum \hat{u}_t \cdot x_t = 0$.

מן המשוואות הנורמליות נובעות:

1. התכונות הגיאומטריות:

   א. $\sum \hat{u}_i = 0$.

   ב. $\sum x_i \hat{u}_i = 0$.

- ברגרסיה <u>ללא שיפוע</u> מתקיימת רק התכונה הגיאומטרית הראשונה.
  ברגרסיה <u>ללא חותך</u> מתקיימת רק התכונה הגיאומטרית השנייה.

2. התכונות האלגבריות:

   א. $\operatorname{cov}(x_i, \hat{u}_i) = 0$.

   ב. $\operatorname{cov}(\hat{y}_i, \hat{u}_i) = 0$.

   ג. $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} = \bar{\hat{y}}$.

- התכונות האלגבריות תקפות עבור קו הרגרסיה הקלאסי (עם חותך ושיפוע) במדגם בלבד.

ההנחות הקלאסיות של מודל הרגרסיה:

1. קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.

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

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

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

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

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

7. ההפרעות האקראיות מתפלגות נורמלית: $u_t\approx N$.

תכונות האומדים:

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

1. לינאריות:
אר"פ ניתנים להצגה כטרנספורמציה לינארית של Y_t.
כדי ש-$\hat{\beta}$ למשל, יהיה אומד ליניארי צריך להתקיים: $\hat{\beta}=\sum W_t\cdot Y_t$.
כאשר W_t היא קומבינציה של ערכי X בדרך כלל. למשל: $\hat{\beta}=\dfrac{\sum X_t\cdot Y_t}{\sum X_t^2}$.
כדי להביא את האומד לצורה: $\tilde{\beta}=\sum w_t\cdot y_i$ נעזר בשוויון: $\dfrac{\sum 0}{\sum 0}=\sum\dfrac{0}{\sum 0}$.
אומד זה ניתן להצגה בצורה הבאה:
$$\hat{\beta}=\sum\frac{X_t}{\sum X_t^2}Y_t=\sum W_t\cdot Y_t$$
$$W_t=\frac{X_t}{\sum X_t^2}$$
לפיכך מדובר באומד ליניארי.

- שימו לב כי:
W_t אסור שיכלול את Y_t.
Y_t אסור שיהיה במכנה או בשורש/חזקה (אלא אם כן במודל הנתון הוא מצוי בשורש/חזקה).

2. חוסר הטיה:

אומד $\hat{\vartheta}$ מסוים יהווה אח"ה לפרמטר ϑ אותו הוא אומד באוכלוסייה אם מתקיים: $E(\hat{\vartheta})=\vartheta$.

כיצד יודעים אם אומד הוא חסר הטיה?

1. בשלב הראשון יש לבצע עבודת הכנה – מבטאים את האומד באמצעות הפרמטר האמיתי – מציבים במקום ה- Y_t את המודל ומפתחים אלגברית.

- יש לזכור כי:

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

x_t איננו משתנה מקרי (על פי הנחה מס' 4) ⇐ יוצא מחוץ לתוחלת ולשונות אך נשאר בתוך ה- $\sum$ ו- α β קבועים ⇐ יוצאים מחוץ לתוחלת, לשונות ול- $\sum$.

2. בשלב השני מפעילים תוחלת על האומד המפותח ואם התוחלת שווה לפרמטר האמיתי אז האומד חסר הטייה.

- חוסר הטייה מחייב את התקיימותן של הנחות (3) $E(u_t)=0$ לכל t ו- (4) $\operatorname{cov}(X_t,u_t)=0$.

3. יעילות:

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

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

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

כיצד מחשבים שונות של אומד?

חייבות להתקיים הנחות (4) $\operatorname{cov}(X_t,u_t)=0$, (5) $V(u_t)=\sigma_u^2$ לכל t

ו-(6) $\operatorname{cov}(u_t,u_s)=0$ לכל $t\neq s$. אם הן מתקיימות, מחשבים את השונות של האיברים המכילים את u_t מהפיתוח הקודם (לפי כללי הסיגמא והשונות).

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 האמיתי באוכלוסייה: $\begin{pmatrix}\hat{\theta}\to\vartheta\end{pmatrix}$ $(T\to\infty)$.
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 באוכלוסייה.

סיכום: השלבים להוכחת התכונות:

1. הוכחת ליניאריות.
2. הכנת האומד ⇐ להציב במקום Y_t את המודל האמיתי.
   במודל עם חותך: $Y_t=\alpha+\beta X_t+u_t$.
   במודל ללא חותך: $Y_t=\beta X_t+u_t$.
3. פיתוח האלגברה.
4. חישוב תוחלת, שונות, עקיבות.

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

שאלות:

גזירת ארפ:

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

א. נסחו את בעיית ה-OLS.

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

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

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

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

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

א. נסחו את בעיית ה-OLS.

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

ג. מצאו נוסחה לקבלת $\hat{\beta}$.

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

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

3) חוקר רצה לחקור האם ציוני IQ משפיעים על הציון באקונומטריקה ולכן אסף תצפיות מ-5 סטודנטים:

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

איזה מבין המודלים הבאים נאמד?

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

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

ג. $\hat{score}_i = \hat{\alpha}$.

ד. $\hat{score}_i = \bar{y}$.

4) נבדק הקשר שבין שכר לשעה שעובד מסוים מרוויח אצל מעסיק מסוים (X) לבין כמות העובדים שמועסקים אצל אותו מעסיק (Y) (הניחו שכר שווה בין העובדים אצל אותו המעסיק).
לשם כך נדגמו 10 מעסיקים באופן מקרי ונתקבלו התוצאות הבאות:

$$\bar{x}=35$$
$$\bar{y}=5.8$$
$$\sum_{i=1}^{10} x_i^2 = 19{,}100$$
$$\sum_{i=1}^{10} y_i^2 = 440$$
$$\sum_{i=1}^{10} x_i y_i = 2858.85$$

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

5) כלכלן החליט לאמוד את המודל: $y_i=\alpha+u_i$.
א. נסחו את בעיית ה-OLS.
ב. מצאו את תנאי סדר ראשון של בעיית ה-OLS.
ג. מצאו נוסחה לקבלת $\hat{\alpha}$.

6) חוקר רצה לבדוק את המודל: $y_i=\hat{\beta}x_i+u_i$ כאשר המשתנה התלוי הוא הציון במקרו והב״ת הוא ציוני IQ. לשם כך אסף תצפיות של 5 סטודנטים:

SCORE	IQ	ציון חזוי	e_i
80	100		
90	110		
95	110		
92			5
	102		3

מאמידת הרגרסיה התקבל כי: $\hat{\beta}=0.85$.
השלם את התאים הריקים בטבלה.

הנחות המודל:

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

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

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

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

ii. $E(u_i) = 0$

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

iv. $\bar{e} = 0$

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

vi. $\sum u_i = 0$

vii. $V(u_i) = \sigma_i$

viii. $S_s^{\,2} \neq 0$

ix. $\text{cov}(s, e) = 0$

x. $\text{cov}(\hat{y}_i, e) = 0$

8) חוקר מעוניין לאמוד את ההשפעה של נוכחות בתרגולים על הציון בקורס אקונומטריקה. לשם כך אמד את המשוואה: $score = \alpha + \beta attendance + u$
הועלתה הטענה כי מודל זה אינו מקיים את הנחה מס' 4 של אי תלות בין המשתנה הב"ת לטעויות $\left(\text{cov}(x_i, u_i) = 0\right)$. חווה דעתך על טענה זו.

ליניאריות:

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

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

11) כלכלן החליט לאמוד את המודל: $y_i = \alpha + \beta x_i + u_i$.
מי מהאומדים הבאים הוא ליניארי ומהן המשקולות:

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

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

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

ד. $\tilde{\beta} = \sum \frac{Y_i}{n}$.

ה. $\tilde{\beta} = \sum \frac{X_t}{Y_i}$.

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

ז. $\tilde{\beta} = \frac{\bar{Y}}{\bar{X}}$.

חוסר הטיה:

12) 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?
א. בדוק במודל עם חותך.
ב. בדוק במודל ללא חותך.

13) נתון המודל הבא: $y_i = \alpha + \beta\, x_i + u_i$ נתון בנוסף כי האומד ל-β הינו ליניארי וחסר הטיה.
איזה מן הטענות מתקיימת בהכרח:
א. $\sum w_i x_i = 1$.
ב. $\sum w_i x_i = 0$.
ג. $\sum w_i = 0$.

יעילות ועקיבות:

14) כלכלן הציע את האומד הבא עבור β : $\tilde{\beta}=\dfrac{y_9-y_5+y_2}{x_9-x_5+x_2}$.

א. בדוק האם האומד חסר הטיה עבור המודל הקלאסי.

ב. האם תשתנה תשובתך אם מדובר באומד ללא חותך?

ג. חשב את שונות האומד עבור מודל ללא חותך.

תרגול ממבחנים:

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

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

א. האומד $\tilde{\beta}$ הינו אומד חסר הטיה ל- β .	נכון / לא נכון
ב. האומד $\tilde{\beta}$ הינו אומד עקיב ל- β .	נכון / לא נכון
ג. האומד $\tilde{\beta}$ הינו אומד ליניארי ל- β .	נכון / לא נכון
ד. האומד $\tilde{\beta}$ הינו אומד יעיל ל- β .	נכון / לא נכון

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

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

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

א. האומד $\tilde{\beta}$ הינו אומד מוטה ל- β :	נכון / לא נכון / אי אפשר לדעת
ב. על סמך משפט גאוס-מרקוב ניתן להסיק כי $\tilde{\beta}$ איננו אומד יעיל יותר מאומד הריבועים הפחותים:	נכון / לא נכון / לא ניתן לדעת

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

17) נתון המודל: $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות.
(יש לשים לב המודל ללא חותך).
נתון האומד: $\tilde{\beta} = \dfrac{\sum X_t Y_t}{\sum \left(X_t - \bar{X} \right)^2}$.

א. מהי התוחלת של $\tilde{\beta}$?

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

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

ד. מהי השונות האמיתית של האומד: $\dfrac{\sum X_t Y_t}{\sum X_t^2}$?

18) בכל השאלות ההנחות הקלאסיות מתקיימות.
האומדים הם אר״פ, והמודל הוא: $Y_t = \alpha + \beta X_t + u_t$.

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

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

ג. אמידת המודל בשיטת
הריבועים הפחותים תיתן את
התוצאה: $\sum_{t=1}^{T} u_t = 0$.   נכון / לא נכון / אי אפשר לדעת

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

i. הוא בהכרח שלילי.

ii. הוא בהכרח חיובי.

iii. הוא בהכרח שווה לאפס.

iv. לא ניתן לקבוע את סימנו על סמך הנתונים הקיימים.

ה. סמן את הטענה הנכונה בהכרח:

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

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

iii. $\sum_{t=1}^{T} X_t u_t = 0$.

iv. אף אחת מהטענות הנ״ל אינה נכונה בהכרח.

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

ז. אומד חסר הטיה הוא אינו בהכרח גם אומד עקיב. נכון / לא נכון / אי אפשר לדעת

תשובות סופיות:

1) א. $\min_{\hat{\alpha}\hat{\beta}} \sum\left[y_t-\left(\hat{\alpha}+\hat{\beta}x_t\right)\right]^2$ ב. $\sum_{i=1}^{n}\hat{u}_i x_i=0 \Rightarrow \hat{\beta}$, $\sum_{i=1}^{n}\hat{u}_i=0 \Rightarrow \hat{\alpha}$.

ג. $\bar{y}-\hat{\beta}\bar{x}=\hat{\alpha}$, $\frac{S_{xy}}{S_{xx}}=\hat{\beta}$. ד. הוכחה. ה. ראה סרטון.

2) א. $\min_{\hat{\beta}} \sum\left[y_t-\left(\hat{\beta}x_t\right)\right]^2$ ב. $\sum_{i=1}^{n}\hat{u}_i x_i=0 \Rightarrow \hat{\beta}$.

ג. $\frac{\sum y_i x_i}{\sum x_i^2}=\hat{\beta}$. ד. הוכחה. ה. ראה סרטון.

3) אי'.

4) $\hat{y}_i=4.59$.

5) א. $\min_{\hat{\alpha}} \sum\left[y_t-\left(\hat{\alpha}\right)\right]^2$. ב. $\sum_{i=1}^{n}\hat{u}_i=0$. ג. $\hat{\alpha}=\bar{y}$.

6)

e_i	ציון חזוי	IQ	SCORE
-5	85	100	80
-3.5	93.5	110	90
1.5	93.5	110	95
5	97	114	92
3	86.7	102	89.7

7) א. ראה סרטון.
ב.i. הנחה, לא בהכרח מתקיים.
ii. הנחה, לא בהכרח מתקיים.
iii. הנחה, לא בהכרח מתקיים.
iv. נגזר מהמשוואה, מתקיים בהכרח.
v. נגזר מהמשוואה, מתקיים בהכרח.
vi. אף אחד מהשניים.
vii. אף אחד מהשניים.
viii. הנחה, לא בהכרח מתקיים.
ix. נגזר מהמשוואה, מתקיים בהכרח.
x. נגזר מהמשוואה, מתקיים בהכרח.

8) ראה סרטון.

9) כן.

10) לא.

11) א. ליניארי, $W_i=\frac{x_i}{\sum x_i^2}$.
ב. ליניארי, $W_i=\frac{(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$.
ג. לא ליניארי.
ד. ליניארי, $W_i=\frac{1}{n}$.
ה. לא ליניארי.
ו. ליניארי, $W_i=\frac{x_i}{\sum(x_i-\bar{x})^2}$.
ז. ליניארי, $W_i=\frac{1}{\sum x_i}$.

12) א. לא.
ב. כן.

13) א' ו-ג'.

14) א. מוטה.
ב. חסר הטיה.
ג. $\frac{1}{(x_9-x_5+x_2)^2}\sigma_u^2$.

15) א. נכון.
ב. לא נכון.
ג. נכון.
ד. לא נכון.
ה. $V(\tilde{\beta})=\frac{100\sigma_u^2}{\left(\sum_{t=51}^{100}X_t-\sum_{t=1}^{50}X_t\right)^2}$.

16) א. לא נכון.
ב. נכון.
ג. $V(\tilde{\beta})=\frac{T\sigma_u^2}{\left(\sum X_t\right)^2}$.

17) א. $E(\tilde{\beta})=\frac{\beta\sum X_t^2}{\sum(X_t-\bar{X})^2}$.
ב. לא נכון.
ג. לא נכון.
ד. $\frac{\sigma_u^2}{\sum X_t^2}$.

18) א. נכון.
ב. לא נכון.
ג. לא נכון.
ד. ii.
ה. i.
ו. לא נכון.
ז. נכון.

אקונומטריקה למדעי הנתונים

פרק 3 - מודלים לא ליניאריים

תוכן העניינים

מודלים לא ליניאריים:

רקע:

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

- המשתנה שיש בו LN השינוי בו יהיה באחוזים.

תזכורת של חוקי לוגים:

$$LN\left(e^{x}\right)=X$$
$$LN\left(X^{Y}\right)=Y\cdot LN\left(X\right)$$
$$LN\left(X\cdot Y\right)=LN\left(X\right)+LN\left(Y\right)$$
$$LN\left(\frac{X}{Y}\right)=LN\left(X\right)-LN\left(Y\right)$$

שאלות:

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

1. $MWAGE_t = 139.547 + 118.628 \cdot SCL_t$.
2. $MWAGE_t = -1445.08 + 1239.60 \cdot LN(SCL)_t$.
3. $LN(MWAGE)_t = 5.244 + 0.778 \cdot LN(SCL)_t$.
4. $LN(MWAGE)_t = 6.292 + 0.070 \cdot SCL_t$.

א. הסבירו את המשמעות של β בכל אחד מהמודלים.

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

2) נתונים תוצאות האמידה של המודלים הבאים:

1. $\hat{Y} = e^{4.5} \cdot X^{0.05}$.
2. $\hat{Y} = e^{4.5+0.05X}$.
3. $\hat{Y} = 4.5 + \dfrac{0.05}{X}$.
4. $\hat{Y} = \dfrac{1}{1+e^{4.5+0.05X}}$.

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

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

3) נתונים המודלים הבאים עבור התוצר במשק:

1. $Q_i = AK_i^{\beta_1} e^{u_i}$.
2. $Q_i = Ae^{\beta_1 L_i + u_i}$.
3. $Q_i = A + K_i^{\beta_1} + e^{u_i}$.
4. $Q_i = A + \dfrac{\beta_1}{L_i} + u_i$.
5. $Q_i = A + \beta_1 \sqrt{K_i} + u_i$.
6. $Q_i = e^{A+\beta_1 Ki + u_i}$.
7. $Q_i = A\left(\dfrac{K_i}{2} + 7\right)^{\beta_1} e^{u_i}$.
8. $Q_i = A + \beta_1 L_i + u_i$.
9. $Q_i = A + \beta_1 \left(\dfrac{K_i}{L_i}\right) + u_i$.

כאשר:

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

A - הוצאות צריכה על המוצר בהינתן רמת הכנסה אפסית.

K - הכנסת הפרט.

L - שנות לימוד.

א. מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?

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

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

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

4) נתון המודל הבא: $Q_i = \frac{A}{K_i^{\beta_1}} e^{u_i}$.

א. האם ניתן לאמוד את המודל בשיטת OLS?

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

ג. נאמד המודל הבא: $\ln(Q_i) = \alpha_0 + \alpha_1 \ln(K_i) + u_i$, והתקבלו התוצאות הבאות: $\hat{\alpha}_0 = 3$, $\hat{\alpha}_1 = 0.8$.

מהם האומדנים עבור: β_1 , A ?

5) נתון כי הקשר באוכלוסייה בין X ל- Y נתון על ידי המודל הבא: $\ln Y = \alpha + \beta \ln X + u$. נתון גם כי עבור המודל הנ"ל כל ההנחות הקלאסיות מתקיימות.

כלכלן הציע את האומד הבא עבור β: $\tilde{\beta} = \frac{\sum_{t=1}^{T}\left(\ln X_t - \ln \bar{X}\right)\ln Y_t}{\sum_{t=1}^{T}\left(\ln X_t - \ln \bar{X}\right)^2}$.

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

ג. האם האומד *blue*?

ד. מהי שונותו?

תשובות סופיות:

1) א.1. השינוי השולי. 2. אין משמעות כלכלית. 3. גמישות.
4. שיעור השינוי השולי.
ב.1. 0.912. 2. 0.77. 3. 0.778. 4. 0.861.

2) א.1. $\hat{\ln}(Y)=4.5+0.05\cdot\ln(X)$. 2. $\hat{\ln}(Y)=4.5+0.05X$.
3. אין צורך. 4. $\ln\left(\frac{1-\hat{Y}}{\hat{Y}}\right)=4.5+0.05X$.
ב.1. 98.45. 2. 121.51. 3. 4.50833. 4. 0.00816.

3) א. מודלים: 4, 5, 8 ו-9.
ב. מודלים: 1, 2, 6 ו-7.
ג.1. מסביר: $\ln(K_i)$, מוסבר: $\ln(Q_i)$. 2. מסביר: L_i, מוסבר: $\ln(Q_i)$.
3. אינו ליניארי. 4. מסביר: $\frac{1}{L_i}$, מוסבר: Q_i.
5. מסביר: $\sqrt{K_i}$, מוסבר: Q_i. 6. מסביר: K_i, מוסבר: $\ln(Q_i)$.
7. מסביר: $K_i=\frac{K_i}{2}+7$, מוסבר: $\ln(Q_i)$. 8. מסביר: L_i, מוסבר: Q_i.
9. מסביר: $\frac{K_i}{L_i}$, מוסבר: Q_i.
ד. מודלים: 1 ו-7.

4) א. לא. ב. $\ln(Q_i)=\ln(A)-\beta_1\ln(K_i)+u_i$.
ג. $A=20$, $\beta_1=-0.8$.

5) א. כן. ב. כן. ג. כן. ד. $V(\hat{\beta})=\frac{\sigma_u^2}{SS\ln x}$.

אקונומטריקה למדעי הנתונים

פרק 4 - רגרסיה מרובה ומולטיקוליניאריות

תוכן העניינים

רגרסיה מרובה ומולטיקולינאריות:

רקע:

מודל הרגרסיה המרובה:

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

כאשר:

Y_i = משתנה תלוי.

$X_{1i} \ldots X_{Ji}$ = משתנים ב"ת.

U_i = טעות מקרית המקיימת את כל ההנחות הקלאסיות.

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

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

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

אמידת מודל הרגרסיה המרובה:

1. שיטת הריבועים הפחותים:

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

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

2. המשוואות הנורמאליות:

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

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

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

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

דוגמא:

מקרה פרטי, מודל עם שני משתנים מסבירים:

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

הנוסחאות הנורמאליות הן:

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

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

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

מפתרון מערכת המשוואות נקבל את הנוסחאות הבאות לחישוב האומדים:

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

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

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

הערה:

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

שיפועי הרגרסיה המרובה זהים לשיפועי הרגרסיה הפשוטה:

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

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

מולטיקולינאריות:

מולטיקולינאריות מתייחסת למתאם בין המשתנים המסבירים במודל.

מולטיקולינאריות מלאה:
מתאם מלא בין המשתנים המסבירים במודל.
הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני: $x_1 = a + bx_2$ (x_1 הוא קומבינציה ליניארית מלאה של x_2)
מכאן ש: $r_{12} = 1$.

- שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל $x_1 = x_2^2$), אז בהכרח $r_{12} \neq 1$.
- מולטיקולינאריות מלאה יכולה להיווצר גם כאשר קבוצה של משתנים מסבירים מהווה קומבינציה ליניארית מלאה של אחד המשתנים המסבירים: $x_1 + x_2 = a + bx_3$.

במצב של מולטיקולינאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני ולא ניתן לאמוד את המודל שכן אר"פ אינם מוגדרים.
פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

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

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

מכיוון שיש מתאם גבוה בין המשתנים הב"ת לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי.
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.

שאלות:

רגרסיה מרובה:

1) כלכלן החליט לאמוד מודל ליניארי עם שלושה משתנים מסבירים: x_1, x_2, x_3.
$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.

א. מהי בעיית ה-OLS שעליו לפתור?

ב. מצאו את תנאי סדר ראשון של הבעיה.

2) כלכלן החליט לבחון מה משפיע על שער הדולר בישראל.
לכן אסף מדגם בין ארבע תצפיות חודשיות.
להלן טבלה מסכמת:

חודש	X_2 השקעות זרים בישראל (במיליוני דולרים)	X_1 שער הריבית	Y דולר	טעות (e^i)
אוגוסט	100	3	3.2	-5
ספטמבר	95	3.5	3.6	6
אוקטובר	90	3.5	3.8	0
נובמבר	100	3	3.5	-2

מהו המודל אשר אותו אמד הכלכלן?

3) הניחו כי הקשר באוכלוסייה בין X ל- Y נתון ע"י המשוואה הבאה: $Y_i = 2 + \beta_1 X_{1i} + 5X_{2i} + u_i$ וכל ההנחות הקלאסיות מתקיימות.

נתון האומד: $\tilde{\beta}_1 = \dfrac{\sum X_{1i}\left((Y_i - 8X_{2i} - 2) - \overline{(y - 8X_2 - 2)}\right)}{\sum X_{1i}^2}$.

א. חשבו את תוחלת האומד.

ב. חשבו את שונות האומד.

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

4) הניחו כי הקשר באוכלוסייה בין X ל- Y נתון ע"י המשוואה הבאה: $y_i = \alpha + \beta_1 x_{1i} + 8x_{2i} + u_i$.
כל ההנחות הקלאסיות מתקיימות וכן: $\sum x_{1i} = 0$.
אומדים את β_1 באופן הבא: $b_1 = \dfrac{\sum (x_{1i} - \bar{x})\left(y_i - 8x_{2i} - \left(\overline{y - 8x_2}\right)\right)}{\sum (x_{1i} - \bar{x})^2}$.

א. האם האומד חסר הטיה?
ב. מהי שונות האומד?

מולטיקולינאריות:

5) נתון המודל: $Y_i = \alpha + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + U_i$.
חוו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

א. בהנחה כי מתקיים: $X_{1i} - 2X_{2i} = 1$,
לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים. נכון/לא נכון/ לא ניתן לדעת.

ב. בהנחה כי מתקיים: $x_{1i} = x_{2i}^2$,
לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים. נכון/לא נכון/ לא ניתן לדעת.

ג. הוכיחו תשובותיכם לסעיפים ב' ו-ג'.

ד. בהנחה כי מתקיים: $r_{12} = 0.98$,

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

ii. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

6) כלכלן אמד את המודל: $y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.
בשל החשש ממולטיקוליניאריות בחן הכלכלן את המתאם בין כל זוג של משתנים מסבירים וקיבל: $r_{x_1,x_2} = 0.9$, $r_{x_1,x_3} = 0.99$, $r_{x_3,x_2} = 0.5$.
לכן הסיק כי אין בעיה של מולטיקוליניאריות מושלמת במודל.
האם הוא צודק?

7) כלכלן אמד את המודל הבא: $\ln(Q_i) = \alpha + \beta_1 \ln(K_i) + \beta_2 \ln(K_i^2) + \beta_3 L_i^{0.5} + u_i$.
האם קיימת בעיה של מולטיקוליניאריות במודל?

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

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

בנוסף למשתנים במשוואה, החליט החוקר להוסיף גם את משתנה הוותק : EXP_i.

מכיוון שלא היו בידו נתונים על הוותק, החליט החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים (מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).

להלן משוואה מס' 2 :

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

חווה דעתך על המשוואה השנייה.

תשובות סופיות:

1) א. $\min \sum \left(y_i - \hat{\alpha} - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}\right)^2$.

ב. $\sum_{i=1}^{n} e_i = 0$, $\sum_{i=1}^{n} e_i X_{1i} = 0$, $\sum_{i=1}^{n} e_i X_{2i} = 0$, $\sum_{i=1}^{n} e_i X_{3i} = 0$.

2) $y_i = \beta_1 x_{1i} + u_i$.

3) א. לא ניתן לחשב. ב. $Var = \frac{\sigma^2}{\sum X_{1i}^2}$. ג. לא ניתן לדעת.

4) א. כן. ב. $V(b_1) = \frac{\sigma^2}{\sum x_{1i}^2}$.

5) א. נכון. ב. לא נכון. ג. הוכחה. ד.i. לא נכון.
ii. מולטיקולינאריות חלקית.

6) הכלכלן לא יכול להיות בטוח.

7) כן.

8) ראו סרטון.

אקונומטריקה למדעי הנתונים

פרק 5 - מבחן t

תוכן העניינים

מבחן t:

רקע:

המבחן הסטטיסטי למובהקות מקדמי הרגרסיה.

המבחן הסטטיסטי	ניסוי	השערות	סטטיסטי המבחן	כלל הכרעה לדחיית H_0	מסקנה
מובהקות השיפוע	האם משתנה מסביר מסוים רלוונטי למודל / משפיע על התלוי?	$H_0:\beta=0$ $H_1:\beta\neq 0$	$t_{\beta=0}=\frac{\hat{\beta}-\beta_0}{S_{\hat{\beta}}}$	שימוש בטבלת T: $\left\|t_{\beta=0}\right\|>t_{\left(n-K,\frac{\alpha}{2}\right)}$ K=**מספר מקדמים (כולל חותך)	יש/אין עדות לכך שהמשתנה הב"ת מובהק באוכ'
מבחן חד צדדי לשיפוע	האם מקדם השיפוע חיובי/שלילי באוכ'?	$H_0:\beta=0$ $H_1:\beta>/<0$		שימוש בטבלת T: $t_{\beta=0}>t_{(n-K,\alpha)}$ $t_{\beta=0}<-t_{(n-K,\alpha)}$	יש/אין עדות לכך שהשיפוע חיובי/שלילי באוכ'
השיפוע = ערך מסוים באוכ'	האם מקדם השיפוע = לערך מסוים (למשל ל-2)?	$H_0:\beta=2$ $H_1:\beta\neq 2$	$t_{\beta=2}=\frac{\hat{\beta}-2}{S_{\hat{\beta}}}$	שימוש בטבלת t	יש/אין עדות לכך שהשיפוע = ל-2 באוכ'.
מבחן למובהקות החותך **ניתן לבצע גם מבחן חד צדדי ושהחותך = לערך מסוים באוכ'	האם קו הרגרסיה יוצא מראשית הצירים?	$H_0:\alpha=0$ $H_1:\alpha\neq 0$	$t_{\alpha=0}=\frac{\hat{\alpha}-\alpha_0}{S_{\hat{\alpha}}}$	נדחה את H_0 אם: שימוש בטבלת T: $\left\|t_{\alpha=0}\right\|>t_{\left(n-K,\frac{\alpha}{2}\right)}$	יש/אין עדות לכך שקו הרגרסיה עובר דרך ראשית הצירים

רב"ס ל- β : $P\left(\hat{\beta}-t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\beta}}\le\beta\le\hat{\beta}+t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\beta}}\right)=1-\alpha$.

רב"ס ל- α : $P\left(\hat{\alpha}-t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\alpha}}\le\alpha\le\hat{\alpha}+t_{\left(n-K,\frac{\alpha}{2}\right)}\cdot S_{\hat{\alpha}}\right)=1-\alpha$.

- ניתן לבדוק השערות באמצעות הרב"ס.
  צריך לבדוק האם הרב"ס מכיל את הערך המבוקש לפי השערת האפס.
  אם כן – נקבל את H_0 ואם לא – נדחה אותה.

תחזית:

המטרה של קו הרגרסיה הוא ביצוע תחזיות:
<u>תחזית נקודתית</u> מחושבת על פי קו הרגרסיה שאמדנו.
נציב במקום ה- X ים ערכים נתונים ונקבל למה שווה ה- Y המנובא.

אמידת התחזית באוכלוסייה עבור ערך מסוים של X (ברגרסיה פשוטה):

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

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

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

התחזית מדויקת יותר (שונות התחזית קטנה יותר) כאשר:

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

מבחן t מורכב (בחינת קשרים ליניאריים בין הפרמטרים):

משמש לבדיקת השערות העוסקות בקשרים בין הפרמטרים.

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

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

ונחשב את סטטיסטי המבחן t: $t_{\hat{\alpha}-5\hat{\beta}}=\frac{\left(\hat{\alpha}-5\hat{\beta}\right)-0}{S_{\hat{\alpha}-5\hat{\beta}}}$ או $t_{\hat{\beta}_1-2\hat{\beta}_2}=\frac{\left(\hat{\beta}_1-2\hat{\beta}_2\right)-0}{S_{\hat{\beta}_1-2\hat{\beta}_2}}$.

כאשר את טעות התקן של המבחן מחשבים תוך שימוש בנוסחאות:

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

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

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

ואחר כך מוציאים לשונות שורש כדי לקבל את סטית התקן.

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

שאלות:

מובהקות מקדמי הרגרסיה:

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

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

(0.08953) (0.01622).

סטיות התקן של האומדים נתונות בסוגריים.

א. מהי המשמעות הכלכלית של β ושל α?

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

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

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס ברמת מובהקות של 5% וברמת מובהקות של 1%.

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

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

2) חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר ($SALARY$) לפי המודל: $\ln(SALARY_t) = \alpha + \beta \cdot EXP_t + u_t$. הוא אסף 403 תצפיות, ואמד את הפרמטרים. להלן תוצאות האמידה:

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

(0.068) (0.0026).

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

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

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

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

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

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

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

ב. בנו רווח בר סמך להשפעת X על Y.

תחזית:

4) נתונה משוואת הרגרסיה הבאה: $\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i}$.
כאשר y_i הינו סה"כ הוצאות משק בית i לחינוך לשבוע, x_{ji} הינו גילו של הילד j.
מה יהיה סה"כ הוצאות משק הבית אם גיל הילד הראשון הוא 2 שנים, של השני 4.5 שנים, השלישי הוא בן 5 ואילו הרביעי בן 8?

5) במדגם של 30 דירות המושכרות לסטודנטים ברדיוס של עד 2 ק"מ מסביב למכללה נחקר הקשר בין שכר דירה למספר הסטודנטים הגרים בדירה.
להלן תוצאות האמידה: $\hat{Y}_i = 686.207 + 233.52 \cdot X_i$.
נתון בנוסף כי:

$S_x^2 = 1.313^2$

$S_u^2 = 414.055^2$

$\bar{x} = 3$

א. חשבו אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.
ב. אמוד את שכר הדירה שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.

t מורכב:

6) נתון המודל: $Y_t = \alpha + \beta X_t + u_t$ שלצורך אמידתו נאספו 240 תצפיות ונתקבל ש:

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

(0.12) (0.25).

נתון בנוסף כי: $\text{cov}\left(\hat{\alpha}, \hat{\beta}\right) = -0.003$.
יש לבדוק את ההשערה: $H_0: \alpha = 5\beta$.

7) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$.
להלן תוצאות האמידה של המשוואה הנ"ל:

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

(0.4) (0.0678).

נתון גם ש: $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.009$.
יש לבדוק את ההשערה שהנטייה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון.

תרגול מסכם:

8) כלכלן בנה עבור מכבי ת"א מודל החוזה את השכר שיש לשלם לשחקן כדורסל לחוזה של שנה: $\hat{Y}_i = \hat{\alpha} + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + u_i$.

כאשר:

Y: שכר השחקן באלפי $.

X_1: מס' נקודות שקולע השחקן בממוצע למשחק.

X_2: מס' האסיסטים שיש לשחקן בממוצע למשחק.

X_3: מס' הדקות שיושב שחקן על הספסל בממוצע למשחק.

הכלכלן דגם 34 משחקים וקיבל את התוצאות הבאות:

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

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

**הערכים שבסוגריים הם ערכי t.

התקבל בנוסף כי: $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = 4, \text{cov}(\hat{\beta}_1, \hat{\beta}_3) = -3, \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -6$.

א. תנו פירוש למקדמי הרגרסיה.

ב. איזה מהמשתנים הב"ת רלוונטי למודל?

ג. בנו רב"ס למשתנים המובהקים.

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

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

9) כלכלן אמד את המודל הבא: $\ln(Q_i) = \alpha + \beta_1 \ln(K_i) + u_i$ שמתאר את הקשר שבין צריכת מוצר מסוים להכנסת הפרט (עקומת אנג'ל):

K - הכנסה חודשית באלפי שקלים.

Q - צריכה שנתית באלפי שקלים.

לשם כך אסף 60 נתונים והריץ רגרסיה.

התוצאות אשר קיבל הן: $\hat{\beta} = -2$, $\hat{\alpha} = 4$, $t_{\hat{\beta}} = -7$, $t_{\hat{\alpha}} = 3$, $\text{cov}(\hat{\alpha}, \hat{\beta}) = -0.05$

, $S_Q = 0.05$, $S_K = 1.5$.

נקודת הממוצעים הינה: (6.7,0.4).

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

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

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

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

תשובות סופיות:

1) א. ראה סרטון. ב. כן. ג. אין עדות לכך. ד. יש עדות לכך.
ה. $P(0.12 \le \beta \le 0.184) = 0.95$. ו. ניתן לדחות את השערת האפס.

2) א. לא. ב. אין עדות לכך. ג. $\hat{Y} = 1404$.

3) א. כן. ב. $P(0.13 \le \beta \le 1.81) = 0.95$.

4) 142.5 ₪ לשבוע.

5) א. 1153.247. ב. $P(282.94 \le Y_{X=2} \le 2023.55) = 0.95$.

6) אין עדות לכך.

7) אין עדות לכך.

8) א. ראה סרטון. ב. כל שלושת המשתנים.
ג. $P(6 \le \beta_1 \le 30) = 0.95$, $P(4.364 \le \beta_2 \le 11.636) = 0.95$, $P(-30.8 \le \beta_3 \le -13.2) = 0.95$.
ד. 940 אלף $. ה. כן.

9) א. אין עדות לכך. ב. יש עדות לכך. ג. 545 ₪.
ד. $P(-6.205 \le Q_i \le 7.295) = 0.95$.

אקונומטריקה למדעי הנתונים

פרק 6 - מבחן F ו R בריבוע

תוכן העניינים

מבחן F ו-R ברביבוע:

רקע:

מדד R^2 לטיב הרגרסיה:

מדד לפרופורציית השונות המוסברת:

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

מתבסס על הנוסחה לפירוק השונות של קו הרגרסיה:

$$TSS=RSS+ESS$$

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

תכונות R^2:

- נע בין 0 ל-1: $0\le R^2\le 1$.

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

  כאשר $R^2=0$ הכל טעות ואין שום הסבר במודל.
- אר״פ מביא למקסימום את R^2.
- לא ניתן להשוות במדד בין מודלים שבהם אין את אותו משתנה מוסבר.
- בהוספת משתנים מסבירים נוספים למודל, R^2 יכול רק לעלות או להישאר ללא שינוי. זהו למעשה החיסרון הגדול של המדד.

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

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

  K = מס׳ הפרמטרים במודל (כולל החותך).
- המדד המתוקנן לוקח בחשבון את מספר המשתנים הב״ת שיש במודל ויכול לרדת בהוספת משתנים למודל לכן מתקיים תמיד ש: $\bar{R}^2<R^2$.
- המדד המתוקנן – $\bar{R}^2$ עדיף על המדד – R^2 בכדי לבחון האם כדאי לנו להוסיף משתנים ב״ת למודל.

זהויות שכדאי לדעת לגבי R^2 :

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

1. $R^2 = r_{yx}^2$.
2. $r_{yx} = \hat{\beta}\frac{S_x}{S_y}$.
3. $R^2 = \hat{\beta}^2\frac{S_x^2}{S_y^2}$.
4. במודלים: $\begin{matrix} y_i = \alpha_1 + \beta_1 x_i + u_i \\ x_i = \alpha_2 + \beta_2 y_i + \varepsilon_i \end{matrix}$ מתקיים:
   i. הם בעלי אותו R^2 .
   ii. $R^2 = \beta_1 \cdot \beta_2$.

שימו לב:

1. במודל ללא שיפוע: $y_i = \alpha + u_i$, ה- R^2 שווה ל-0 כי אין מקדם הסבר לרגרסיה.
2. במודל ללא חותך: $y_i = \beta x_i + u_i$ אין משמעות ל- R^2 כיוון שלא מתקיימת המשוואה הנורמלית הראשונה: $\sum \hat{u}_i = 0$ ולכן גם $\bar{\hat{y}} \neq \bar{y}$ ולכן גם לא מתקיים: $SST = SSR + SSE$.

מבחן F:

משמש לבדיקת:

1. הגבלות שונות המתקיימות במודל (מבחן WALD).
2. מובהקות מודל הרגרסיה כולו.

מבחן WALD:

לבדיקת השערת אפס שיש בה מספר שוויונים (במבחן t היה רק שוויון אחד בהשערת האפס).

1. אומדים את המודל המקורי – הלא-מוגבל (Unrestricted) ומקבלים את סכום ריבועי הסטיות של הטעויות $\left(\sum e_{i\ UR}^2\right)$.
2. מגדירים את כל השוויונים של השערת האפס.
3. מציבים את השוויונים של השערת האפס במודל המקורי לקבלת המודל המוגבל (Restricted).
4. אומדים את המודל המוגבל ומקבלים את סכום ריבועי הסטיות של הטעויות $\left(\sum e_{i\ R}^2\right)$.
5. חישוב הסטטיסטי: $\frac{\left(\sum e_{\text{מ}}^2 - \sum e_{\text{מל}}^2\right)/m}{\sum e_{\text{מ}}^2/(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

(כש- m מספר המגבלות ו- k מס' הפרמטרים במודל הלא מוגבל).

- כאשר לשתי הרגרסיות (המוגבלת והלא מוגבלת) אותו משתנה מוסבר ניתן להשתמש גם בנוסחה הבאה: $\frac{\left(R_{\text{מ}}^2 - R_{\text{מ}}^2\right)/m}{\left(1-R_{\text{מ}}^2\right)/(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

כלל הכרעה לדחיית H_0: $F_{stat} > F_{(m,n-k;1-\alpha)}$.

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

מבחן F למובהקות המודל:

משמש לבדיקה האם מודל הרגרסיה שלנו לניבוי משתנה תלוי מסוים על ידי המשתנים הב"ת, מובהק באוכלוסייה.

השערות: $\begin{matrix} H_0: \beta_1 = \beta_2 = \ldots = \beta_K = 0 \\ H_1: OTHERWISE \end{matrix}$.

המודל הלא מוגבל יהיה: $Y_t = \alpha + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + u_t$ U:

המודל המוגבל יהיה: $Y_t = \alpha + u_t$ R:

מאחר ו- $R_{\text{מ}}^2 = 0$ ו- $m = k-1$: $F = \dfrac{\dfrac{R_U^2 - R_R^2}{m}}{\dfrac{1-R_U^2}{n-k}} = \dfrac{\dfrac{R_U^2}{k-1}}{\dfrac{1-R_U^2}{n-k}}$.

הערה :
בדיקת מובהקות המודל ברגרסיה מרובה ניתנת לביצוע רק על ידי מבחן F מאחר ויש יותר ממגבלה אחת בהשערת האפס.
לעומת זאת בדיקת מובהקות המודל ברגרסיה חד משתנית ניתנת לביצוע גם על ידי מבחן t שכן יש רק מגבלה אחת בהשערת האפס : $F = t^2$.

לסיכום :

1. מתי נשתמש במבחן t ומתי במבחן F?
   - רק t : השערות חד צדדיות (סימן אי שוויון בהשערות).
   - t או F (כאשר : $F = t^2$) : מגבלה אחת (שוויון אחד בלבד) בהשערת האפס.
   - רק F : כאשר יש כמה מגבלות (שוויונים) בהשערת האפס.
2. מצב של סתירה בין מבחן F למבחני t :
   כאשר המודל מובהק אולם אף אחד מהשיפועים לא יוצא מובהק – בעיה של מולטיקוליניאריות חלקית במודל (מתאמים גבוהים בין המשתנים הב״ת).

שאלות:

R בריבוע:

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

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

2) על סמך מדגם של 100 תצפיות נאמדו המודלים הבאים:

1. $\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i}$.
2. $\hat{y}_i = \hat{\delta}_0 + \hat{\delta}_1 x_{1i} + \hat{\delta}_2 x_{2i}$ $R^2 = 0.70$.
3. $\hat{y}_i = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2i}$ $R^2 = 0.65$.

א. שלושה חוקרים העלו טענה לגבי מקדם R^2 של משוואה מסי (1):

1. אי אפשר לדעת מהנתונים המובאים לעיל אם R^2 של משוואה (1) הוא גדול או קטן מ-0.70.
2. אי אפשר לדעת מהנתונים המובאים לעיל אם R^2 של משוואה (1) הוא גדול או קטן מ-0.65.
3. ניתן לצפות כי R^2 של משוואה (1) יהיה גדול מ-0.70.

בהתייחס לטענות החוקרים ניתן לומר:

i. רק הטענה של חוקר 1 נכונה.
ii. רק הטענה של חוקר 2 נכונה.
iii. רק הטענה של חוקר 3 נכונה.
iv. כל הטענות שגויות.

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

i. ניתן לצפות ש- $\bar{R}^2$ של משוואה (1) יהיה גדול מ-0.7.   נכון/לא נכון/ לא ניתן לדעת
ii. ניתן לצפות כי $\bar{R}^2$ של משוואה (2) יהיה קטן מ-0.7.   נכון/לא נכון/ לא ניתן לדעת
iii. ניתן לצפות כי $\bar{R}^2$ של משוואה (3) יהיה קטן מ-0.7.   נכון/לא נכון/לא ניתן לדעת

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

1. $\hat{y}_i = 5 + 2x_{1i} + 2x_{2i}$ $R^2 = 0.77$.
2. $\hat{y}_i = 24 + 0.8x_{1i}$ $R^2 = 0.62$.
3. $\hat{y}_i = 14 + 0.7x_{2i}$ $R^2 = 0.25$.
4. $\hat{y}_i = 4 + 0.5w_i$ $R^2 = 0.30$.
5. $\hat{\ln}(y)_i = 7 + 0.9x_{1i} + 0.6x_{2i}$ $R^2 = 0.45$.
6. $\hat{\ln}(y)_i = 11 + 0.7x_{1i} + 0.9x_{2i} + 0.6x_{3i}$.
7. $\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i} + 9x_{4i}$.

כאשר y_i הינו סה"כ הוצאות משק בית i, x_{ji} הינו גילו של הילד j, ונתון כי: $w_i = 2x_{3i} + x_{1i} - x_{2i}$.
דרגו את הרגרסיות לפי קריטריון R^2 (הימני עדיף על השמאלי).

4) נתונות שתי המשוואות הבאות: $y_i = 58 + b_1x_i + e_{1i}$ ו- $x_i = a_2 - 0.2y_i + e_{2i}$,
כאשר: $\bar{y} = \bar{x} = 40$. למה שווה מקדם המתאם של פירסון בין X ל- Y?

א. 0.09.
ב. 0.69.
ג. 0.3.
ד. 0.72.
ה. אף תשובה לא נכונה.

5) נתון מודל רגרסיה: $y_i = \alpha + \beta x_i + u_i$.
הוכיחו כי: $SST = SSR + SSE$.

מבחן F:

6) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבל כי: $\sum e^2 = 620.1683$ וכי: $R^2 = 0.99$.
הועלתה ההשערה כי ההשפעה על Y של משתנה S היא פי 3 מזו של משתנה Z, וכן כי החותך הוא 5.
א. מהי השערת האפס?
ב. מהו המודל המוגבל שאותו צריך לאמוד?

מאמידת המודל המוגבל התקבל כי: $\sum e^2 = 623.99$ וכי: $R^2 = 0.99$.
ג. חשב את הסטטיסטי של WALD.
ד. כמה דרגות חופש יש במונה וכמה במכנה?
ה. האם דוחים או מקבלים את השערת האפס?

7) במדגם של 82 תצפיות התקבל: $y_i = 12 + 3x_{1i} + 4x_{2i} + e_i$ $R^2 = 0.73$.
א. בחנו את ההשערה כי: $\begin{matrix} H0: \beta_2 = 0 \\ H1: \beta_2 \neq 0 \end{matrix}$.
כאשר נתון כי לאחר אמידת המודל המוגבל התקבל כי: $R^2 = 0.6$.
ב. חשבו את $S_{\hat{\beta}_2}$.

8) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot W_i + \beta_2 \cdot P_i + u_i$.
מתוצאות האמידה של המשוואה הנ"ל התקבל כי: $\sum e^2 = 52968$.
על מנת לבדוק את ההשערה שהנטייה השולית לצרוך (נש"צ) מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון נאמדה גם המשוואה הבאה: $C_i = \alpha + \beta_1 \cdot Y_i + u_i$ כאשר: $Y_i =$ סה"כ ההכנסה של משק בית t $(W_i + P_i)$.
התקבל: $\sum e^2 = 54156$.
א. בדקו את ההשערה.
ב. חשבו את סטטיסטי t לבדיקת ההשערה.

מבחן F למובהקות המודל:

9) נתון המודל: $y = A \frac{x_{1i}^{\beta_1}}{x_{3i}^{\beta_3}} e^{\beta_2 x_2} e^{u_i}$.
באמידת מדגם של 58 נבדקים התקבל: $R^2 = 0.56$.
האם המודל מובהק?

תרגול מסכם:

10) נאמדו חמשת המודלים הבאים על 70 תצפיות:

1. $ESS=130 \quad I_i = 12+0.13\cdot \exp_i + 0.08\cdot scl_i + 2\cdot workh_i + u_i$.
2. $ESS=150 \quad I_i = 11+0.1\cdot scl + 0.1\cdot workh_i + u_i$.
3. $ESS=151 \quad I_i = 9+0.22\cdot scl + u_i$.
4. $ESS=152 \quad I_i = 15+0.15\cdot workh_i + u_i$.
5. $ESS=200 \quad I_i = 25+u_i$.

המשתנה המוסבר הוא הכנסה מעבודה (I) והמשתנים המסבירים שבחנו הם מספר שנות הלימוד (scl), מספר שעות עבודה ($workh$) וותק בעבודה (exp)

הערה: הניחו כי ערך F הקריטי הוא 4.

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

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

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

ד. האם השפעת הוותק יכול להיות 0.15?

ה. כלכלן נוסף הציע להריץ את המודל:
$I_i + \exp_i = 2 - 3(scl_i - workh_i) + u_i \ , \ ESS = 145$.
איזו השערה ניתן לבחון באמצעות מודל זה?
כמה דרגות חופש יש לסטטיסטי שנקבל? בחנו אותה.

11) על סמך מדגם של 40 תצפיות נאמדו המשוואות הבאות:

1. $y_i = 2+3X_{1i}+4X_{2i}+e_i \quad R^2 = 0.76$.
2. $y_i = 3+5D_i+e_i \quad R^2=0.60$.
3. $D_i = 0.2X_{1i}+X_{2i}$.

כאשר Y הינו הציון בתואר ראשון, X_1 ציוני הבגרות ו-X_2 ציוני הפסיכומטרי.

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

ב. בדקו את ההשערה כי רגרסיה 2 מובהקת.

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

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

1. $\hat{y}_i = 5 + 2X_{1i} + 2X_{2i}$ $R^2 = 0.6$.
2. $\hat{y}_i = 11 + 0.9x_{2i} + 0.6x_{3i}$ $R^2 = 0.45$.
3. $\hat{y}_i = 13 + 8x_{1i} + 7x_{2i} + 2x_{3i}$ $R^2 = 0.78$.

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

חשבו את האומדן לסטיית התקן של המקדם X_3 ברגרסיה 3.

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

א. מהי המשוואה לאמידת המקדמים של המודל?

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

ג. מהן דרגות החופש במונה ובמכנה?

ד. רשמו את הנוסחה לחישוב סטטיסטי המבחן.

14) המודל הבא מתאר את פונקציית הייצור של מוצר P:

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

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

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

תשובות סופיות:

1) $4>5>1=3>2$.

2) א. iii. ב. i. לא ניתן לדעת. ii. נכון. iii. נכון.

3) 7 , 1 , 2 , 4 , 3 ובאופן נפרד: 6 , 5.

4) ג'.

5) הוכחה.

6) א. $H_0: \alpha = 5\ ,\ \beta_s = 3\beta_z$. ב. $Y_t - 5 = \beta_x X_t + \beta_z (Z_t + 3S_t) + \beta_w W_t + u_t$.
ג. $F = 0.6145$. ד. מונה: 2, מכנה: 199. ה. מקבלים.

7) א. יש עדות לכך. ב. $S_{\hat{\beta}_2} = 0.645$.

8) א. אין עדות לכך. ב. $t = 0.934$.

9) יש עדות לכך.

10) א. אין עדות לכך. ב. אין עדות לכך. ג. יש עדות לכך.
ד. כן. ה. $H0;\ \beta_{\exp} = -1\ ;\ \beta_{work} = -\beta_{scl}$.

11) א. יש עדות לכך. ב. יש עדות לכך. ג. $\beta_1 = 0.2\beta_2$.

12) $S.E = 0.25$.

13) א. $\ln(y_i) = \ln(A) + \beta_1 \ln(X_{1i}) + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$.
ב. $\ln(y_i) = \ln(A) + \beta_3 \left(\ln(X_{1i}) + 3X_{2i} + X_{3i}\right) + u_i$.
ג. מונה: $m=2$, מכנה: $n-k = n-4$.
ד. $F = \dfrac{\dfrac{R_U^2 - R_R^2}{m}}{\dfrac{1-R_U^2}{n-k}}$.

14) $\ln\left(\dfrac{P_i}{S_i}\right) = \alpha + \beta_J \ln\left(\dfrac{J_i}{S_i}\right) + \varepsilon_i$.

אקונומטריקה למדעי הנתונים

פרק 7 - שינוי יחידות מדידה

תוכן העניינים

שינוי יחידות מדידה:

רקע:

טרנספורמציה ליניארית: הוספה/החסרה של קבוע ו/או הכפלה/חילוק של קבוע של אחד או שני המשתנים (התלוי והב"ת).

- טרנספורמציה ליניארית של המשתנים לא תשפיע על: R^2, F, $t_{\hat{\beta}}$ ו- PF.

השינויים מסוכמים בטבלה הבאה:

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

- $t_{(\hat{\beta}'=0)}=t_{(\hat{\beta}=0)}$ תמיד.
- $t_{(\hat{\alpha}'=0)}=t_{(\hat{\alpha}=0)}$ רק בהכפלות.

שאלות:

1) חוקר ביקש לאמוד את הקשר בין שכר ב-₪ ($MWAGE$) לבין שנות לימוד (SCL) באמצעות 2 מודלים שונים.
להלן תוצאות האמידה:

א. $MWAGE_t = 139.54 + 118.62 \cdot SCL_t$.

ב. $MWAGE_t = -1445.08 + 1239.60 \cdot LN\left(SCL\right)_t$.

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

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

2) בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:
החוקר טען כי יש לבדוק את הקשר בין שכר לוותק ע"י שימוש בשכר נטו (NET) ולא בשכר ברוטו ($SALARY$). (קיים שיעור מס קבוע של 20%).
המודל הוא: $\ln\left(NET_t\right) = \alpha' + \beta' \cdot EXP_t + v_t$.
מה יהיו ערכי האומדים, סטיות התקן שלהם וטיב ההתאמה באמידת מודל זה?

תשובות סופיות:

1) 1. א. $\hat{\alpha}' = \alpha = 139.59$, $\hat{\beta}' = 98.85$, סטטיסטי F לא משתנה.
ב. $\hat{\alpha}' = -1671$, $\hat{\beta}' = 1239.6$, סטטיסטי F לא משתנה.
2. לא ניתן לדעת.

2) $S_{\hat{\alpha}'} = S_{\hat{\alpha}} = 0.0688935$, $S_{\hat{\beta}'} = S_{\hat{\beta}} = 0.0026235$, $\hat{\alpha}' = 7.11161$, $\hat{\beta}' = \hat{\beta} = -0.00874$,
$R^2 = 0.0269$.

אקונומטריקה למדעי הנתונים

פרק 8 - המודל הריבועי

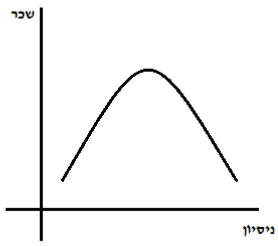
תוכן העניינים

המודל הריבועי:

רקע:

משמש עבור משתנים שתרומתם לניבוי המשתנה התלוי איננה ליניארית אלא פרבולית עם נקודת מינימום או מקסימום.

למשל:



מודל הרגרסיה הריבועית מניח כי בשלב מסוים התרומה השולית משנה את סימנה (מחיובי לשלילי או משלילי לחיובי).

המודל הריבועי: $Y_i = \alpha + \beta_1 X_i + \beta_2 X_i^2 + u_i$.

התרומה השולית של X בניבוי Y: $\frac{dy}{dx} = \beta_1 + 2\beta_2 X_i$.

התרומה השולית במקרה זה <u>איננה קבועה אלא תלויה ב-X</u>.

הנגזרת מתאפסת בנקודה: $X^* = -\frac{\beta_1}{2\beta_2}$.

אם β_2, המקדם של X^2, הוא <u>חיובי</u>: מדובר <u>בנקודת מינימום</u> שעד אליה התרומה השולית שלילית וממנה ואילך חיובית.

אם β_2, המקדם של X^2, הוא <u>שלילי</u>: מדובר <u>בנקודת מקסימום</u> שעד אליה התרומה השולית חיובית וממנה ואילך שלילית.

שאלות:

1) במחקר על השפעת הגיל על מספר הדקות שפרט משוחח בטלפון הנייד נאמד המודל הבא: $Y_i = \beta_0 + \beta_1 A_i + \beta_2 A_i^2 + u_i$.
מה צריכים להיות סימני המקדמים שייתנו את התוצאה הבאה:
בגילאים מבוגרים ובגילאים צעירים מדברים יותר מאשר בגילאי הביניים?

2) חוקר החליט להשוות בין שני מודלים:
מודל רגרסיה פשוטה של X = הוצאות פרסום (באלפי שקלים לשנה)
על Y = ציון לחוזק המותג (בציון 0-10).
מודל רגרסיה מרובה הכולל בנוסף את המשתנה X^2 = הוצאה עבור פרסום בריבוע.
א. תארו את המודל באופן אלגברי.
להלן תוצאות האמידה של המודלים על סמך מדגם של 33 חברות:
1. $S_{\hat{\beta}} = 0.097$, $\hat{Y}_i = 22.163 + 0.363X_i$ $R^2 = 0.4676$.
2. $S_{\hat{\beta}_1} = 0.37$, $S_{\hat{\beta}_2} = 0.002$, $\hat{Y}_i = 7.059 + 1.085X_i - 0.004X_i^2$ $R^2 = 0.53$.

ב. מהו גודל השינוי השולי בכל אחד מהמודלים?
אמדו את גודלו והסבירו את משמעותו.
ג. איזה מודל עדיף? מהם המבחנים הסטטיסטיים המתאימים? בצעו אותם.
ד. בדקו האם לאחר רמת הוצאה מסוימת כבר לא משתלם לפרסם.

3) בכדי לאמוד את הקשר בין הישגיהם של תלמידים שסיימו את בית הספר התיכון באמצעות ציון של מבחן כניסה לאוניברסיטה (G שנמדד בנקודות) לגודל בית הספר (HS שנמדד במאות תלמידים) נאמד מודל ריבועי על בסיס מדגם של 400 תלמידים מתוך כלל התלמידים שניגשו לבחינת הכניסה.
להלן המשוואה הנאמדת (סטיות התקן נתונות בסוגריים):

$$\hat{G} = \underset{(0.55)}{997.8} + \underset{(3.99)}{19.81}HS - \underset{(6.20)}{2.13}HS^2 \quad , \quad R^2 = 0.076.$$

א. הסבירו את המשמעות של המודל הריבועי (לווה את תשובתך בחישוב של הגודל האופטימאלי של בית הספר ובתיאור גרפי של המודל).
ב. מה יהיה השינוי במבחן בין תלמיד שלמד בבית ספר עם 300 תלמידים לבין תלמיד שלמד בבית ספר עם 330 תלמידים?
ג. מה יהיה הגודל האופטימאלי של בית הספר בהנחה שהמשתנה HS נמדד בעשרות תלמידים ולא במאות תלמידים?
ד. האם מספר התלמידים בריבוע תורם להסבר של המודל? נסח את ההשערה ובדוק אותה.
ה. הועלתה הטענה כי המודל איננו מצליח כלל להסביר את התנהגות הציונים במבחן. נסח את ההשערה המתאימה ובדוק אותה.

תשובות סופיות:

1) $\beta_1 < 0\ ;\ \beta_2 > 0$.

2) א. המודל הליניארי הפשוט: $Y_i = \alpha + \beta X_i + u_i$.

המודל הריבועי: $Y_i + \alpha + \beta_1 X_i + \beta_2 X_i^2 + u_i$.

ב.1. גודל השינוי הוא β , האומד הינו: $b = 0.363$.

2. גודל השינוי הוא: $\beta_1 + 2\beta_2 X_i$, האומד הינו: $1.085 - 0.008 \cdot X_i$.

ג. המודל הריבועי עדיף עפ"י מבחנים t ו-WALD.

ד. בשלב מסוים השינוי השולי הופך מחיובי לשלילי.

3) א. ראה סרטון.

ב. השינוי יהיה: 1.278 נקודות.

ג. $X^{*}{}' = 46.5$.

ד. כן.

ה. יש עדות לכך.

אקונומטריקה למדעי הנתונים

פרק 9 - מבחן 1 ללא פלטים

תוכן העניינים

מבחן 1 ללא פלטים:

שאלות:

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

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

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

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

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

ד. האם אומד זה הוא blue?

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

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

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

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

1. $y_i = 15 + 0.7x_{1i} + 0.35x_{2i} + 0.20x_{3i} \quad R^2 = 0.85$.
2. $y_i = 2 + 0.1z_i \quad R^2 = 0.25$.
3. $z_i = x_{1i} - x_{2i} + 2x_{3i}$.

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

א. ההשערה שניתן לבדוק באמצעות המשוואות הנתונות הינה:

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

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

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

iv. לא ניתן לדעת.

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

i. 56.

ii. 57.

iii. 112.

iv. 74.66.

3) כלכלן הציע את המודלים הבאים:

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

האם ניתן לאמוד את המודלים בשיטת OLS?

א. אין בעיה לאמוד את שני המודלים.
ב. לא ניתן לאמוד את המודל הראשון בלבד.
ג. לא ניתן לאמוד את המודל השני בלבד.
ד. לא ניתן לאמוד את שני המודלים.

4) כלכלן אמד את המודל הבא: $y_i = \alpha_0 + \alpha_1 \ln(x_i) + u_i$,
וקיבל את האומדנים: $\hat{\alpha}_0 = 10$ ו- $\hat{\alpha}_1 = 6$.
על אותו המדגם אמד חברו את המודל הבא: $y_i = \beta_0 + \beta_1 \ln(x_i^2) + u_i$.
מכאן ש:

א. $\hat{\beta}_0 = 5$ ו- $\hat{\beta}_1 = 3$.
ב. $\hat{\beta}_0 = 10$ ו- $\hat{\beta}_1 = 3$.
ג. $\hat{\beta}_0 = 5$ ו- $\hat{\beta}_1 = 6$.
ד. כל התשובות שגויות.

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

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

(2) (1)

הערכים שבסוגריים הם סטיות התקן של המקדמים.

א. בדוק האם המודל מובהק.
ב. בדוק האם מקדמי השיפוע מובהקים.
ג. מה תוכל להסיק מסעיפים א ו-ב?

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

1. $y_i = 4 + 0.1x_{1i} + 0.8x_{2i}$ $\quad R^2 = 0.84$.
2. $y_i = 2 + 0.8x_{1i}$ $\quad R^2 = 0.7$.
3. $y_i = 7 + 0.23x_{2i}$ $\quad R^2 = 0.25$.
4. $y_i = 3 + 0.23z_i$ $\quad R^2 = 0.55$.

כאשר x_{1i} ו- x_{2i} הם השכלת הבעל והאישה בהתאמה במשפחה i ו- y_i הכנסת משק בית i. כמו כן נתון כי: $z_i = x_{1i} + 2.2x_{2i}$.

א. בדוק את ההשערה כי להשכלה אין השפעה על הכנסות המשפחה
ב. איזה השערה ניתן לבדוק תוך שימוש במשוואות (1) ו-(4)? בדוק אותה.
ג. חשב את סטית התקן של המקדם x_{1i} ברגרסיה (1).

7) חוקר מעוניין לאמוד את המודל: $y_i = \alpha + u_i$.
א. חשב את נוסחת אומד הריבועים הפחותים ל- α על ידי פתרון בעיית המינימיזציה של סכום ריבועי הסטיות.
ב. חשב את נוסחת שונותו של האומד.

8) על סמך מדגם של 45 תצפיות נאמדו המודלים הבאים:
1. $y_i = 5.4 + 1.2x_{2i} + 4.4x_{3i} + u_i \quad R^2 = 0.75$.
2. $y_i = 6.3 + 5.8x_{3i} + u_i \quad R^2 = 0.65$.
3. $y_i = 5.7 + 1.2x_{2i} + u_i \quad R^2 = 0.70$.
4. $y_i = 3.9 + 3.4\ln(x_{2i}) + u_i \quad R^2 = 0.56$.
5. $\ln(y_i) = 2.4 + 1.8x_{2i} + 2.7x_{3i}^2 + 4.2x_{4i}^2 + u_i$.
6. $y_i = 1.3 + 3.1x_{2i} + 0.5x_{3i} + 4.8x_{4i}^2 + 1.5x_{5i}^2 + u_i$.

א. דרג את הרגרסיות על פי מדד ההסבר (מהנמוך לגבוה)
ב. בדוק את ההשערות של משתנים X_2 ו- X_3 ביחד אין השפעה על Y במודל (1).
ג. בדוק בהסתמך על מודל (2) האם המשתנה X_2 מובהק ברגרסיה (1).
ד. ברגרסיה (1) נתונים כעת אומדי הטעויות הסטנדרטיות (סטיות התקן) של מקדמי X_2 ו- X_3 0.5 ו-2.5 בהתאמה. בדוק עבור כל אחד מהמקדמים הנ״ל האם מובהק ומה אפשר ללמוד מרגרסיה (1).
ה. איזו השערה ניתן לבדוק תוך שימוש במשוואות (6) ו-(3)?

תשובות סופיות:

1) א. לינארי. ב. מוטה. ג. אי אפשר לדעת. ד. לא.
ה. נכון. ו. $V\left(\tilde{\beta}\right)=\frac{n\sigma^2}{S^2{}_{XX}}$. ז. לא נכון.

2) א. iii. ב. iii.

3) ד'.

4) ב'.

5) א. מובהק. ב. לא מובהקים. ג. קיימת בעיה מולטיקולינאריות חלקית.

6) א. מובהק. ב. $\beta_2=2.2\beta_1$. ג. $S.E=0.00743$.

7) ראה סרטון.

8) א. $6>1>3>2>4$. ב. יש עדות לכך. ג. יש עדות לכך.
ד. X_2 מובהק, X_3 אינו מובהק. ה. $HO:\beta_2=\beta_4=\beta_5=0$.

אקונומטריקה למדעי הנתונים

פרק 10 - מבחן 2 ללא פלטים

תוכן העניינים

מבחן 2 ללא פלטים:

שאלות:

אם לא נאמר אחרת בנתוני השאלה, התבסס על ההנחות הבאות:

1. ערך t קריטי הוא 2.
2. ערך F קריטי הוא 4.

1) הנח כי הקשר באוכלוסייה בין X לבין Y נתון ע"י המשוואה הבאה: $\sqrt{y_i} = \beta \ln(x_i) + u_i$.

נתון גם כי עבור המודל הנ"ל כל ההנחות הקלאסיות מתקיימות.

כלכלן הציע את האומד הבא עבור β: $\hat{\beta} = \sum_{i=1}^{n} \frac{\sqrt{y_i}\ln(x_i)}{\sum \ln(x_i)^2}$.

א. מהי הטענה הנכונה:

i. האומד חסר הטיה ובעל שונות מינימאלית.

ii. האומד מוטה.

iii. האומד לא ליניארי.

iv. האומד מוטה אך יש לו שונות נמוכה מאומד OLS.

v. כל התשובות האחרות אינן נכונות.

ב. שונות האומד הוא:

i. $\frac{\sigma^2}{\sum [\ln(x_i)]^2}$.

ii. $\sigma^2 \sum (\frac{\sqrt{y_i}}{\ln(x_i)})^2$.

iii. $\frac{\sigma^2}{\sum \ln(x_i)}$.

iv. לא ניתן לחשב את האומד שכן הוא לא ליניארי.

v. כל התשובות האחרות אינן נכונות.

ג. מה בהכרח מתקיים עבור אומד זה:

.i $\sum_{i=1}^{n}[\frac{\ln(x_i)}{\sum[\ln(x_i)]^2}\times\ln(x_i)]=1$.

.ii $\sum_{i=1}^{n}\frac{\ln(x_i)}{\sum[\ln(x_i)]^2}=0$.

.iii $\sum_{i=1}^{n}\frac{1}{\sum\ln(x_i)^2}=1$.

.iv $\sum_{i=1}^{n}[\frac{\ln(x_i)}{\sum\ln(x_i)^2}\times\ln(x_i)]=0$.

.v כל התשובות האחרות אינן נכונות.

2) נתונים שני מודלים:

1. $\ln(y_i)=\alpha+\beta_1\ln(x_i)+\beta_2\ln(x_i^{\ 2})+u_i$.
2. $y_i=\alpha+\beta_1x_i+\beta_2x_i^{\ 2}+u_i$.

להלן שלוש טענות:

1. במודל 1 יש מולטיקוליניאריות מושלמת ולכן הוא לא ניתן לאמידה בשיטת OLS.
2. במודל 2 יש מולטיקוליניאריות מושלמת ולכן הוא לא ניתן לאמידה בשיטת OLS.
3. במודל 1 יש מולטיקוליניאריות חלקית ולכן הוא ניתן לאמידה בשיטת OLS.

א. רק טענה 1 נכונה.
ב. רק טענה 2 נכונה.
ג. רק טענות 2 ו-3 נכונות.
ד. רק טענה 3 נכונה.
ה. כל התשובות האחרות אינן נכונות.

3) אסף הוא כלכלן צעיר שמתעניין מאוד בביתר ירושלים. לאור אכזבות חוזרות ונשנות להבאת שחקנים טובים. החליט אסף לפנות למאמן הקבוצה ולהסביר לו את הפרמטרים החשובים לשחקן כדורגל.
אסף הריץ את הרגרסיה הבאה: $\hat{y}_i=\alpha+\beta_1\cdot x_{1i}+\beta_2\cdot x_{2i}+\beta_3\cdot x_{3i}$ על סמך 500 תצפיות וקיבל את התוצאות הבאות:

$\hat{y}_i=10+2\cdot x_{1i}+1.5\cdot x_{2i}+2.5\cdot x_{3i}$,

$S^2_u=10$ $S_{\hat{\alpha}}=3$, $S_{\hat{\beta}1}=0.5$, $S_{\hat{\beta}2}=0.75$, $S_{\hat{\beta}3}=1$,

$\text{cov}(\beta_3,\beta_2)=-3$, $\text{cov}(\beta_1,\beta_3)=1$, $\text{cov}(\beta_1,\beta_2)=-0.6$.

כאשר:

y - טיב השחקן (על סמך דירוג הפרשנים).

x_{1i} - מהירות השחקן.

x_{2i} - קשיחות השחקן.

x_{3i} - הרמה הטכנית של השחקן.

א. מהו רווח בר סמך ל- β_1 והאם היא מובהקת?

i. $[1,3]$, לכן ה- β_1 מובהקת.

ii. $[1.5,2.5]$, לכן ה- β_1 מובהקת.

iii. $[1,3]$, לכן ה- β_1 לא מובהקת.

iv. $[1.5,2.5]$, לכן ה- β_1 לא מובהקת.

v. כל התשובות האחרות אינן נכונות.

ב. המאמן טוען כי השפעת הרמה הטכנית על טיב השחקן היא כפולה מזו של הקשיחות. T סטטיסטי לבחינת ההשערה הוא (מעוגל ובערך מוחלט):

i. 0.128.

ii. 1.255.

iii. 0.125.

iv. 0.156.

v. כל התשובות האחרות אינן נכונות.

4) על סמך מדגם של 50 תצפיות נאמדו המשוואות הבאות:

1. $\hat{Y}_i = 2+4X_{1i}+2X_{2i}{}^2-4X_{3i}$ $\quad R^2=0.70$.

2. $\hat{Y}_i = 4+5X_{1i}-2X_{3i}$ $\quad R^2=0.65$.

3. $\hat{X}_{2i} = 3+5.2Y_i$ $\quad R^2=0.40$.

4. $\hat{Y}_i = 5+2X_{1i}-1.2X_{2i}$.

מה ניתן לדעת על R^2 ברגרסיה (4)?

א. $R^2 > 0.4$.

ב. לא ניתן לדעת.

ג. $R^2 > 0.65$.

ד. $R^2 < 0.7$.

ה. כל התשובות האחרות אינן נכונות.

5) על סמך מדגם בגודל 30 תצפיות אמדו יצחק וטל את המודל הבא: $Y_i = \beta \cdot X_i + u_i$ והתקבל: $\hat{Y}_i = 3X_i \quad R^2 = 0.75$.

כעת הגיע מנדי (כלכלן חדש) והציע את המודל הבא: $Y_i = \alpha + \beta \cdot X_i^{\,2} + u_i$.

מה ניתן להסיק על R^2 של המודל החדש על סמך R^2 של המודל המקורי?

א. לא ניתן להסיק על R^2 של המודל החדש על סמך R^2 של המודל המקורי.

ב. $R^2 > 0.75$.

ג. $R^2 = 0.75$.

ד. $R^2 < 0.75$.

ה. כל התשובות האחרות אינן נכונות.

6) ערן החליט לבדוק את אהבת הסטודנטים לאקונומטריקה א'. לכן הריץ רגרסיה בה בדק השפעת שעות הלימוד של הסטודנט על הציון בבחינה.

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

לשם כך ערן אסף 51 תצפיות והריץ רגרסיה.

התוצאות אשר קיבל הן: $\hat{\alpha} = 1, \;\; \hat{\beta} = 5$.

מספר סטודנטים קטני אמונה, טענו כי ההשפעה של שעת לימוד על הציון צריכה להיות 3 (ולא יותר). הם בדקו זאת ע"י בחינת T סטטיסטי וקיבלו $\mathrm{T} = 1$.

כמו כן ידוע כי השונות של X היא 10. מכאן סכום הטעויות בריבוע הינו:

א. 98,000.

ב. 49,000.

ג. 24,500.

ד. אין מספיק נתונים כדי לפתור את השאלה.

ה. כל התשובות האחרות אינן נכונות.

7) נתון המודל הבא: $y_i = \beta_1 \cdot x_{1i} + \beta_2 \cdot x_{2i} + u_i$ במודל זה בהכרח מתקיים:

א. $\sum (x_{1i} + x_{2i}) e_i = 0$.

ב. $\sum (1 + x_{1i} + x_{2i}) e_i = 0$.

ג. $\sum (1 + x_{1i}) e_i = 0$.

ד. $\sum e_i = 0$.

ה. כל התשובות האחרות לא נכונות.

8) נתון המודל: $Y_i = AX_{1i}^{\beta_1}X_{2i}^{\beta_2}X_{3i}^{\beta_3}e^{\beta_4 X_{4i}^3}X_{5i}e^{u_i}$ שהורץ על המדגם בן 45 תצפיות.
מהי המשוואה לאמידת המקדמים של המודל?

א. $\ln(\frac{y_i}{x_5}) = \alpha + \beta_1\ln(x_{1i}) + \beta_2\ln(x_{2i}) + \beta_3\ln(x_{3i}) + \beta_4 X_{4i}^3 + u_i$.

ב. $\ln(\frac{y_i}{K5}) = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + 3x_{4i} + u_i$.

ג. $\ln(y_i) - \ln(x_{5i}) = \alpha + \beta_1\ln(x_{1i}) + \beta_2\ln(x_{2i}) + \beta_3 x_{3i} + \beta_4 x_{4i} + u_i$.

ד. $\ln(\frac{y_i}{x_5}) = \alpha + \beta_1\ln(x_{1i}) + \beta_2\ln(x_{2i}) + \beta_3\ln(x_{3i}) + \beta_4\ln(X_{4i}^3) + u_i$.

ה. כל התשובות האחרות אינן נכונות.

9) נתון המודל: $Y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$.

א. מהו המודל המוגבל עבור ההשערה: $\beta_1 = \beta_2 + 1 \,,\, \beta_3 = 2$?

i. $Y_i - x_{1i} - 2x_{3i} = \alpha + \beta(x_{1i} + x_{2i}) + u_i$.

ii. $Y_i - x_{1i} + 2x_{3i} = \alpha + \beta(x_{1i} + x_{2i}) + u_i$.

iii. $Y_i - 2x_{3i} = \alpha + \beta(x_{1i} + x_{2i} + 1) + u_i$.

iv. $Y_i + 2x_{3i} = \alpha + \beta(x_{1i} + x_{2i} + 1) + u_i$.

v. כל התשובות האחרות אינן נכונות.

ב. מהו סטטיסטי המבחן?

i. רק מבחן $\frac{\left(\sum e_{מ}^2 - \sum e_{מל}^2\right)/m}{\sum e_{מ}^2 /_{ל}(n-k)}$.

ii. רק מבחן $\frac{R^2/(k-1)}{\left(1-R^2\right)/\left(n-k\right)}$.

iii. רק מבחן $\frac{\left(R_{מ}^2 - R_{מ}^2\right)/_{ל}\, m}{\left(1-R_{מ}^2\right)/_{ל}(n-k)}$.

iv. מבחנים $\frac{\left(R_{מ}^2 - R_{מ}^2\right)/_{ל}\, m}{\left(1-R_{מ}^2\right)/_{ל}(n-k)}$ או $\frac{\left(\sum e_{מ}^2 - \sum e_{מל}^2\right)/m}{\sum e_{מ}^2 /_{ל}(n-k)}$.

v. כל התשובות האחרות אינן נכונות.

10) על סמך מדגם של 50 תצפיות נאמדו המשוואות הבאות :

1. $\hat{Y}_i = 3 + 45.X_{1i} + 5X_{2i}$, $R^2 = 0.65$.
2. $\hat{Y}_i = 5.2X_{1i}$, $R^2 = 0.30$.
3. $\hat{Y}_i = 4.5 + 5.9X_{1i}$, $R^2 = 0.40$.

בדוק את ההשערה שהשפעת המשתנה X_2 מובהקת ברגרסיה (1), ומהו סטיית התקן של β_2 .

א. מובהקת. וסטיית התקן של β_2 היא 0.86 (בקירוב).
ב. אינה מובהקת. וסטיית התקן של β_2 היא 0.72 (בקירוב).
ג. אינה מובהקת. וסטיית התקן של β_2 היא 0.86 (בקירוב).
ד. מובהקת. וסטיית התקן של β_2 היא 0.72 (בקירוב).
ה. כל התשובות האחרות אינן נכונות.

11) על סמך מדגם של 50 תצפיות נאמדו המשוואות הבאות :

1. $\hat{Y}_i = 3 + 3X_{1i} + 5X_{2i} + 2X_{3i}$.
2. $\hat{Y}_i = 9.3 + 0.6W_i$.

נתון גם כי : $W_i = X_{1i} - 2X_{2i} + X_{3i}$.
איזו השערה ניתן לבדוק תוך שימוש במשוואות (1) ו-(2)?

א. ברגרסיה (1) $\beta_1 = \beta_3 = -0.5\beta_2$.
ב. ברגרסיה (1) $\beta_1 = \beta_3 = 0.5\beta_2$.
ג. ברגרסיה (1) $\beta_1 = \beta_3 = 2\beta_2$.
ד. ברגרסיה (1) $\beta_1 = \beta_3 = \beta_2$.
ה. כל התשובות האחרות לא נכונות.

תשובות סופיות:

1) א. i.　　ב. ii.　　ג. i.
2) א׳.
3) א. i.　　ב. i.
4) א׳.
5) א׳.
6) א׳.
7) א׳.
8) א׳.
9) א. i　　ב. i.
10) א׳.
11) א׳.

אקונומטריקה למדעי הנתונים

פרק 11 - מבחן 3 ללא פלטים

תוכן העניינים

מבחן 3 ללא פלטים:

שאלות:

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

1) הקשר באוכלוסייה בין X ל-Y מוגדר על ידי המודל הבא: $Y_t^2 = \alpha + X_{1t}^2 + \beta \ln X_{2t} + u_t$.
נתון כי עבור המודל הנ"ל כל ההנחות הקלאסיות מתקיימות.
אומדים את המקדם β לפי הנוסחה: $\hat{\beta} = \frac{\sum (Y_t^2 - X_{1t}^2)\ln X_{2t}}{\sum (\ln X_{2t})^2}$.

א. האומד ליניארי אבל מוטה.
ב. האומד לא ליניארי ומוטה.
ג. האומד ליניארי וחסר הטיה.
ד. האומד לא ליניארי אך חסר הטיה.
ה. כל התשובו האחרות אינן נכונות.

2) שונות האומד הנ"ל הינה:

א. $\frac{\sigma^2 (\ln X_{2t})}{\sum (\ln X_{2t})^2}$.
ב. $\frac{\sigma^2}{4X_t^2}$.
ג. $\frac{\sigma^2}{\sum (\ln X_{2t})^2}$.
ד. $\sigma^2 \frac{1}{2} \sum \frac{1}{X_t^2}$.
ה. כל התשובות האחרות אינן נכונות.

3) נתון המודל הבא: $Y_t = \beta_1 X_{1t} + \beta_2 X_{2t} + u_t$, נתון בנוסף כי מקדם המתאם בין שני המשתנים הבלתי תלויים הינו מושלם ($\rho_{12} = 1$). להלן 3 טענות:

1. בהכרח קיימת מולטיקוליניאריות מושלמת במודל.
2. ייתכן כי ברגרסיה אין מולטיקוליניאריות מושלמת.
3. אם היה חותך במודל, בהכרח לא ניתן היה לאמוד את המודל.

מכאן ש:
א. רק טענות 1 ו-3 נכונות.
ב. רק טענה 2 נכונה.
ג. רק טענה 3 נכונה.

ד. כל התשובות האחרות אינן נכונות.
ה. רק טענות 2 ו-3 נכונות.

4) נאמד המודל הבא: $Y_t = \alpha + \beta_1 X_{1t} + u_t$. ידוע כי במדגם: $\hat{\alpha} = 2, \hat{\beta} = 0.5$, $SSY = SSX$.
מכאן ניתן להסיק כי R^2 של המודל הוא:
א. 0.5.
ב. 0.25.
ג. 1.
ד. 0.75.
ה. כל התשובות האחרות אינן נכונות.

5) נאמד המודל הבא: $Y_t = \beta X_t + u_t$.
אם מתקיים: $E(u_t) \neq 0$ ומלבד זאת כל ההנחות הקלאסיות מתקיימות. אזי:
א. $\hat{\beta}$ יהיה חסר הטיה.
ב. $\sum X_t \hat{u}_t < 0$.
ג. $\sum X_t \hat{u}_t > 0$.
ד. $\hat{\beta}$ יהיה מוטה.
ה. כל התשובות האחרות אינן נכונות.

6) נתון המודל הבא: $Y_t = \alpha + \beta_1 \ln(x_t^2) + \beta_2 \ln(2x_t) + u_t$.
א. יש במודל מולטיקוליניאריות מ ושלמת ולכן לא ניתן לאמוד את המודל.
ב. אין במודל מולטיקוליניאריות מושלמת ולכן ניתן לאמוד את המודל.
ג. יש במודל מולטיקוליניאריות מושלמת אבל ניתן לאמוד את המודל.
ד. יתכן ויש במודל מולטיקוליניאריות חלקית ולכן לא ניתן לאמוד את המודל.
ה. כל התשובות האחרות אינן נכונות.

7) הנח כי הקשר באוכלוסייה בין X ל-Y נתון על ידי המודל הבא: $Y_t = \alpha + \beta_1 X_{1t} + u_t$.
אומד OLS במודל ל-β יהיה:
א. עם שונות קטנה יותר ככל ששונות ההפרעות (Ui) באוכלוסייה תהיה גדולה יותר.
ב. עם שונות קטנה יותר ככל ששונות Y במדגם תהיה גדולה יותר.
ג. עם שונות קטנה יותר ככל ששונות X במדגם תהיה גדולה יותר.
ד. עם שונות גדולה יותר ככל ששונות X במדגם תהיה גדולה יותר.
ה. כל התשובות האחרות אינן נכונות.

8) סטודנט אמד מודל מסוים וקיבל את התוצאות הבאות:

$\hat{Y}_i = 2 - 3\ln X_{1i} + 2X_{2i} + 6X_{2i} \cdot X_{1i}$.

מה יכול להיות המודל אותו אמד הסטודנט:

א. $Y_i = AX_{1i}^{\beta_1}(X_{1i} \cdot X_{2i})^{\beta_3} e^{\beta_2 X_{1i} + u_i}$.

ב. $Y_i = X_{1i}^{\beta_1}(X_{1i} \cdot X_{2i})^{\beta_3} e^{\beta_2 X_{1i} + u_i}$.

ג. $e^{Y_i} = X_{1i}^{\beta_1} e^{\beta_2 X_{2i} + \beta_3 X_{1i} \cdot X_{2i} + u_i}$.

ד. כל התשובות האחרות אינן נכונות.

ה. $e^{Y_i} = AX_{1i}^{\beta_1} e^{\beta_2 X_{2i} + \beta_3 X_{1i} \cdot X_{2i} + u_i}$.

9) על סמך מדגם של 50 תצפיות נאמדו המשוואות הבאות:

1. $\hat{Y}_i = 2 + 4X_{1i} + 2X_{2i} - 4X_{3i}$.
2. $\hat{Y}_i = 5 + 2X_{1i} - 1.2X_{2i}$.
3. $\hat{Y}_i = 5 + 4\ln(X_{1i}) + 12X_{2i} + 3X_{3i} + 2X_{4i}$.
4. $\hat{Y}_i = 5 + 2X_{4i} - 1.2X_{2i}$.

מה מתקיים בהכרח:

א. R בריבוע של משוואה 3 גדול מ-R בריבוע של משוואה 1.

ב. R בריבוע של משוואה 3 גדול מ-R בריבוע של משוואה 4.

ג. R בריבוע של משוואה 3 גדול מ-R בריבוע של משוואה 2.

ד. R בריבוע של משוואה 2 גדול מ-R בריבוע של משוואה 4.

ה. כל התשובות האחרות אינן נכונות.

10) על סמך מדגם של 50 תצפיות נאמדו המשוואות הבאות:

1. $ESS = 200 \quad \hat{Y}_i = 4 + 2.8X_{1i} + 2X_{2i}$.
2. $ESS = 320 \quad \hat{Y}_i = 2 + 2.5X_{2i}$.

ידוע כי R בריבוע של משוואה 1 הוא 0.75. מה הוא R בריבוע של משוואה 2?

א. 0.7.

ב. 0.5.

ג. 0.4.

ד. 0.6.

ה. כל התשובות האחרות אינן נכונות.

11) על סמך מדגם של 43 תצפיות נאמדו המשוואות הבאות:

1. $\hat{Y}_i = 1.8 + 3.4X_{1i} + 0.9X_{2i}$.
2. $\hat{Y}_i = 3.3 + 3.2X_{1i} + 2.4X_{3i}$ $\quad R^2 = 0.8$.
3. $X_{1i} = 2.7 + 3Y_i$ $\quad R^2 = 0.6$.

על פי נתונים אלו ניתן להסיק כי:

א. מדד טיב הרגרסיה ברגרסיה (1) בהכרח גדול מ-0.6.
ב. מדד טיב הרגרסיה ברגרסיה (1) בהכרח גדול מ-0.8.
ג. מדד טיב הרגרסיה ברגרסיה (1) בהכרח קטן מ-0.8.
ד. מדד טיב הרגרסיה ברגרסיה (1) בהכרח קטן מ-0.6.
ה. כל התשובות האחרות אינן נכונות.

12) נתון המודל הבא: $Y_t = \alpha + \beta_1 X_{1t} + u_t$.

כלכלן א' אמד את β על ידי האומד הבא: $\dfrac{Y_{10} - Y_1}{2X_{10} - 2X_1}$.

כלכלן ב' אמד את β על ידי האומד הבא: $\dfrac{Y_{10} - Y_1}{X_{10} - X_1}$.

א. האומדנים של שני הכלכלנים הינם חסרי הטיה.
ב. אין הבדל בין שני האומדנים כי שני האומדנים הינם אומדנים ליניאריים.
ג. לאומדן של כלכלן א' יש שונות נמוכה יותר.
ד. האומדנים של שני הכלכלנים הינם מוטים.
ה. כל התשובות האחרות אינן נכונות.

13) נתון המודל: $Y_i = \alpha + \beta_1 \ln(X_{1i}) + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$.

מהו סטטיסטי המבחן עבור בחינת ההשערה הבאה: $\beta_3 = 0, \beta_1 = \beta_2$?

א. רק מבחן: $\dfrac{R^2/(k-1)}{1-R^2/(n-k)}$.
ב. רק מבחן: $\dfrac{(R^2_{\text{מ}} - R^2_{\text{ל}})/m}{(1-R_{\text{מ}})/(n-k)}$.
ג. רק מבחן: $\dfrac{(\sum e_n^2 - \sum e_{\text{מ}}^2)/m}{\sum e_{\text{מ}}^2/(n-k)}$.
ד. כל התשובות האחרות אינן נכונות.
ה. מבחנים: $\dfrac{(R^2_{\text{מ}} - R^2_{\text{ל}})/m}{(1-R_{\text{מ}})/(n-k)}$ או $\dfrac{(\sum e_n^2 - \sum e_{\text{מ}}^2)/m}{\sum e_{\text{מ}}^2/(n-k)}$.

14) נתון המודל הבא: $\ln Y_i = \alpha + \beta_1 \ln(X_{1i}) + \beta_2 \ln(X_{2i}) + \beta_3 X_{3i} + u_i$.
מה המודל המוגבל עבור ההשערה: $\beta_1 = -\beta_3, \beta_2 = -1$.
א. $\ln(Y_i + X_{2i}) = \alpha + \beta_1(\ln(X_{1i}) - X_{3i}) + u_i$.
ב. $\ln(Y_i \cdot X_{2i}) = \alpha + \beta_1(\ln(X_{1i}) - X_{3i}) + u_i$.
ג. $\ln Y_i + X_{2i} = \alpha + \beta_1 \ln(X_{1i}) + \beta_3 X_{3i} + u_i$.
ד. $\ln Y_i + X_{3i} = \alpha + \beta_1[\ln(X_{1i}) - \ln(X_{2i}] + u_i$.
ה. כל התשובות האחרות אינן נכונות.

תשובות סופיות:

1) א'.
2) ג'.
3) ה.
4) ב'.
5) ד'.
6) א'.
7) ג'.
8) ה.
9) ב'.
10) ד'.
11) א'.
12) ג'.
13) ה.
14) ב'.

אקונומטריקה למדעי הנתונים

פרק 12 - מבחני המובהקות וקריאת פלטים - תוכנת SAS

תוכן העניינים

מבחני המובהקות וקריאת פלטים – תוכנת SAS:

רקע:

פלט ניתוח שונות (Analysis of Variance):

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 מקדמי הרגרסיה (Parameter Estimates):

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	$\hat{\beta}$	$s_{\hat{\beta}}$	$\frac{\hat{\beta}}{s_{\hat{\beta}}} = t_{(\hat{\beta}=0)}$	$Pt_{\hat{\beta}}$

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	$s_{\hat{\alpha}}^2$	$\operatorname{cov}(\hat{\alpha},\hat{\beta})$
X	$\operatorname{cov}(\hat{\alpha},\hat{\beta})$	$s_{\hat{\beta}}^2$

עריכת תחזית וקריאת פלטים (תוכנת SPSS):

אמידה נקודתית:

אמידה נקודתית עבור X_0 מסוים (תחזית).

מחושבת על פי קו הרגרסיה במדגם: $\hat{Y}_0=\hat{\alpha}+\hat{\beta}\cdot X_0$.

אמידה מרווח ל- $E(Y)$:

אמידת התחזית באוכלוסייה 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: $\hat{Y}\pm t_{n-2;1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$.

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

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

אמידה מרווח ל- Y :

אמידת ערך בודד של Y באוכלוסייה 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: $p(\_\_\_\leq Y\leq\_\_\_)=1-\alpha$.

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

שאלות:

פלט ניתוח שונות:

1) 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:

2) בהמשך לדוגמא הקודמת – בדיקת השפעת ההכנסה על גודל המס, התקבלו גם התוצאות הבאות:

Parameter Estimates

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

א. אמדו את המודל: $TAX = \alpha + \beta \cdot INCOME + U$.
מהי המשמעות הכלכלית של β?

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

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

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

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

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

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

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

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

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

פלט שונויות משותפות:

3) נתון פלט האמידה של המודל: $Y_t = \alpha + \beta X_t + u_t$, שלצורך אמידתו נאספו 240 תצפיות:

Parameter Estimates

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

Covariance of Estimates

	INTERCEP	X
INTERCEP	0.0625	-0.003
X	-0.003	0.0144

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

שאלה מסכמת:

4) חוקר רצה לבדוק את השפעת הותק בעבודה (*EXP*) על השכר (*SALARY*) לפי המודל: $\ln(SALARY_t)=\alpha+\beta\cdot EXP_t+u_t$. 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: $EXP=22$.

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

ב. שיעור התשואה בשכר לשנת ותק הוא?

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

ביצוע תחזיות:

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

Descriptive Statistics

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

Model Summary[b]

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

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

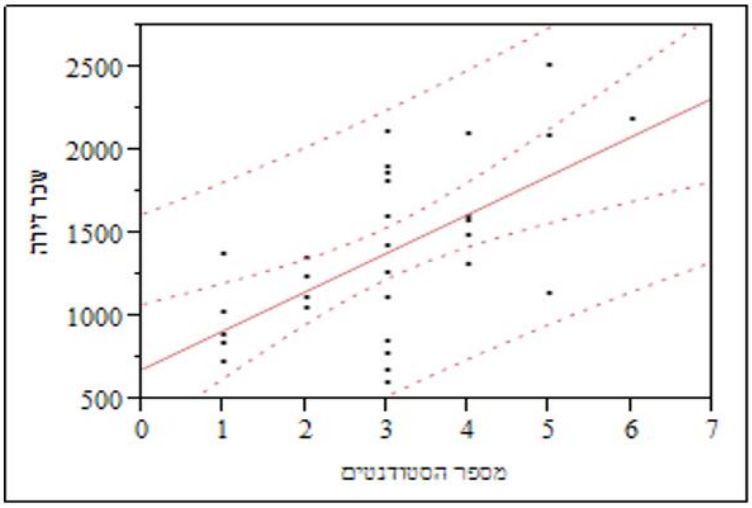
ANOVA[b]

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

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

Coefficients[a]

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



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

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

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

תשובות סופיות:

1) יש עדות לכך.

2) א. ראה סרטון. ב. יש עדות לכך. ג. $Pt_{\hat{\beta}} = 0.0001$.

ד. יש עדות לכך. ה. $P\left(0.1488 \le \beta \le 0.1554\right) = 0.95$.

ו. יש עדות לכך.

3) אין עדות לכך.

4) א. לא נכון. ב. -0.87%. ג. 7.24735.

5) א. 1153.247. ב. $p\left(957.4 \le \mu_{Y_{X=2}} \le 1349.08\right) = 0.95$.

ג. $p\left(282.94 \le Y_{X=2} \le 2023.55\right) = 0.95$.

אקונומטריקה למדעי הנתונים

פרק 13 - רגרסיה מרובה תוך שימוש בפלטים של SAS

תוכן העניינים

רגרסיה מרובה:

רקע:

מבחן T ו-F:

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.

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

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

מבחן F למובהקות המודל:

השערות: $\begin{matrix} H_0 = \beta_1 = \beta_2 = \ldots = \beta_K = 0 \\ H_1 : OTHERWISE \end{matrix}$.

סטטיסטי המבחן F וכלל ההכרעה:

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

מבחן t למובהקות ה-β טות:

מבחן לבדיקת מובהקות β ספציפית:

השערות: $\begin{matrix} H_0 = \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{matrix}$.

סטטיסטי המבחן t וכלל ההכרעה:

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

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

בכדי להחליט האם כדאי לנו להוסיף למודל משתנה ב"ת מסוים:
נשווה את פרופורציית השונות המוסברת המתוקנת $\bar{R}^2$ בין המודל ללא המשתנה המסביר לבין המודל עם המשתנה המסביר שהוספנו.

- ניתן להשתמש גם באומד המוטה - R^2 להשוואה בין מודלים אם מתקיימים שני התנאים הבאים:
  1. מספר המשתנים זהה.
  2. המשתנה המוסבר זהה.

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

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

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

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

שאלות:

מבחן T ו-F:

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

Analysis of Variance

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

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

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	-------------	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-------------	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

א. השלם את הנתונים החסרים בפלט.
ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05.
ג. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.

השוואה בין מודלים:

2) 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?

מבחן Wald ו-T מורכב:

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

Analysis of Variance

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

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

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

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

א. מהי השערת האפס?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

להלן אמידת המודל המוגבל:

Analysis of Variance

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

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

Parameter Estimates

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

ג. חשב את הסטטיסטי של WALD.
ד. כמה דרגות חופש יש במונה וכמה במכנה?
ה. האם דוחים או מקבלים את השערת האפס?

4) 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 שהנטייה השולית לצרוך מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון, נאמדה גם המשוואה הבאה:
$C_t = \alpha + \beta_1 \cdot Y_t + u_t$, כאשר: Y_t = סה"כ ההכנסה של משק בית t.
התקבל: $ESS = 0.4566$.
בדקו את ההשערה בשתי דרכים.

תרגיל מסכם:

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

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

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

ההפרעות האקראיות מקיימות את כל ההנחות הקלאסיות התקבל הפלט הבא:

Analysis of Variance

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

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

Parameter Estimates

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

א. מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?

ב. מהו אחוז השונות בתצרוכת המוסבר ע"י ההכנסה?

ג. מהו אומדן לתצרוכת ההתחלתית של משק בית?

ד. האם אומדן זה מובהק?

ה. על עוזר מחקר הטיל החוקר לבדוק את ההשערה כי על כל 1000 ₪ נוספים בהכנסה צורך הפרט 700 ₪, כנגד ההשערה כי הוא צורך יותר מ-700 ₪. נסח את השערת האפס ואת ההשערה האלטרנטיבית.

ו. מהו הסטטיסטי t לבדיקת ההשערה?

ז. מהו הסטטיסטי WALD לבדיקת ההשערה?

ח. התברר כי הייתה טעות בנתונים, וכי יש להוסיף 1000 ₪ לתצרוכת של כל משק בית:

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

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

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

iv. 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:

$$EXPENSE_t = \alpha + \beta_1 \cdot INCOME_t + \beta_2 \cdot SAVINGS_t + \beta_3 \cdot NE_t + u_t \text{ ,}$$

וקיבל כי סכום ריבועי הסטיות של הטעויות הוא 121.

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

י. מהו הסטטיסטי WALD לבדיקת ההשערה?

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

יא. מהי השערת האפס לבדיקה זו?

יב. 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.

תשובות סופיות:

1) א.

Analysis of Variance

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

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

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

ב. יש עדות לכך. ג. יש עדות לכך.

2) ירד.

3) א. $H_0: \alpha = 5 \ , \ \beta_s = 3\beta_z$. ב. $Y_t - 5 = \beta_x X_t + \beta_z (Z_t + 3S_t) + \beta_w W_t + u_t$.

ג. $WALD_{stat} = 0.6145$. ד. מונה: 2 , מכנה: 199-.

ה. מקבלים.

4) בדיקה ע"י מבחן WALD ו-t: אין עדות לכך.

5) א. $PF = 0.000$. ב. 92%. ג. $\hat{\alpha} = 0.04195$. ד. לא.

ה. $H_0: \beta = 0.70 \ , H_1: \beta > 0.70$. ו. $t_{\hat{\beta}} = 1.583$. ז. $WALD_{stat} = 2.505$.

ח. i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ט. $H_0: \beta_2 = \beta_3 = 0 \ , H_1: OTHERWISE$. י. $WALD_{stat} = 68.32$.

יא. $H_0: \beta_3 = 2 \cdot \beta_2 \ , \ \beta_1 = 0.6$.

יב. $W_t = SAVINGS_t + 2 \cdot NE_t \ , \ Z_t = EXPENCE_t - 0.6 \cdot INCOME_t$.

אקונומטריקה למדעי הנתונים

פרק 14 - מבחן 1

תוכן העניינים

מבחן 1:

שאלות:

1) 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, וההרעה 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	---- – ----

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

i. המקדם של $\ln Y$ לא ישתנה. נכון / לא נכון / לא ניתן לדעת

ii. החותך לא ישתנה. נכון / לא נכון / לא ניתן לדעת

iii. הסטטיסטי t לבדיקת המובהקות של β לא ישתנה. נכון / לא נכון / לא ניתן לדעת

iv. הסטטיסטי F לבדיקת מובהקות המודל לא ישתנה. נכון / לא נכון / לא ניתן לדעת

v. R^2 לא ישתנה. נכון / לא נכון / לא ניתן לדעת

החוקר טען כי גם גודל האוכלוסייה, P, משפיע על ההשקעה לפי המודל הבא: $\ln I_t = \alpha + \beta_1 \ln Y_t + \beta_2 \ln P_t + u_t$.

ח. מהי השערת האפס לבדיקת הטענה?

התקבל הפלט הבא:

Parameter Estimates

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

ט. באיזו רמת מובהקות נקבל את טענת החוקר?

י. 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.

יא. מהי השערת האפס לבדיקת הטענה?

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

יג. האם ניתן לדחות את השערת האפס?

2) ענה על הסעיפים הבאים:

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

ב. אם הערך 0 נמצא בתוך רווח הסמך ל-β, אזי β מובהקת. נכון / לא נכון / לא ניתן לדעת

ג. בהוספת משתנה לא רלוונטי למודל האומד המתוקן לפרופורציית השונות המוסברת ירד בהכרח. נכון / לא נכון / לא ניתן לדעת

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

ה. אם דוחים H0 ברמת מובהקות מסוימת, אזי דוחים H0 בכל רמות המובהקות הקטנות יותר. נכון / לא נכון / לא ניתן לדעת

ו. אומד חסר הטיה הוא אינו בהכרח אומד עקיב. נכון / לא נכון / לא ניתן לדעת

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

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

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

ג. האומד $\tilde{\beta}$ הוא אומד ליניארי. נכון / לא נכון / לא ניתן לדעת

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

ה. מהי השונות של $\tilde{\beta}$?

4) נתון מודל ללא חותך: $Y_t = \beta X_t + u_t$, ונתון האומד: $\hat{\beta} = \dfrac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$.

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

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

ג. האומד $\hat{\beta}$ הוא אומד ליניארי. נכון / לא נכון / לא ניתן לדעת

ד. מהי השונות של $\hat{\beta}$?

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

תשובות סופיות:

1) א. $PF=0.0001$. ב. 0.57% ג. $p(1.79\le\alpha\le5.15)=0.95$,
$p(0.026\le\beta\le1.11)=0.95$. ד. $\begin{matrix}H_0:\beta=0.4\\H_1:\beta\ne0.4\end{matrix}$.
ה. $\ln I_t-0.4\ln Y_t=\alpha+u_t$. ו. $WALD_{stat}=7.054$.
ז.i. לא נכון. ii. לא נכון. iii. נכון. iv. נכון. v. נכון.
ח. $H_0:\beta_2=0$. ט. $Pt_{\hat{\beta}}=0.0736$. י. לא נכון.
יא. $H_0:\beta_1+\beta_2=0$. יב. $t=-1.089$. יג. אין סיבה מספקת.

2) א. לא נכון. ב. לא נכון. ג. לא נכון. ד. לא נכון. ה. לא נכון.
ו. נכון.

3) א. לא נכון. ב. לא נכון. ג. נכון. ד. לא ניתן לדעת.
ה. $V\left(\tilde{\beta}\right)=\frac{\sum X_t^2\sigma^2}{S^2{}_{XX}}$.

4) א. נכון. ב. נכון. ג. נכון. ד. $V\left(\tilde{\beta}\right)=\frac{\sigma_u^2}{\sum X_t^2}$.

ה. נכון.

אקונומטריקה למדעי הנתונים

פרק 15 - מבחן 2

תוכן העניינים

מבחן 2:

שאלות:

1) 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: $S_{XX} = 35079$, $\bar{X} = 46.040873$:

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

א. מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?
ב. מהו האומדן לשכר התחלתי?

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

ג. מהן ההשערות לבדיקת הטענה?
ד. מהו הסטטיסטי t למבחן?
ה. מהו הסטטיסטי WALD למבחן?
ו. מהי התחזית לשכר של עובד העובד 55 שעות בשבוע?

ז. 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:

i. ישנה את הסטטיסטי t לבדיקת המובהקות של α. — נכון / לא נכון / לא ניתן לדעת

ii. יכפיל את האומד של β ב-0.23. — נכון / לא נכון / לא ניתן לדעת

iii. יכפיל את סטית התקן של $\hat{\beta}$ ב-0.23. — נכון / לא נכון / לא ניתן לדעת

iv. 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:

$$SALARY_t = \alpha + \beta_1 \cdot HOURS_t + \beta_2 \cdot AGE_t + \beta_3 \cdot SCL_t + u_t .$$

ח. מהי השערת האפס לבדיקת הטענה?

ט. מהו הנתון הנדרש כדי לחשב את הסטטיסטי של WALD לבדיקת טענת החוקר?

י. בפלט האמידה של המשוואה החדשה לא היה ברור אם ערכו של נתון זה הוא 315968434 או 515968434 (בשל בעיה במדפסת). מהו הסטטיסטי של WALD לבדיקת טענת החוקר?

יא. 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

יב. הנתונים בפלט אינם מספיקים לבדיקת ההשערה לפי מבחן t.
מהו הנתון החסר? באיזה פלט של SAS ניתן למצוא אותו?

יג. בהנחה שנתון זה הוא 8.3969, חשב את הסטטיסטי t לבדיקת הטענה.
מהי מסקנתך לגבי נכונות הטענה?

יד. אם תרצה לבדוק את הטענה לפי מבחן WALD, יהיה המודל המוגבל:

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

$Z_0 =$ __________

$Z_1 =$ __________

$Z_2 =$ __________

טו. אם יש מספיק נתונים, חשב את הסטטיסטי של WALD לבדיקת הטענה?

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

נתון האומד: $\tilde{\beta} = \dfrac{S_{XY}}{\sum_{t=1}^{T} X_t^2}$.

א. אומד זה הוא הפתרון של המשוואות הנורמליות: $\sum_{t=1}^{T} \hat{u}_t = 0$, $\sum_{t=1}^{T} \hat{u}_t X_t = 0$. נכון / לא נכון / לא ניתן לדעת

ב. התוחלת של $\tilde{\beta}$ היא:

i. β.

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

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

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

v. כל התשובות אינן נכונות.

ג. הטענה כי: $E\left(\tilde{\beta}\right) < \beta$:

i. תמיד נכונה.

ii. אינה נכונה.

iii. נכונה אם ורק אם: $\bar{X} > 0$.

iv. נכונה אם ורק אם: $\bar{X} \neq 0$.

v. כל התשובות אינן נכונות.

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

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

.ii $\dfrac{\sigma^2}{S_{XX}^2}$.

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

.iv $\dfrac{\sigma^2}{S_{XX}}$.

.v כל התשובות אינן נכונות.

ה. אם $\bar{X}=O$, אז $\tilde{\beta}$ הינו האומד הלינארי חסר ההטיה בעל השונות הקטנה ביותר. נכון / לא נכון / לא ניתן לדעת

3) נתון המודל: $Y_t=\alpha+\beta\cdot X_t+u_t$. ידוע כי כל ההנחות הקלאסיות מתקיימות.
נתון כי $\tilde{\beta}$ הוא אומד לינארי וחסר הטיה ל- β, אך איננו אומד עקיב ל- β.
מאחר ש- $\tilde{\beta}$ אינו אומד עקיב, לא נוכל להשתמש במשפט גאוס מרקוב ולקבוע
כי: $\hat{\beta}=\dfrac{S_{XY}}{S_{XX}}$ (אר"פ) הינו אומד יעיל יותר.

נכון / לא נכון / לא ניתן לדעת

תשובות סופיות:

1) א. $PF = 0.00$.
ב. $\hat{\alpha} = -80.5246$.
ג. $H_0: \beta = 40$, $H_1: \beta \neq 40$.
ד. $t_{\hat{\beta}} = -0.75$.
ה. $WALD_{stat} = 0.5625$.
ו. $SALARY_t = 1903.16$.
ז. i. לא נכון. ii. נכון. iii. נכון. iv. לא נכון.
ח. $H_0: \beta_2 = \beta_3 = 0$.
ט. SSE.
י. $WALD_{stat} = 54.49$.
יא. לא ניתן לחשב.
יב. $S^2_{\hat{\beta}_2,\hat{\beta}_3}$, Covariance of Estimates.
יג. נכונה.
יד. $Z_0 = SALARY_t$, $Z_1 = HOURS_t$, $Z_2 = AGE_t + 8 \cdot SCL_t$.
טו. $WALD_{stat} = 0.000324$.

2) א. לא נכון. ב. ה. ג. ה. ד. ה. ה. לא נכון.

3) לא נכון.

אקונומטריקה למדעי הנתונים

פרק 16 - מבחן 3

תוכן העניינים

מבחן 3:

שאלות:

1) על מנת לאמוד את פונקציית הייצור נאספו נתונים על 150 פירמות בשנת 2007 ונאמדה המשוואה הבאה:

1. $\ln(Y_t)=\alpha+\beta_1\cdot\ln(L)_t+U_t$.

כאשר:

$\ln(Y)_t$ - תפוקה שנתית באלפי ₪ בלוגים.

$\ln(L)_t$ - מספר העובדים בלוגים.

U_t - הטעות המקרית המקיימת את כל ההנחות הקלאסיות.

משוואה מס' 1 נאמדה בפלט 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>ITI
INTERCEP	1	4.389949	0.21003743	20.901	0.0001
lnL	1	0.257487	0.04767276		0.0001

א. סטטיסטי F לבדיקת מובהקות המודל:

i. לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ________________

ב. סטטיסטי t לבדיקת מובהקות המודל:

i. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה

ii. לא ניתן לחשבו בעזרת הנתונים הקיימים.

iii. ניתן לחשבו וערכו הוא: ________________.

הועלתה הטענה כי עליה ב-1% במסי' העובדים תגדיל את התפוקה בפחות מ-1%.

ג. ההשערות לבדיקת הטענה הן: $H_0 :$ ______ . $H_1 :$ ______

ד. הסטטיסטי לבדיקת הטענה הינו:
i. לא ניתן לחשבו בנתונים הקיימים.
ii. 5.5.
iii. -5.5.
iv. -15.5.
v. 15.5.

ה. הסטטיסטי של WALD לבדיקת הטענה:
i. לא ניתן לחשבו בעזרת הנתונים הקיימים.
ii. ניתן לחשבו וערכו הוא: ______________.

ו. לאור התשובות לסעיפים הקודמים, אחוז התפוקה קטן ככל שאחוז מסי' העובדים גדל: נכון/לא נכון/אי אפשר לדעת

החוקרת טענה כי יש משתנים נוספים המסבירים את תפוקת הפירמה ואמדה את המשוואה הבאה:

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:

$\ln(K)_t$ - מלאי ההון של הפירמה באלפי ₪ בלוגים.

$\ln(PY)_t$ - הוצאות למחקר ופיתוח באלפי ₪ בלוגים.

משוואה מסי' (2) נאמדה בפלט מסי' 2 .

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	3	15.63370	5.21123	22.825	0.0001
Error	146	33.33425	0.22832		
C Total	149	48.96795			

Root MSE	0.47783	R-square	0.3193
Dep Mean	5.54003	Adj R-sq	0.3053
C. V.	8.62496		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	0.542062	1.66317350	0.326	0.7450
lnL	1	0.267771	0.08146608	3.287	0.0013
lnK	1	0.405694	0.09700769	4.182	0.0001
lnPY	1	0.406149	0.30781185	1.319	0.1891

ז. ההשערות לבדיקת הטענה הינן: $H_0 :$ ________ . $H_1 :$ ________

ח. הסטטיסטי של WALD לבדיקת הטענה הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. ניתן לחישוב וערכו: ________________________________.

ט. הסטטיסטי של t לבדיקת הטענה הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. לא ניתן לחשב סטטיסטי t לטענה מסוג זה

iii. ניתן לחישוב וערכו: ________________________________.

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

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

כאשר:

משוואה מסי (3) נאמדה בפלט מסי 3.

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	15.23620	7.61810	33.199	0.0001
Error	147	33.73175	0.22947		
C Total	149	48.96795			

Root MSE	0.47903	R-square	0.3111
Dep Mean	5.54003	Adj R-sq	0.3018
C. V.	8.64667		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>ITI
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

י. ההשערות לבדיקת הטענה הינן : $H_0:$ _________ $H_1:$ _________.

יא.הסטטיסטי של WALD לבדיקת הטענה הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים

ii. ניתן לחישוב וערכו: ______________________________________.

הועלתה הטענה כי גמישות התפוקה ביחס להון גדולה פי 2 מגמישות התפוקה ביחס לעבודה.
בדקו את הטענה במשוואה (3).

יב. השערת האפס לבדיקת הטענה היא : $H_0:$ _________.

יג. הסטטיסטי t לבדיקת הטענה הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחישוב וערכו: ____________________________.

יד. הרגרסיה המוגבלת כאשר H_0 נכונה ("תחת H_0") למבחן WALD הינה: $Z_0=\gamma_0+\gamma_1\cdot Z_1+V$.
כאשר:
$Z_1=$ ________
$Z_0=$ ________.

טו.הסטטיסטי של WALD לבדיקת הטענה (חשבי ישירות) :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחישוב וערכו: ____________________________.

טז.נטען כי אם נמדוד את המשתנים הב"ת במודל בדולרים במקום בשקלים, האומדים ל- β ול- α יישארו ללא שינוי (הנח כי שער הדולר הוא 3.5 ₪): נכון/לא נכון/אי אפשר לדעת

יז. נטען שאם נוריד את משתנה PY מהמודל ה- $\overline{R}^2$ יעלה: נכון/לא נכון/אי אפשר לדעת

2) ענו על כל השאלות הבאות. כל שאלה בפני עצמה. בכל השאלות מונח המודל: $Y=\alpha+\beta X+U$ (ומתקיימות כל ההנחות הקלאסיות).

א. במודל לוגריתמי כפול β מייצגת את שיעור השינוי השולי: נכון/לא נכון/ אי אפשר לדעת

ב. במודל ללא חותך מתקיימת המשוואה הנורמאלית: $\sum\hat{u}_t x_t=0$ בלבד: נכון/לא נכון/אי אפשר לדעת

ג. כאשר מוסיפים משתנה ב"ת למודל, עליה ב- $\bar{R}^2$ מעידה על כך שהמשתנה שהוסף מובהק באוכלוסייה: נכון/לא נכון/אי אפשר לדעת

ד. אם הנחה מס' 3 ($E(\hat{u})=0$ לכל t) איננה מתקיימת, האומדים של המודל לא יהיו יעילים: נכון/לא נכון/אי אפשר לדעת

ה. ככל ש- S_{XX} גדול יותר, קל יותר לדחות את H_0 למובהקות ה- β: נכון/לא נכון/אי אפשר לדעת

ו. $R^2 > \bar{R}^2$ מתקיים תמיד: נכון/לא נכון/אי אפשר לדעת

ז. מבחן F למובהקות המודל מהווה מקרה פרטי של מבחן t למובהקות ה- β: נכון/לא נכון/אי אפשר לדעת

ח. ככל שגודל המדגם גדל כך האומד יהיה יעיל יותר לפרמטר באוכלוסייה: נכון/לא נכון/אי אפשר לדעת

ט. ה-PVALUE גדל ביחס הפוך לרמת המובהקות של המבחן (ה- α): נכון/לא נכון/אי אפשר לדעת

י. אם דחינו את H_0 במבחן t למובהקות ה- β כאשר האומד חיובי, נדחה אותה בהכרח גם ביחס להשערה כי מקדם השיפוע חיובי באוכלוסייה: נכון/לא נכון/אי אפשר לדעת

יא. אם ידוע כי הקשר בין X ל-Y מובהק באוכלוסייה, הדבר מעיד בהכרח על מובהקות המודל: נכון/לא נכון/אי אפשר לדעת

3) נתון המודל: $Y_t = \beta X_t + U_t$.

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

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

ב. שונותו של האומד: ______________________.

ג. על סמך משפט גאוס מרקוב ניתן להסיק כי אר"פ הינו אומד יעיל יותר מ- $\tilde{\beta}$: נכון/לא נכון/אי אפשר לדעת

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

ה. אם נתון ש: $\bar{X} = 0$ אזי $\tilde{\beta}$ הינו אומד הריבועים הפחותים: נכון/לא נכון/אי אפשר לדעת

תשובות סופיות:

1) א. ii , $F = 31.273$.
ב. iii , $t = 5.5$.
ג. $\begin{matrix} H_0 : \beta = 1 \\ H_1 : \beta < 1 \end{matrix}$.
ד. iv.
ה. i.
ו. לא נכון.
ז. ראו סרטון.
ח. ראו סרטון.
ט. ראו סרטון.
י. $\begin{matrix} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 \neq 0 \end{matrix}$.
יא. ii , $WALD_{stat} = 1.74$.
יב. $H_0 : \beta_2 = 2 \cdot \beta_1$.
יג. ii , $t = 0.1417$.
יד. $\begin{matrix} Z_0 = \ln(Y)_t \\ Z_1 = \ln(L)_t + 2\ln(K)_t \end{matrix}$.
טו. ii , $WALD_{stat} = 0.585$.
טז. לא נכונה.
יז. לא נכון.

2) א. לא נכון.
ב. נכון.
ג. לא נכון.
ד. לא נכון.
ה. נכון.
ו. נכון.
ז. נכון.
ח. נכון.
ט. לא נכון.
י. נכון.
יא. נכון.

3) א. נכון.
ב. $V(\tilde{\beta}) = \frac{\sigma^2}{S_{XX}}$.
ג. נכון.
ד. לא נכון.
ה. נכון.

אקונומטריקה למדעי הנתונים

פרק 17 - מבחן 4

תוכן העניינים

מבחן 4:

שאלות:

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

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

כאשר:

$CREDIT_t$ - סך הפעילות בכרטיסי אשראי ב-₪.

$SAVINGS_t$ - סך הפעילות בחשבונות חיסכון ב-₪.

U_t - סטיה מקרית המקיימת את כל ההנחות הקלאסיות.

משוואה (1) נתונה בפלט מס' 1.

Dependent Variable: credit

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	---	-----	------	------	<0.0001
Error	---	-----	------		
C Total	---	-----			

Root MSE	43859	R-square	0.0106
Dep Mean	7433.60809	Adj R-sq	0.0106
C. V.	589.99662		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0:	Prob>lTl	95% Confidence	
INTERCEP	1	11151.91516	394.35144	2.92	0.0035	378.97	1924.8
savings	1	0.56719	0.02884	19.67		0.51	0.623

א. סטטיסטי F לבדיקת מובהקות המודל הינו:

i. לא ניתן לחשב את סטטיסטי F בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ____________________.

ב. PVALUE של סטטיסטי t לבדיקת מובהקות ה- β:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.

iii. ניתן לחשבו וערכו: ____________________.

הבנק טען שאם יגדילו לקוחותיו את הפעילות בחשבונות חיסכון שלהם אפילו בשקל אחד, הפעילות בכרטיסי אשראי תגדל ביותר מ 40 אגורות.

ג. ההשערות לבדיקת הטענה הינן: H_0 : ______ . H_1 : ______

ד. הסטטיסטי לבדיקת טענת הבנק הינו:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. הסטטיסטי לבדיקת הטענה צריך להיות שלילי.

iii. 19.67.

iv. 5.797.

ה. הסטטיסטי של WALD לבדיקת טענת הבנק:

i. לא ניתן לחשב בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו: ______________________.

ו. ברמת ביטחון של 95% מהו טווח הגידול בפעילות בכרטיסי אשראי, על כל שקל נוסף בפעילות בחשבונות חיסכון?

ז. ברמת ביטחון של 95% מהו האומד לתוחלת פעילות בכרטיסי אשראי עבור סך פעילות בחשבונות חיסכון של 50,000 ₪?

ח. אם פעילות כרטיסי האשראי של כל לקוח תגדל ב- 1000 ₪:

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

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

iii. 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:

2. $CREDIT_t = \alpha + \beta_1 \cdot PIKADON1_t + \beta_2 \cdot PIKADON2_t + U_t$.

כאשר:

$PIKADON1_t$ - סה"כ הפקדה לפקדונות יומיים ב-₪.

$PIKADON2_t$ - סה"כ הפקדה לפקדונות חודשיים ב-₪.

משוואה (2) נאמדה בפלט מס' 2.

Dependent Variable: lnY

Analysis of Variance

Source	DF	Sum of Squares	Mean Squares	F Value	Prob>F
Model	2	1.00791E12	5.003955E11	261.10	0.0001
Error	35968	6.893195E13	1916479937		
C Total	35970	6.993274E13			

Root MSE	43778	R-square	0.0143
Dep Mean	7433.68809	Adj R-sq	0.0143
C. V.	588.90847		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>lTl
INTERCEP	1	1259.36230	379.00751	3.32	0.0009
Pikadon1	1	0.07552	0.05539	1.36	0.1728
Pikadon2	1	0.72350	0.03199	22.62	0.0001

Covariance of Estimates

COVB	INTERCEP	Pikadon1	Pikadon2
INTERCEP	143646.69097	-8.178835194	-9.154578973
Pikadon1	-8.176835154	0.0030678685	0.0003564263
Pikadon2	-9.15457897	0.0003564263	0.0010231462

ט. השערת האפס לבדיקת הטענה הינה : _______ : H_0 .

י. הסטטיסטי של WALD לבדיקת הטענה :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו :________________________.

יא. הסטטיסטי של t לבדיקת הטענה :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. לא ניתן להשתמש בסטטיסטי t בהשערה מסוג זה.

iii. 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.

יב. השערת האפס לבדיקת הטענה הינה : _______ : H_0 .

יג. הסטטיסטי t לבדיקת הטענה הינו :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא :____________________.

יד. PVALUE של סטטיסטי t מהסעיף הקודם :

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

ii. ניתן לחשבו וערכו הוא :____________________________.

טו. הרגרסיה המוגבלת כאשר H_0 נכונה למבחן WALD
הינה: $D_0 = \gamma_0 + \gamma_1 \cdot D_1 + \gamma_2 \cdot D_2 + v$.
D_0 : ______
כאשר: D_1 : ______ .
D_2 : ______

טז. על פי משוואה מס' 2, כל שקל שיועבר
לפיקדון הראשון יוסיף כ-0.07552 ₪
לסה"כ הפעילות בכרטיסי אשראי: נכון/לא נכון/אי אפשר לדעת

2) ענו על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל: $Y = \alpha + \beta \cdot X + U$ ומתקיימות כל ההנחות הקלאסיות).

א. אם המודל מובהק אזי שיפוע
הרגרסיה מובהק בהכרח: נכון/לא נכון/אי אפשר לדעת

ב. הגמישות במודל חצי לוגרתמי
היא קבועה: נכון/לא נכון/אי אפשר לדעת

ג. אם X_2 מהווה קומבינציה ליניארית
של X_1 לא ניתן לאמוד את הרגרסיה
המרובה: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + U$: נכון/לא נכון/אי אפשר לדעת

ד. $\bar{R}^2 > R^2$ רק בתנאי שהמודל מובהק: נכון/לא נכון/אי אפשר לדעת

ה. ליניאריות וחוסר הטיה של האומדים
מהווים תנאי הכרחי לעקיבותם: נכון/לא נכון/ אי אפשר לדעת

ו. 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

ז. ככל שפיזור U_t גדול יותר כך קשה יותר
לדחות את H_0 למובהקות המודל: נכון/לא נכון/ אי אפשר לדעת

ח. מודלים לא ליניאריים מתארים קשרים שאינם
ליניאריים בין המשתנה המסביר למוסבר: נכון/לא נכון/אי אפשר לדעת.

ט. אם הנחה 5 (שונות קבועה) לא מתקיימת, אומדי
הריבועים הפחותים אינם חסרי הטיה: נכון/לא נכון/אי אפשר לדעת

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

3) נתון המודל: $Y_t = \beta \cdot X_t + U_t$, כאשר כל ההנחות הקלאסיות מתקיימות.

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

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

תשובות סופיות:

1) א. ii , $F = 386.9089$. ב. iii , $PF < 0.0001 = Pt$. ג. $\begin{matrix} H_0 : \beta = 0.4 \\ H_1 : \beta > 0.4 \end{matrix}$.

ד. iv. ה. i. ו. $p\left(0.51 \le \beta \le 0.623\right) = 0.95$.

ז. $p\left(-32,387,174.83 \le E\left(Y\right) \le 32,458,197.67\right) = 0.95$.

ח. i. נכון. ii. לא נכון. iii. נכון. ט. $H_0 : \beta_1 = \beta_2 = 0$.

י. i. יא. ii. יב. $H_0 : \beta_2 = 10 \cdot \beta_1$. יג. ii , $t = -0.0574$.

יד. ii , $PVALUE > 0.1$. טו. $\begin{matrix} D_0 : CREDIT_t \\ D_1 : SAVINGS_t \\ D_2 : PIKADON1_t + 10 \cdot PIKADON2_t \end{matrix}$.

טז. נכון.

2) א. נכון. ב. לא נכון. ג. נכון. ד. לא נכון.
ה. לא נכון. ו. נכון. ז. נכון. ח. לא נכון.
ט. לא נכון. י. לא נכון.

3) א. $E\left(\tilde{\beta}\right) = \frac{\beta \sum X_t}{S_{XX}}$. ב. לא ניתן לדעת. ג. נכון.

ד. $V\left(\tilde{\beta}\right) = \frac{T\sigma^2}{S_{XX}^2}$. ה. לא נכון.

אקונומטריקה למדעי הנתונים

פרק 18 - מבחן 5

תוכן העניינים

מבחן 5:

שאלות:

1) על מנת לאמוד את הקשר בין רמת המחירים במשק (P) לכמות הכסף (M), נאספו נתונים חודשיים בשנים 86-94 (סה"כ 105 תצפיות) ונאמדה המשוואה הבאה:

1. $M_t = e^{\alpha} + p^{\beta} + e^{u}$.

כאשר:

m - כמות הכסף במשק לחודש (מזומנים + עו"ש).

p - מדד המחירים לצרכן במשק.

U_t - סטיה מקרית המקיימת את כל ההנחות הקלאסיות.

משוואה מס' (1) נאמדה בפלט מס' 1.

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: Paramete r=0	Prob>ITI
INTERCE P	1	1.49372	0.09862	15.15	<.0001
lnp	1	1.69267	0.02360		<.0001

א. כתבו את המשוואה בצורה ליניארית בעזרת הטרנספורמציה המתאימה.

ב. האומדן למשוואה (1) הינו: ________________________.

ג. המשמעות הכלכלית של β היא: _______________

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

גבול תחתון: ______________________.

גבול עליון: ______________________.

ה. ערך t לחישוב מובהקות ה-β הינו:

i. לא ניתן לחשב ערך זה בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ______________________.

ו. אם נגדיל את מדד המחירים לצרכן ביחידה אחת, כמות הכסף במשק תגדל ב:

i. 71.7233.

ii. 1.69267.

iii. 169.267.

iv. 1.69267%.

v. אף תשובה איננה נכונה.

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

ז. ההשערות לבדיקת הטענה: ____________________.

ח. סטטיסטי t לבדיקת הטענה הינו:

i. לא ניתן לחשב באמצעות הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ________________________________.

ט. על פי התשובות לסעיפים הקודמים ניתן להסיק כי ערכו של סטטיסטי F לבדיקת מובהקות המודל הינו:

i. לא ניתן לחשב את ערכו של סטטיסטי F על סמך סטטיסטי t.

ii. 861.4225.

iii. 5144.23.

iv. 71.7233 4.

י. אם נוציא שורש ריבועי למדד המחירים לצרכן במשק:

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

ii. האומד של β יעלה: נכון/לא נכון/ אי אפשר לדעת

iii. 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:

2. $LN(M)_t = \alpha + \beta_1 \cdot LN(P)_t + \beta_2 \cdot LN(Y)_t + U_t$.

משוואה (2) נתונה בפלט 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>ITI
INTERCEP	1	0.78242	0.59739	1.31	0.1932
lnp	1	1.63491	0.05332	30.66	<.0001
lny	1	0.20001	0.16568	-----	0.2302

Covariance of Estimates

COVB	INTERCEP	lnp	lny
INTERCEP	0.35687	0.025884	-0.09762
lnp	0.02588	0.002843	-0.00792
lny	-0.09762	-0.00792	0.02745

יא. סטטיסטי t לבדיקת הטענה הינו :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ________________________.

יב. על פי התשובה לסעיף הקודם, ניתן להסיק את ערכו של סטטיסטי F למובהקות המודל. נכון/ לא נכון/ אי אפשר לדעת

יג. על פי התשובה לסעיף יא' ניתן להסיק את ערכו של סטטיסטי WALD לבדיקת הטענה. נכון/לא נכון/ אי אפשר לדעת

הועלתה הטענה כי הגמישות ביחס למחיר גבוהה פי 10 מהגמישות ביחס לפעילות הכלכלית במשק.

יד. סטטיסטי WALD לבדיקת הטענה הינו :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ____________________________.

טו. הרגרסיה המוגבלת כאשר H_0 נכונה למבחן WALD הינה: __________

כאשר: D_0 : _______

D_1 : _______ .

טז.

i. איזה מבין המודלים המוצעים במשוואות 1 ו-2 עדיף?

משוואה 1/משוואה2/אין הבדל בין המודלים

ii. אם משתנה רמת המחירים במשק היה מובהק במשוואה מס' 1, הוא יהיה מובהק בהכרח גם במשוואה מס' 2 : נכון/לא נכון/אי אפשר לדעת

2) ענו על השאלות הבאות (כל שאלה בפני עצמה, בכל שאלה מונח המודל: $Y=\alpha+\beta\cdot X+U$ ומתקיימות כל ההנחות הקלאסיות).

א. $\bar{R}^2 < R^2$ מתקיים תמיד: נכון/לא נכון/אי אפשר לדעת.

ב. 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.

ג. אם ערך האומד ל- β גבוה, השערת האפס למובהקות השיפוע תידחה בוודאות : נכון/לא נכון/אי אפשר לדעת.

ד. הוספת משתנה מסביר למשוואת הרגרסיה עשויה להקטין את R^2 : נכון/לא נכון/אי אפשר לדעת.

ה. 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.

ו. אם רווח בר סמך לשיפוע כולל את הערך אפס, ניתן לומר כי השערת האפס למובהקות השיפוע מתקבלת בהכרח: נכון/לא נכון/אי אפשר לדעת.

ז. האומדים היעילים ביותר לפרמטרים באוכלוסייה יהיו בהכרח אומדי הריבועים הפחותים : נכון/לא נכון/אי אפשר לדעת.

ח. בהוספת משתנה מסביר מובהק למודל, ערך $\bar{R}^2$ יעלה בהכרח. נכון/לא נכון/אי אפשר לדעת.

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

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

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

נתון כי אר"פ למודל זה הינו: $\hat{\beta} = \frac{S_{XY}}{S_{XX}}$.

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

ב. חשבו את $VAR\left(\hat{\beta}\right)$.

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

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

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

תשובות סופיות:

1) א. $LN(M)_t = \alpha + \beta \cdot LN(P)_t + U_t$. ב. $LN(M)_t = 1.49372 + 1.69267 \cdot LN(P)_t$.

ג. גמישות. ד. גבול תחתון : 1.64527 , גבול עליון : 1.73987.

ה. ii, $t_{\beta=0} = 71.7233$. ו. v. ז. $\begin{matrix} H_0 : \beta = 1 \\ H_1 : \beta > 1 \end{matrix}$.

ח. ii, $t = 29.35$. ט. i. י. i. לא ניתן לדעת.

ii. אי אפשר לדעת. iii. אי אפשר לדעת. יא. ii, $t = 1.2$.

יב. לא נכון. יג. נכון. יד. ii, $WALD = 0.048$.

טו. $\begin{matrix} D_0 = LN(M)_t \\ D_1 = 10 \cdot LN(P)_t + LN(Y)_t \end{matrix}$. טז. i. משוואה 1.

ii. לא נכון.

2) א. נכון. ב. לא נכון. ג. לא נכון. ד. לא נכון.

ה. לא נכון. ו. לא נכון. ז. לא נכון. ח. נכון.

ט. לא נכון. י. נכון.

3) א. הוכחה. ב. $V(\hat{\beta}) = \frac{\sigma^2}{S_{XX}}$. ג. הוכחה.

ד. ראו סרטון.

אקונומטריקה למדעי הנתונים

פרק 19 - מבחן LM

תוכן העניינים

מבחן LM:

רקע:

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: $H_0 : \beta_3 = \beta_4 = 0$, $H_1 : OTHERWISE$.

המודל המוגבל (RESTRICTED): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t}$.
במבחן LM אומדים את המודל המוגבל ומקבלים עבור כל תצפית את הסטייה מקו הרגרסיה: $Y_t - \hat{Y}_t = \hat{u}_t$.
כעת אומדים את רגרסיית העזר שבה מנסים לנבא את הסטייה מקו הרגרסיה עבור כל תצפית: $\hat{u}_t = \delta_0 + \delta_1 x_{1t} + \delta_2 x_{2t} + \delta_3 x_{3t} + \delta_4 x_{4t} + \omega_t$.
חישוב הסטטיסטי: (R^2 של רגרסיית העזר * מספר התצפיות) $LM_{stat} = R^2 \cdot T$.
כלל הכרעה: אם: $LM_{stat} > \chi^2_m$ נדחה את H_0 (m = מס' ההגבלות ב- H_0).

- 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.

שאלות:

1) נניח מודל הכולל 4 משתנים מסבירים (UNRESTRICTED):

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

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

א. בדוק את הטענה כי לפחות אחד מן המשתנים הנוספים רלוונטי למודל בשתי דרכים.

ב. איזה מן המשתנים הנוספים רלוונטי למודל?

ג. הסבירו את הקשרים בין שלוש המשוואות: R, U ועזר ואת הקשר בין מבחן WALD ומבחן LM.

U: $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{\beta}_4 x_{4t} + \hat{v}_t$.

R: $\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{u}_t$.

עזר: $\hat{u}_t = \hat{\delta}_0 + \hat{\delta}_1 x_{1t} + \hat{\delta}_2 x_{2t} + \hat{\delta}_3 x_{3t} + \hat{\delta}_4 x_{4t} + \hat{\omega}_t$.

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

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

תשובות סופיות:

1) א. מבחן LM ומבחן WALD, יש עדות לכך.

ב. $pt_{\hat{\beta}_3} = pt_{\hat{\beta}_4} = 0.00$.

ג. i. U=R+עזר.

ii. $ESS_U = ESS_{ע}$.

iii. $ESS_R = TSS_{ע}$.

iv. $R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{ESS_U}{ESS_R} = \frac{ESS_R - ESS_U}{ESS_R}$.

ד. $LM_{stat} = 43.489$.

ה. $WALD_{stat} = 26.962$.

אקונומטריקה למדעי הנתונים

פרק 20 - בעיות ספציפיקציה

תוכן העניינים

בעיות ספציפיקציה:

רקע:

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

1. הוספת משתנה לא רלוונטי:
למשל, המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$.
המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon$.
אם נקבל את H_0 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.

2. השמטת משתנה רלוונטי:
למשל, המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$.
המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$.
בהיעדר x_2, בדיקות ההשערות לפרמטרים של המודל הטעותי אינן תקפות:

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

אקונומטריקה למדעי הנתונים

פרק 21 - תיאוריה מולטיקוליניאריות

תוכן העניינים

מולטיקוליניאריות:

רקע:

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

נבחין בין מולטיקוליניאריות מלאה לחלקית.

מולטיקוליניאריות מלאה:

מתאם מלא בין המשתנים המסבירים במודל.
הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני: $x_1 = a + bx_2$ (x_1 הוא קומבינציה ליניארית מלאה של x_2)
מכאן ש: $r_{12} = 1$.

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

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני.
מדוע זה בעייתי?
כיוון שלא ניתן לאמוד את המודל שכן אר״פ אינם מוגדרים.
פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקוליניאריות חלקית:

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

מכיוון שיש מתאם גבוה בין המשתנים הב״ת לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי.
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 מולטיקולינאריות 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 הב"ת, באופן שלא מאפשר לדחות את השערת האפס למובהקות השיפועים: $S^2_{\hat{\beta}_1}=\frac{MSE}{SSX_1(1-r_{12})}$, $t=\frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$.

2. רגישות לספציפיקציה – הורדת משתנה ב"ת שאיננו מובהק תהפוך משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקולינאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים הב"ת האחרים.

3. סימנים הפוכים – כאשר השיפועים של המשתנים הב"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.

השלכות של מולטיקולינאריות חלקית:
מולטיקולינאריות 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 למולטיקולינאריות מלאה).
במובן הזה, בעיה של מולטיקולינאריות חלקית דומה לבעיה של הוספת משתנה ב"ת שאיננו רלוונטי.

פתרונות למולטיקולינאריות 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 הכל יש עליה ב- $AdjR^2$ (לפי חוק חיטובסקי).

2. ניתן לעיתים לאחד את שני המשתנים למשתנה אחד.

שלבי בדירת ההשערות:

1. מבצעים מבחן F לבדיקת מובהקות המודל.
2. במידה והמודל מובהק, מבצעים מבחן t למובהקות כל אחד מהשיפועים.
3. ביצוע מבחן WALD לבדיקת כל השיפועים שלא יצאו מובהקים:

א. אם מקבלים את H_0: אין סתירה בין מבחן WALD למבחני t - אין בעיה של מולטיקוליניאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.

ב. אם דוחים את H_0: יש סתירה בין מבחן WALD למבחני t - קיימת בעיה של מולטיקוליניאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALD בלעדיו, עד שמזהים את המשתנה / משתנים שיש להוריד מהמודל.

אקונומטריקה למדעי הנתונים

פרק 22 - סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות

תוכן העניינים

סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות:

רקע:

<table>
<tr><th>הבעיה</th><th>הגדרה</th><th>זיהוי</th><th>השלכות</th><th>פתרון</th></tr>
<tr>
<td>הוספת משתנה לא רלוונטי</td>
<td>המודל האמיתי:
$Y_t=\alpha+\beta_1x_{1t}+\varepsilon$
המודל הנאמד (הטעותי):
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$</td>
<td>קבלת H_0 במבחן t למובהקות β_2</td>
<td>ניתן לבצע בדיקת השערות
אר"פ $\left(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\right)$ חסרי הטיה
אומדי השונות $\left(S^2_{\hat{\alpha}},S^2_{\hat{\beta}_1},S^2_{\hat{\beta}_2}\right)$ חסרי הטיה</td>
<td>*הורדת המשתנה</td>
</tr>
<tr>
<td>השמטת משתנה רלוונטי</td>
<td>המודל האמיתי:
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
המודל הנאמד (הטעותי):
$Y_t=\alpha+\beta_1x_{1t}+\varepsilon$</td>
<td>דחיית H_0 במבחן t למובהקות β_2</td>
<td>
<table>
<tr><th>בהיעדר x_2:</th><th>אומד ל-β_1</th><th>אומד ל-α</th><th>אומד לשונות הפרמטרים</th></tr>
<tr><td>$S_{12}=0$</td><td>חסר הטיה</td><td>מוטה
אלא אם:
$\bar{x}_2=0$</td><td>מוטה חיובית</td></tr>
<tr><td>$S_{12}\neq 0$</td><td>מוטה חיובית:
S_{12} ו- β_2 שווי סימן
מוטה שלילית:
S_{12} ו- β_2 מנוגדי סימן</td><td>מוטה</td><td>מוטה חיובית</td></tr>
</table>
לא ניתן לבצע בדיקת השערות
</td>
<td>הוספת המשתנה</td>
</tr>
<tr>
<td>מולטיקוליניאריות מלאה</td>
<td>מתאם מלא בין המשתנים המסבירים במודל
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
כאשר: $r_{12}=\pm 1$</td>
<td>אם:
$x_1=a+bx_2$
אז: $r_{12}=1$</td>
<td>לא ניתן לבצע בדיקת השערות
אר"פ $\left(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\right)$ בלתי מוגדרים.</td>
<td>הורדת אחד המשתנים</td>
</tr>
<tr>
<td>מולטיקוליניאריות חלקית</td>
<td>מתאם חזק בין המשתנים המסבירים במודל
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
כאשר: $0.7<\left|r_{12}\right|<1$</td>
<td>א. סתירה בין מבחן F ל-t
ב. רגישות לספציפיקציה
ג. סימנים הפוכים</td>
<td>ניתן לבצע בדיקת השערות
אין פגיעה בתכונות אר"פ ושונותם</td>
<td>**הורדת אחד המשתנים
או
איחוד</td>
</tr>
</table>

*במידה והמובהקות גבולית ($1<t_{\hat{\beta}}<2$) 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).

שאלות:

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

1. $W_t = \alpha + \beta \cdot S_t + u_t$.

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

2. $W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + v_t$.

כל האומדים חיוביים ומובהקים וקיים קשר שלילי בין גיל להשכלה.

א. $\hat{\beta}_1$ במשוואה (1) הוא:

i. אומד חסר הטיה.

ii. אומד מוטה שלילית.

iii. אומד מוטה חיובית.

iv. 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:

3. $W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + \beta_3 \cdot EXP_t + w_t$.

חווה דעתך על המשוואה השלישית.

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

1. $X_{2t} = \lambda + \delta \cdot X_{1t} + V_t$. כאשר התקבל: $\sum \hat{V}_t^2 = \sum \left(X_{2t} - \bar{X}_{2t}\right)^2$.

2. $$\underset{(10.3)}{Y_t = \alpha + \beta_1 \cdot X_{1t}} + \underset{(19.8)}{\beta_2 \cdot X_{2t}} + U_t$$.

3. $$Y_t = \alpha + \underset{(9.9)}{\beta_1 \cdot X_{1t}} + \underset{(17.3)}{\beta_2 \cdot X_{2t}} + \underset{(0.37)}{\beta_3 \cdot X_{3t}} + W_t$$.

4. $$Y_t = \alpha + \underset{(6.3)}{\beta_1 \cdot X_{1t}} + \sum_t$$.

(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 השונויות 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).

3) נתון המודל: $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$.
חוו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

א. בהנחה כי מתקיים: $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t \qquad R^2 = 0.92$
(0.3) (0.5)
הערכים בסוגריים הם ערכי t.
למובהקות הבטות יש טעות במודל
כי המודל מובהק והמקדמים לא: נכון/לא נכון/לא ניתן לדעת

ב. בהנחה כי מתקיים: $X_{1t} - 2X_{2t} = 1$ לא ניתן לאמוד
את המודל בשיטת הריבועים הפחותים: 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: $r_{12} = 0.98$.

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

ii. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

iii. בהנחה שהמודל יצא מובהק אולם הבטות אינן מובהקות וערכי t למובהקות הבטות הן כדלקמן: $t_{\hat{\beta}_2} = 1.45$, $t_{\hat{\beta}_1} = 1.31$, מה יהיה הפתרון הטוב ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף ii)?
1. להוריד את x_1.
2. להוריד את x_2.
3. להוריד את שני המשתנים.
4. להותיר את שני המשתנים.

תשובות סופיות:

1) א. ii. ב. לא נכון. ג. קיימת בעיית מולטיקוליניאריות מלאה.
2) א. לא נכון. ב. לא נכון. ג. נכון. ד. נכון. ה. נכון.
ו. לא נכון. ז. נכון. ח. לא נכון. ט. נכון.
3) א. לא נכון. ב. נכון. ג. לא נכון. ד. הוכחה. ה.i. לא נכון.
ii. מולטיקוליניאריות חלקית. iii. 4.

אקונומטריקה למדעי הנתונים

פרק 23 - משתנה דמי

תוכן העניינים

משתנה דמי:

רקע:

הכנסת משתנים ב"ת איכותיים למודל הרגרסיה.
למשל, נתונה משוואת הרגרסיה: $W_t = \alpha + \beta \cdot S_t$.
W_t = השכר (התלוי).
S_t = שנות לימוד (הב"ת) שניהם כמותיים.
נניח שאנו סבורים שגם משתנה המגדר (משתנה איכותי) משפיע על השכר.
כדי להכניסו למשוואת הרגרסיה יש להגדיר משתני דמי (dummy variable):
נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר".
ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

1. משתנה דמי לחותך – המגדר משפיע על השכר ההתחלתי בלבד.
2. משתנה דמי לשיפוע – המגדר משפיע על התוספת לשכר בגין שנות הלימוד.
3. משתנה דמי לכל הפונקציה – המגדר משפיע גם על החותך וגם על השיפוע.

משתנה דמי לחותך:

המין משפיע על השכר ההתחלתי בלבד.
המודל: $W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t$ החותך מייצג כאן את השכר ההתחלתי.
שכר ההתחלתי של אישה: α_0.
שכר התחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר בין נשים וגברים: α_1 (הפרש בין החותכים).
בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש החותכים: $H_0: \alpha_1 = 0$.

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

פונקציית 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 זהה למבחן t להבדל בין ממוצעים).

משתנה דמי לשיפוע:

המגדר משפיע על התוספת לשכר בגין שנות הלימוד: $W_t = \alpha + \beta_0 S_t + \beta_1 DS_t + u_t$.
השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.
אצל אישה: התוספת לשכר בגין שנות לימוד: β_0.
אצל גבר: התוספת לשכר בגין שנות לימוד: $\beta_0 + \beta_1$.
הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: β_1 (הפרש השיפועים).
בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש השיפועים: $H_0 : \beta_1 = 0$.

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

משתנה דמי לכל הפונקציה:

המין משפיע גם על החותך וגם על השיפוע – גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.
המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 DS_t + u_t$.
השכר ההתחלתי של אישה: α_0.
השכר ההתחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר ההתחלתי בין המינים: α_1 (הבדל בחותכים).
אצל אישה: התוספת לשכר בגין שנות הלימוד: β_0.
אצל גבר: התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$.
הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: β_1 (הבדל בשיפועים).

2 דרכים לבדיקה האם יש השפעה למשתנה האיכותי:

1. בדיקת השערות למשתני הדמי:
   באמצעות מבחן WALD יש לבדוק: $H_0 : \alpha_1 = \beta_1 = 0$.
   לפחות אחד הפרמטרים שונה מ-0: H_1.
   אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים בנפרד: $H_0 : \alpha_1 = 0$ ו- $H_0 : \beta_1 = 0$.

2. 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:
   נשים: $W_t = \alpha_f + \beta_f X_t + u_t$.
   גברים: $W_t = \alpha_m + \beta_m X_t + u_t$.
   השערות: $H_0 : \alpha_f = \alpha_m ; \beta_f = \beta_m$.

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALD):

המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכלול את המדגם המאוחד: $W_t = \alpha + \beta X_t + u_t$.

המודל הלא מוגבל (U) כולל את שני חלקי המדגם: $ESS_U = ESS_f + ESS_m$, $DF_U = DF_f + DF_m$.

סטטיסטי המבחן: $WALD_{stat} = \dfrac{\dfrac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\dfrac{ESS_f + ESS_m}{DF_f + DF_m}} = CHOW_{stat}$.

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. אם דחינו את H_0 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.

סיכום ביניים:

	משתנה דמי לחותך	משתנה דמי לשיפוע	משתנה דמי לכל הפונקציה
המודל	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$
ההשערה במילים	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).	קיים הבדל בין הקטגוריות בתוספת ל- Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).
בדיקת ההשערה	מבחן t להפרש החותכים: $H_0: \alpha_1 = 0$	מבחן t להפרש השיפועים: $H_0: \beta_1 = 0$	מבחן WALD להפרש בין הפונקציות (החותכים והשיפועים): $H_0: \alpha_1 = \beta_1 = 0$ **ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW. אם דוחים את H_0 יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALD): $H_0: \alpha_1 = 0$, $H_0: \beta_1 = 0$

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים:

כאשר המשתנה האיכותי כולל יותר משני ערכים/קטגוריות נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

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

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

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

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

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס.
נניח שאנו רוצים לבדוק עונתיות במחירי הירקות :

$V_t=$ מדד מחירי הירקות.

$p_t=$ מדד המחירים לצרכן.

1. משתני דמי לחותך :
הטענה : יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות.
המודל : $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$.
כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- 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, $\alpha_0+\alpha_1$ - החותך בקטגוריה i.

בדיקת השערות :

השערות : $H_0:\alpha_1=\alpha_2=\alpha_3=0$, $H_1:OTHERWISE$.

המבחן הסטטיסטי – מבחן WALD :

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

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

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

אם נדחה את H_0 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:
$H_0: \alpha_1 = 0$.

2. האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:
$H_0: \alpha_2 = 0$.

3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:
$H_0: \alpha_3 = 0$.

2. משתני דמי לשיפוע:

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

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

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: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, $H_1: OTHERWISE$.

המבחן הסטטיסטי – מבחן WALD:

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

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

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

אם נדחה את 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 בפונקציית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.
המודל:

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

בדיקת השערות:
השערות: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$.
המבחן הסטטיסטי - מבחן WALD:
(U):

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

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

אם דוחים את H_0, יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים
או בין השיפועים: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.
אם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחני t:
$H_0 : \alpha_j = 0$, $H_0 : \beta_j = 0$.

משתני דמי עבור שני משתנים איכותיים:

לדוגמא – 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 $\left(S_t\right)$.

1. הבדל בחותך ללא אינטראקציה:
המודל: $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$.
במודל זה – אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.
ניתן לבדוק השערות על כל אחד מהמשתנים הב"ת האיכותיים בנפרד:

1. הבדל בשכר ההתחלתי בין גברים לנשים: $H_0 : \alpha_1 = 0$.

2. הבדל בשכר ההתחלתי בין שחורים ללבנים: $H_0 : \alpha_2 = 0$.

2. הבדל בחותך עם אינטראקציה:
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.
במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

3. $H_0: \alpha_3 = 0$.

3. 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.
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:

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

ההשערות לבדיקת קיום האינטראקציה: $H_0: \gamma_1 - \gamma_3 = \gamma_2$ או $H_0: \gamma_1 - \gamma_2 = \gamma_3$
התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת: $\begin{matrix} WALD = t^2 \\ PF = Pt \end{matrix}$.

שאלות:

משתנה דמי לחותך:

1) 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 באוכלוסייה?

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.

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

2) על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים.

תוצאות האמידה: $W_t = 5200 + 1120 \cdot D$.

נתון: $S_{\hat{\alpha}_1} = 63$.

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

משתנה דמי לשיפוע:

3) על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים לנשים בתוספת לשכר בגין שנות הלימוד.

תוצאות האמידה נתונות להלן:

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

(68) (23) (25)

בדוק את ההשערה.

משתנה דמי לכל פונקציה:

4) 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:

1. $NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t}$ כבישים מהירים בלבד.
2. $NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t}$ כבישים לא מהירים בלבד.
3. $NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t}$ שני סוגי הכביש (כל המדגם).
4. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t$

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 מופיעות בהמשך השאלה.

א. בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.
ב. חשבו את הערכים המספריים עבור אומדני משוואה (4).
ג. 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.

ד. מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?
ה. מהי הרגרסיה "תחת H_0" למבחן WALD?

משוואה (1) - כבישים מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	344
Number of Observations Used	344

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4700.81174	4700.81174	89.12	<.0001
Error	342	18039	52.74684		
Corrected Total	343	22740			

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

Parameter Estimates

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

משוואה (2) - כבישים לא מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	410
Number of Observations Used	410

Analysis of Variance

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

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

Parameter Estimates

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

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

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

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

Root MSE	5.27990	R-Square	0.2775
Dependent Mean	3.08355	Adj R-Sq	0.2765
Coeff Var	171.22758		

Parameter Estimates

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

משוואה (4):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

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

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

Parameter Estimates

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

משתנה איכותי עם יותר משתי קטגוריות:

5) ענה על הסעיפים הבאים:

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

i. מהי השערת האפס לבדיקת הטענה?

ii. 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.

i. מהי השערת האפס לבדיקת הטענה?

ii. מהו המבחן הסטטיסטי המתאים? פרטו.

משתנה דמי עבור שני משתנים איכותיים:

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

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

$\ln\left(Y\right)$ - לוג השכר.

EXP - שנות ניסיון.

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

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

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

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	---------	------	------	------
Error	300	140	------		
Corrected Total	305	210			

Root MSE	-------	R-Square	-------
Dependent Mean	-------	Adj R-Sq	-------
Coeff Var	-------		

Parameter Estimates

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

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. מהי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?

ה. לבדיקת ההשערה של הסעיף הקודם בוצע מבחן WALD.
הרגרסיה המוגבלת תחת השערת האפס הינה:
$Z_0=\gamma_0+\gamma_1 Z_1+\gamma_2 Z_2+\gamma_3 Z_3+\gamma_4 Z_4+v$.
מהם ה-Z-ים?

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

ז. החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה:
$$\ln(Y)_t=\lambda_0+\lambda_1 S+\lambda_2 E+\lambda_3(S\cdot E)+\delta_1 EXP+\delta_2 EXP^2+\omega_t$$
כאשר:
S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים).
E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).
מה הקשר בין המקדמים של שני המודלים?

ח. אם יאמוד החוקר את המשוואה:
$\ln(Y)_t=\lambda_0+\lambda_1 S+\lambda_2 E+\delta_1 EXP+\delta_2 EXP^2+\omega_t$, האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

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

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

כאשר:

S משתנה דמי: 1= עבור נשים, 0=גברים.

E משתנה דמי: 1= עבור השכלה גבוהה $(scl>12)$, 0=השכלה נמוכה.

א. רשמו את הפונקציה לחישוב:

i. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.

ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.

iii. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

i. אין השפעה של מגדר והשכלה על השכר.

ii. השפעת ההשכלה אינה תלויה במגדר.

iii. אין השפעות השכלה אצל גברים.

iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

תשובות סופיות:

1) א. $W_t = 7971$. ב. 1,043 ₪. ג. כן. ד. יש עדות לכך.
ה. יש עדות לכך.

2) יש עדות לכך.

3) יש עדות לכך.

4) א. יש עדות לכך, מבחן CHOW : 1 , 2 ו-3, משתנה דמי: 3 ו-4.
ב. $\hat{\alpha} = 0.14978$, $\hat{\beta}_1 = 1.40311$, $\hat{\beta}_2 = 0.002877$, $\hat{\beta}_3 = -0.008$.
ג. $NUM_t = 1.532398$. ד. $H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2$, $H_0 : \beta_3 = \beta_2$.
ה. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot \left(AVGD_t + AVGD \cdot TYPE\right)_t + U_t$.

5) א.i. $H_0 : \alpha_1 = \alpha_2$. ii. WALD ו-t. ב.i. $H_0 : \alpha_1 = \alpha_2 , \alpha_3 = 0$.
ii. WALD.

6) א. נכון. ב. יש עדות לכך. ג. יש עדות לכך.
ד. $H_0 : \alpha_2 = \alpha_1 - \alpha_3$ או $H_0 : \alpha_3 = \alpha_1 - \alpha_2$.
ה. $Z_0 = \ln\left(Y\right)_t$, $Z_1 = D_1 + D_3$, $Z_2 = D_2 - D_3$, $Z_3 = EXP_t$, $Z_4 = EXP_t^2$
ו. אין עדות לכך. ז. $\lambda_0 = \alpha_0$, $\lambda_1 = \alpha_2$, $\lambda_2 = \alpha_3$, $\lambda_3 = \alpha_1 - \alpha_2 - \alpha_3$.
ח. לא.

7) א.i. $\hat{\ln}\left(MWAGE\right) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$.
ii. $\hat{\ln}\left(MWAGE\right) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$. iii. $EXP_t = \dfrac{-\left(\alpha_1 + \alpha_3\right)}{\beta_1 + \beta_3}$.
ב.i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$. ii. $H_0 : \alpha_3 = \beta_3 = 0$.
iii. $H_0 : \alpha_2 = \beta_2 = 0$. iv. $H_0 : \beta_2 + \beta_3 = 0$.

אקונומטריקה למדעי הנתונים

פרק 24 - תיאוריה הפרת ההנחות קלאסיות

תוכן העניינים

הפרת ההנחות הקלאסיות:

רקע:

ארבעת הנושאים הבאים עוסקים במצב של הפרת אחת ההנחות הקלאסיות הדרושות לאמידת הפרמטרים בשיטת OLS:

- הטרוסקדסטיות (הפרת הנחה מס' 5) – שונות קבועה ויחידה לאורך קו הרגרסיה: $V(u_t)=\sigma^2$.
- מתאם סידרתי (הפרת הנחה מס' 6) – אי תלות בין הטעויות: $\text{cov}(u_t,u_s)=0$.
- מודלים דינמיים ו-משוואות סימולטניות (הפרת הנחה מס' 4) – אי תלות בין המשתנים הב"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 ההפרה.
- כיצד נתקן את משוואת הרגרסיה כך שניתן יהיה לאמוד את הפרמטרים בשיטת OLS.

אקונומטריקה למדעי הנתונים

פרק 25 - הטרוסקדסטיות

תוכן העניינים

הטרוסקדסטיות:

רקע:

הטרוסקדסטיות הוא מצב שבו מופרת הנחת ההומוסקדסטיות, הגורסת כי שונות הטעויות 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 שונות הטעויות של כל תצפית היא שונה: $V(u_t)=\sigma_t^2$.

ההשלכות של הטרוסקדסטיות על אומדי OLS:

בהינתן הטרוסקדסטיות מופרת תכונת היעילות של אומדי הריבועים הפחותים שכן בכדי לחשב שונות יעילה של האומדים השתמשנו בהנחה של שונות קבועה.

מבחנים לזיהוי הטרוסקדסטיות:

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות – באיזה אופן השונות של הטעויות משתנה בין תצפית לתצפית.
שיטות לזיהוי הטרוסקדסטיות: מבחן GQ (Goldfeld-Qauandt) ומבחן White.
מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונויות 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 שונויות 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: $F_{stat}=\dfrac{ESS_1/T_1-K-1}{ESS_2/T_2-K-1}$ (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: $\begin{matrix} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 > \sigma_2^2 \end{matrix}$.

2. מבחן White:
ההנחה העומדת בבסיס מבחן זה כי לכל תצפית ותצפית שונות שונה של טעויות. הביטוי המתמטי של הנחה זו היא היותה של השונות פונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:

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

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

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 בתוכנה הסטטיסטית RES-SAS).
- 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 = מסי המשתנים ברגרסיית העזר.

השערות: $\begin{matrix} H_0 : \alpha_j = \beta_j = \gamma_{jj} = 0 \\ H_1 : OTHERWISE \end{matrix}$.

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: $W_t = \dfrac{1}{\sqrt{Z_t}}$.

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

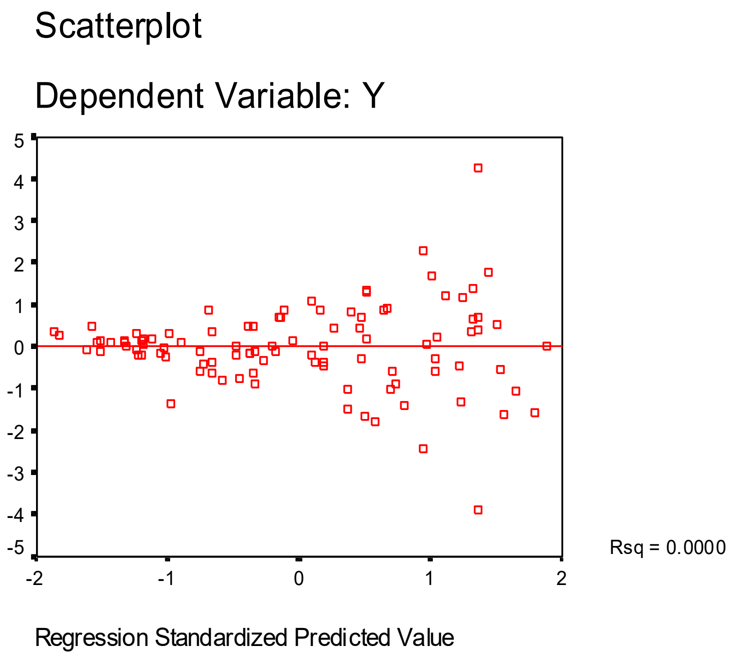
המשוואה המקורית: $Y_t = \alpha + \beta X_t + u_t$
$Y_t W_t = \alpha \cdot W_t + \beta\left(X_t W_t\right) + u_t W_t$.

בצורתה המפורשת המשוואה החדשה נראית כך: $\frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$.

שאלות:

מבחן GQ:

1) נאמד הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$.
גרף השאריות של הרגרסיה הנ"ל נתון להלן:



מגרף זה אנו למדים כי השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה – זהו מצב של הטרוסקדסטיות.
בכדי לבצע מבחן GQ:

- התצפיות של משתנה ההכנסה 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

                              Sum of             Mean
Source           DF          Squares           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

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

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

2) על מנת לבחון את פונקציית הייצור בענף מסוים נאספו נתונים על 150 פירמות.
נסמן:
Q - תפוקה שנתית באלפי שקלים.
L - מספר עובדים.
המודל הנאמד: $\ln(Q)=\alpha+\beta\cdot\ln(L)$.
החוקר חשש שההפרעה המקרית איננה הומוסקדסטית. לשם כך הוא מיין את התצפיות בסדר עולה של מספר העובדים, השמיט $\frac{1}{3}$ מהתצפיות האמצעיות והריץ שתי רגרסיות נפרדות עם מספר שווה של תצפיות:
ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא

קיבל: $R^2 = 0.403$, $ESS = 279.3$.

ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא

קיבל: $R^2 = 0.238$, $ESS = 493.8$.

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

מבחן WHITE:

3) על אותו הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$.

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

בכדי לבצע את מבחן WHITE:

- נחשב את השאריות של הרגרסיה: $\hat{u}_t = Y_t - \hat{Y}_t$.
- נעלה את השאריות בריבוע: $\hat{u}_t^2$.
- נאמוד את המשוואה: $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + v_t$.

תוצאות האמידה מוצגות להלן:

The REG Procedure

Model: MODEL1

Dependent Variable: RES^2

Number of Observations Read	48
Number of Observations Used	48

Analysis of Variance

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

Root MSE		R-Square	0.390763
Dependent Mean		Adj R-Sq	

בצעו מבחן White (השערות, סטטיסטי המבחן, כלל הכרעה ומסקנה).

4) חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות והאפשרות של מכירת עיתונים.

נסמן:

Y_SALES - מכירות חודשיות (₪).

$X1\_SQUARES$ - 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

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

                  Parameter Estimates
                        Parameter      Standard
Variable         DF      Estimate       Error     t Value     Pr > |t|
Intercept
X1_SQUARE
X1_SQUARE^2
X1_SQUARE*X2_RENT
X2_RENT
X2_RENT^2
X2_RENT*PAPERS
PAPERS
```

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: ______________________________.

שיטת WLS:

5) נתון המודל:

1. $Y_t = \alpha + \beta \cdot X_t + U_t$.

ונתון כי: $VAR(U_t) = \frac{\sigma^2}{Z_t^2}$ (Z_t 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:

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

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:

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

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

6) ענו על השאלה הקודמת, כאשר נתון כי: $VAR(U_t) = \sigma^2 \cdot X_t^2$.

7) 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. פתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה: $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t$.

כאשר: $W_t =$ ______.

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

האם ישתנו תשובותיכם לסעיפים א' ו-ב': כן/לא/לא ניתן לדעת

תשובות סופיות:

1) יש עדות לכך, השערות: $\begin{matrix} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{matrix}$, חישוב סטטיסטי: 17.62, כלל הכרעה: 2.48.

2) יש עדות לכך, השערות: $\begin{matrix} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{matrix}$, חישוב סטטיסטי: 1.77, כלל הכרעה: 1.69.

3) יש עדות לכך, השערות: $\begin{matrix} H_0: \alpha_1 = \beta_1 = 0 \\ H_1: OTHERWISE \end{matrix}$, חישוב סטטיסטי: 18.75,
כלל הכרעה: 5.991.

4) קיום הטרוסקדסטיות בנתונים.
$(LM)WHITE$.
RES^2.
$X1\_SQUARE$, $X1\_SQUARE^2$, $X1\_SQUARE * X2\_RENT$,
$X2\_RENT$, $X2\_RENT^2$, $X2\_RENT * PAPERS$, $PAPERS$.
$LM_{stat} = 1.73$.
אין עדות לכך.

5) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t = \frac{1}{\sqrt{\frac{1}{Z_t^2}}} = Z_t$.
ד. $\sigma^2 = \frac{ESS}{T-K}$. ה. לא, המודל השני. ו. i לא נכון. ii לא נכון.

6) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t = \frac{1}{X_t}$.
ד. $S^2 = \frac{ESS}{T-k-1}$. ה. ראו סרטון. ו. ראו סרטון.

7) א. הטרוסקדסטיות. ב. נכון.
ג. $t \geq 50$ $W_t = \frac{1}{\sqrt{X_t}}$, $t \leq 50$ $W_t = \frac{1}{\sqrt{X_t^2}}$. ד. לא.

אקונומטריקה למדעי הנתונים

פרק 26 - מתאם סדרתי

תוכן העניינים

מתאם סדרתי:

רקע:

מתאם סדרתי עוסק במצב שבו מופרת ההנחה (מס' 6) של אי תלות בין הטעויות: $\text{cov}(u_t,u_s)=0$ ונוצרת 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. בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.

השלכות על אומדי הריבועים הפחותים (OLS):

מבין התכונות של אר״פ (ליניאריות, חוסר הטיה, 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: R^2, F ו-t מוטים כלפי מעלה.

מבנה המתאם הסדרתי:

מתאם סדרתי מסדר ראשון:
ההנחה היא כי יש מתאם בין הטעויות במרחק אחד, כלומר u_t תלוי ישירות רק ב-u_{t-1}: $\text{cov}(u_t,u_{t-1})\neq 0$.
את המתאם בין הטעויות מסדר ראשון ניתן לנסח באופן הבא: $u_t=\rho\cdot u_{t-1}+\varepsilon_t$

כך ש:

1. $\rho\neq 0$ (כי אם $\rho=0$ אין מתאם סידרתי).
2. $1<\rho<-1$ (כי אם חורג מ-1 הטעות הולכת וגדלה עם הזמן).
3. ρ חיובי פירושו מתאם סדרתי חיובי ואילו ρ שלילי פירושו מתאם סדרתי שלילי (לא נפוץ).

4. ε_t מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין (בניגוד ל- u_t) כך ש:

$$E(\varepsilon_t)=0$$
$$V(\varepsilon_t)=\sigma_\varepsilon^2 .$$
$$\operatorname{cov}(\varepsilon_t,\varepsilon_{t-s})=0$$

המודל יכיל 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:

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

u_t מושפע מתקופות שונות בעבר.

תכונות המתאם הסדרתי:

מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן: $r_{(u_t,u_{t-s})}=\rho^s$

המתאם של u_t הולך ופוחת עם הזמן: $\rho_{u_tu_{t-1}}>\rho^2_{u_tu_{t-2}}>\rho^3_{u_tu_{t-3}}>\ldots>\rho^s_{u_tu_{t-s}}$.

בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t)=0$$
$$V(u_t)=\sigma_u^2=\frac{\sigma_\varepsilon^2}{1-\rho^2}$$
$$COV(u_t,u_{t-s})=\rho^s\sigma_{u_t}^2=\rho^s\frac{\sigma_s^2}{1-\rho^2}$$

מבחנים לזיהוי מתאם סדרתי:

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 = מס׳ המשתנים הב״ת במודל ו-T = מס׳ התצפיות במדגם נשלוף שני ערכים: d_U ו- d_L.
- נחלק את הטווח שבין 0 ל-4 באופן הבא:

$$0 - \rho > 0 - d_L - d_U - \rho = 0 - 4 - d_U - 4 - d_L - \rho < 0 - 4$$

- 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 באדום.

השערות: $\begin{array}{l} H_0 : \rho = 0 \\ H_1 : \rho > 0, \rho < 0 \end{array}$.

חישוב הסטטיסטי: $DW_{stat} \cong 2 \cdot (1 - \hat{\rho})$.

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

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

מבחן LM:

לעומת מבחן DW מבחן LM מתאים גם לבחינת קיומו של מתאם סדרתי מסדרים גבוהים יותר מסדר ראשון: $\begin{matrix} Y_t = \alpha + \beta \cdot X_t + u_t \\ u_t = \rho \cdot u_{t-1} + \varepsilon_t \end{matrix}$.

המודל הלא מוגבל:

(U): $Y_t = \alpha + \beta \cdot X_t + \rho \cdot u_{t-1} + \varepsilon_t$.

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: $\hat{u}_t / x_1, x_2, ..., x_k; \hat{u}_{t-1}$.
- נחשב סטטיסטי LM: $LM_{stat} = T \cdot R^2$.
- נדחה את H_0 כאשר: $LM_{stat} > \chi^2_m$ כאשר m = סדר המתאם הסדרתי.

אם נדחה את H_0 נדע את סימנו של המתאם הסדרתי לפי המקדם של $\hat{u}_{t-1}$ ברגרסיית העזר ששווה ל- $\hat{\rho}$.

שימו לב כי אם נרצה לבדוק מתאם סדרתי מסדרים גבוהים יותר:

ההשערות: $\begin{matrix} H_0: \rho_1 = \rho_2 = ... = \rho_s = 0 \\ H_1: OTHERWISE \end{matrix}$.

רגרסיית העזר: $\hat{u}_t / x_1, x_2, ..., x_k; \hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-s}$.

פתרון 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.

משוואה (1): המודל בזמן t: $Y_t = \alpha + \beta X_t + u_t$.

משוואה (2): המודל בזמן $t-1$ מוכפל ב- ρ: $\rho \cdot Y_{t-1} = \rho \cdot \alpha + \rho \cdot X_{t-1} + \rho \cdot u_t$.

החסרת משוואה (2) ממשוואה (1):

$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$.

כדי לאמוד את הפרמטרים של רגרסיית ההפרשים נגדיר:

$$Y_t^* = Y_t - \rho Y_{t-1}$$
$$\alpha^* = \alpha(1-\rho)$$
$$\beta^* = \beta$$
$$X_t^* = X_t - \rho X_{t-1}$$
$$\varepsilon_t = u_t - \rho u_{t-1}$$

כך "נטרלנו" את המתאם הסדרתי: $u_t = \rho \cdot u_{t-1} + \varepsilon_t \Leftarrow \varepsilon_t = u_t - \rho \cdot u_{t-1}$.

ε_t מקיים את כל ההנחות הקלאסיות ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.
המשוואה "המתוקנת" אותה נאמוד: $Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$.
לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים: $\hat{\beta}$, $\hat{\alpha}$.
מאחר ש- ρ איננו ידוע יש צורך לאמוד אותו.

אמידת ρ בשיטת קוקרן אורקוט:
שיטת קוקרן אורקוט לאמידת ρ היא שיטה איטראטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות.
התהליך הממוחשב נקרא אוטורגרסיה (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 אוטורגרסיה מסדר שני) וכו'. אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

שאלות:

תכונות המתאם הסדרתי:

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

נתון כי: $\rho = 0.9$ וכי: $V(\varepsilon_t) = \sigma_\varepsilon^2 = 1$.

מצאו את:

א. המתאם בין u_t ל- u_{t-1}.

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

ג. השונות σ_u^2.

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

מבחן DW:

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

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

כאשר:

$CLOSE_t$ = מחיר סגירה של מניה ב-$ ביום t.

$TIME_t$ = משתנה זמן שמקבל את הערכים: $1,2,3\ldots$.

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

Analysis of Variance

F Value	Prob>F	Mean Square	Sum of Squares	DF	Source
	55.78		-------	1	Model
			-------	151	Error
			-------	152	C Total
		0.181	R-square	---- Root MSE	
		------	Adj R-sq	---- Dep Mean	
				---- C.V.	

Parameter Estimates

T for H0: Parameter=0	Prob$>\vert T\vert$	Standard Error	Parameter Estimate	DF	Variable
91.047	0.0000	0.0148	1.3474	1	INTERCEP
0.0000	-7.468	0.0001	-0.00075	1	TIME

Durbin-Watson D 0.150

האם קיים מתאם סדרתי?

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 $d_L \le d_\alpha \le d_U$, 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

3) המודל הינו: $\begin{matrix} Y_t = \alpha + \beta X_t + \varepsilon_t \\ u_t = \rho u_{t-1} + \varepsilon_t \end{matrix}$.

u_t - הפרעה מקרית קלאסית.

וידוע כי: $T = 100$

$u_t = 0.9 u_{t-1}$

$d_L = 1.57$

$d_U = 1.65$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

מבחן LM:

4) עבור הדוגמא הקודמת – ניבוי מחיר סגירה של מניה כפונקציה של הזמן: $CLOSE_t = \alpha + \beta \cdot TIME + u_t$.

נבחן את קיומו של מתאם סדרתי מסדר ראשון באמצעות מבחן LM.

נאמוד את רגרסיית העזר: $u_t = \gamma_0 + \gamma_1 \cdot TIME_t + \rho \cdot u_{t-1} + \varepsilon_t$.

תוצאות האמידה שהתקבלו:

Dependent variable: RES

Analysis of Variance

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

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

Parameter Estimates

Parameter Estimate	DF	Variable
-.000096	1	INTERCEP
1.16331E-05	1	TIME
0.927172	1	RES1

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסייה?

תיקון המתאם הסדרתי:

5) 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

6) 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			
(1)AR	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?

תרגול מסכם:

7) נאמד הקשר שבין הכנסה לתצרוכת לתקופה ינואר 1994 עד דצמבר 1997 (T=48).
המודל הינו: $C_t=\alpha+\beta Y_t+u_t$.
בניסיון לבדוק האם מתקיים קשר מהסוג הבא: $u_t=\rho_1u_{t-1}+\rho_2u_{t-2}+\rho_3u_{t-3}+\varepsilon_t$.
נאמדה המשוואה הבאה: $\hat{u}_t=\gamma_1\hat{u}_{t-1}+\gamma_2\hat{u}_{t-2}+\gamma_3\hat{u}_{t-3}+\gamma_4Y_t+\omega_t$.
תוצאות האמידה מוצגות להלן:

Depended Variable: RES

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-54.709	710.85	-0.076	0.939
RESID1	1	0.705	0.152	4.631	0.000
RESID2	1	-0.0066	0.188	-0.035	0.972
RESID3	1	-0.337	0.167	-2.012	0.051
Y	1	0.0027	0.032	0.085	0.932

Durbin-Watson D 1.954

נתון בנוסף כי: $R^2 = 0.479$.

א. הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: ____________________.
על ידי מבחן: ______________________.
ההשערות הינן: _________________________.
גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): _______________.
המסקנה המתקבלת היא: __________________________.

בהנחה כי קיים מתאם סדרתי מסדר שלישי בנתונים נאמד מחדש הקשר שבין ההכנסה לתצרוכת בהתאם לשיטתם של קוקרן ואורקוט.
תוצאות האמידה מוצגות להלן:

Depended Variable:C

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: 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.

8) 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: $\ln\hat{Y}_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$, $\hat{\rho} = 0.2$.
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?

תשובות סופיות:

1) א. $r_{(u_t,u_{t-1})}=0.9$. ב. $r_{(u_t,u_{t-4})}=0.6561$. ג. $\sigma_u^2=5.263$.

ד. $\sigma_u^2=1.19$, $r_{(u_t,u_{t-4})}=0.0256$, $r_{(u_t,u_{t-1})}=0.4$.

2) יש עדות לכך.

3) יש עדות לכך.

4) א. יש עדות לכך. ב. $\hat{\rho}=0.927$. ג. חיובי.

5) לא נכונה.

6) ראו סרטון.

7) א. קיומו של מתאם סדרתי מסדר שלישי בנתונים.

LM.

$$\begin{aligned}&H_0:\rho_1=\rho_2=\rho_3=0\\&H_1:OTHERWISE\end{aligned}.$$

$LM_{stat}=22.99$.

יש עדות לכך.

ב. $\hat{C}_t=-2103.53+0.719\cdot Y_t+0.707\cdot\hat{u}_{t-1}-0.0064\cdot\hat{u}_{t-2}-0.328\cdot\hat{u}_{t-3}$.

8) א. $\begin{aligned}&H_0:\rho=0\\&H_1:\rho\neq 0\end{aligned}$, יש עדות לכך. ב. $Y_{t+1}=5,568$.

ג. $u_t=\gamma_0+\gamma_1\ln P_t+\gamma_2\ln P_{t-1}+\gamma_3 u_{t-1}+\gamma_4 u_{t-2}+\omega_t$. ד. אין עדות לכך.

אקונומטריקה למדעי הנתונים

פרק 27 - סיכום מתאם סדרתי והטרוקדסטיות

תוכן העניינים

סיכום מתאם סדרתי והטרוקדסטיות:

רקע:

	מתאם סדרתי	הטרוסקדסטיות
המשוואה העיקרית של המודל	למשל: $Y_t = \alpha + \beta X_t + u_t$	
ההנחה הקלאסית המופרת	$\text{cov}(u_t, u_{t-s}) = 0$	$V(u_t) = \sigma^2$
המצב לאחר ההפרה	$\text{cov}(u_t, u_{t-s}) \neq 0$	$V(u_t) = \sigma_t^2$
המשוואה המאפיינת את ההפרה	$u_t = \rho u_{t-1} + \varepsilon_t$	$V(u_t) = W_t \sigma^2$
מה קורה אם אומדים ב-OLS	מתקבלים אומדים חסרי הטיה ועקיבים, אך תכונת היעילות נפגעת.	
זיהוי הבעיה	מבחן DW מבחן LM	מבחן GQ מבחן White
פתרון הבעיה	שיטת קוקרן – אורקוט (רגרסיית ההפרשים) הכנסת משתנה מוסבר בפיגור (מודל דינמי)	שיטת WLS

אקונומטריקה למדעי הנתונים

פרק 28 - מודלים דינאמיים

תוכן העניינים

מודלים דינאמיים:

רקע:

מודל דינמי הוא מודל שיש בו משתנה מוסבר בפיגור, כלומר Y היום מושפע מ-Y של אתמול: $Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + u_t$.

המכפילים הדינמיים:

שלוש סוגים של השפעות בהקשר של המודל הדינמי (מכפלים):

1. מכפל לטווח קצר (מיידי):
   איך X היום משפיע על Y היום: $\frac{\partial Y_t}{\partial X_t}$.
2. מכפל ביניים מסדר j (מכפיל דינמי):
   איך X מלפני j תקופות משפיע על Y היום: $\frac{\partial Y_t}{\partial X_{t-j}}$.
3. מכפל טווח ארוך (מצב עמיד):
   איך X משפיע על Y לאורך P תקופות: $\frac{\partial Y^*}{\partial X^*}$.
   כאשר X ו-Y נותרים קבועים על פני הזמן (מצב עמיד):
   $$\begin{matrix} Y_t = Y_{t-1} = \ldots = Y_{t-p} = Y^* \\ X_t = X_{t-1} = \ldots = X_{t-p} = X^* \end{matrix}.$$

הקשר בין מתאם סדרתי למודלים דינמיים

המתאם הסדרתי נובע מהשמטה של דינמיות מבנית במודל. המודל המקורי היה צריך להיות מודל דינמי אך נאמד בטעות מודל סטאטי. הדינמיות תבוא אז לידי ביטוי בטעויות, כלומר במתאם הסדרתי.
רגרסיית ההפרשים, המהווה פתרון למתאם הסדרתי, היא למעשה מודל דינמי.

לסיכום:
בכדי לפתור את בעיית המתאם הסדרתי יש לאמוד מלבד את המשתנה המוסבר בזמן t גם את המשתנה המוסבר והמסביר בזמן $t-1$.
המשתנה בפיגור Y_{t-1} נועד לפתור מתאם סדרתי מסדר ראשון, Y_{t-2} משמש לפתירת מתאם סדרתי מסדר שני וכך הלאה.
בכדי לבדוק קיומו של מתאם סדרתי במודל דינמי לא נוכל לבצע מבחן DW אלא רק מבחן LM.

השלכות על אר"פ של משתנה מוסבר בפיגור כמשתנה מסביר:

בניגוד למשתנה מסביר רגיל (X), Y_{t-1} הינו משתנה מקרי.
משום כך אר"פ ברגרסיה הכוללת משתנים כאלה הם מוטים (להזכירכם בהוכחת חוסר הטיה של האומדים השתמשנו בהנחה מס' 4 הגורסת כי המשתנים המסבירים אינם משתנים מקריים).
בנוסף לכך העקיבות של האומדים תלויה בקיום מתאם סדרתי:

$$\hat{\beta} \to \beta + \frac{COV(Y_{t-1}, u_t)}{V(Y_{t-1})}.$$

אם אין מתאם סדרתי: $COV(Y_{t-1}, u_t) = 0$ $\Leftarrow$ האומד עקיב.

אם יש מתאם סדרתי: $COV(Y_{t-1}, u_t) \neq 0$ $\Leftarrow$ האומד איננו עקיב.

לסיכום – ההשלכות על אר"פ:

1. האומדים מוטים ולכן ניתן לבצע בדיקת השערות רק במדגמים גדולים (T>30).
2. אם אין מתאם סדרתי $\Leftarrow$ האומדים עקיבים ויעילים (ניתן לבצע בדיקת השערות במדגמים גדולים).
   אם יש מתאם סדרתי $\Leftarrow$ האומדים אינם עקיבים ואינם יעילים (לא ניתן לבצע בדיקת השערות גם במדגמים גדולים).

שאלות:

חישוב מכפלים:

1) חשבו את שלושת סוגי המכפלים של המודלים הדינמיים הבאים:

א. $Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + u_t$.

ב. $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + u_t$.

תרגול מסכם:

2) המודל הבא הורץ ב-SAS עם מדגם בעל 100 תצפיות: $Y_t = \alpha + \beta X_t + u_t$.

א. מהפלט עולה DW=0.195 לפיכך:

i. לא קיים מתאם סדרתי.

ii. קיים מתאם סדרתי והוא: ______________.

iii. לא ניתן לקבוע אם המתאם הסדרתי מובהק.

ב. לפי תשובתך לסעיף א' חווה דעתך על תכונות האומדים:

i. מוטים נכון/ לא נכון/ לא ניתן לדעת

ii. ליניאריים נכון/ לא נכון/ לא ניתן לדעת

iii. יעילים נכון/ לא נכון/ לא ניתן לדעת

iv. עקיבים נכון/ לא נכון/ לא ניתן לדעת

ג. אמידה של איזו משוואה תפתור באופן מלא את הבעיה שנוצרה במודל:

i. $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + u_t$.

ii. $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-1} + u_t$.

iii. $Y_t = \alpha + \beta_1 X_t + \beta_2 Y_{t-1} + u_t$.

ד. בדוק את ההשערה כי לפי מודל (3) השפעת X על Y הולכת ופוחתת עם הזמן. מצורף החלק הרלוונטי מהפלט:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	0.42	0.06	7.00	0.000
X	1	0.25	0.03	8.33	0.000
Y1	1	0.85	0.05	17.00	0.000

ה. מהו המכפיל הדינמי בתקופה t-8?

3) הקשר בין כמות הכסף לבין רמת האינפלציה במשק נאמד בסדרה עתית על ידי המשוואה הבאה:

1. $M_t = \alpha + \beta \cdot P_t + U_t$.

כאשר:

M_t - כמות הכסף במשק בחודש t.

P_t - מדד המחירים לצרכן במשק בחודש t.

משוואה (1) נאמדה בפלט מס' 1.

א. לפי מבחן על הסטטיסטי DW , נראה כי ב- U_t:
   i. לא ניתן לחשב את הסטטיסטי DW מהנתונים הקיימים.
   ii. קיים מתאם סדרתי שלילי.
   iii. קיים מתאם סדרתי חיובי.
   iv. לא קיים מתאם סדרתי.
   v. לא ניתן לקבוע אם המתאם הסדרתי מובהק.

ב. סמנו את התשובה הנכונה בהכרח:
   i. האומדים ליניאריים חסרי הטיה, עקיבים אך לא יעילים.
   ii. האומדים ליניאריים חסרי הטיה, עקיבים ויעילים.
   iii. האומדים מוטים אך עקיבים.
   iv. האומדים חסרי הטיה, אך לא עקיבים.
   v. כל התשובות אינן נכונות.

ג. אומד השונות מוטה ובדיקת השערות לא תקפה: נכון/לא נכון/אי אפשר לדעת

חוקר טען כי הבעיה שנוצרה במשוואה (1) תיפתר ע"י אמידת המשוואה הבאה:

2. $M_t = \alpha_1 + \beta_1 \cdot P_t + \beta_2 \cdot M_{t-1} + \varepsilon_t$.

כאשר:

M_{t-1} - כמות הכסף בשנה הקודמת.

משוואה (2) נאמדה בפלט מס' 2, כמו כן נאמדה על ידי החוקר המשוואה המופיעה בפלט מס' 3.

ד. הרגרסיה המופיעה בפלט מס' 3 נועדה לבדיקת: ____________________.
במשוואה: ______________________________.
על ידי מבחן: ______________________________.
גודל הסטטיסטי למבחן הינו (רשום תוצאה מספרית): ________________.

ה. לאור תשובתך לסעיף ד' טענת החוקר: נכונה/לא נכונה/אי אפשר לדעת

ו. האומד ל- β_1 במשוואה (2) הוא מוטה אך עקיב: נכון/לא נכון/אי אפשר לדעת

ז. ניתן להשתמש בסטטיסטי DW לבדיקת מתאם סדרתי במשוואה (2): נכון/לא נכון/אי אפשר לדעת.

ח. חשבו את המכפיל הדינמי לשינוי P בתקופה $t-1$.

ט. בדקו את הטענה כי המכפיל הדינמי לשינוי P בתקופה $t-1$ $\left(\frac{\partial M_t}{\partial P_{t-1}}\right)$ הינו 90% מהמכפיל המיידי בטווח הקצר.

י. רשמו את השערת האפס עבור הטענה כי המכפיל בט״א שווה ל-1 .
מהו המבחן הסטטיסטי המתאים לבחינת ההשערה?

פלט מס׳ 1 – משוואה (1):

pendent Variable: m

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	44.03828	44.03828	5145.80	<.0001
Error	103	0.88148	0.00856		
Corrected Total	104	44.91976			

Root MSE	0.09251	R-Square	0.9804
Dependent Mean	8.53854	Adj R-Sq	0.9802
Coeff Var	1.08344		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.49372	0.09862	15.15	<.0001
p	1	1.69267	0.02360	71.73	<.0001

Durbin-Watson D 0.208

פלט מס׳ 2 – משוואה (2):

pendent Variable: m

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42.19946	21.09973	15988.1	<.0001
Error	101	0.13329	0.00132		
Corrected Total	103	42.33275			

Root MSE	0.03633	R-Square	0.9969
Dependent Mean	8.55393	Adj R-Sq	0.9968
Coeff Var	0.42469		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.40374	0.06790	5.95	<.0001
m1	1	0.81811	0.03857	21.21	<.0001
p	1	0.28127	0.06633	4.24	<.0001

$M1 = M_{t-1}$

פלט מס' 3:

```
Dependent Variable: res Residual
```

$RES = \hat{\varepsilon}_t$ = פלט אמידה משוואה (ב)

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.00032896	0.00010965	0.08	0.9695
Error	99	0.13173	0.00133		
Corrected Total	102	0.13206			

Root MSE	0.03648	R-Square	0.0025
Dependent Mean	0.00033853	Adj R-Sq	-0.0277
Coeff Var	10775		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.03298	0.08192	0.40	0.6881
p		1	0.02640	0.07979	0.33	0.7415
m1		1	-0.01672	0.04705	-0.36	0.7230
res1		1	-0.01304	0.11039	-0.12	0.9062

$RES1 = \hat{\varepsilon}_{t-1}$

תשובות סופיות:

1) א. מכפיל מיידי: $\frac{\partial Y_t}{\partial X_t}=0$, מכפיל טווח ביניים: $\frac{\partial Y_t}{\partial X_{t-j}}=\beta_1\beta_2^{j-1}$, מכפיל של הטווח הארוך: $\frac{\partial Y^*}{\partial X^*}=\beta_1\cdot\frac{1}{1-\beta_2}$.

ב. מכפיל מיידי: $\frac{\partial Y_t}{\partial X_t}=\beta_0$, מכפיל טווח ביניים: $\frac{\partial Y_t}{\partial X_{t-j}}=\beta_2^{j-1}\left(\beta_1+\beta_0\beta_2\right)$,

מכפיל של הטווח הארוך: $\frac{\partial Y^*}{\partial X^*}=\frac{\beta_1+\beta_0\beta_2}{1-\beta_2}$.

2) א. ii, חיובי. ב. i. לא נכון. ii. נכון. iii. לא נכון.
iv. נכון. ג. ii. ד. יש עדות לכך.
ה. $\frac{\partial Y_t}{\partial X_{t-j}}=0.068$.

3) א. iii. ב. i. ג. נכון.
ד. קיום מתאם סדרתי במשוואה (2).
מבחן LM.
$LM_{stat}=0.2575$.

ה. נכונה. ו. נכון. ז. לא נכון. ח. $\frac{\partial M_t}{\partial P_{t-i}}=\beta_1\beta_2^i$.

ט. יש עדות לכך. י. $H_0:\beta_1=1-\beta_2$, WALD.

אקונומטריקה למדעי הנתונים

פרק 29 - משוואות סימולטניות

תוכן העניינים

משוואות סימולטניות:

רקע:

עוסקות בהפרת ההנחה של אי תלות בין המב"ת לטעויות בניבוי: $\operatorname{cov}(x,u)=0$.
ה- X ים במשוואה נחשבו משתנים אקסוגניים – משפיעים על Y אך לא מושפעים ממנו בחזרה לעומת זאת משתנים אנדוגניים – משפיעים על Y אך גם מושפעים ממנו בחזרה. מאחר ומשתנים אלו הם גם מסבירים וגם מוסברים, הם נחשבים כמשתנים מקריים, המתואמים עם הטעויות במודל: $\operatorname{cov}(x,u)\neq 0$.

משוואות המבנה (משוואות סימולטניות):

מערכת משוואות הכוללות משתנים מסבירים אנדוגניים ואקסוגניים.
בד"כ מדובר בשתי משוואות אשר המשתנה המוסבר בראשונה הוא משתנה מסביר בשנייה והמשתנה המוסבר בשנייה הוא משתנה מסביר בראשונה.
משתנים המופיעים באחת המשוואות כמוסברים ובאחרת כמסבירים הם משתנים אנדוגניים. יתר המשתנים במשוואות הם אקסוגניים.
המטרה היא לאמוד בצורה יעילה את הפרמטרים (אלפות ובטות) ולבצע בדיקת השערות.

השלכות על אר"פ:

הנחת אי תלות בין המשתנה הב"ת והטעויות שימשה אותנו להוכחת ליניאריות, חוסר הטיה ועקיבות.
לכן הפרתה משמעה פגיעה בכל תכונות אר"פ.
האומדים לא ליניאריים, מוטים לא עקיבים ולכן גם לא יעילים (לפי גאוס מרקוב).
אומד השונות מוטה גם הוא ובדיקת ההשערות לא תקפה (ללא תלות בגודל המדגם).

הצורה המצומצמת של מודל עם משוואות סימולטניות:

משוואות הצורה המצומצמת הן פתרון עבור המשתנים האנדוגניים במערכת:
הגדרת המשתנים האנדוגניים כפונקציה של המשתנים האקסוגניים במערכת בלבד.
מספר המשוואות המצומצמות הוא כמספר המשתנים האנדוגניים במערכת (במקרה זה שניים).

תכונות המשוואות מהצורה המצומצמת:

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

אמידת הפרמטרים של משוואות המבנה באמצעות משוואות הצורה המצומצמת:
משוואות הצורה המצומצמת מאפשרות לאמוד את הפרמטרים בשיטת OLS אבל אנחנו מעוניינים למעשה לאמוד את הפרמטרים של המשוואות המקוריות – משוואות המבנה. מתוך הפרמטרים של הצורה המצומצמת נחלץ את הפרמטרים של משוואות המבנה.

בתהליך החילוץ של הפרמטרים המבניים ייתכנו 3 מצבים:

1. אין זיהוי: לא ניתן לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת.
2. זיהוי מדויק: יש רק דרך אחת לחלץ את הפרמטרים המבניים מהפרמטרים של הצורה המצומצמת.
3. זיהוי יתר: יש יותר מדרך אחת לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת.

בכדי להקל על בעיית הזיהוי מומלץ לאמץ את הכלל הבא:
עבור כל אחת מהמשוואות המבניות יש לחשב:

1. $g-1$: מס' אנדוגניים במשוואה הספציפית פחות 1
ולהשוות עם:
2. $K-k$: מספר אקסוגניים סה"כ בשתי המשוואות כולל חותך (K) פחות מספר אקסוגניים במשוואה הספציפית כולל חותך (k).

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

שיטות לפתרון משוואות סימולטניות:

1. שיטת ריבועים פחותים עקיפה (ILS):
   א. יש להציג את מערכת משוואות המבנה בצורתה המצומצמת.
   ב. יש לאמוד בשיטת OLS את הפרמטרים של המשוואות בצורה המצומצמת.
   ג. יש לחלץ מן הפרמטרים של המערכת המצומצמת את הפרמטרים של הצורה המבנית.

משום שתהליך החילוץ איננו ליניארי האומדים המבניים המתקבלים הם מוטים אך עקיבים.
כאשר הזיהוי מדויק: האומדים יהיו גם אסימפטוטית יעילים (במדגמים גדולים).
כאשר הזיהוי הוא יתר: האומדים לא יהיו יעילים.

2. שיטת ריבועים פחותים בשני שלבים (2SL2):
   א. אמידת משוואות הצורה המצומצמת בשיטת OLS ושימוש בתוצאות האמידה כדי לחשב את המשתנים האנדוגניים (המסבירים).
   ב. הצבת המשתנים האנדוגניים שהתקבלו במשוואות המבנה ואומדתם ב-OLS.

אם משוואות המבנה מזוהות בדיוק או ביתר – האומדים שיתקבלו יהיו אמנם מוטים אבל עקיבים ויעילים אסימפטוטית. האומדים שיתקבלו יהיו זהים לאומדים שהתקבלו בשיטת הריבועים הפחותים העקיפה.
כאשר אין זיהוי: אין אקסוגניים ולכן אין משתנים מסבירים בצורה המצומצמת או שכל האקסוגניים בצורה המצומצמת כבר קיימים במשוואה המקורית ולכן החלפת x ב- $\hat{x}$ תיצור בעיה של מולטיקוליניאריות מלאה.

3. שיטת משתני העזר (IV):
   משתנה עזר הוא משתנה שיחליף את המשתנה המסביר האנדוגני במשוואת המבנה ויעזור לאמוד את הקשר בינו לבין התלוי.
   משתנה העזר צריך להיות:
   א. משתנה אקזוגני או פונקציה ליניארית של משתנים אקזוגניים: $\operatorname{cov}(Z,u)=0$.
   ב. מתואם עם המשתנה האנדוגני אותו הוא מחליף: $\operatorname{cov}(Z,X)\neq 0$.

ככל שהמתאם גבוה יותר, האומד שיתקבל באמצעותו יהיה טוב יותר.
הבעיה: אומדני OLS שיתקבלו יהיו מוטים, לא עקיבים ולא יעילים.
הפתרון בשיטת IV: אמידת השפעה של Y על X עם משתנה אקסוגני שלא קיים במערכת שמתואם עם Y (אותו הוא מחליף) אך לא עם u.

אם יש יותר ממשתנה עזר אחד המקיימים את התנאים הנ"ל, האומדים שיתקבלו יהיו כולם מוטים אך עקיבים (ניתן להשתמש בהם במדגמים גדולים).
משתנה העזר היחיד שיניב אומד יעיל יהיה בעל המתאם הגבוה ביותר עם המשתנה האנדוגני אותו הוא בא להחליף. משתנה עזר זה יהיה אומדן לאנדוגני שהתקבל מאמידת משוואת הצורה המצומצמת בשלב הראשון של 2SLS.

משתנה לא יוכל לשמש כמשתנה עזר:
אם נוסחתו מכילה רק משתנים אקזוגניים המצויים במשוואת המבנה בה הוא משמש כמשתנה עזר, שכן אז תיווצר בעיית מולטיקוליניאריות מלאה.
במילים אחרות, נוסחת משתנה העזר צריכה להיות מורכבת מלפחות משתנה אקזוגני אחד שלא מופיע במשוואה כדי שהמשתנה יוכל לשמש כמשתנה עזר.

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

סיכום תוצאות אמידה של משוואות סימולטניות:

מסי האומדים שיתקבלו בשלושת השיטות ותכונותיהם תלויים בזיהוי של המשוואה:
אם המשוואה לא מזוהה: לא ניתן להשתמש באף אחת מהשיטות.
כאשר המשוואה מזוהה (בדיוק או ביתר): האומדים שיתקבלו בשלושת השיטות יהיו תמיד מוטים אך עקיבים.

תכונת היעילות ומסי האומדים האפשרי מסוכמים בטבלה הבאה:

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

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

משתנים בפיגור ומשוואות סימולטניות:

אם X_t אקסוגני אז גם המשתנים בפיגור X_{t-p} בוודאות אקסוגניים.
אם Y_t אנדוגני אז מעמדם של המשתנים בפיגור תלוי בקיומו של מתאם סדרתי:
אם יש מתאם סדרתי: $\text{cov}\left(Y_{t-1},u_t \neq 0\right)$, אז Y_{t-1} אנדוגני.
אם אין מתאם סדרתי: $\text{cov}\left(Y_{t-1},u_t = 0\right)$, אז Y_{t-1} אקסוגני.

מבחנים סטטיסטיים לבחינת אנדוגניות ולחוזק משתנה עזר:

מבחן האוזמן (Hausman Test):
מבחן המשמש אותנו לבחינת אנדוגניות של משתנה מסוים.

- השלב הראשון לביצוע מבחן האוזמן הוא הרצת המשוואה המצומצמת – כלומר, המשתנה שחושדים שהוא אנדוגני כתלוי על כל האקסוגניים.
- מאמידה זו נשמור את סידרת השאריות הנאמדות (vhat).
- כעת נאמוד את המודל המקורי (משוואת המבנה) ונוסיף לו את (vhat) כמשתנה מסביר חדש.
- לפי תוצאות האמידה – אם המקדם של vhat מובהק נסיק כי המשתנה הוא אכן משתנה אנדוגני במודל.

מבחן לחוזק IV:
מבחן שמתבצע על המשוואה המצומצמת שבה נעשה שימוש במשתני העזר.
בודקים:

א. האם משתנה העזר לניבוי המשתנה התלוי מובהק באופ' באמצעות מבחן t למובהקות מקדם הרגרסיה. אם כן- ניתן להסיק כי המשתנה האקסוגני, המשמש כמשתנה עזר, מתואם עם האנדוגני אותו הוא אמור להחליף.

ב. אולם בכדי לבדוק האם משתני העזר חזקים מספיק נבצע מבחן F למובהקות כל משתני העזר המוצעים במשוואה המצומצמת. כלל אצבע-רק אם:
$F_{stat} > 10$ נוכל להסיק כי משתני העזר חזקים מספיק בכדי שנוכל לקבל תוצאות אמינות כאשר אנו משתמשים בהם.

שאלות:

זיהוי משוואות המבנה:

1) חוקר רצה לאמוד את פונקציית הביקוש ואת פונקציית ההיצע לתות שדה.
הוא אסף נתונים עבור 30 תקופות:
Pt - מחיר קופסא בש״ח בתקופה t.
Qt - כמות נקנית בק״ג בתקופה t.
Zt - מחיר פרי תחליפי ב-₪ בתקופה t.
INCOMEt - הכנסת הצרכנים באלפי ₪ בתקופה t.
Lt - מחיר שעת עבודה ב-₪ בתקופה t.

א. החוקר מניח שהכמות המבוקשת היא פונקציה של מחיר התות שדה, של מחיר הפרי התחלפי ושל הכנסת הצרכנים, והכמות המוצעת היא פונקציה של מחיר התות שדה ושל מחיר העבודה.
נסחו את המודל הסימולטני, תחת ההנחה שהגמישויות קבועות.
הציגו גם את תנאי הסדר וקבעו עבור כל משוואה אם היא מזוהה במדויק, ביתר או בחסר.

ב. עיינו במודל 1 שבדפי הפלט (ראו סרטון) והשיבו: איזו פונקציה נאמדה, והאם תוצאות האמידה שהתקבלו מתיישבות עם התיאוריה הכלכלית? נמקו.

ג. עיינו בדפי הפלט המתאימים (ראו סרטון) והשיבו: אם העלות של שעת עבודה תעלה באחוז אחד, מהם השינויים הצפויים בכמות ובמחיר של שווי משקל?

ד. בתקופה מסוימת אנו צופים שמחיר המוצר התחלפי יהיה 10 ₪, ההכנסה תהיה 50 אלף ₪, מחיר שעת עבודה 25 ₪.
מה יהיה מחיר שווי המשקל של תות השדה?
האם ניתן גם לאמוד את כמות שווי המשקל?

להלן הפלטים:

Model 1: TSLS estimates using the 30 observations 1-30

Dependent variable: l_Q

Instruments: l_L

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.83485	1.14385	-1.604	0.10869
l_P	-1.34898	0.645690	-2.089	0.03669 **
l_Z	1.72145	0.467875	3.679	0.00023 ***
l_income	0.984145	0.483543	2.035	0.04182 **

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 2.67757

Standard error of residuals = 0.32091

Unadjusted R-squared = 0.222881

Adjusted R-squared = 0.133214

F-statistic (3, 26) = 2.48564 (p-value = 0.0829)

Model 3: OLS estimates using the 30 observations 1-30

Dependent variable: l_Q

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	0.499595	0.630065	0.793	0.43500
l_L	-0.611731	0.163450	-3.743	0.00091 ***
l_income	0.395076	0.142590	2.771	0.01019 **
l_Z	0.937441	0.197381	4.749	0.00007 ***

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 0.834357

Standard error of residuals = 0.179139

Unadjusted R-squared = 0.681008

Adjusted R-squared = 0.644201

F-statistic (3, 26) = 18.5023 (p-value < 0.00001)

Log-likelihood = 11.1662

(Log-likelihood for Q = -75.1618)

Model 4: OLS estimates using the 30 observations 1-30

Dependent variable: l_P

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.73053	0.419481	-4.125	0.00034 ***
l_L	0.453478	0.108821	4.167	0.00030 ***
l_income	0.436678	0.0949326	4.600	0.00010 ***
l_Z	0.581185	0.131411	4.423	0.00015 ***

Mean of dependent variable = 2.99427

Standard deviation of dep. var. = 0.314761

Sum of squared residuals = 0.369832

Standard error of residuals = 0.119266

Unadjusted R-squared = 0.871281

Adjusted R-squared = 0.856428

F-statistic (3, 26) = 58.6632 (p-value < 0.00001)

Log-likelihood = 23.3704

(Log-likelihood for P = -66.4576)

שיטת ILS:

2) נניח שאנו מתכוונים לאמוד את המשוואות:

$$C_t = \alpha + \beta Y_t + u_t$$
$$Y_t = C_t + Z_t$$

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

א. מהי הבעיה באמידת המשוואות בשיטת הריבועים הפחותים? מהן תכונות אר"פ?

ב. האם המשוואות מזוהות?

ג. אמדו את מערכת המשוואות בצורתה המצומצמת באופן ידני.

ד. מהו הפתרון של המשוואות המצומצמות בשיטת ILS?

להלן תוצאות אמידת מערכת המשוואות בצורה המצומצמת:

Dependent Variable: C

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	14848.38	2568.8	5.78027	0.000
Z	1	-0.087066	0.3036	-0.2867	0.776

Dependent Variable: Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	14848.38	2568.8	5.78027	0.0000
Z	1	0.912934	0.3036	3.00699	0.0049

ה. חשבו את האומדים המבניים.

שיטת 2SLS:

3) תאר את תהליך האמידה בשני שלבים (2SLS) של משוואות המבנה:

$$C_t = \alpha + \beta Y_t + u_t$$
$$Y_t = C_t + Z_t$$

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

א. מה ניתן יהיה לומר על האומדים שהתקבלו בשיטה זו?

ב. מה יהיה ערכם של האומדים $\hat{\beta}$ ו- $\hat{\alpha}$?

להלן תוצאות האמידה בשיטת 2 השלבים:

Dependent variable: C

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	16264.47	8221.233	1.978349	0.0520
y	1	-0.095370	0.364274	-0.261808	0.7943

Dependent variable: Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-9.95E-09	3.52E-09	-2.828212	0.0062
C	1	1.00000	2.08E-13	4.80E+12	0.0000
Z	1	1.00000	1.99E-13	5.04E+12	0.0000

4) לפניך המודל הסימולטני הבא:

משוואת הביקוש: $Q_t^D = \alpha_0 + \alpha_1 P_t + \alpha_2 Z_t + u_t$.

משוואת ההיצע: $Q_t^S = \beta_0 + \beta_1 P_t + v_t$.

P_t - מחיר המוצר בתקופה t.

Q_t^D - כמות מבוקשת בתקופה t.

Q_t^S - כמות מוצעת בתקופה t.

Z_t - מחיר המוצר התחלפי בתקופה t.

Z_t הוא משתנה אקסוגני.

א. רשום את המשוואות המצומצמות וקבע את התכונות של אומדי OLS למשוואות אלה.

ב. היעזר בשיטת ILS לאמידת הפרמטרים של המשוואה שניתן לזהות, אם התקבלו המשוואות המצומצמות הבאות:

$$\hat{Q}_t = 2 + 3Z_t$$
$$\hat{P}_t = 1 + 4Z_t$$

ג. באם ננסה לאמוד את משוואת הביקוש בשיטת TSLS:
האם ניתן יהיה לאמוד מספרית את המשוואה של השלב הראשון? נמק.
האם ניתן יהיה לאמוד מספרית את המשוואה של השלב השני? נמק.

ד. החוקר מנסה לאמוד את משוואת ההיצע בשיטת TSLS.
למה שווה האומדן שיתקבל ל-β_1?

שיטת IV:

5) נתונות המשוואות הבאות:

1. $Y_t = \alpha_0 + \alpha_1 \cdot X_t + \alpha_2 \cdot Z_{1t} + \alpha_3 \cdot Z_{2t} + \varepsilon_t$.

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

נתון כי: T_t , X_t משתנים אנדוגניים ו-Z_{2t} , Z_{1t} משתנים אקסגוניים.

חוו דעתכם על כל אחת מהטענות הבאות, והסבירו:

א. ניתן להשתמש ב-Z_{1t} כמשתנה עזר לאמידת משוואה מס' 1.

ב. ניתן להשתמש ב-$\frac{Z_{1t}+Z_{2t}}{2}$ כמשתנה עזר לאמידת משוואה מס' 2.

ג. יתכנו מספר אומדים עקיבים שונים זה מזה ל-β_2 במשוואה מס' 2.

ד. שימוש ב-Z_2 כמשתנה עזר לאמידת משוואה מס' 2 יניב אומדים עקיבים וגם יעילים.

ה. משתנה העזר $Z_{1t} + Z_{2t}$ יניב אומדים זהים לאלו שהתקבלו בסעיף ב'.

ו. משתנה העזר $3Z_{1t} + 5Z_{2t}$ יניב אותם אומדים כמו משתנה העזר בסעיף ד'.

ז. משתנה העזר $7Z_{1t} + 5Z_{2t}$ יניב אומדים זהים לאלו שהתקבלו בסעיף ב'.

מבחן האוזמן:

6) נניח שאתם מעוניינים לאמוד את הקשר בין כלכלה לפיתוח פוליטי.
לכל מדינה i נסמן ב- Y_i – את הרמה הנאמדת של ההכנסה, נסמן ב- s_i את שיעור החיסכון במדינה i וב- D_i את האיתנות הנאמדת של השלטון הדמוקרטי. אתם שוקלים לאמוד מודל ליניארי של הכנסה ושיעור החיסכון על איתנות הממשל דמוקרטי: $D_i = \beta_1 + \beta_2 Y_i + \beta_3 s_i + \varepsilon_i$.
אבל אתם חוששים ש- $\operatorname{cov}(Y_i, \varepsilon_i) \neq 0$.
הסבירו כיצד תשתמשו ב – *Hausman Test* כדי לבחון את ההשערה:
$H_0 : \operatorname{cov}(Y_i, \varepsilon_i) = 0$?

מבחן לחוזק IV:

7) נתונה מערכת המשוואות הסימולטניות הבאות:

$$Y_{1i} = \gamma Y_{2i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$
$$Y_{2i} = \delta Y_{1i} + \beta_3 X_{3i} + v_i$$

כאשר: X_1, X_2, X_3 הינם משתנים אקסוגניים.
להלן מערכת המשוואות של הצורה המצומצמת:

$$Y_{1i} = \pi_{11} X_{1i} + \pi_{12} X_{2i} + \pi_{13} X_{3i} + \tilde{u}_i$$
$$Y_{2i} = \pi_{21} X_{1i} + \pi_{22} X_{2i} + \pi_{23} X_{3i} + \tilde{v}_i$$

תארו כיצד בודקים ש- X_{1i} ו- X_{2i} אינם משתני עזר חלשים ל- Y_{1i} במשוואה השנייה?

תרגילים מסכמים:

1) נתונות המשוואות הבאות:

1. $Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_{1t} + \alpha_3 Z_{2t} + \alpha_4 Z_{3t} + \alpha_5 Z_{4t} + u_t$.
2. $X_t = \beta_0 + \beta_1 Y_t + \beta_2 Z_{1t} + \beta_3 Z_{2t} + \beta_4 Z_{5t} + v_t$.

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

א. אמידת כל אחת מהמשוואות תניב אומדים:

i. מוטים — נכון/לא נכון/לא ניתן לדעת.

ii. עקיבים — נכון/לא נכון/לא ניתן לדעת.

ב. משוואה 1 — מזוהה בדיוק/ מזוהה ביתר/בלתי מזוהה

משוואה 2 — מזוהה בדיוק/ מזוהה ביתר/ בלתי מזוהה

ג. חווה דעתך על הטענות הבאות:

i. תוך שימוש בשיטת ILS ניתן לאמוד את משוואה 1 באופן עקיב וחד ערכי: — נכון/לא נכון/לא ניתן לדעת

ii. תוך שימוש בשיטת ILS ניתן לאמוד את משוואה 2 באופן עקיב וחד ערכי: — נכון/לא נכון/לא ניתן לדעת

ד. משוואות הצורה המצומצמת הן:

$Y_t = \lambda_0 + \lambda_1 Z_{1t} + \lambda_2 Z_{2t} + \lambda_3 Z_{3t} + \lambda_4 Z_{4t} + \lambda_5 Z_{5t} + \varepsilon_{1t}$.

$X_t = \mu_0 + \mu_{1t} + \mu_2 Z_{2t} + \mu_3 Z_{3t} + \mu_4 Z_{4t} + \mu_5 Z_{5t} + \varepsilon_{2t}$.

נכון/לא נכון/לא ניתן לדעת.

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

ו. להלן רשימה של משתני עזר פוטנציאליים:

i. Z_5.

ii. $\frac{Z_1 + Z_5}{2}$.

iii. $2Z_1 + 3Z_2 + Z_3$.

iv. $Z_3 + Z_4$.

v. $3Z_3 + 4Z_4$.

vi. $3Z_3 + 3Z_4$.

vii. Z_1.

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

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

i. i ו-ii.

ii. iv ו-vi.

iii. v ו-vi.

iv. iv ו-v.

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

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

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

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

יב. הנח כי הוטלו המגבלות הבאות על הפרמטרים המבניים: $\alpha_2 = \beta_2 = 0$.
האם ניתן כעת לזהות את יתר הפרמטרים במודל?

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

$$HOURS = \beta_1 + \beta_2 WAGE + \beta_3 EDUC + \beta_4 AGE + \beta_5 KIDSL6 + \beta_6 KIDS618 + \beta_7 NWIFEINC + \varepsilon$$

כאשר:

$HOURS$ - היצע העבודה בשעות.

$WAGE$ - שכר לשעה.

$EDUC$ - מספר שנות הלימוד.

AGE - גיל.

$KIDSL6$ - מספר הילדים בבית מתחת לגיל 6.

$KIDS618$ - מספר הילדים בגיל 6-18.

$NWIFEINC$ - הכנסת משק הבית ממקורות שאינם בעבודתה של האישה.

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

ב. הסבירו מדוע לא ניתן לאמוד את משוואת ההיצע הנ"ל בשיטת הריבועים הפחותים.

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

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

3) נתונה מערכת המשוואות הסימולטניות הבאה:

$$Y_{1t} = \gamma Y_{2i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$
$$Y_{2i} = \delta Y_{1i} + \beta_3 X_{3i} + v_i$$

כאשר X_1 , X_2 , X_3 הינם משתנים אקסוגניים.

א. חלצו את מערכת המשוואות המצומצמת (Reduced Form Equations) של Y_1 ו- Y_2 (ז"א פתרו את המערכת המבנית עבור שני המשתנים האנדוגניים Y_1 ו- Y_2 על מנת לקבל את הצורה המצומצמת. כתבו את המקדמים והשאריות במערכת המצומצמת למטה כפונקציות של הפרמטרים והשאריות במערכת המבנית).

ב. הראו שבהינתן אומדים עקיבים ל- $\pi_{11},...,\pi_{23}$ ניתן למצוא אומד עקיב ל- γ.

ג. האם γ ניתן לזיהוי כאשר $\beta_3 = 0$?

ד. אילו תנאים צריכים X_{1i} ו- X_{2i} לקיים בכדי להיות משתני עזר ל- Y_{1i} במשוואה השנייה?

ה. תארו כיצד בודקים ש- X_{1i} ו- X_{2i} אינם משתני עזר חלשים ל- Y_{1i} במשוואה השנייה?

4) נניח שאתם מעוניינים לאמוד את הקשר בין כלכלה לפיתוח פוליטי.
לכל מדינה i נסמן ב- Y_i - את הרמה הנאמדת של ההכנסה וב- D_i את האיתנות הנאמדת של השלטון הדמוקרטי. אתם שוקלים לאמוד מודל ליניארי של הכנסה על ממשל דמוקרטי: $D_i = \beta_1 + \beta_2 Y_i + \varepsilon_i$, אבל אתם חוששים ש- $\text{cov}(Y_i, \varepsilon_i) \neq 0$.

א. הסבירו מדוע החשש שההכנסה מתואמת עם השגיאה במשוואה הנ"ל הגיוני?

ב. האם אומד הריבועים הפחותים של β_2 הינו חסר הטיה?

ג. נסמן ב- S_i את שיעור החיסכון במדינה i. הסבירו אלו תנאים צריך משתנה עזר (iv) לקיים. נמקו מדוע S_i מתאים או לא מתאים לשמש כמשתנה עזר.

ד. הסבירו כיצד תשתמשו בשיטת 2SLS כדי לאמוד את β_2. האם האומד המתקבל עקיב?

ה. הסבירו כיצד תשתמשו ב- *Hausman Test* כדי לבחון את ההשערה: $H_0 : \text{cov}(Y_i, \varepsilon_i) = 0$.

תשובות סופיות:

1) א. $\ln Q_t^D = \alpha_0 + \alpha_1 \ln P_t + \alpha_2 \ln Z_t + \alpha_3 \ln INCOME_t + u_t$

$\ln Q_t^S = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln L_t + v_t$

$Q_t^D = Q_t^S$

משוואת הביקוש מזוהה במדויק.

משוואת ההיצע מזוהה ביתר.

ב. פונקציית הביקוש, התוצאות מתיישבות.

ג. הכמות תרד ב-0.61173%, המחיר יעלה ב-0.453478%.

ד. $\hat{Q} = 9.34$, $\hat{P} = 16.05$.

2) א. ראו סרטון. ב. מזוהות בדיוק. ג. $\hat{Y}_t = \hat{\gamma}_3 + \hat{\gamma}_4 Z_t$, $\hat{C}_t = \hat{\gamma}_0 + \hat{\gamma}_1 Z_t$.

ד. $\hat{\beta} = \frac{\hat{\gamma}_1}{\hat{\gamma}_4}$, $\hat{\alpha} = \frac{\hat{\gamma}_3}{\hat{\gamma}_4} = \frac{\hat{\gamma}_0}{\hat{\gamma}_4}$. ה. $\hat{\beta} = -0.09537$, $\hat{\alpha} = 16,264.46$.

3) א. מוטים אך עקיבים ויעילים במדגמים גדולים.

ב. $\hat{\alpha} = 16,264.47$, $\hat{\beta} = -0.09537$.

4) א. $Q_t = \beta_0 + \beta_1 \cdot \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\beta_1 \cdot \alpha_2}{\alpha_1 - \beta_1} \cdot Z_t + \frac{\beta_1 (u_t - v_t)}{\alpha_1 - \beta_1} + v_t$, BLUE.

ב. $\hat{\beta}_1 = 0.75$, $\hat{\beta}_0 = 1.25$. ג. שלב ראשון: ניתן, שלב שני: לא ניתן.

ד. 0.75.

5) א. לא נכון. ב. נכון. ג. נכון. ד. נכון. ה. נכון.

ו. נכון. ז. לא נכון.

6) ראו סרטון.

7) ראו סרטון.

תרגילים מסכמים:

1) א.i. נכון. ii. לא נכון.

ב. משוואה 1: מזוהה בדיוק, משוואה 2: מזוהה ביתר.

ג.i. נכון. ii. לא נכון. ד. נכון. ה. נכון.

ו.i. 1. ii. 1. iii. 2. iv. 2. v. 2.

vii. 2. ז.i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ח. כן. ט. אקסוגניים. י. לא. יא. כן.

יב. משוואה 1 מזוהה בדיוק ומשוואה 2 מזוהה ביתר.

2) א. מקדם wage חיובי, מקדם educ לא ניתן לדעת, מקדם age יכול להיות חיובי או שלילי, מקדם kidsl6 שלילי, מקדם kids618 חיובי, מקדם nwifeinc שלילי.

ב. ראו סרטון. ג. ראו סרטון. ד. ראו סרטון.

3) א. $\pi_{13} = \frac{\beta_3 \gamma}{1 - \delta\gamma}$, $\pi_{12} = \frac{\beta_2}{1 - \delta\gamma}$, $\pi_{11} = \frac{\beta_1}{1 - \delta\gamma}$,

$\pi_{23} = \frac{\beta_3}{1 - \delta\gamma}$, $\pi_{22} = \frac{\beta_2 \delta}{1 - \delta\gamma}$, $\pi_{21} = \frac{\beta_1 \delta}{1 - \delta\gamma}$,

$\tilde{u}_i = \dfrac{u_i + \gamma v_i}{1 - \delta\gamma}$, $\tilde{v}_i = \dfrac{v_i + \delta u_i}{1 - \delta\gamma}$.

ב. מכיוון ש- $\gamma = \dfrac{\pi_{13}}{\pi_{23}}$, ניתן לקבל אומד עקיב ל- γ ע״י $\hat{\gamma} = \dfrac{\hat{\pi}_{13}}{\hat{\pi}_{23}}$.

ג. לא. ד. צריכים להיות מתואמים עם y_{1i} ובלתי מתואמים עם v_i .

ה. ראו סרטון.

4) א. טעות מדידה במשתנה המוסבר, משתנה מושמט, משוואות סימולטניות.

ב. לא. ג. ראו סרטון. ד. עקיב. ה. ראו סרטון.

אקונומטריקה למדעי הנתונים

פרק 30 - סיכום תכונות ארפ

תוכן העניינים

סיכום תכונות אר"פ:

רקע:

אם המודל שאותו אנו אומדים מנסח נכון את הקשר המתמטי בין המשתנים וכולל את כל המשתנים הרלוונטיים להסבר התופעה ורק אותם ולא קיים קשר חזק או מלא בין המשתנים הבלתי תלויים במודל וכל ההנחות הקלאסיות מתקיימות, אזי האומדים שאותם נקבל הם א.ח.ה, הם BLUE (יעילים) והם עקיבים.

הבעיה במודל	כולל את כל המשתנים הרלוונטיים	לא כולל משתנים שאינם רלוונטיים	אין קשר חזק או מלא בין הב"ת	כל ההנחות הקלאסיות מתקיימות	פגיעה בתכונות אר"פ
אין בעיה	√	√	√	√	אין פגיעה
השמטת משתנים רלוונטיים	×	√	√	√	פגיעה בכל התכונות - *מוטים, לא יעילים ולא עקיבים
הוספת משתנים לא רלוונטיים	√	×	√	√	אין פגיעה
מולטיקולינאריות חלקית	√	√	×	√	אין פגיעה
מולטיקולינאריות מלאה	√	√	×	√	אר"פ אינם מוגדרים
הטרוסקדסטיות	√	√	√	× $V(u_t)=\sigma_t^2$	יעילות
מתאם סדרתי	√	√	√	× $\operatorname{cov}(u_t,u_s)\neq 0$	יעילות
מודלים דינמיים	√	√	√	× $\operatorname{cov}(x,u)\neq 0$	מוטים היעילות והעקיבות תלויים בקיום מתאם סדרתי
משוואות סימולטניות	√	√	√	× $\operatorname{cov}(x,u)\neq 0$	**אומדים מוטים אך עקיבים. היעילות תלויה בשיטה.

*אלא אם אין מתאם בין המשתנה הב"ת שהושמט לאלו המצויים במודל.
**בהנחה שהמשוואות מזוהות.

אקונומטריקה למדעי הנתונים

פרק 31 - הדרכה בקריאת פלטים רלוונטיים בתוכנת GRTL

תוכן העניינים

קריאת פלטים רלוונטיים בתוכנת GRTL:

שאלות:

רגרסיה מרובה:

1) המחקר עוסק בחקירת פונקציית הצריכה של סיגריות בתורכיה, בשנים 1960-1988. המחקר גם מתייחס לכך שבשנת 1982 חלה עלייה משמעותית בתל"ג לנפש ובמחירי הסיגריות, ולכן סביר להניח שלאורך התקופה הנחקרת חלו שינויים בפונקציית הצריכה.
המודלים נאמדו לפי המשתנים הבאים:
Q - צריכת סיגריות ממוצעת, בק״ג.
Y - תל״ג לנפש במחירים של 1982, בלירות תורכיות.
P - מחיר ריאלי של סיגריות, בלירות תורכיות לק״ג.
D82 - משתנה דמי השווה ל-1 החל משנת 1982, ואחרת -אפס.

בשל טרנספורמציות לוגריתמיות הוגדרו גם המשתנים:
1_Q - הלוג של Q.
1_Y - הלוג של Y.
1_P - הלוג של P.
המשתנה D82 1_Y : 1_Y כפול המשתנה D82.
המשתנה D82 1_P : 1_P כפול המשתנה D82.
להלן תוצאות האמידה:

Model 1: OLS estimates using the 29 observations 1960-1988
Dependent variable: l_Q

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-4.58987	0.724913	-6.332	<0.00001
l_Y	0.688498	0.0947276	7.268	<0.00001
l_P	-0.485683	0.101394	-4.790	0.00006

Mean of dependent variable = 0.784827
Standard deviation of dep. var. = 0.108499
Sum of squared residuals = 0.0949108
Standard error of residuals = 0.0604187
Unadjusted R-squared = 0.712058
Adjusted R-squared = 0.689908
(F-statistic (2, 26) = 32.148 (p-value < 0.00001
Durbin-Watson statistic = 1.00057
First-order autocorrelation coeff. = 0.489867
Log-likelihood = 41.8214
(Log-likelihood for Q = 19.0615)
Akaike information criterion (AIC) = -77.6429
Schwarz Bayesian criterion (BIC) = -73.541
Hannan-Quinn criterion (HQC) = -76.3582

Model 2: OLS estimates using the 29 observations 1960-1988

Dependent variable: l_Q

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-5.02489	0.541262	-9.284	<0.00001
D82	13.8623	2.85197	4.861	0.00007
l_Y	0.735837	0.0726077	10.134	<0.00001
D82l_Y	-1.68927	0.345466	-4.890	0.00006
l_P	-0.381857	0.103289	-3.697	0.00119
D82l_P	0.490526	0.153989	3.185	0.00412

Mean of dependent variable = 0.784827

Standard deviation of dep. var. = 0.108499

Sum of squared residuals = 0.0261724

Standard error of residuals = 0.0337332

Unadjusted R-squared = 0.920598

Adjusted R-squared = 0.903336

(F-statistic (5, 23) = 53.3328 (p-value < 0.00001

Durbin-Watson statistic = 2.02153

First-order autocorrelation coeff. = -0.0136939

Log-likelihood = 60.5008

(Log-likelihood for Q = 37.7408)

Akaike information criterion (AIC) = -109.002

Schwarz Bayesian criterion (BIC) = -100.798

Hannan-Quinn criterion (HQC) = -106.432

א. לפי תוצאות האמידה של מודל (1) ושל מודל (2), האם האבחנה בין השנים שלפני ואחרי 1982 תורמת הסבר משמעותי לשונות צריכת הסיגריות? נסחו ובחנו השערה זו בר"מ 0.05.

ב. חשבו, לפי מודל (2), המבחין בין השנים שלפני ואחרי 1982, רווח סמך לגמישות הביקוש לסיגריות לפי המחיר, ברמת בטחון 95%, ועבור השנים 1960-1981 בלבד.

ג. התייחסו למודל (2) בלבד והשיבו : מהי גמישות הביקוש ביחס למחיר לפני 1982? ומהי גמישות הביקוש ביחס למחיר אחרי 1982. כמו כן, בחנו את השערה, בר״מ 0.05, שההבדל בין גמישויות אלה שווה ל-0.4.

ד. האם קיים מתאם סדרתי מסדר ראשון לפי מודל (1), והאם קיים מתאם סדרתי מסדר ראשון לפי מודל (2)? בחנו השערות אלה בר״מ 0.05. האם האבחנה בין השנים שלפני ואחרי 1982 גורמת למתאם סדרתי בהפרעה המקרית?

מבחן וולד:

2) המחקר התמקד במתכנתים אקדמאים בעלי תואר ראשון ושני.
לאמידת התשואה של כל שנת ותק לשכר נאספו הנתונים על ההכנסה (INCOME) ושנות הוותק (VETEK). כמו כן נעשתה אבחנה בין התארים האקדמיים בעזרת משתני הדמי M כאשר M מקבל את הערך 1 אם לנבדק יש תואר שני – MA , ואחרת את הערך 0.
מצורף הפלט הבא:

LS // Dependent Variable is INCOME
Included observations: 26

Variable	Coefficient	Std. Error	t-Statistic	Prob.
M	335.3201	213.8745	1.567835	0.1312
VETEK	31.50026	6.113796	5.152324	0
M*VETEK	10.78545	13.80593	0.781219	0.443
C	3459.347	92.51852	37.39085	0

R-squared	0.752005	Mean dependent var	3990.769
Adjusted R-squared	0.718187	S.D. dependent var	390.1787
S.E. of regression	207.1305	Akaike info criterion	10.80734
Sum squared resid	943867	Schwarz criterion	11.00089
Log likelihood	-173.3878	F-statistic	22.2371
Durbin-Watson stat	1.945103	Prob(F-statistic)	0.000001

==================

INCOME = C(1)*M + C(2)*VETEK + C(3)*(M*VETEK) + C(4)

Wald Test:
Equation: Untitled

Null Hypothesis:
C(1)=0
C(3)=0

F-statistic	12.89398	Probability	0.000197
Chi-square	25.78797	Probability	0.000003

מהי השערת האפס שנבחנה ומהן תוצאות המבחן הסטטיסטי שנעשה?

הטרוסקדסטיות:

3) חוקר ביקש לאמוד מודל המקשר בין מחירי בתים חד משפחתיים המוסברים על ידי המשתנים המסבירים: שטח המגורים בבית ספר ומספר חדרי המגורים והאמבטיה.

להלן המשתנים:

Prices - מחירי הבתים באלפי דולרים.

Sqft - שטח המגורים בפיט רבוע.

Bedrm - מספר חדרים בבית.

Baths - מספר חדרי אמבטיה בבית.

החוקר חשד כי יש בעיית הטרוסקדסטיות בנתונים וביצע מבחן White.

תוצאות המבחן נתונות בפלט הבא:

```
White's test for heteroskedasticity

OLS using observations 1-14
Dependent variable: uhat^2
Omitted due to exact collinearity : sq_bdrms
              Coefficients     std.error     t-ratio     p-value
-------------------------------------------------------------------------

Const
Sqft
Bedrms
Baths
Sq_sqft
X2_X3
X2_X4
X3_X4
Sq_baths

Unadjusted R-squared=0.736608
Test statistic: TR^2=10.312512
With p-value=p(chi-square(8)>10.312512)=0.243773
```

רשמו את המשוואה שנאמדה ואת השערת האפס.

על סמך הפלט המצורף, האם יש בעיה של הטרוסקדסטיות בנתונים?

מתאם סדרתי:

4) ענו על הסעיפים הבאים:

א. האם במודל 1 קיים מתאם סדרתי מסדר ראשון? בחנו את ההשערה לפי מבחן דרבין ווטסון.

DATA7-19: Data on demand for cigarette consumption in Turkey
Source: "Cigarett demand, health scares, and education in Turkey," by Aysit Tansel, APPLIED ECONOMICS, 1993, pp. 521-529.
Q Cigarette consumption per adult (kg), Range 1.86 - 2.723.
Y Per capita real GNP in 1968 prices (in Turkish liras), Range 2560 - 5723
P Real price of cigarettes in Turkey liras per kg, Range 1.361 - 3.968

Model 1: OLS, using observations 1960-1988 (T = 29)
Dependent variable: Q

	coefficient	std. error	t-ratio	p-value
const	1.656542	0.123678	13.39398	3.53E-13
Y	0.000344	5.28E-05	6.517848	6.56E-07
P	-0.4233	0.096944	-4.36639	0.000179

Mean dependent var	2.204655	S.D. dependent var	0.24319
Sum squared resid	0.595167	S.E. of regression	0.151298
R-squared	0.640589	Adjusted R-squared	0.612942
F(2, 26)	23.17031	P-value(F)	1.67E-06
Log-likelihood	15.20081	Akaike criterion	-24.4016
Schwarz criterion	-20.2997	Hannan-Quinn	-23.117
rho	0.536727	Durbin-Watson	0.911596

ב. הסבירו את מבחן ההשערה שבפלט הבא:

Performing iterative calculation of rho...

ITER	RHO	ESS
1	0.53673	0.34981
2	0.70446	0.324784
3	0.74270	0.323411
4	0.75026	0.323356
5	0.75173	0.323354
6	0.75201	0.323354

Model 2: Cochrane-Orcutt, using observations 1961-1988 (T = 28)
Dependent variable: Q

	coefficient	std. error	t-ratio	p-value	
const	1.82344	0.392439	4.646	9.32e-05	***
Y	0.000181935	9.82015e-05	1.853	0.0758	*
P	-0.171091	0.0826337	-2.070	0.0489	**

Statistics based on the rho-differenced data:

Mean dependent var	2.216964	S.D. dependent var	0.238275
Sum squared resid	0.323354	S.E. of regression	0.113728
R-squared	0.789097	Adjusted R-squared	0.772225
F(2, 25)	2.577408	P-value(F)	0.096004
rho	-0.087820	Durbin-Watson	2.171288

ג. הסבירו כיצד ניתן היה לאמוד את ערכו של Q בשנת 1989.

5) נאמד הקשר שבין הכנסה לתצרוכת לתקופה: ינואר 1994 עד דצמבר 1997 (T=48). המודל הינו: $C_t = \alpha + \beta Y_t + u_t$.
בניסיון לבדוק האם מתקיים קשר מהסוג הבא: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$.
נאמדה המשוואה הבאה: $u_t = \gamma_1 u_{t-1} + \gamma_2 u_{t-2} + \gamma_3 u_{t-3} + \gamma_4 Y_t + \omega_t$.

```
Dependent variable: uhat
              Coefficients      std.error       t-ratio        p-value
-----------------------------------------------------------------------------------------

Const
Uhat1
Uhat2
Uhat3

Y

Unadjusted R-squared=0.45002
Test statistic: TR^2=22.99
With p-value=p(chi-square(3)<22.99)=0.0432
```

הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: ________________.
על ידי מבחן: ________________.
ההשערות הינן: __________________.
גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): ___________.
המסקנה המתקבלת היא: ____________________________________.

משוואות סימולטניות:

6) חוקר רצה לאמוד את פונקציית הביקוש ואת פונקציית ההיצע לתות שדה.
הוא אסף נתונים עבור 30 תקופות:
Pt - מחיר קופסא ב-₪ בתקופה t.
Qt - כמות נקנית בק״ג בתקופה t.
Zt - מחיר פרי תחליפי ב-₪ בתקופה t.
INCOMEt - הכנסת הצרכנים באלפי ₪ בתקופה t.
Lt - מחיר שעת עבודה ב-₪ בתקופה t.
משוואות הביקוש וההיצע שנאמדו הינן:

$$\ln Q_t^D = \alpha_0 + \alpha_1 \ln P_t + \alpha_2 \ln Z_t + \alpha_3 \ln INCOME_t + u_t$$

$$\ln Q_t^S = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln L_t + v_t$$

$$Q_t^D = Q_t^S$$

מצורפים הפלטים הבאים שמתארים ארבעה מודלים:

Model 1: TSLS estimates using the 30 observations 1-30
Dependent variable: l_Q
Instruments: l_L

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.83485	1.14385	-1.604	0.10869
l_P	-1.34898	0.645690	-2.089	0.03669 **
l_Z	1.72145	0.467875	3.679	0.00023 ***
l_income	0.984145	0.483543	2.035	0.04182 **

Mean of dependent variable = 2.8776
Standard deviation of dep. var. = 0.300322
Sum of squared residuals = 2.67757
Standard error of residuals = 0.32091
Unadjusted R-squared = 0.222881
Adjusted R-squared = 0.133214
F-statistic (3, 26) = 2.48564 (p-value = 0.0829)

Model 2: TSLS estimates using the 30 observations 1-30
Dependent variable: l_Q
Instruments: l_Z l_income

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	3.07410	0.263717	11.657	<0.00001 ***
l_P	1.24927	0.115149	10.849	<0.00001 ***
l_L	-1.27242	0.129253	-9.844	<0.00001 ***

Mean of dependent variable = 2.8776
Standard deviation of dep. var. = 0.300322
Sum of squared residuals = 0.377808
Standard error of residuals = 0.118291
Unadjusted R-squared = 0.855763
Adjusted R-squared = 0.845079
F-statistic (2, 27) = 80.096 (p-value < 0.00001)

Model 3: OLS estimates using the 30 observations 1-30

Dependent variable: l_Q

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	0.499595	0.630065	0.793	0.43500
l_L	-0.611731	0.163450	-3.743	0.00091 ***
l_income	0.395076	0.142590	2.771	0.01019 **
l_Z	0.937441	0.197381	4.749	0.00007 ***

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 0.834357

Standard error of residuals = 0.179139

Unadjusted R-squared = 0.681008

Adjusted R-squared = 0.644201

F-statistic (3, 26) = 18.5023 (p-value < 0.00001)

Log-likelihood = 11.1662

(Log-likelihood for Q = -75.1618)

Model 4: OLS estimates using the 30 observations 1-30

Dependent variable: l_P

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.73053	0.419481	-4.125	0.00034 ***
l_L	0.453478	0.108821	4.167	0.00030 ***
l_income	0.436678	0.0949326	4.600	0.00010 ***
l_Z	0.581185	0.131411	4.423	0.00015 ***

Mean of dependent variable = 2.99427

Standard deviation of dep. var. = 0.314761

Sum of squared residuals = 0.369832

Standard error of residuals = 0.119266

Unadjusted R-squared = 0.871281

Adjusted R-squared = 0.856428

F-statistic (3, 26) = 58.6632 (p-value < 0.00001)

Log-likelihood = 23.3704

(Log-likelihood for P = -66.4576)

על סמך הפלטים הנ״ל, איזה פונקציה נאמדה בכל מודל?

תשובות סופיות:

1) א. כן. ב. $[-0.5955,-0.168]$. ג. לפני: 0.381857-, 0.108669.

ד. מודל (1): יש - חיובי, מודל (2): אין. אין בעיה של מתאם סדרתי.

2) השערת האפס: $H_0 : \beta_M = \beta_{M \cdot Vetek} = 0$.

תוצאות מבחן WALD: $P.V = 0.000197 < \alpha = 0.001$.

3) המשוואה שנאמדה: $e_i^2 = \hat{\gamma}_0 + \hat{\gamma}_1 sqft_i + \hat{\gamma}_2 bedrms_i + \hat{\gamma}_3 baths_i + \hat{\gamma}_4 sqft_i^2 + \hat{\gamma}_5 bedrms_i^2 +$

$\hat{\gamma}_6 baths_i^2 + \hat{\gamma}_7 sq*bed + \hat{\gamma}_8 sq*bath + \hat{\gamma}_9 bed*bath$

השערת האפס: $H_0 : \hat{\gamma}_1 = \hat{\gamma}_2 = ... = \hat{\gamma}_9$.

תוצאות מבחן White: $p-value = 0.243773 > \alpha = 0.005$.

4) א. כן. ב. ראו סרטון. ג. ראו סרטון.

5) קיומו של מתאם סדרתי מסדר שלישי בנתונים.

מבחן LM.

$$H_0 : \rho_1 = \rho_2 = \rho_3 = 0$$
$$H_1 : OTHERWISE$$

$LM_{stat} = T \cdot R^2 = 22.99$.

$pvalue < \alpha$, יש עדות לכך.

6) מודל 1: $\ln Q_t^D = \alpha_0 + \alpha_1 \ln P_t + \alpha_2 \ln Z_t + \alpha_3 \ln INCOME_t + u_t$.

מודל 2: $\ln Q_t^S = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln L_t + v_t$.

מודל 3: $\ln Q_t = \pi_0 + \pi_1 \ln L_t + \pi_2 \ln INCOME_t + \pi_3 \ln Z_t + \varepsilon_t$.

מודל 4: $\ln P_t = \lambda_0 + \lambda_1 \ln L_t + \lambda_2 \ln INCOME_t + \lambda_3 \ln Z_t + \varpi_t$.

אקונומטריקה למדעי הנתונים

פרק 32 - רגרסיה לוגיסטית

תוכן העניינים

רגרסיה לוגיסטית:

רקע:

מתי נבצע רגרסיה לוגיסטית?

כאשר המשתנה המנובא הוא דיכוטומי (Binary Logistic):
יכול לקבל ערכים של 0 או 1.
הפונקציה הלוגיסטית מתארת את הסיכויים לקבל " 1" במשתנה התלוי כתלות במשתנים הב"ת.

הלוגיקה בניתוח רגרסיה לוגיסטית:

השוואת ניבוי Y ללא המשתנים המנבאים במודל לניבוי Y במודל הכולל את המשתנים המנבאים (סטטיסטי $\chi 2$).

טיב מודל הרגרסיה ("Goodness of fit"):

1. מובהקות המודל:

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	12.225	4	.016
	Block	12.225	4	.016
	Model	12.225	4	.016

מבחן $\chi 2$ - תחת שורת ה-model נמצא את חי בריבוע ואת מובהקות המודל.

2. אחוז שונות מוסברת:

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	96.524	.139	.189

Nagelkerke R^2 – מקביל ל- R^2 כללי ברגרסיה. אחוז שונות Y המוסברת ע"י כל המנבאים יחד (בטווח מוכר של 0-1).

3. דיוק בניבוי:

Classification Table[a]

Observed			Predicted: whether mom believes course will help: no	Predicted: whether mom believes course will help: yes	Percentage Correct
Step 1	whether mom believes course will help	no	46	5	90.2
		yes	17	14	45.2
	Overall Percentage				73.2

a. The cut value is .500

סגוליות (true negative) – ביחס ל- Y =0 במדגם, כמה המודל דייק בניבוי (90.2%).

רגישות (true positive) – ביחס ל- Y=1 במדגם, כמה המודל דייק בניבוי (45.2%).

אחוז הניבוי הכללי – בכמה בסה"כ המודל מדייק בניבוי (73.2%).

מושגים חשובים להבנת טבלת המקדמים:

ODDS:

"הסיכוי להתרחשות אירוע מסוים"- ההסתברות שהאירוע יקרה לעומת ההסתברות שאותו אירוע לא יקרה: $ODDS = \frac{p}{1-p}$

ODDS=1 – הסיכוי שהאירוע יתרחש שווה לסיכוי שהוא לא יתרחש ($\frac{0.5}{0.5}$).

ODDS>1 – הסיכוי שהאירוע יתרחש גבוה מהסיכוי שלא יתרחש (למשל- $\frac{0.75}{0.25}$).

ODDS<1 – הסיכוי שהאירוע יתרחש נמוך מהסיכוי שלא יתרחש (למשל- $\frac{0.25}{0.75}$).

ODDS RATIO (OR):

יחס בין סיכויים - $\text{OR} = \frac{ODDS_{(A)}}{ODDS_{(B)}}$

כיצד משתנה ההסתברות במעבר מקבוצה A לקבוצה B.

OR=1 – הסיכוי להתרחשות האירוע שווה בין שתי הקבוצות- אין קשר בין המב"ת למ"ת.

OR>1 – הסיכוי להתרחשות האירוע בקבוצה A גבוה מאשר בקבוצה B – קשר חיובי.

OR<1 – הסיכוי להתרחשות האירוע בקבוצה A נמוך מאשר בקבוצה B – קשר שלילי.

טבלת המקדמים – תרומות ייחודיות של כל מנבא:

(מקביל לטבלת Coefficients ברגרסיות ליניאריות)

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	EDU_YRS	-.107	.138	.603	1	.438	.898
	AGE	-.029	.020	2.078	1	.149	.971
	SATISFAC	.118	.175	.457	1	.499	1.126
	BIRTH#	.882	.321	7.530	1	.006	2.415
	Constant	.001	1.796	.000	1	.999	1.001

a. Variable(s) entered on step 1: EDU_YRS, AGE, SATISFAC, BIRTH#.

1. מבחן WALD למובהקות המשתנים:
מבטא את מובהקות המשתנה מבחינת תרומתו הייחודית לניבוי Y.

2. B – מקדמי המשתנים ב-log odds:
בטא חיובית – עלייה ב-log odds של Y כפונקציה של עליה ביחידה אחת של X.
בטא שלילית – ירידה ב-log odds של Y כפונקציה של עליה ביחידה אחת של X.

3. משוואת הרגרסיה:

$$\log\left(\frac{p}{1-p}\right) = \widehat{\alpha} + \widehat{\beta_1}\, x_{1i} + \widehat{\beta_2} x_{2i} + \widehat{\beta_3} x_{3i} + \widehat{\beta_4} x_{4i}$$

חישוב הניבוי במונחי הסתברות (p): $p = \frac{1}{1+e^{-\log odds}}$

חישוב הניבוי במונחי סיכויים (ODDS): $ODDS = e^{\log odds}$

4. Exp (B) - יחס הסיכויים (Odds Ratio):
מבטא את העלייה (אם גדול מ-1) או את הירידה (אם קטן מ-1) בסיכויים להיות בעלי ערך '1' ב-Y כאשר הערך במשתנה המנבא גדל ביחידה אחת.

5. היחס בין B ל-Exp (B): $\log Exp(B) = B$; $e^{B} = Exp(B)$

שאלות:

1) חוקרת בחוג למגדר ביקשה לבדוק האם מגדר משפיע על תעסוקה.
היא התבססה על סקר של הלמ"ס שדגם 826 מבוגרים בגילאי העבודה המרכזיים (25-55).
היא הגדירה את המשתנים באופן הבא:
WOMEN - "1" = אישה ; "0" = גבר.
WORKING - "1" = כן ; "0" = לא.
מהצלבה של שני המשתנים התקבלה הטבלה הבאה:

working * women Crosstabulation
Count

		women .00	women 1.00	Total
working	.00	13	130	143
	1.00	338	345	683
Total		351	475	826

על סמך הטבלה חשבו:
א. מה ההסתברות של אישה לעבוד?
ב. מה הסיכוי של אישה לעבוד?
ג. מה ההסתברות של גבר לעבוד?
ד. מה הסיכוי של גבר לעבוד?
ה. מה יחס הסיכויים (OR) של נשים לעבוד לעומת גברים?
ו. מה הלוגריתם של יחס הסיכויים?
ז. מה יהיה ערך מקדם השיפוע B ברגרסיה הלוגיסטית לניבוי תעסוקה על פי מגדר ומה משמעותו?
ח. מה יהיה ערך Exp(B) ברגרסיה הלוגיסטית ומה משמעותו?

2) במחקר ביקשו לבדוק כיצד מצב משפחתי וגובה המשכורת משפיעים על בעלות על דירה.
משתני המחקר:
apartm - בעלות על דירה: "1"- כן; "0"- לא.
status - מצב משפחתי: status (0) - רווק; status (1) - בזוגיות;
status (2) - בזוגיות עם ילדים; status(3) - פרוד או גרוש.
incom - הכנסה (בעשרות אלפי שקלים).
התקבלו הממצאים הבאים:

Classification Table[a]

Observed			Predicted apartm .00	Predicted apartm 1.00	Percentage Correct
Step 1	apartm	.00	22	11	66.7
		1.00	10	22	68.8
	Overall Percentage				67.7

a. The cut value is .500

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	10.218	4	.037
	Block	10.218	4	.037
	Model	10.218	4	.037

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	79.876[a]	.145	.194

a. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Status			.682	3	.877	
	Status(1)	-.498	.713	.487	1	.485	.608
	Status(2)	-.520	.784	.441	1	.507	.594
	Status(3)	-.180	.748	.058	1	.810	.835
	income	.000	.000	8.580	1	.003	2.536
	Constant	-2.734	1.079	6.417	1	.011	.065

a. Variable(s) entered on step 1: Status, income.

א. האם ניתן לדחות את השערת האפס הטוענת כי אין קשר בין בעלות על דירה להכנסה ולסטטוס משפחתי?
ב. כמה אחוזים מצליחים המשתנים הב"ת להסביר מהשונות של המשתנה "בעלות על דירה"?
ג. באיזה אחוז מצליח המודל לנבא באופן מדויק בעלות על דירה מתוך כלל המקרים?
ד. באיזה מידה מצליח המודל לנבא בהצלחה בעלות על דירה מתוך בעלי הדירה במדגם? כיצד נקרא המדד המתאים?
ה. באיזה מידה מצליח המודל לנבא בהצלחה אי-בעלות על דירה מתוך אלו שאינם בעלי דירה במדגם? כיצד נקרא המדד המתאים?
ו. מהי המשוואה לניבוי בעלות על דירה על סמך המשתנים הב"ת?
ז. לאיזה מהמשתנים הב"ת יש תרומה ייחודית מובהקת לניבוי בעלות על דירה? מהי משמעות מקדם B ו-Exp(B) של משתנה זה?
ח. על כל עליה ב-10,000 ₪ בהכנסה, בכמה אחוזים יעלה הסיכוי לבעלות על דירה?

i. 53.6%.

ii. 253.6%.

iii. 153.6%.

iv. 93%.

ט. על כל עליה של 20,000 ₪ בהכנסה, בכמה אחוזים יעלה הסיכוי לבעלות על דירה?

i. 307%.

ii. 423%.

iii. 542%.

iv. 642%.

י. מה ההסתברות של רווק המשתכר 20,000 ₪ להיות בעלים של דירה?
יא. האם ההסתברות של אותו רווק להיות בעל דירה גבוהה / שווה / קטנה מההסתברות שלו לא להיות בעל דירה?
יב. מהם הסיכויים (ODDS) שלו להיות בעל דירה?
יג. עבור איזה משכורת הסיכוי (הסתברות) של רווק להיות בעל דירה עולה על הסיכוי שלו לא להיות בעל דירה?
יד. במידה ומשתנה ההכנסה היה נמדד באלפי שקלים (ולא בעשרות אלפי שקלים), כיצד הדבר היה משפיע על ההשפעה השולית של מקדם ההכנסה, אם בכלל?

3) חוקרים בחנו את המאפיינים שעשויים לנבא את הביצוע של חניכים במבחן הסיום של קורס פקחי טיסה.
הביצוע במבחן נמדד על סולם של הצלחה/כשלון והמשתנים הבלתי תלויים כללו מין (1-זכר 0-נקבה), השכלה קודמת (0- ריאלית, 1- לא ריאלית) וביצוע במהלך הקורס (1-7).
להלן תוצאות ניתוח הרגרסיה:

Omnibus Tests of Model Coefficients

		Chi-square	Df	Sig.
Step 1	Step	20.982	3	.000
	Block	20.982	3	.000
	Model	20.982	3	.000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	17.209[a]	.503	.699

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	מין	4.445	2.611	2.897	1	.089	85.161
	השכלה קודמת	-.146	2.054	.005	1	.943	.864
	ביצוע במהלך הקורס	2.283	.944	5.846	1	.016	9.810
	Constant	-19.284	8.056	5.731	1	.017	.000

a. Variable(s) entered on step 1: מין, השכלה, ביצוע במהלך הקורס

א. האם למודל הכולל את שלושת המנבאים יכולת הסבר משמעותית?
ב. כמה אחוזים מתוך השונות של Y מצליח המודל להסביר?
ג. מהי משוואת הניבוי?
ד. לאיזה מן המשתנים הב"ת תרומה מובהקת לניבוי?
ה. הסבירו את משמעות המקדמים (b) שהתקבלו עבור המשתנים הב"ת: מגדר, השכלה קודמת והביצוע במהלך הקורס.
ו. בטאו את המקדמים במונחי הסיכויים להצלחה בקורס (odds) והסבירו אותם.
ז. הועלתה הטענה כי ההסתברות ההצלחה של נשים בקורס היא נמוכה ביותר, גם אם הן בעלות השכלה ריאלית ושביצוען במהלך הקורס מקסימאלי. אנא בדקו את הטענה.
ח. עבור זכר, בעל השכלה ריאלית, מהי ההשפעה השולית של עליה ביחידה אחת בדירוג הביצוע במהלך הקורס על הסיכוי להצליח בקורס?

4) לפי מדגם של 20 זוגות נשואים, נאספו נתונים על המשתנה Y השווה ל-1 אם הזוג נוהג לצאת למסעדה לפחות פעם בשבוע ו-0 אחרת.

נאמד המודל: $p=\frac{1}{1+e^{-z}}$ כאשר $p=P(Y=1)$.

התקבלו התוצאות הבאות: $z=-9.456+0.368INCOM-1.207BABY$.

INCOM - ההכנסה של שני בני הזוג (באלפים). ההכנסה במדגם נעה בין 17 אלף ל- 44 אלף.

BABY - משתנה דמי המקבל את הערך '1' אם הזוג צריך להיעזר בשמרטפית ו-'0' אחרת.

ענה נכון/לא נכון:

א. זוג הנעזר בשמרטפית ומשתכר 30.5 אלף, יוצא למסעדה לפחות פעם בשבוע בהסתברות גבוהה מ-0.5.

ב. עבור זוג שאינו נעזר בשמרטפית, עליה של אלף שח בהכנסה, מעלה את ההסתברות לצאת למסעדה ב-0.368.

ג. כל אחד מערכי P הנאמדים כאן איננו גבוה יותר מ-0.99.

ד. הסיכוי של זוג, שהכנסתו עלתה ב-3000 ₪, לצאת למסעדה יעלה ב-200% בערך.

ה. המשכורת שצריך להרוויח זוג, אשר אינו נעזר בשמרטפית, כדי שהסיכוי שלו לצאת למסעדה יהיה שווה לסיכוי שלא לצאת למסעדה הוא 27,000.

ו. זוג, שלא נעזר בשמרטפית, צריך להרוויח יותר מ-28,000 ₪ כדי שהסיכוי שלו לצאת למסעדה יהיה גבוה פי 3 מהסיכוי שלו לא לצאת למסעדה.

ז. עבור odds ratio של משתנה "שמרטפות" התקבל רווח בר סמך הבא: $[0.123\ ;\ 1.01]$ ברמת ביטחון של 95%.
לפיכך ניתן לומר כי למשתנה "שמרטפות" תרומה מובהקת לניבוי הסיכוי לצאת למסעדה.

5) בשנה מסוימת הוגשו 750 בקשות לקבלת משכנתא ורק חלק מהן אושר.
המשתנה התלוי Y=1 אם הבקשה למשכנתא אושרה ול-0 אם נדחתה.
המנבאים:
S משתנה דמי השווה ל-1 אם מבקש המשכנתא הוא רווק ול-0 אחרת.
AGE = גיל בשנים.

המודל הנאמד הינו: $p=\frac{1}{1+e^{-z}}$ כאשר $p=P(Y=1)$.

$z=\alpha+\beta_1 age+\beta_2 age^2+\beta_3 S$.

תוצאות אמידת המודל: $z=-9.3+0.52age-0.006age^2-0.314S$

א. הסבירו את השפעת הגיל והמצב המשפחתי על ההסתברות לאישור המשכנתא.

ב. מה ההסתברות שתאושר משכנתא לרווק בן 30?

ג. עבור איזה גיל ההסתברות של אדם נשוי לקבל משכנתא היא מקסימאלית?

6) משרד הקבלה של האוניברסיטה רצה לבדוק באיזה מידה ניתן לחזות את ההצלחה של הסטודנט בקורס בסטטיסטיקה על סמך נתונים של מבחן פסיכומטרי, ציון ממוצע של תעודת בגרות וסוג תעודת הבגרות: ריאלית או לא ריאלית.
במדגם של 50 סטודנטים נאספו נתונים על המשתנה Y השווה ל-1 אם הסטודנט הצליח במבחן בסטטיסטיקה ו-0 אם נכשל.
כמו כן נרשמו עבור כל סטודנט ציון הפסיכומטרי, ממוצע הבגרות וסוג הבגרות (1 - בגרות ריאלית, 0 - לא ריאלית).
להלן התוצאות שהתקבלו:

	B	S.E.	Wald	df	Sig.
פסיכומטרי	.090	.046	3.723	1	.054
ציון בגרות		2.070	1.089	1	.297
בגרות_ריאלי	4.535	2.519	3.241	1	.072
Constant	-84.892	42.858	3.923	1	.048

א. באיזה שיטת ניתוח הייתם ממליצים להשתמש ומדוע?
ב. נתון כי ההסתברות להצליח בקורס בסטטיסטיקה עבור סטודנט שעשה בגרות הומנית, קיבל 690 בפסיכומטרי וציון 9 בבגרות הינה: 0.034.
ההסתברות של סטודנט שקיבל אותו ציון בפסיכומטרי, עם בגרות הומנית אבל ציונו בבגרות הוא 10 הינה: 0.233
על סמך הנתונים הללו השלם את הערך החסר בפלט המקדמים.
ג. לאיזה משתנים השפעה מובהקת על הסיכוי להצליח במבחן לסטטיסטיקה? (אלפא 10%)
ד. מה ההסתברות של סטודנט להצליח במבחן אם קיבל 680 בפסיכומטרי, ציון 10 בבגרות ולמד במגמה ריאלית?
ה. מהו השינוי בסיכויים (odds) להצליח במבחן בסטטיסטיקה כפונקציה של שינוי ביחידה אחת בפסיכומטרי?
ו. מהי ההשפעה השולית של נקודה נוספת בציון הבגרות על הסיכוי להצליח במבחן בסטטיסטיקה עבור סטודנט שקיבל 640 בפסיכומטרי ולמד במגמה ריאלית?
ז. רותי שיפרה את הפסיכומטרי שלה ב-20 נקודות.
בכמה יעלה הסיכוי שלה להצליח בקורס בסטטיסטיקה?
ח. אם החוקר היה מחליט לקודד בגרות שאינה ריאלית כ-1 ובגרות ריאלית כ-0, האם הדבר היה משפיע על ערכו של Exp(b) של סוג בגרות ועל המשמעות שלו?

תשובות סופיות:

1) א. 0.73. ב. 2.7. ג. 0.96. ד. 24. ה. 0.11.
ו. -2.207. ז. $B=-2.207$. ח. $\text{Exp}(B)=0.11$.

2) א. כן. ב. 19.4%. ג. 67.7%. ד. רגישות = 68.8%.
ה. סגוליות = 66.7%.
ו. $In(odds)=-2.734-0.498status(1)-0.52status(2)-0.18status(3)+0.93\cdot incom$.
ז. משתנה "הכנסה". ח. 3. ט. 3. י. 0.3.
יא. קטנה. יב. 0.42. יג. 29,400. יד. 0.093.

3) א. כן. ב. 69.9%. ג. $\log(odds)=-19.284+4.445x_{1i}-0.146x_{2i}+2.283x_{3i}$.
ד. המשתנה – "ביצוע במהלך הקורס". ה. ראו סרטון.
ו. מגדר - $\text{Exp}(b)=85.19$, השכלה קודמת - $\text{Exp}(b)=0.864$,
ביצוע במהלך הקורס - $\text{Exp}(b)=9.81$. ז. הטענה נכונה $(p=0.035)$.
ח. $\text{Exp}(b)=9.81$.

4) א. נכון. ב. לא נכון. ג. לא נכון. ד. נכון. ה. לא נכון.
ו. נכון. ז. לא נכון.

5) א. ראו סרטון. ב. 0.533. ג. 40.

6) א. רגרסיה לוגיסטית. ב. $B=2.16$. ג. "פסיכומטרי" ו-"בגרות ריאלית".
ד. 0.914. ה. 1.09. ו. 8.67. ז. 504%. ח. ראו סרטון.

אקונומטריקה למדעי הנתונים

פרק 33 - מבחן לדוגמה מס' 1

תוכן העניינים

מבחן לדוגמא מס' 1:

שאלות:

לשם החישובים כשצריך הנח כי (אם לא נאמר אחרת): $\chi^2_{(2)}=3.0$, $\chi^2_{(5)}=11.05$,
$F=4$, $t=2$.

1) חוקר בדק את השפעת השכר ההתחלתי של עובד, עמדה ניהולית ומגדר על השכר הנוכחי של העובד. לכן אמד את המודל הבא בשיטת הריבועים הפחותים (OLS):

$$y_t=\beta_0+\beta_1G_t+\beta_2M_t+\beta_3(G_tM_t)+\beta_5(X_t,G_t)+\beta_6(X_tM_t)+\beta_7(X_tG_tM_t)+\varepsilon_t.$$

Y - השכר הנוכחי של העובד (באלפי שקלים).

G - משתנה דמי למגדר. $G=1$ עבור גברים ו-$G=0$ עבור נשים.

M - משתנה דמי לעמדה ניהולית. $M=1$ מחזיק בעמדה ניהולית ו-$M=0$ לא מחזיק בעמדה ניהולית.

X - משתנה המתאר את השכר ההתחלתי של העובד (באלפי שקלים).

א. מהי ההשערה שבוחנת כי השכר ההתחלתי של העובד משפיע באופן זהה עבור גברים ונשים (כלומר, השפעת השכר ההתחלתי אצל גבר המחזיק בעמדה ניהולית שווה לזה של אישה המחזיקה בעמדה ניהולית והשפעת השכר ההתחלתי אצל גבר שאינו מחזיק בעמדה ניהולית זהה לזה של אישה שאינה מחזיקה בעמדה ניהולית):

i. $H_0:\beta_3=\beta_5=0$.

ii. $H_0:\beta_5=\beta_7=0$.

iii. $H_0:\beta_5=0$.

iv. $H_0:\beta_1=\beta_3=\beta_5=\beta_7=0$.

v. כל התשובות לעיל אינן נכונות.

ב. מהי ההשערה אשר בוחנת כי שכרו הנוכחי של גבר שלא מחזיק בעמדה ניהולית זהה לו של אישה המחזיקה בעמדה ניהולית כאשר לשניהם יש שכר התחלתי של 1000 ₪.

i. $H_0:\beta_5=\beta_6$.

ii. $H_0:\beta_1=\beta_5=\beta_2=\beta_6$.

iii. $H_0:\beta_1=\beta_2$.

iv. $H_0:\beta_1+\beta_5=\beta_2+\beta_6$.

v. כל התשובות לעיל אינן נכונות.

2) נתון המודל הבא: $y_t = \alpha + \beta_0 X_t + \beta_1 X_t^2 + u_t$ המקיים את כל ההנחות הקלאסיות פרט להומוסקדסטיות. ידוע כי מתקיים: $\operatorname{var}(u_t) = \frac{x_t^4}{2}$.

חוקר זיהה שקיימת בעיה של הומוסקדסטיות במודל זה במודל וביצע תיקון למודל כך שכעת לא קיימת הבעיה.

א. מי מבין האפשרויות הבאות יכול להיות החותך במודל המתוקן:

i. β_0.

ii. β_1.

iii. x_t.

iv. 2.

v. כל התשובות לעיל אינן נכונות.

ב. מהי השונות במודל החדש:

i. 2.

ii. $\frac{1}{2}$.

iii. x_t^4.

iv. $\frac{x_t^2}{2}$.

v. לא ניתן לדעת שכן יש בעיית מולטיקוליניאריות מלאה.

3) חוקר מעוניין לאמוד את המודל הבא: $y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 z_t + u_t$.

עקב חשש להטרוסקדסטיות חילק החוקר את המדגם לשלושה חלקים כאשר:

חלק ראשון בן 60 תצפיות כאשר התוצאות שהתקבלו: $ESS_1 = 1400$.

חלק שני בן 40 תצפיות כאשר התוצאות שהתקבלו: $ESS_2 = 900$.

סך כל התצפיות במדגם היו 120.

א. תוצאות הסטטיסטי F=GQ (Goldfeld-Quandt) ותוצאת בדיקת הטרוסקדסטיות היא (הניחו כי F קריטי במבחן זה הוא 2):

i. $GQ = 1$ קיימת הטרוסקדסטיות.

ii. $GQ = 1.55$ לא קיימת הטרוסקדסטיות.

iii. $GQ = 1$ לא קיימת הטרוסקדסטיות.

iv. $GQ = 1.55$ קיימת הטרוסקדסטיות.

v. כל התשובות לעיל אינן נכונות.

ב. אם החוקר היה מבצע את מבחן WHITE לבדיקת קיום הטרוסקדסטיות בנתונים, מספר דרגות החופש של המבחן הוא:
i. 2.
ii. 4.
iii. 5.
iv. 6.
v. אף אחת מהתשובות לעיל איננה נכונה.

4) נתון המודל הבא: $y_t = \alpha + \beta_0 X_t + \beta_1 X_t^2 + u_t$ שנאמד על סמך מדגם של 100 נבדקים. נתון בנוסף כי סטטיסטי המבחן של DW שווה ל-1.5. חוקר חשד שיש מתאם סידרתי מסדר ראשון במודל.

א. איזה מבין האפשרויות הנ"ל נכונה:
i. יש מתאם סידרתי שלילי מובהק בנתונים.
ii. יש מתאם סידרתי חיובי מובהק בנתונים.
iii. אין עדות למתאם סידרתי בנתונים.
iv. לא ניתן לדעת האם יש מתאם סדרתי מובהק בנתונים.
v. כל התשובות לעיל אינן נכונות.

ב. בהנחה כי קיים מתאם סידרתי מובהק בנתונים מהצורה הבאה: $u_t = \rho u_{t-1} + \varepsilon_t$. האומד ל- ρ הוא:
i. -0.25.
ii. 1.5.
iii. -1.5.
iv. 0.25.
v. אף אחת מהתשובות לעיל איננה נכונה.

ג. ידוע כי: $\operatorname{var}(u_t) = 2$. מהו האומד ל- σ_ε^2:
i. 2.133.
ii. 1.875.
iii. 1.5.
iv. 2.66.
v. אף אחת מהתשובות לעיל איננה נכונה.

ד. בשל החשש ממתאם סידרתי מסדר שני $(u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t)$ הוצע לבצע את מבחן LM. רגרסיית העזר לביצוע המבחן הינה:
i. $\hat{u}_t = \alpha + \beta_0 X_t + \beta_1 X_t^2 + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t$.
ii. $\hat{u}_t = \alpha + \beta_0 X_t + \beta_1 X_t^2 + \beta_3 \hat{u}_{t-1} + \beta_4 \hat{u}_{t-2} + \varepsilon_t$.
iii. $\hat{u}_t^2 = \alpha + \beta_0 X_t + \beta_1 X_t^2 + \beta_3 \hat{u}_{t-1} + \beta_4 \hat{u}_{t-2} + \varepsilon_t$.
iv. $\hat{u}_t^2 = \alpha + \beta_0 X_t + \beta_1 X_t^2 + \rho_1 \hat{u}_{t-1} + \rho_2 \hat{u}_{t-2} + \varepsilon_t$.

v. אף אחת מהתשובות לעיל אינה נכונה.

ה. אם ידוע כי קיים מתאם סדרתי בנתונים ואומדים אותו בשיטת הריבועים הפחותים (OLS) ללא ביצוע תיקון אזי:
i. האומדים יהיו מוטים, לא עקיבים ולא יעילים.
ii. האומדים יהיו חסרי הטיה, עקיבים ויעילים.
iii. האומדים יהיו חסרי הטיה, עקיבים אך לא יעילים.
iv. האומדים יהיו מוטים אך עקיבים.
v. אף אחת מהתשובות לעיל איננה נכונה.

5) מאמן כושר אמד את הרגרסיה הבאה: $y_t = 1 + 3x_t + 0.7y_{t-1} + \hat{u}_t$. כאשר:
y_t - סך ק"ג ירידה במשקל בחודש t.
x_t - סך שעות האימונים של אדם בחודש t.

א. מהן תכונות האומדים של המשוואה:
i. האומדים מוטים אך עקיבים.
ii. האומדים חסרי הטיה, עקיבים אך אינם יעילים.
iii. האומדים מוטים ואינם עקיבים.
iv. האומדים חסרי הטיה אך אינם עקיבים.
v. אף אחת מהתשובות לעיל אינה נכונה.

ב. אם רמת אימוני הכושר לפני ארבעה חודשים עלו ב-3 שעות, כיצד צפויה להשתנות רמת הירידה במשקל היום:
i. 2.16.
ii. 3.15.
iii. 0.2401.
iv. 3.087.
v. כל התשובות לעיל אינן נכונות.

ג. אילו אדם זה מעולם לא התעמל, מה צפויה להיות רמת הירידה במשקל שלו היום:
i. 1.4.
ii. 10.
iii. 3.3.
iv. כל התשובות לעיל אינן נכונות.

ד. אם רמת האימונים עלתה היום בשעה, מהי סך ההשפעה הצפויה על רמת הירידה במשקל?
i. 10.
ii. 3.
iii. 2.1.

iv. אף אחת מהתשובות לעיל איננה נכונה.

6) נתונה מערכת המשוואות הנ"ל:

1. $y_i = \alpha_0 + \alpha_1 x_i + \alpha_2 z_i + u_i$.
2. $x_i = \beta_0 + \beta_1 s_i + \beta_2 z_i + v_i$.
3. $z_i = \delta_0 + \delta_1 y_i + \delta_2 k_i + \varepsilon_i$.

א. מה ניתן לומר על תכונות אומדי הריבועים הפחותים של שלושת המשוואות הנתונות:

i. אומדי OLS של המשוואות הם חסרי הטיה עקיבים אך אינם יעילים.

ii. אומדי OLS של המשוואות הם חסרי הטיה, עקיבים ויעילים.

iii. אומדי OLS של המשוואות הם מוטים, לא עקיבים ולא יעילים.

iv. אומדי OLS של המשוואה הראשונה בלבד הם חסרי הטיה, עקיבים ויעילים.

v. אף אחת מהתשובות לעיל איננה נכונה.

ב. איזה משיטות האמידה: ILS (שיטת הריבועים הפחותים העקיפה), TSLS (שיטת הריבועים הפחותים בשני שלבים) ו-IV (שיטת משתני העזר) מתאימה לכל אחת משלושת המשוואות הנתונות:

i. שלושת המשוואות ניתנות לאמידה בכל שלושת השיטות.

ii. רק המשוואה הראשונה ניתנת לאמידה בכל שלושת השיטות. המשוואות השנייה והשלישית ניתנות לאמידה בשיטת ILS ו-IV בלבד.

iii. שלושת המשוואות אינן ניתנות לאמידה באף אחת מהשיטות.

iv. שלושת המשוואות ניתנות לאמידה בשיטת ILS ו-IV בלבד.

v. אף אחת מהתשובות לעיל איננה נכונה.

ג. לשם אמידת הפרמטרים של המשוואות המבניות על פי שיטת ILS נאמדו המשוואות המצומצמות עבור Y ו-Z והתקבל ש:

$$y_i = 2 + 3k_i$$
$$z_i = 4 - 2k_i$$

i. האומד ל- α_0 שווה ל-2 ואילו האומד ל- δ_0 שווה ל-4.

ii. האומד ל- α_0 שווה ל-8 ואילו האומד ל- α_2 שווה ל-1.5-.

iii. האומד ל- α_0 שווה לאומד ל- β_0 .

iv. לא ניתן לאמוד את α_0 ואת α_2 .

v. כל התשובות לעיל אינן נכונות.

תשובות סופיות:

1)	א. ii.	ב. iv.			
2)	א. ii.	ב. ii.			
3)	א. iii.	ב. iii.			
4)	א. ii.	ב. iv.	ג. ii.	ד. ii.	ה. iii.
5)	א. v.	ב. i.	ג. iii.	ד. i.	
6)	א. iii.	ב. i.	ג. ii.		

אקונומטריקה למדעי הנתונים

פרק 34 - מבחן לדוגמה מס' 2

תוכן העניינים

מבחן לדוגמא מס' 2:

שאלות:

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

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

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

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

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

ד. האם אומד זה הוא blue?

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

ו. חשבן את השונות של β עבור מודל שבו $\alpha \neq 0$.

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

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

1. $\hat{y}_i = 15 + 0.7x_{1i} + 0.35x_{2i} + 0.20x_{3i}$ $R^2 = 0.85$.
2. $y_i = 2 + 0.1z_i$ $R^2 = 0.25$.
3. $z_i = x_{1i} - x_{2i} + 2x_{3i}$.

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

א. ההשערה שניתן לבדוק באמצעות המשוואות הנתונות הינה:

i. $H_0: \beta_1 = \beta_2; \beta_1 = 2\beta_3$.

ii. $H_0: \beta_1 = -\beta_2 = 2\beta_3$.

iii. $H_0: \beta_2 = -\beta_1; \beta_3 = 2\beta_1$.

iv. לא ניתן לדעת.

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

i. 56.

ii. 57.

iii. 112.

iv. 74.66.

3) כלכלן הציע את המודלים הבאים :

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

האם ניתן לאמוד את המודלים בשיטת OLS?

א. אין בעיה לאמוד את שני המודלים.
ב. לא ניתן לאמוד את המודל הראשון בלבד.
ג. לא ניתן לאמוד את המודל השני בלבד.
ד. לא ניתן לאמוד את שני המודלים.

4) כלכלן אמד את המודל הבא : $y_i = \alpha_0 + \alpha_1 \ln(x_i) + u_i$ וקיבל את האומדנים :
$\hat{\alpha}_0 = 10$ ו-$\hat{\alpha}_1 = 6$.

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

א. $\hat{\beta}_0 = 5$ ו-$\hat{\beta}_1 = 3$.
ב. $\hat{\beta}_0 = 10$ ו-$\hat{\beta}_1 = 3$.
ג. $\hat{\beta}_0 = 5$ ו-$\hat{\beta}_1 = 6$.
ד. כל התשובות שגויות.

5) על סמך מדגם של 95 תצפיות נאמד המודל הבא : $y_i = 2 + 0.5x_{1i} + 0.3x_{2i}$ $R^2 = 0.73$.
(2) (1)

הערכים שבסוגריים הם סטיות התקן של המקדמים.

א. בדוק האם המודל מובהק.
ב. בדוק האם מקדמי השיפוע מובהקים.
ג. מה תוכל להסיק מסעיפים א' ו-ב'?

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

1. $\hat{y}_i = 4 + 0.1x_{1i} + 0.8x_{2i}$ $R^2 = 0.84$.
2. $\hat{y}_i = 2 + 0.8x_{1i}$ $R^2 = 0.7$.
3. $\hat{y}_i = 7 + 0.23x_{2i}$ $R^2 = 0.25$.
4. $\hat{y}_i = 3 + 0.23z_i$ $R^2 = 0.55$.

כאשר x_{1i} ו-x_{2i} הם השכלת הבעל והאישה בהתאמה במשפחה i ו-y_i הכנסת משק בית i.

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

א. בדוק את ההשערה כי להשכלה אין השפעה על הכנסות המשפחה.
ב. איזה השערה ניתן לבדוק תוך שימוש במשוואות (1) ו-(4)? בדוק אותה.
ג. חשב את סטיית התקן של המקדם x_{1i} ברגרסיה (1).

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

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

ב. חשב את נוסחת שונותו של האומד.

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

1. $y_i = 5.4 + 1.2x_{2i} + 4.4x_{3i} + u_i \quad R^2 = 0.75$.
2. $y_i = 6.3 + 5.8x_{3i} + u_i \quad R^2 = 0.65$.
3. $y_i = 5.7 + 1.2x_{2i} + u_i \quad R^2 = 0.70$.
4. $y_i = 3.9 + 3.4\ln(x_{2i}) + u_i \quad R^2 = 0.56$.
5. $\ln(y_i) = 2.4 + 1.8x_{2i} + 2.7x_{3i}^2 + 4.2x_{4i}^2 + u_i$.
6. $y_i = 1.3 + 3.1x_{2i} + 0.5x_{3i} + 4.8x_{4i}^2 + 1.5x_{5i}^2 + u_i$.

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

ב. בדוק את ההשערות של משתנים X_2 ו- X_3 ביחד אין השפעה על Y במודל (1).

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

ד. ברגרסיה (1) נתונים כעת אומדי הטעויות הסטנדרטיות (סטיות התקן) של מקדמי X_2 ו- X_3 0.5 ו-2.5 בהתאמה. בדוק עבור כל אחד מהמקדמים הנ"ל האם מובהק ומה אפשר ללמוד מרגרסיה (1).

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

9) על מנת לאמוד השפעת מגדר ומצב משפחתי על השכר, נאמדה המשוואה הבאה:

$$WAGE = \alpha_0 + \alpha_1 \cdot GENDER + \alpha_2 \cdot FS + \alpha_3 \cdot (GENDER \cdot FS) + \beta_1 \cdot EDUC + \beta_2 \cdot AGE + U$$

כאשר:

$GENDER$ = מגדר: 1=גבר, 0=אישה.

FS = מצב משפחתי: 1=נשואים, 0=לא נשואים.

$EDUC$ = מס' שנות לימוד של העובד.

AGE = גיל העובד.

$WAGE$ = שכר העובד.

משוואה (1) נאמדה בפלט מס' 1.

בנוסף נאמד גם פלט מס' 2.

א. החוקרת הניחה כי פערי השכר, באים לידי ביטוי בשכר ההתחלתי בלבד: נכון/ לא נכון/ לא ניתן לדעת.

ב. החוקרת הניחה כי הפערים בין נשים לגברים בשכר אינם תלויים בגיל: נכון/ לא נכון/ לא ניתן לדעת.

ג. השערת האפס לבדיקת הטענה היא: ________________.

ד. המשתנה המוסבר ברגרסיה מס' 2 הינו : ___________________.
(כתבו את המודל שבו מחושב המשתנה המוסבר).

ה. הסטטיסטי של LM לבדיקת הטענה :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ___________________.

ו. המקדם של GENDER בפלט מס' 2 הינו : __________.

הועלתה הטענה כי הפערים בין גברים לנשים בקרב העובדים הנשואים גבוהים ביותר מ-1500 ₪ מאשר הפערים בקרב העובדים שאינם נשואים.

ז. ההשערות לבדיקת הטענה הינן: ___________________.

ח. הסטטיסטי לבדיקת הטענה :

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: _______________.

ט. ההשערות לבדיקת הטענה הן: __________________.

י. המודל המוגבל לבדיקת הטענה הוא: ________________.

יא. הסטטיסטי לבדיקת הטענה :

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחישובו וערכו הוא: __________.

פלט מס' 1 - משוואה 1:

Dependent variable: WAGE
Number of observations used: 17495
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	1.504815E11	30096294654	646.42	<.0001
Error	17489	8.142567E11	46556220		
C Total	17494	9.647382E11			

Root MSE	6823.35843	R-square	0.1560
Dep Mean	7286.58004	Adj R-sq	0.1557
C.V.	93.64281		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-3642.10108	260.72351	-13.97	<.0001
GENDER	1	2006.13583	187.64224	10.69	<.0001
FS	1	899.68055	159.19316	5.65	<.0001
GENDER*FS	1	1964.31810	227.43348	8.64	<.0001
EDUC	1	428.20041	12.45434	34.38	<.0001
AGE	1	64.72379	4.43948	14.58	<.0001

פלט מס' 2 - מבחן LM:

Dependent variable :
Number of observations used: 17495
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	66653745252	13330749050	286.32	<.0001
Error	17489	8.142567E11	46558220		
C Total	17494	8.809105E11			

Root MSE	6823.35843	R-square	0.0757
Dep Mean	2.29222E-12	Adj R-sq	0.0754
C.V.	2.97675E17		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-1244.40187	260.72351	-4.77	<.0001
GENDER	1				
FS	1	899.68055	159.19316	5.65	<.0001
GENDER*FS	1	1964.31810	227.43348	8.64	<.0001
EDUC	1	23.18457	12.45434	1.86	0.0627
AGE	1	-38.13257	4.43948	-8.59	<.0001

10) ענה על הסעיפים הבאים:

א. במודל דינאמי עם מתאם סדרתי האומדים מוטים אך עקיבים: נכון/לא נכון/לא ניתן לדעת

ב. בהשמטת משתנה רלוונטי, אם נתון שהמתאם בין המשתנים המסבירים חיובי, והשפעת המשתנה שהושמט שלילית, אזי האומד ל- β במודל הנאמד בעל הטיה חיובית: נכון/לא נכון/לא ניתן לדעת

ג. המודל הבא נאמד במדגם של 1000 תצפיות והתקבלו התוצאות הבאות:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U_t$$

(0.2) (0.8) (0.6)

$R^2 = 0.4$.

(המספרים בסוגריים הם ערכי t של המקדמים).

לאור הנתונים שהתקבלו נראה כי במשוואה:

i. אין כל בעיה סטטיסטית ולכן האומדים המתקבלים הם חסרי הטיה.

ii. יש בעיה של: ____________, אך האומדים המתקבלים עדיין חסרי הטיה.

iii. יש בעיה של: ____________ והאומדים המתקבלים מוטים.

ד. המודל הבא נאמד במדגם של 1000 תצפיות והתקבלו התוצאות הבאות:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + U_t$$
$$(-6.3) \quad (0.2) \quad (7.8)$$

לאור הנתונים שהתקבלו נראה כי במשוואה:

i. אין כל בעיה סטטיסטית ולכן האומדים המתקבלים הם חסרי הטיה.

ii. יש בעיה של: ______________, אך האומדים המתקבלים עדיין חסרי הטיה.

iii. יש בעיה של: _______________ והאומדים המתקבלים מוטים.

11) נתון המודל: $Y = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t$.

א. אם קיים מתאם סדרתי מסדר ראשון, אומדי הריבועים הפחותים חסרי הטיה ועקיבים אך לא יעילים: נכון/לא נכון/לא ניתן לדעת

ב. הוספת המשתנה Y_{t-1} למשוואה, יכולה לפתור את בעיית המתאם הסדרתי מסדר ראשון: נכון/לא נכון/לא ניתן לדעת

ג. רגרסיית עזר למתאם סדרתי מסדר שלישי הינה: ____________.

ד. דרגות החופש של הערך הקריטי לבדיקת ההשערה בסעיף ג' הינן: ______.

12) נתון המודל:

1. $Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + U_t$.

א. האומדים המתקבלים בשיטת הריבועים הפחותים הם מוטים, אך עקיבים: נכון/לא נכון/אי אפשר לדעת

פתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה: $(Y_t \cdot W_t) = \delta_0 \cdot W_t + \delta_1 \cdot (X_{1t} \cdot W_t) + \delta_2 \cdot (X_{2t} \cdot W_t) + V_t$.

ב. כאשר W_t הינו: ________________.

ג. משוואה (2) בצורתה המפורשת הינה:

$$\_\_\_\_ = \delta_0 \cdot \_\_\_\_ + \delta_1 \cdot \_\_\_\_ + \delta_2 \cdot \_\_\_\_ + \_\_\_\_ .$$

מלאו את החסר במונחי המשתנים של משוואה (1).

ד. האומד היעיל ל- σ^2 הינו: _________.

תשובות סופיות:

1) א. כן. ב. מוטה. ג. אי אפשר לדעת. ד. לא.
ה. נכון. ו. $V\left(\tilde{\beta}\right)=\frac{n\sigma^2}{S_{XX}^2}$. ז. לא נכון.

2) א. iii. ב. iii.

3) ד'.

4) ב'.

5) א. מובהק. ב. אינם מובהקים. ג. ראו סרטון.

6) א. מובהק. ב. $H_0:\beta_2=2.2\beta_1$. ג. $S.E=0.00743$.

7) א. $\hat{\alpha}=\bar{y}$. ב. $V\left(\hat{\alpha}\right)=\frac{\sigma_e^2}{n}$.

8) א. $6>1>3>2>4$. ב. יש עדות לכך. ג. יש עדות לכך.
ד. X_2 מובהק, X_3 אינו מובהק. ה. $HO:\beta_2=\beta_4=\beta_5=0$.

9) א. נכון. ב. נכון. ג. $H_0:\alpha_1=\alpha_2=\alpha_3=0$.
ד. ראו סרטון. ה. ראו סרטון. ו. ראו סרטון.
ז. $\begin{matrix} H_0:\alpha_3=1500 \\ H_1:\alpha_3>1500 \end{matrix}$. ח. ii, $t_{stat}=2.04$. ט. $\begin{matrix} H_0:\alpha_1=\alpha_2 \\ H_1:\alpha_1\neq\alpha_2 \end{matrix}$.
י. $WAGE=\alpha_0+\alpha_2\cdot\left(GENDER+FS\right)+\alpha_3\cdot\left(GENDER\cdot FS\right)+\beta_1\cdot EDUC+\beta_2\cdot AGE+U$.
יא. i.

10) א. לא נכון. ב. לא נכון. ג. ii, בעיה של מולטיקולינאריות חלקית.
ד. ii, בעיה של הוספת משתנה לא רלוונטית.

11) א. נכון. ב. נכון. ג. $\hat{u}_t / X_{1t}, X_{2t};\hat{u}_{t-1},\hat{u}_{t-2},\hat{u}_{t-3}$.
ד. 3.

12) א. לא נכון. ב. $W_t=\frac{X_{2t}}{X_{1t}}$.
ג. $\frac{Y_t\cdot X_{2t}}{X_{1t}}=\delta_0\cdot\frac{X_{2t}}{X_{1t}}+\delta_1\cdot X_{2t}+\delta_2\cdot\frac{X_{2t}^2}{X_{1t}}+\frac{u_tX_{2t}}{X_{1t}}$. ד. $S^2=\frac{ESS}{T-k}$.