# מבוא לאקונומטריקה

# תוכן העניינים

# תוכן העניינים

# מבוא לאקונומטריקה

## 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 ( $\mu_X$ או $E(X)$).
**השונות** – מייצגת את מידת הפיזור של ההתפלגות ( $\sigma_X^2$ או $V(X)$).
**סטית התקן** – היא השורש של השונות ( $\sigma_X$).
**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:

חוקי הסיגמה:

1. $\sum_{t=1}^{T}X_t=X_1+X_2+X_3+\ldots+X_T$.
2. סכום של קבוע: $\sum_{t=1}^{T}a=Ta$.
3. סכום של קבוע כפול משתנה = לקבוע כפול הסכום: $\sum_{t=1}^{T}aX_i=a\sum_{t=1}^{T}X_t$.
4. סכום של סכום/הפרש = לסכום/הפרש הסכומים: $\sum_{t=1}^{T}\left(X_t\pm Y\right)_t=\sum_{t=1}^{T}X_t\pm\sum_{t=1}^{T}Y_t$.
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 $\neq$ לכפל/חילוק התוחלות:
$$E(X\cdot Y)\neq E(X)\cdot E(Y)$$
$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}.$$
$$E\left(X^2\right)\neq\left[E(X)\right]^2$$
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 $\neq$ 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: $V(aX+b)=a^2V(X)$.

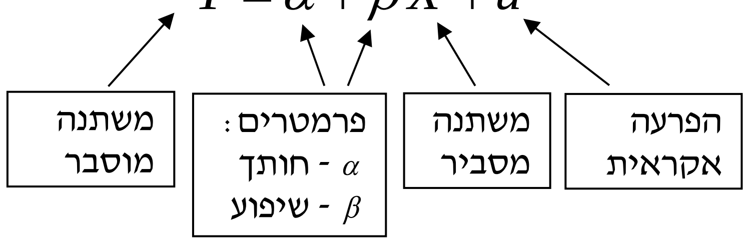
- חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).
חוקי הסכימה מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

<u>חוקי השונות המשותפת</u>:

1. שונות משותפת בין משתנה לקבוע = 0: $\operatorname{cov}(X,a)=0$.
2. שונות משותפת של משתנים המוכפלים בקבוע: $\operatorname{cov}(aX,bY)=ab\cdot\operatorname{cov}(X,Y)$.
3. שונות משותפת של משתנה עם עצמו = שונות המשתנה: $\operatorname{cov}(X,X)=V(X)$, $\operatorname{cov}(Y,Y)=V(Y)$.

**המודל האקונומטרי:**

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



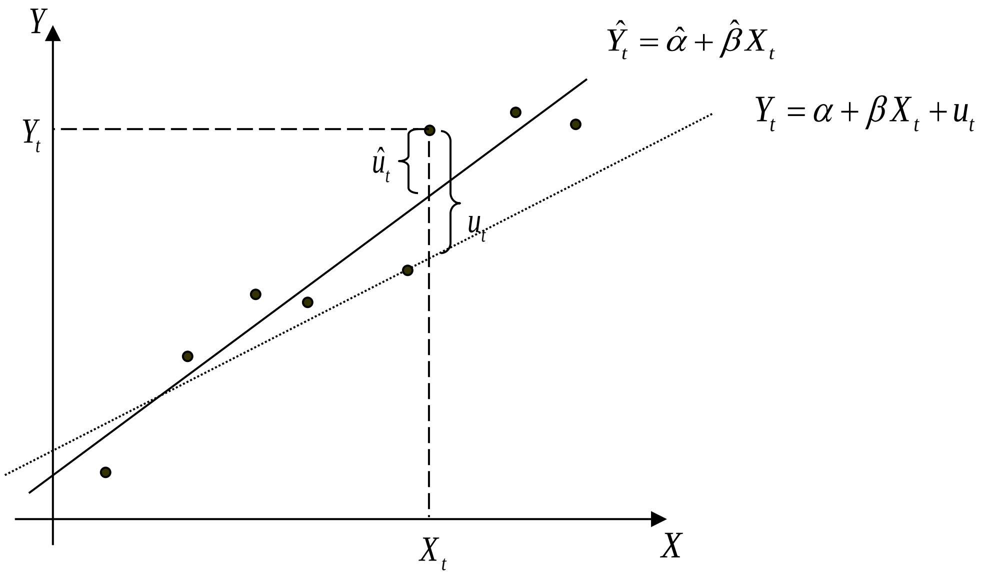

1. במודל: $Y_t=\alpha+\beta X_t+u_t$, $\alpha$ ו- $\beta$ הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).
2. $\hat{\alpha}$ הוא האומד ל- $\alpha$ ו- $\hat{\beta}$ הוא האומד ל- $\beta$.
3. אומדי ריבועים פחותים (אר"פ) הם האומדים שחושבו בשיטת הריבועים הפחותים. מסומנים בד"כ ע"י 'כובע' - $\hat{\beta}$.
אומדים אחרים מסומנים בד"כ ע"י 'תלתל' - $\tilde{\beta}$.

4. בעוד $\alpha$ ו-$\beta$ הם קבועים, $\hat{\alpha}$ ו-$\hat{\beta}$ הם משתנים מקריים כיוון שבכל מדגם מתקבלים $\hat{\alpha}$ ו-$\hat{\beta}$ אחרים.

5. את $\alpha$ ו-$\beta$ 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 $\left(\hat{Y}_t\right)$ המתקבל לפי הרגרסיה הוא: $\hat{Y}_t = \hat{\alpha} + \hat{\beta} X_t$.
- הסטיה של התצפית $(Y_t)$ מהערך הצפוי לפי הרגרסיה $\left(\hat{Y}_t\right)$ 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: $\text{var}(y),\text{var}(x),\text{cov}(x,y)$ 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. $a^2\,\text{var}(x)$. ב. $\text{var}(x)+\text{var}(y)+2\text{cov}(x,y)$. ג. $a^2\,\text{var}(x)$. ד. $a\,\text{cov}(x,y)$.

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

# מבוא לאקונומטריקה

## פרק 2 - אומדי הריבועים הפחותים

תוכן העניינים

# אומדי הריבועים הפחותים:

## רקע:

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

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

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

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

עבור מודל ללא חותך:

בגזירת $\beta$ 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 ללא שיפוע מתקיימת רק התכונה הגיאומטרית הראשונה. ברגרסיה ללא חותך 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 ליניאריים, חסרי הטיה, יעילים ועקיבים.

1. ליניאריות:
אר"פ ניתנים להצגה כטרנספורמציה ליניארית 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 $\vartheta$ 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 $\Leftarrow$ נשארים בתוך התוחלת, השונות וה- $\sum$.

$x_t$ איננו משתנה מקרי (על פי הנחה מס' 4) $\Leftarrow$ יוצא מחוץ לתוחלת ולשונות אך נשאר בתוך ה- $\sum$ ו- $\begin{matrix}\alpha\\ \beta\end{matrix}$ קבועים $\Leftarrow$ 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: $\begin{matrix}(\hat{\theta}\rightarrow\vartheta)\\(T\rightarrow\infty)\end{matrix}$.

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

ii. $E(u_i) = 0$

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

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

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

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 $\left(\operatorname{cov}(x_i, u_i) = 0\right)$. 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- $\beta$ 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 $\beta$ : $\tilde{\beta} = \frac{y_9 - y_5 + y_2}{x_9 - x_5 + x_2}$.

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

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

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

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

ד. האומד $\tilde{\beta}$ הינו אומד יעיל ל- $\beta$. 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: $\tilde{\beta} = \frac{\sum Y_t}{\sum X_t}$.

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

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

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.

# מבוא לאקונומטריקה

## פרק 3 - מודלים לא ליניאריים

### תוכן העניינים

# מודלים לא ליניאריים:

## רקע:

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

- 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 $\beta$ 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 אנגל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנגל?

**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: $\beta_1$ , $A$?

**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 $\beta$: $\tilde{\beta} = \frac{\sum_{t=1}^{T} \left(\ln X_t - \ln \bar{X}\right) \ln Y_t}{\sum_{t=1}^{T} \left(\ln X_t - \ln \bar{X}\right)^2}$.

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. $V(\hat{\beta})=\frac{\sigma_u^2}{SS\ln x}$.

# מבוא לאקונומטריקה

## 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.

$\alpha$ = 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 $\beta_j$: 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:

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

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

**מולטיקוליניאריות:**

מולטיקוליניאריות מתייחסת למתאם בין המשתנים המסבירים במודל.

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 $r_{12} \neq 1$.
- מולטיקוליניאריות מלאה יכולה להיווצר גם כאשר קבוצה של משתנים מסבירים מהווה קומבינציה ליניארית מלאה של אחד המשתנים המסבירים: $x_1 + x_2 = a + bx_3$.

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

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

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

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 $\beta_1$ 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.

# מבוא לאקונומטריקה

## פרק 5 - 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 <u>שווה</u> לסיכוי שהוא לא יתרחש ($\frac{0.5}{0.5}$).

ODDS>1 – הסיכוי שהאירוע יתרחש <u>גבוה</u> מהסיכוי שלא יתרחש (למשל- $\frac{0.75}{0.25}$).

ODDS<1 – הסיכוי שהאירוע יתרחש <u>נמוך</u> מהסיכוי שלא יתרחש (למשל- $\frac{0.25}{0.75}$).

ODDS RATIO (OR):
יחס בין סיכויים - $OR = \frac{ODDS_{(A)}}{ODDS_{(B)}}$

כיצד משתנה ההסתברות במעבר מקבוצה A לקבוצה B.

OR=1 – הסיכוי להתרחשות האירוע <u>שווה</u> בין שתי הקבוצות- אין קשר בין המב"ת למ"ת.

OR>1 – הסיכוי להתרחשות האירוע בקבוצה A <u>גבוה</u> מאשר בקבוצה B – קשר חיובי.

OR<1 – הסיכוי להתרחשות האירוע בקבוצה A <u>נמוך</u> 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 <u>העלייה</u> (אם גדול מ-1) או את <u>הירידה</u> (אם קטן מ-1) בסיכויים להיות בעלי ערך '1' ב-Y כאשר הערך במשתנה המנבא גדל ביחידה אחת.
5. <u>היחס בין B ל-Exp (B):</u> $e^{B} = Exp(B)$ ; $\log Exp(B) = B$

## 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.

# מבוא לאקונומטריקה

## פרק 6 - 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? | $H_0:\beta=0$<br>$H_1:\beta\neq 0$ | $t_{\beta=0}=\dfrac{\hat{\beta}-\beta_0}{S_{\hat{\beta}}}$ | שימוש בטבלת T:<br>$\left\|t_{\beta=0}\right\|>t_{\left(n-K,\frac{\alpha}{2}\right)}$<br>K=**מספר מקדמים (כולל חותך) | יש/אין עדות לכך שהמשתנה הב"ת מובהק באוכ' |
| **מבחן חד צדדי לשיפוע** | האם מקדם השיפוע חיובי/שלילי באוכ'? | $H_0:\beta=0$<br>$H_1:\beta>/<0$ | | שימוש בטבלת T:<br>$t_{\beta=0}>t_{(n-K,\alpha)}$<br>$t_{\beta=0}<-t_{(n-K,\alpha)}$ | יש/אין עדות לכך שהשיפוע חיובי/שלילי באוכ' |
| **השיפוע = ערך מסוים באוכ'** | האם מקדם השיפוע = לערך מסוים (למשל ל-2)? | $H_0:\beta=2$<br>$H_1:\beta\neq 2$ | $t_{\beta=2}=\dfrac{\hat{\beta}-2}{S_{\hat{\beta}}}$ | שימוש בטבלת t | יש/אין עדות לכך שהשיפוע = ל-2 באוכ'. |
| **מבחן למובהקות החותך<br>**ניתן לבצע גם מבחן חד צדדי ושהחותך = לערך מסוים באוכ'** | האם קו הרגרסיה יוצא מראשית הצירים? | $H_0:\alpha=0$<br>$H_1:\alpha\neq 0$ | $t_{\alpha=0}=\dfrac{\hat{\alpha}-\alpha_0}{S_{\hat{\alpha}}}$ | נדחה את $H_0$ אם:<br>שימוש בטבלת T:<br>$\left\|t_{\alpha=0}\right\|>t_{\left(n-K,\frac{\alpha}{2}\right)}$ | יש/אין עדות לכך שקו הרגרסיה עובר דרך ראשית הצירים |

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

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

- 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: $t_{\hat{\alpha}-5\hat{\beta}}=\dfrac{(\hat{\alpha}-5\hat{\beta})-0}{S_{\hat{\alpha}-5\hat{\beta}}}$ או $t_{\hat{\beta}_1-2\hat{\beta}_2}=\dfrac{(\hat{\beta}_1-2\hat{\beta}_2)-0}{S_{\hat{\beta}_1-2\hat{\beta}_2}}$.
כאשר את טעות התקן של המבחן מחשבים תוך שימוש בנוסחאות:

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

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

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

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 $\beta$ ושל $\alpha$?
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:
$\hat{y}_i = 5.06 + 0.97x_i + 3z_i - 5.02w_i + 8.97s_i$.
(0.29) (0.7) (0.08) (0.42) (0.456).

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:

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

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

**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. $P(-6.205 \le Q_i \le 7.295) = 0.95$.

# מבוא לאקונומטריקה

## פרק 7 - מבחן F ו R בריבוע

### תוכן העניינים

# מבחן F ו-R ברגרסיה:

## 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: $\dfrac{\left(\sum e_{\text{מ}}^2 - \sum e_{\text{לא מ}}^2\right)/m}{\sum e_{\text{לא מ}}^2 /(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

(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: $\dfrac{\left(R_{\text{לא מ}}^2 - R_{\text{מ}}^2\right)/m}{\left(1-R_{\text{לא מ}}^2\right)/(n-k)} \sim F_{(m,n-k,1-\alpha)}$.

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

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- $R_{מ}^2 = 0$ 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: $y = A\dfrac{x_{1i}^{\beta_1}}{x_{3i}^{\beta_3}} e^{\beta_2 x_2} e^{u_i}$.
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 תק"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. $Y_t-5=\beta_x X_t+\beta_z\left(Z_t+3S_t\right)+\beta_w W_t+u_t$.
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. $\ln\left(y_i\right)=\ln\left(A\right)+\beta_1\ln\left(X_{1i}\right)+\beta_2 X_{2i}+\beta_3 X_{3i}+u_i$.
ב. $\ln\left(y_i\right)=\ln\left(A\right)+\beta_3\left(\ln\left(X_{1i}\right)+3X_{2i}+X_{3i}\right)+u_i$.
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) $\ln\left(\dfrac{P_i}{S_i}\right)=\alpha+\beta_J\ln\left(\dfrac{J_i}{S_i}\right)+\varepsilon_i$.

# מבוא לאקונומטריקה

## פרק 8 - 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-** $X$ **:** $Y=\alpha'+\beta'\left(X+d\right)+v$ | $\hat{\beta}'=\hat{\beta}$ | $\hat{\alpha}'=\hat{\alpha}-\hat{\beta}d$ | $s_{\hat{\beta}'}=s_{\hat{\beta}}$ | $s_{\hat{\alpha}'}\neq s_{\hat{\alpha}}$ |
| **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:** $Y=\alpha'+\beta'\left(dX\right)+v$ | $\hat{\beta}'=\frac{\hat{\beta}}{d}$ | $\hat{\alpha}'=\hat{\alpha}$ | $s_{\hat{\beta}'}=\frac{s_{\hat{\beta}}}{d}$ | $s_{\hat{\alpha}'}=s_{\hat{\alpha}}$ |
| **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)** $S_{\hat{\alpha}'} = S_{\hat{\alpha}} = 0.0688935$ , $S_{\hat{\beta}'} = S_{\hat{\beta}} = 0.0026235$ , $\hat{\alpha}' = 7.11161$ , $\hat{\beta}' = \hat{\beta} = -0.00874$ ,
$R^2 = 0.0269$.

# מבוא לאקונומטריקה

## פרק 9 - המודל הריבועי

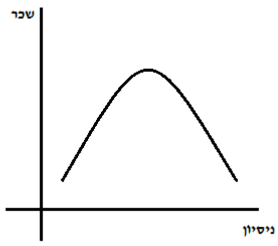
### תוכן העניינים

# 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 $\beta_2$, המקדם של $X^2$, הוא <u>חיובי</u>: מדובר <u>בנקודת מינימום</u> שעד אליה התרומה השולית שלילית וממנה ואילך חיובית.

אם $\beta_2$, 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. $S_{\hat{\beta}_2} = 0.002$ , $S_{\hat{\beta}_1} = 0.37$ , $\hat{Y}_i = 7.059 + 1.085X_i - 0.004X_i^2$ $R^2 = 0.53$.

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):

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

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 $\beta$ , 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.

# מבוא לאקונומטריקה

## פרק 10 - 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: $\tilde{\beta} = \dfrac{\sum Y_t}{S_{XX}}$.

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. 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- $\alpha$ 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. $HO:\beta_2=\beta_4=\beta_5=0$.

# מבוא לאקונומטריקה

## פרק 11 - 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 $\beta$: $\hat{\beta} = \sum_{i=1}^{n} \frac{\sqrt{y_i}\ln(x_i)}{\sum \ln(x_i)^2}$.

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 מולטיקולינאריות מושלמת ולכן הוא לא ניתן לאמידה בשיטת OLS.
2. במודל 2 יש מולטיקולינאריות מושלמת ולכן הוא לא ניתן לאמידה בשיטת OLS.
3. במודל 1 יש מולטיקולינאריות 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- $\beta_1$ והאם היא מובהקת?

i. $[1,3]$, לכן ה- $\beta_1$ מובהקת.

ii. $[1.5,2.5]$, לכן ה- $\beta_1$ מובהקת.

iii. $[1,3]$, לכן ה- $\beta_1$ לא מובהקת.

iv. $[1.5,2.5]$, לכן ה- $\beta_1$ 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 $T = 1$.

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: $Y_i = AX_{1i}^{\beta_1}X_{2i}^{\beta_2}X_{3i}^{\beta_3}e^{\beta_4X_{4i}^3}X_{5i}e^{u_i}$ שהורץ על המדגם בן 45 תצפיות.

מהי המשוואה לאמידת המקדמים של המודל?

א. $\ln(\frac{y_i}{x_5}) = \alpha + \beta_1\ln(x_{1i}) + \beta_2\ln(x_{2i}) + \beta_3\ln(x_{3i}) + \beta_4X_{4i}^3 + u_i$.

ב. $\ln(\frac{y_i}{K5}) = \alpha + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + 3x_{4i} + u_i$.

ג. $\ln(y_i) - \ln(x_{5i}) = \alpha + \beta_1\ln(x_{1i}) + \beta_2\ln(x_{2i}) + \beta_3x_{3i} + \beta_4x_{4i} + u_i$.

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: $Y_i = \alpha + \beta_1x_{1i} + \beta_2x_{2i} + \beta_3x_{3i} + u_i$.

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 $\frac{\left(\sum e_{\text{מ}}^2 - \sum_{\text{מל}} e^2\right)/m}{\sum e_{\text{מ}}^2/(n-k)}$.

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

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

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

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 $\beta_2$.

א. מובהקת. וסטיית התקן של $\beta_2$ היא 0.86 (בקירוב).
ב. אינה מובהקת. וסטיית התקן של $\beta_2$ היא 0.72 (בקירוב).
ג. אינה מובהקת. וסטיית התקן של $\beta_2$ היא 0.86 (בקירוב).
ד. מובהקת. וסטיית התקן של $\beta_2$ 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'.

# מבוא לאקונומטריקה

## פרק 12 - 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 $\beta$ לפי הנוסחה: $\hat{\beta} = \dfrac{\sum (Y_t^2 - X_{1t}^2)\ln X_{2t}}{\sum (\ln X_{2t})^2}$.

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. $\dfrac{\sigma^2(\ln X_{2t})}{\sum (\ln X_{2t})^2}$.
ב. $\dfrac{\sigma^2}{4X_t^2}$.
ג. $\dfrac{\sigma^2}{\sum (\ln X_{2t})^2}$.
ד. $\sigma^2 \dfrac{1}{2} \sum \dfrac{1}{X_t^2}$.
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-$\beta$ 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. $\hat{Y}_i = 4 + 2.8X_{1i} + 2X_{2i} \quad ESS = 200$.

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

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 $\beta$ על ידי האומד הבא: $\dfrac{Y_{10} - Y_1}{2X_{10} - 2X_1}$.

כלכלן ב' אמד את $\beta$ 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: $\dfrac{(R^2_{\text{מ}} - R^2_{\text{ל}})/m}{(1-R^2_{\text{מ}})/(n-k)}$.
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: $\dfrac{(R^2_{\text{מ}} - R^2_{\text{ל}})/m}{(1-R^2_{\text{מ}})/(n-k)}$ 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'.

# מבוא לאקונומטריקה

## פרק 13 - 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: $\hat{Y}\pm t_{n-2;1-\frac{\alpha}{2}}\hat{\sigma}_u\sqrt{\frac{1}{n}+\frac{(X_0-\bar{X})^2}{\sum(X_i-\bar{X})^2}}$.

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

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 $\beta$?

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 $\beta$ חיובי) ברמת מובהקות של 0.01.

ה. בנו רווח-סמך ברמת סמך של 95% עבור $\beta$.

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-$\beta$:

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

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

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

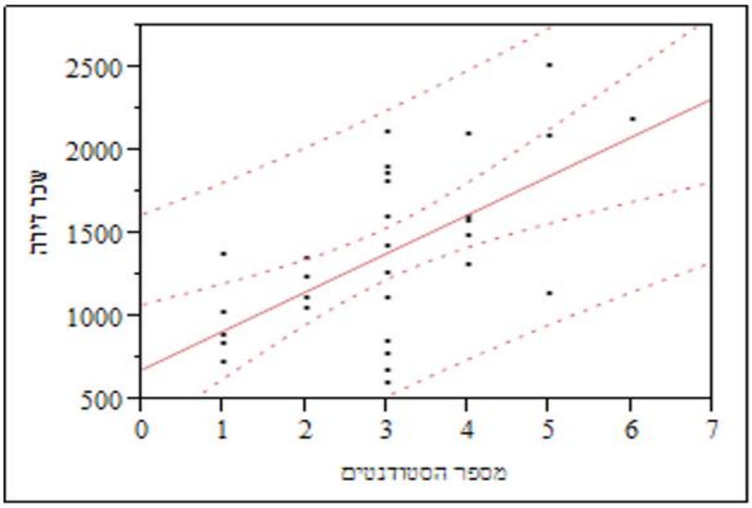
ANOVA[b]

| 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. $P(0.1488\le\beta\le0.1554)=0.95$.
ו. יש עדות לכך.
3) אין עדות לכך.
4) א. לא נכון. ב. -0.87%. ג. 7.24735.
5) א. 1153.247. ב. $p(957.4\le\mu_{Y_{X=2}}\le1349.08)=0.95$.
ג. $p(282.94\le Y_{X=2}\le2023.55)=0.95$.

# מבוא לאקונומטריקה

## פרק 14 - 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 $\alpha$ יש אחד.
- מספר ה- $\beta$ 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- $\beta$ טות:

מבחן לבדיקת מובהקות $\beta$ 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: $|t_{\hat{\beta}}|>1$.

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

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

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

## 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- $\alpha$ : נכון / לא נכון / לא ניתן לדעת.

ii. בעקבות ההוספה האומד ל- $\alpha$ יהיה מובהק: נכון / לא נכון / לא ניתן לדעת.

iii. ההוספה תשנה את האומד ל- $\beta$: 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:

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

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

# מבוא לאקונומטריקה

## פרק 15 - 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 $\beta_3$ 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- $\beta_1$ | אומד ל- $\alpha$ | אומד לשונות הפרמטרים |
|---|---|---|---|
| $S_{12} = 0$ | חסר הטיה | מוטה<br>אלא אם:<br>$\bar{x}_2 = 0$ | מוטה (כלפי מעלה) |
| $S_{12} \neq 0$ | מוטה<br>כיוון ההטיה:<br>חיובי: $S_{12}$ ו- $\beta_2$ שווי סימן<br>שלילי: $S_{12}$ ו- $\beta_2$ מנוגדי סימן | מוטה | |

# מבוא לאקונומטריקה

## פרק 16 - תיאוריה מולטיקוליניאריות

תוכן העניינים

# מולטיקולינאריות:

## רקע:

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

נבחין בין מולטיקולינאריות מלאה לחלקית.

### 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: $r_{12} \neq 1$.

במצב של מולטיקולינאריות 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.

### מולטיקולינאריות חלקית:

כאשר יש מתאם גבוה מאוד בין משתנים מסבירים במודל (אך לא מושלם) עלולה להיווצר בעיה של מולטיקולינאריות 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: $S^2_{\hat{\beta}_1} = \dfrac{MSE}{SSX_1(1-r_{12})}$ , $t = \dfrac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$.

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 מולטיקולינאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.

ב. אם דוחים את $H_0$: יש סתירה בין מבחן WALD למבחני $t$ - קיימת בעיה של מולטיקולינאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALD בלעדיו, עד שמזהים את המשתנה / משתנים שיש להוריד מהמודל.

# מבוא לאקונומטריקה

## פרק 17 - 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:<br>$Y_t=\alpha+\beta_1 x_{1t}+\varepsilon$<br>המודל הנאמד (הטעותי):<br>$Y_t=\alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+\varepsilon$</td>
<td>קבלת $H_0$ במבחן $t$ למובהקות $\beta_2$</td>
<td>ניתן לבצע בדיקת השערות<br>אר"פ $\left(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\right)$ חסרי הטיה<br>אומדי השונות $\left(S^2_{\hat{\alpha}},S^2_{\hat{\beta}_1},S^2_{\hat{\beta}_2}\right)$<br>חסרי הטיה</td>
<td>*הורדת המשתנה</td>
</tr>
<tr>
<td>השמטת משתנה רלוונטי</td>
<td>המודל האמיתי:<br>$Y_t=\alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+\varepsilon$<br>המודל הנאמד (הטעותי):<br>$Y_t=\alpha+\beta_1 x_{1t}+\varepsilon$</td>
<td>דחיית $H_0$ במבחן $t$ למובהקות $\beta_2$</td>
<td>
<table>
<tr><th>בהיעדר $x_2$:</th><th>אומד ל- $\beta_1$</th><th>אומד ל- $\alpha$</th><th>אומד לשונות הפרמטרים</th></tr>
<tr><td>$S_{12}=0$</td><td>חסר הטיה</td><td>מוטה<br>אלא אם:<br>$\bar{x}_2=0$</td><td>מוטה חיובית</td></tr>
<tr><td>$S_{12}\neq 0$</td><td>מוטה חיובית:<br>$S_{12}$ ו- $\beta_2$ שווי סימן<br>מוטה שלילית:<br>$S_{12}$ ו- $\beta_2$ 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<br>$Y_t=\alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+\varepsilon$<br>כאשר: $r_{12}=\pm 1$</td>
<td>אם:<br>$x_1=a+bx_2$<br>אז: $r_{12}=1$</td>
<td>לא ניתן לבצע בדיקת השערות<br>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<br>$Y_t=\alpha+\beta_1 x_{1t}+\beta_2 x_{2t}+\varepsilon$<br>כאשר: $0.7<\left|r_{12}\right|<1$</td>
<td>א. סתירה בין מבחן F ל-t<br>ב. רגישות לספציפיקציה<br>ג. סימנים הפוכים</td>
<td>ניתן לבצע בדיקת השערות<br>אין פגיעה בתכונות אר"פ ושונות</td>
<td>**הורדת אחד המשתנים<br>או<br>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. $$\underset{}{Y_t = \alpha + \underset{(10.3)}{\beta_1} \cdot X_{1t} + \underset{(19.8)}{\beta_2} \cdot X_{2t} + U_t}.$$

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 $\beta_1$ במשוואה (2) הינו חסר הטיה, אך אומד השונות של $\beta_1$ מוטה.

ב. האומד של $\beta_1$ במשוואה (3) הינו חסר הטיה, אך אומד השונות של $\beta_1$ מוטה.

ג. האומד של $\beta_1$ במשוואה (4) הינו חסר הטיה, אך אומד השונות של $\beta_1$ 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- $\alpha$ במשוואה (4) הינו חסר הטיה.

ז. האומד ל- $\alpha$ 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.

# מבוא לאקונומטריקה

## פרק 18 - 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: $\alpha_0$.
שכר התחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר בין נשים וגברים: $\alpha_1$ (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: $\alpha_0$.
שכר הממוצע של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר הממוצע בין נשים וגברים: $\alpha_1$ (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: $\beta_0$.
אצל גבר: התוספת לשכר בגין שנות לימוד: $\beta_0 + \beta_1$.
הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: $\beta_1$ (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: $\alpha_0$.
השכר ההתחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר ההתחלתי בין המינים: $\alpha_1$ (הבדל בחותכים).
אצל אישה: התוספת לשכר בגין שנות הלימוד: $\beta_0$.
אצל גבר: התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$.
הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: $\beta_1$ (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: $$WALD_{stat} = \frac{\dfrac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\dfrac{ESS_f + ESS_m}{DF_f + DF_m}} = CHOW_{stat}.$$

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): $H_0 : \alpha_1 = \beta_1 = 0$<br><br>**ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW.<br><br>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- $\beta$. למחיר זה יתווסף $\alpha_0$ בחורף, $\alpha_0+\alpha_1$ באביב, $\alpha_0+\alpha_2$ בקיץ ו-$\alpha_0+\alpha_3$ בסתיו.

ניתן לראות כי: $\alpha_0$ - החותך בקטגוריה שהושמטה, $\alpha_0+\alpha_1$ - החותך בקטגוריה i.

בדיקת השערות:

השערות: $\begin{matrix} H_0:\alpha_1=\alpha_2=\alpha_3=0 \\ H_1:OTHERWISE \end{matrix}$.

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 $\alpha_0$ 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: $V_t=\alpha+\beta_0P_t+\beta_1(D_{1i}P_t)+\beta_2(D_{2i}P_t)+\beta_3(D_{3i}P_t)+u_t$.

המחיר ההתחלתי של הירקות שווה בין עונות השנה $(\alpha)$ אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב: $\beta_0$ בחורף, $\beta_0+\beta_1$ באביב, $\beta_0+\beta_2$ בקיץ ו- $\beta_0+\beta_3$ בסתיו.

ניתן לראות כי- $\beta_0$: השיפוע בקטגוריה שהושמטה $\beta_0+\beta_i$: השיפוע בקטגוריה i.

בדיקת השערות:

השערות: $\begin{matrix} H_0:\beta_1=\beta_2=\beta_3=0 \\ H_1: OTHERWISE \end{matrix}$.

המבחן הסטטיסטי – מבחן WALD:

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

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

- שימו לב שהשיפוע במשוואה המוגבלת איננו $\beta_0$ 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:
$$V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t .$$

בדיקת השערות:
השערות: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$.
המבחן הסטטיסטי - מבחן WALD:
(U):
$$V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t .$$
(R): $V_t = \alpha + \beta \cdot P_t + u_t$.

אם דוחים את $H_0$, 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 $(S_t)$.

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** | $\gamma_0 + \gamma_2$ | $\gamma_0$ | $\gamma_2$ |
| **הפרש** | $\gamma_1 - \gamma_2$ | $\gamma_3$ | |

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:

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

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

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:

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

(25) (23) (68)

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:

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

$\ln(Y)$ - לוג השכר.

$EXP$ - שנות ניסיון.

$D_1$ - מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (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-Zים?

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.

**תרגול מסכם:**

8) 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.

$EDUC$ = מסי שנות לימוד של העובד.

$AGE$ = גיל העובד.

$WAGE$ = שכר העובד.

משוואה (1) נאמדה בפלט מסי 1.

בנוסף נאמד גם פלט מסי 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 מסי 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 לחישוב 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 |

**9)** במטרה לאמוד את העושר הפיננסי של לקוחות הבנק כפונקציה של ההכנסה, הוותק והתנהגות פיננסית נאמדה המשוואה הבאה:

(1) $OSHER_t = \alpha_0 + \alpha_1 \cdot V + \alpha_2 \cdot G + \alpha_3 \cdot V \cdot G + \beta \cdot INCOME_t + V_t$.

כאשר:

$OSHER_t =$ העושר הפיננסי של הלקוח.

$INCOME_t =$ ההכנסה של הלקוח.

$V =$ משתנה דמי. 1=לקוחות ותיקים. 0=לקוחות חדשים.

$G =$ משתנה דמי. 1= לקוחות בעלי התנהגות פיננסית תקינה. 0=אחרת.

$V_t$=סטייה מקרית המקיימת את כל ההנחות הקלאסיות.

- משוואה (1) נאמדה בפלט מס' 4. כמו כן נאמד פלט מס' 5. התייחסו לפלטים אלו בלבד.

א. החוקרת הניחה כי הנטייה השולית לחסוך (לצבור עושר) מתוך ההכנסה, איננה תלויה בוותק או בהתנהגות הפיננסית של הלקוח:
נכון/ לא נכון/ לא ניתן לדעת

ב. החוקרת הניחה כי ההבדל בין לקוחות חדשים ללקוחות וותיקים, תלוי בהתנהגות הפיננסית של הלקוח:
נכון/ לא נכון/ לא ניתן לדעת

הועלתה הטענה כי אין השפעה של וותק והתנהגות פיננסית על העושר.

ג. השערת האפס לבדיקת הטענה הינה: ____________________

ד. הסטטיסטי של WALD לבדיקת הטענה:

i. לא ניתן לחשב את הסטטיסטי בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ________________

ה. הסטטיסטי של CHOW לבדיקת הטענה שהוותק וההתנהגות הפיננסית אינם משפיעים על פונקצית העושר:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ______________

הועלתה הטענה כי אין השפעה של ותק בקרב לקוחות בעלי התנהגות פיננסית לא תקינה.

ו. השערת האפס לבדיקת הטענה הינה: _____________

ז. הסטטיסטי לבדיקת הטענה:

i. לא ניתן לחשבו בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו הוא: ______________

ח. אם נתון ש: $\alpha_2 = 0$ המשמעות הינה:

i. אין השפעה של התנהגות פיננסית בקרב לקוחות חדשים.

ii. אין השפעה של התנהגות פיננסית

iii. אין השפעה של התנהגות פיננסית בקרב לקוחות וותיקים

iv. כל התשובות אינן נכונות

## פלט מס' 4 - משוואה 1:

4

Model: MODEL1
Dependent Variable: OSHER OSHER

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 4 | 2.337947E14 | 5.844868E13 | 441.20 | <.0001 |
| Error | 58546 | 7.755951E15 | 1.324762E11 | | |
| Corrected Total | 58550 | 7.989746E15 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 363973 | R-Square | 0.0293 |
| Dependent Mean | 10546 | Adj R-Sq | 0.0292 |
| Coeff Var | 3451.27703 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| $\alpha$ = Intercept | Intercept | 1 | 1678.45358 | 1975.42706 | 0.85 | 0.3955 |
| $\alpha_1$ = v | | 1 | 4727.38758 | 3077.65435 | 1.54 | 0.1245 |
| $\alpha_2$ = g | | 1 | 475649 | 18512 | 25.69 | <.0001 |
| $\alpha_3$ = vg | | 1 | -290742 | 26068 | -11.15 | <.0001 |
| $\beta$ = income | income | 1 | 0.25845 | 0.00820 | 31.51 | <.0001 |

## פלט מס' 5:

R

Model: MODEL1
Dependent Variable: OSHER OSHER

Analysis of Variance

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | k= 1 | 1.328125E14 | 1.328125E14 | 989.70 | <.0001 |
| Error | T-k-1 58549 | 7.856933E15 | 1.341941E11 | | |
| Corrected Total | T-1= 58550 | 7.989746E15 | | | |

| | | | |
|---|---|---|---|
| Root MSE | 366325 | R-Square | 0.0166 |
| Dependent Mean | 10546 | Adj R-Sq | 0.0166 |
| Coeff Var | 3473.58313 | | |

Parameter Estimates

| Variable | Label | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|---|
| Intercept | Intercept | 1 | 8057.00227 | 1515.97451 | 5.31 | <.0001 |
| income | income | 1 | 0.25967 | 0.00825 | 31.46 | <.0001 |

## 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. $Z_0 = \ln(Y)_t$ , $Z_1 = D_1 + D_3$ , $Z_2 = D_2 - D_3$ , $Z_3 = EXP_t$ , $Z_4 = EXP_t^2$
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. $H_0 : \beta_2 + \beta_3 = 0$.

**8)** א. נכון. ב. נכון. ג. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$.
ד. ראו סרטון. ה. ראו סרטון. ו. ראו סרטון.
ז. $H_0 : \alpha_3 = 1500$ , $H_1 : \alpha_3 > 1500$. ח. ii, $t_{stat} = 2.04$. ט. $H_0 : \alpha_1 = \alpha_2$ , $H_1 : \alpha_1 \neq \alpha_2$.
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.

# מבוא לאקונומטריקה

## פרק 19 - משתנה דמי - המשך

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.

## תשובות סופיות:

1) א. ii.   ב. iv.

# מבוא לאקונומטריקה

## פרק 20 - 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- $\lambda$ ות וה- $\mu$ 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 ההשפעה 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 מולטיקולינאריות 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:**

מס' האומדים שיתקבלו בשלושת השיטות ותכונותיהם תלויים בזיהוי של המשוואה:
אם המשוואה לא מזוהה: לא ניתן להשתמש באף אחת מהשיטות.
כאשר המשוואה מזוהה (בדיוק או ביתר): האומדים שיתקבלו בשלושת השיטות יהיו תמיד מוטים אך עקיבים.

תכונת היעילות ומס' האומדים האפשרי מסוכמים בטבלה הבאה:

| | **מזוהה בדיוק** | **מזוהה ביתר** |
|---|---|---|
| **שיטת ILS** | אומד אחד לפרמטר<br>**יעיל** | יתכן יותר מאומד אחד לפרמטר<br>**לא יעילים** |
| **שיטת 2SLS** | אומדן אחד למשתנה האנדוגני<br>**יעיל** | |
| **שיטת IV** | אינסוף משתני עזר<br>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: $\operatorname{cov}(Y_{t-1},u_t \neq 0)$, אז $Y_{t-1}$ אנדוגני.
אם אין מתאם סדרתי: $\operatorname{cov}(Y_{t-1},u_t = 0)$, 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 באוכ׳ 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-$\beta_1$?

## 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- $\beta_2$ 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 אקסוגנים.
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: $\text{cov}\left(Z_j, u_t\right) = 0$ 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- $\gamma$.

ג. האם $\gamma$ 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 $\beta_2$ 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 $\beta_2$. 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 היצע מזוהה ביתר.

ב. פונקציית הביקוש, התוצאות מתיישבות.

ג. הכמות תרד ב-0.61173%, המחיר יעלה ב-0.453478%.

ד. $\hat{P} = 16.05$ , $\hat{Q} = 9.34$.

2) א. ראו סרטון. ב. מזוהות בדיוק. ג. $\hat{C}_t = \hat{\gamma}_0 + \hat{\gamma}_1 Z_t$ , $\hat{Y}_t = \hat{\gamma}_3 + \hat{\gamma}_4 Z_t$.

ד. $\hat{\alpha} = \frac{\hat{\gamma}_0}{\hat{\gamma}_4} = \frac{\hat{\gamma}_3}{\hat{\gamma}_4}$ , $\hat{\beta} = \frac{\hat{\gamma}_1}{\hat{\gamma}_4}$. ה. $\hat{\alpha} = 16,264.46$ , $\hat{\beta} = -0.09537$.

3) א. מוטים אך עקיבים ויעילים במדגמים גדולים.

ב. $\hat{\beta} = -0.09537$ , $\hat{\alpha} = 16,264.47$.

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. $\hat{\beta}_0 = 1.25$ , $\hat{\beta}_1 = 0.75$. 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. $\pi_{11} = \frac{\beta_1}{1-\delta\gamma}$ , $\pi_{12} = \frac{\beta_2}{1-\delta\gamma}$ , $\pi_{13} = \frac{\beta_3 \gamma}{1-\delta\gamma}$ ,

$$\pi_{21} = \frac{\beta_1 \delta}{1-\delta\gamma} , \pi_{22} = \frac{\beta_2 \delta}{1-\delta\gamma} , \pi_{23} = \frac{\beta_3}{1-\delta\gamma} ,$$

$\tilde{v}_i = \dfrac{v_i + \delta u_i}{1-\delta\gamma}$ , $\tilde{u}_i = \dfrac{u_i + \gamma v_i}{1-\delta\gamma}$ .

ב. מכיוון ש- $\gamma = \dfrac{\pi_{13}}{\pi_{23}}$ , ניתן לקבל אומד עקיב ל- $\gamma$ 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.

# מבוא לאקונומטריקה

## פרק 21 - הדרכה בקריאת פלטים של SPSS - רגרסיה פשוטה

תוכן העניינים

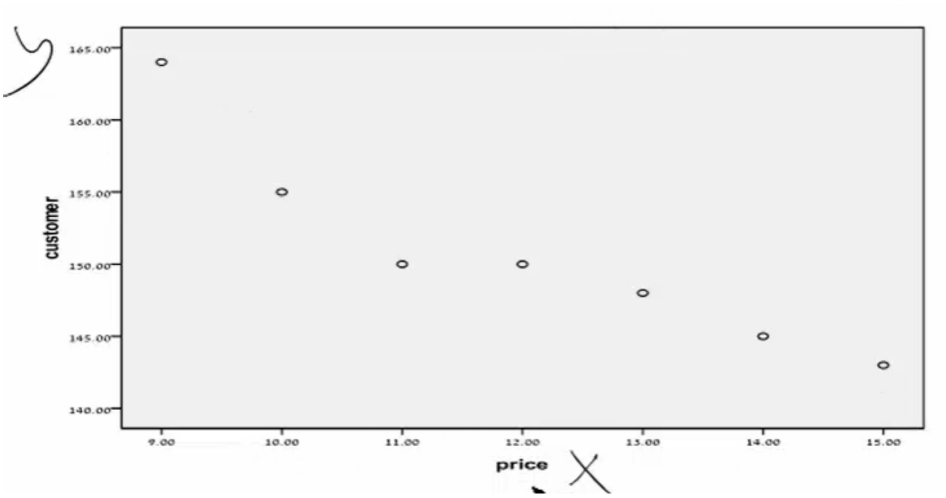
# רגרסיה פשוטה:

## שאלות:

### קריאת פלטים של SPSS:

**1)** על סמך הנתונים של השאלה הקודמת התקבלו הפלטים הבאים:

**דיאגרמת הפיזור (scatter plot):**




**סטטיסטיקה תיאורית (descriptive statistics):**

Descriptive Statistics

| | Mean | Std. Deviation | N |
|---|---|---|---|
| customer | 150.7143 | 7.01699 | 7 |
| Price | 12.0000 | 2.16025 | 7 |

**פלט מקדם המתאם (correlations):**

Correlations

| | | customer | Price |
|---|---|---|---|
| Pearson Correlation | customer | 1.000 | -.935 |
| | price | -.935 | 1.000 |
| Sig. (1-tailed) | customer | . | .001 |
| | price | .001 | . |
| N | customer | 7 | 7 |
| | price | 7 | 7 |

**פלט model summary:**

**Model Summary[b]**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .935[a] | .873 | .848 | 2.73470 |

a. Predictors: (Constant), price

b. Dependent Variable: customer

**פלט ניתוח שונות (ANOVA):**

**ANOVA[b]**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 258.036 | 1 | 258.036 | 34.503 | .002[a] |
| | Residual | 37.393 | 5 | 7.479 | | |
| | Total | 295.429 | 6 | | | |

a. Predictors: (Constant), price

b. Dependent Variable: customer

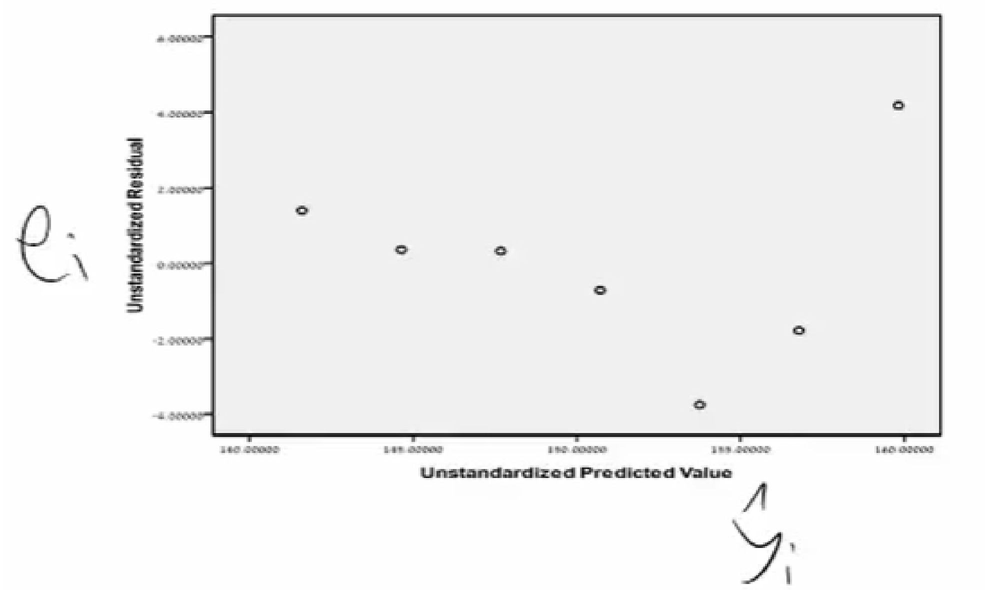
**פלט מקדמי הרגרסיה (coeffitients):**

**Coefficients[a]**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | 187.143 | 6.287 | | 29.765 | .000 |
| | Price | -3.036 | .517 | -.935 | -5.874 | .002 |

a. Dependent Variable: customer

**גרף ניתוח שאריות:**




על סמך הפלטים הנתונים:

א. מהו מודל הרגרסיה שנאמד?

ב. מהו מקדם המתאם $r_{xy}$?

ג. מהי השונות של שארית המודל?

ד. האם נמצא דפוס מיוחד בשאריות?

ה. מהו אחוז השונות המוסברת?

ו. על פי מבחן F: האם קיים קשר בין מחיר הבירה לבין כמות הלקוחות המבקרים בפיצוצייה ברמת מובהקות 5%?

ז. על פי מבחן t: האם קיים קשר בין מחיר הבירה לבין כמות הלקוחות המבקרים בפיצוצייה ברמת מובהקות 5%? השוו את התוצאות.

ח. מה ה-pvalue של המבחנים הסטטיסטיים? מה משמעותו?

ט. בדקו האם קיים קשר חיובי מובהק בין המשתנים ברמת מובהקות 5%?

## תשובות סופיות:

**1)** א. $\hat{y}_i = 187.143 - 3.036x_i$. ב. $r_{xy} = 0.935$. ג. $MSE = 7.479$.

ד. לא. ה. $R^2 = 0.874$. ו. $F = 34.503$.

ז. $t = -5.874$. ח. $pvalue = 0.002$. ט. ראו סרטון.

# מבוא לאקונומטריקה

## פרק 22 - הדרכה בקריאת פלטים של SPSS - רגרסיה מרובה

תוכן העניינים

# רגרסיה מרובה:

## שאלות:

**1)** מעוניינים למצוא קשר בין מחיר הדירה (ב-$) לבין ארבעה משתנים מסבירים:

1. שטח הדירה.
2. גודל שטח האמבטיה (ב-Sqft).
3. מרחק הדירה מהים.
4. מהאוניברסיטה (במיילים).

לשם כך נדגמו מספר דירות והריצו רגרסיה אשר בה המשתנה המוסבר הוא מחיר הדירה.

להלן פלט הרגרסיה שהתקבל:

**Model Summary**

| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
|---|---|---|---|---|
| 1 | .952a | | | |

a. Predictors: (Constant), Sea_Dist, Apartment, Bath

**ANOVAb**

| Model | | Sum of Squares | df | Mean Square | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | | | | | .000a |
| | Residual | | | | | |
| | Total | 1940484.615 | 25 | | | |

a. Predictors: (Constant), Univ_Dist, Bath, Sea_Dist, Apartment

b. Dependent Variable: Price

**Coefficientsa**

| Model | | Unstandardized Coefficients | | Standardized Coefficients | t | Sig. |
|---|---|---|---|---|---|---|
| | | B | Std. Error | Beta | | |
| 1 | (Constant) | -265.514 | 146.673 | | -1.810 | .085 |
| | Apartment | | .449 | .722 | 6.572 | |
| | Bath | 4.256 | | .297 | 2.687 | .014 |
| | Sea_Dist | -32.114 | 11.090 | -.223 | | .009 |
| | Univ_Dist | 11.746 | 9.439 | .095 | 1.244 | .227 |

a. Dependent Variable: Price

ענה על הסעיפים הבאים:

א. מלאו את התאים החסרים בטבלה (אם לא ניתן למלא את כל התאים החסרים באופן מלא ונמקו באופן מפורש מדוע לא ניתן).

ב. כתבו את האומדן למשוואת מחיר הדירה בצורה מפורשת על סמך הפלט הנ"ל. פרשו את מקדמי הרגרסיה.

ג. בדקו האם ארבעת הגורמים ביחד אכן מסבירים את מחיר הדירה. הסבירו את המסקנה שהגעתם אליה. השתמשו ברמת מובהקות 5%.

ד. הסבירו מהו ערך ה- Pvalue ומה ניתן להסיק ממנו לגבי המשתנים המסבירים?

ה. בנו רווח סמך למקדם גודל שטח האמבטיה. השתמשו ברמת מובהקות של 2%.

ו. ברמת מובהקות של 5% יש לבדוק האם המרחק מהאוניברסיטה אכן משפיע על מחיר הדירה.

ז. האם במודל הרגרסיה הנוכחי ניתן לוותר על גורם המרחק מהים? השתמשו ברמת מובהקות 1%.

ח. בדקו את ההשערה כי קיים קשר חיובי בין גודל הדירה למחירה ברמת מובהקות של 5%.

ט. נתונה מטריצת מקדמי המתאם הבאה:

| | *X1* | *X2* | *X3* | *X4* |
|---|---|---|---|---|
| **X1** | **1** | | | |
| **X2** | **0.228579** | **1** | | |
| **X3** | **-0.22413** | **-0.13924** | **1** | |
| **X4** | **-0.24545** | **-0.97295** | **0.029537** | **1** |

מה ניתן ללמוד ממנה ומה משמעותה לגבי המודל?

י. הניחו כי השאריות המתקבלות מניתוח הרגרסיה הן בעלות הערכים הבאים (סדר הקריאה הוא משמאל לימין) הנח כי אלו כל השאריות הקיימות במודל: $7, 9, -3, -7, -3, 12, 18, 6, -5, 10, 5, -4, 6, 3, -12, -1$.

האם משתנים $X_2$ ו- $X_4$ מוסיפים תוספת משמעותית לניבוי?

אם לא ניתן לענות על השאלה, ציין מדוע.

יא. מה יהיו תוצאות מבחן F לבדיקת התוספת לניבוי של המרחק מהאוניברסיטה על פני המשתנים האחרים (ענה ללא חישוב).

## תשובות סופיות:

1) א.1. 0.907.
א.2. 0.89.
א.3. 92.
א.4. 1760019.5.
א.5. 180465.
א.6. 4.
א.7. 21.
א.8. 440004.875.
א.9. 8593.57.
א.10. 51.2.
א.11. 2.95.
א.12. $sig < 0.001$.
א.13. 1.58.
א.14. -2.896.

ב. $\hat{y}_i = -256.514 + 2.95x_{1i} + 4.256x_{2i} - 32.114x_{3i} + 11.746x_{4i}$.

ג. ראו סרטון.
ד. ראו סרטון.
ה. $p(1.016 \le \beta_2 \le 7.496) = 0.98$.
ו. אין עדות לכך.
ז. לא.
ח. יש עדות לכך.
ט. בעיית קוליניאריות.
י. ראו סרטון.
יא. ראו סרטון.

# מבוא לאקונומטריקה

## פרק 23 - 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- $\alpha$? מהו רווח הסמך ל- $\beta$?

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 $\beta$ 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: $\mathrm{cov}\left(\hat{\beta}_1, \hat{\beta}_2\right) = -0.25$).

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 $\beta$. נכון / לא נכון / לא ניתן לדעת

ב. אם הערך 0 נמצא בתוך רווח הסמך ל- $\beta$, אזי $\beta$ 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 לינארי. 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 לינארי. 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. $H_0:\beta=0.4$, $H_1:\beta\ne0.4$.
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. $V(\tilde\beta)=\frac{\sum X_t^2\sigma^2}{S^2{}_{XX}}$.

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

# מבוא לאקונומטריקה

## פרק 24 - 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: $\bar{X} = 46.040873$ , $S_{XX} = 35079$ :

### Analysis of Variance

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

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

### Parameter Estimates

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

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 $\alpha$. נכון / לא נכון / לא ניתן לדעת

ii. יכפיל את האומד של $\beta$ 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. $\beta$.

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- $\beta$, אך איננו אומד עקיב ל- $\beta$.
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. Covariance of Estimates , $S^2_{\hat{\beta}_2,\hat{\beta}_3}$.

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.

# מבוא לאקונומטריקה

## פרק 25 - 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 לחשב בנתונים הקיימים.

ii. 5.5.

iii. -5.5.

iv. -15.5.

v. 15.5.

ה. הסטטיסטי של WALD לבדיקת הטענה:

i. לא ניתן לחשב בעזרת הנתונים הקיימים.

ii. ניתן לחשב 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 מס' 2 .

**Dependent Variable: lnY**

**Analysis of Variance**

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

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

**Parameter Estimates**

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

ז. ההשערות לבדיקת הטענה הינן: $\begin{matrix} H_0 : \_\_\_\_\_\_\_\_ \\ H_1 : \_\_\_\_\_\_\_\_ \end{matrix}$.

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 המשוואה הבאה:

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

כאשר:

משוואה מס' (3) נאמדה בפלט מס' 3.

**Dependent Variable: lnY**

**Analysis of Variance**

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

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

**Parameter Estimates**

| Variable | DF | Parameter Estimate | Standard Error | T for H0: Parameter=0 | Prob>lTl |
|---|---|---|---|---|---|
| INTERCEP | 1 | 2.681787 | 0.37024512 | 7.243 | 0.0001 |
| lnL | 1 | 0.177813 | 0.04470595 | 3.977 | 0.0001 |
| lnK | 1 | 0.465154 | 0.08612163 | 5.401 | 0.0001 |

**Covariance of Estimates**

| COVB | INTERCEP | lnL | lnK |
|---|---|---|---|
| INTERCEP | 0.1370814505 | -0.003289697 | -0.02723683 |
| lnL | -0.003289697 | 0.0019986217 | -0.001270417 |
| lnK | -0.02723683 | -0.001270417 | 0.0074169359 |

י. ההשערות לבדיקת הטענה הינן : $\begin{matrix} H_0: \_\_\_\_\_\_\_\_ \\ H_1: \_\_\_\_\_\_\_\_ \end{matrix}$.

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:

$\begin{matrix} Z_1 = \_\_\_\_\_\_\_\_ \\ Z_0 = \_\_\_\_\_\_\_\_ \end{matrix}$.

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- $\beta$ ול- $\alpha$ 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- $\bar{R}^2$ 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 $\beta$ 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- $\beta$: 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- $\beta$: 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- $\alpha$): נכון/לא נכון/אי אפשר לדעת

י. אם דחינו את $H_0$ במבחן t למובהקות ה- $\beta$ 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- $\beta$: 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. $H_0:\beta=1$, $H_1:\beta<1$.
ד. iv.
ה. i.
ו. לא נכון.
ז. ראו סרטון.
ח. ראו סרטון.
ט. ראו סרטון.
י. $H_0:\beta_3=0$, $H_1:\beta_3\neq 0$.
יא. ii , $WALD_{stat}=1.74$.
יב. $H_0:\beta_2=2\cdot\beta_1$.
יג. ii , $t=0.1417$.
יד. $Z_0=\ln(Y)_t$, $Z_1=\ln(L)_t+2\ln(K)_t$.
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.

# מבוא לאקונומטריקה

## פרק 26 - 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- $\beta$:
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 95% מהו האומד לתוחלת פעילות בכרטיסי אשראי עבור סך פעילות בחשבונות חיסכון של 50,000 ש"ח?

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

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-ש"ח.

$PIKADON2_t$ - סה"כ הפקדה לפקדונות חודשיים ב-ש"ח.

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 לחשב בעזרת הנתונים הקיימים.

ii. ניתן לחשבו וערכו :________________________.

יא. הסטטיסטי של t לבדיקת הטענה :

i. לא ניתן לחשב 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: $D_0=\gamma_0+\gamma_1\cdot D_1+\gamma_2\cdot D_2+v$ .

כאשר:
$D_0$ : _______
$D_1$ : _______ .
$D_2$ : _______

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: $Y=\alpha+\beta_1X_1+\beta_2X_2+U$: 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 $\beta$ 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. $E\left(\tilde{\beta}\right) = \_\_\_\_\_\_\_$.

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

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

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

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

# מבוא לאקונומטריקה

## פרק 27 - 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>lTl |
|---|---|---|---|---|---|
| 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 $\beta$ היא: ____________

ד. גבולות רווח-סמך ברמת סמך של 95% עבור $\beta$ הינם:

גבול תחתון: ________________.

גבול עליון: ________________.

ה. ערך t לחישוב מובהקות ה-$\beta$ 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 $\alpha$ ישתנה: נכון/לא נכון/ אי אפשר לדעת
ii. האומד של $\beta$ 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>lTl |
|---|---|---|---|---|---|
| INTERCEP | 1 | 0.78242 | 0.59739 | 1.31 | 0.1932 |
| lnp | 1 | 1.63491 | 0.05332 | 30.66 | <.0001 |
| lny | 1 | 0.20001 | 0.16568 | ----- | 0.2302 |

Covariance of Estimates

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

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 $\alpha$, אזי בהכרח גם נדחה $H_0$ במבחן הדו צדדי באותה רמת מובהקות: נכון/לא נכון/אי אפשר לדעת.

ג. אם ערך האומד ל- $\beta$ 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 $\alpha$, 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 $\beta$.

ב. חשבו את $VAR(\hat{\beta})$.

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- $\beta$.

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

## 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.

# מבוא לאקונומטריקה

## פרק 28 - פתרון מודרך של מבחן מה- 08.02.2018

תוכן העניינים

# פתרון מודרך של מבחן מה-08.02.2018:

## שאלות:

**1)** בהתייחס למחקר 1. התייחסו לרגרסיה שכוללת את כל המסבירים כרגרסיית הבסיס. בבדיקת ההשערה: $H_0: \beta_{highscl} = \beta_{medinc} = 0$ התקבל (רמת מובהקות 5%):

א. הסטטיסט מתפלג תחת השערת האפס $F$. הערך של הסטטיסט הוא 27.9 ולכן נדחה את השערת האפס.

ב. הסטטיסט מתפלג תחת השערת האפס $F$. הערך של הסטטיסט הוא 2.1 ולכן לא נדחה את השערת האפס.

ג. הסטטיסט מתפלג תחת השערת האפס $t$. הערך של הסטטיסט הוא 2.1 ולכן נדחה את השערת האפס.

ד. הסטטיסט מתפלג תחת השערת האפס $F$. הערך של הסטטיסט הוא 21.24 ולכן נדחה את השערת האפס.

**2)** בהתייחס למחקר 1. התייחסו לרגרסיה שכוללת את כל המסבירים כרגרסיית הבסיס. בבדיקת ההשערה: $H_0: \beta_{highscl} + \beta_{college} = 0$ התקבל (רמת מובהקות 5%):

א. הסטטיסט מתפלג תחת השערת האפס $t$. הערך של הסטטיסט הוא 4.97 ולכן נדחה את השערת האפס.

ב. הסטטיסט מתפלג תחת השערת האפס $F$. הערך של הסטטיסט הוא 2.1 ולכן לא נדחה את השערת האפס.

ג. הסטטיסט מתפלג תחת השערת האפס $t$. הערך של הסטטיסט הוא 2.1 ולכן נדחה את השערת האפס.

ד. הסטטיסט מתפלג תחת השערת האפס $t$. הערך של הסטטיסט הוא 1.4 ולכן לא נדחה את השערת האפס.

**3)** התייחס למחקר 2:

א. משמעות המקדם של ההכנסה היא ששינוי של אחוז ברמת ההכנסה מוביל לשינוי של 0.83 אחוזים בהוצאות על בריאות. המקדם מובהק ברמת מובהקות נמוכה מ-1%.

ב. משמעות המקדם של ההכנסה היא ששינוי של יחידה ברמת ההכנסה מוביל לשינוי של 0.83 אחוזים בהוצאות על בריאות. המקדם מובהק ברמת מובהקות נמוכה מ-1%.

ג. משמעות המקדם של האוכלוסייה היא ששינוי של יחידה באוכלוסייה מוביל לשינוי של יחידה בהוצאות על בריאות. המקדם אינו מובהק ברמת מובהקות של 5%.

ד. משמעות המקדם של ההכנסה היא ששינוי של אחוז ברמת ההכנסה מוביל לשינוי של 0.14 אחוזים בהוצאות על בריאות. המקדם מובהק ברמת מובהקות נמוכה מ-5%.

**4)** חוקרים בדקו את הקשר בין ממוצע הציונים בשנה א' בלימודי כלכלה ($y_i$) והציון הפסיכומטרי ($x_i$) בקרב מדגם אקראי של תלמידי שנה ב' בכלכלה.
על בסיס הנתונים במדגם הם חישבו:
$\bar{Y}=658.75$ , $\bar{X}=82.125$ , $\sum\left(X_i-\bar{X}\right)^2=5687.5$ , $\sum\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)=581.25$.
על בסיס הנתונים הללו החוקרים הסיקו כי:
א. אם ציון הפסיכומטרי עולה ב-50 נקודות אז ממוצע הציונים בשנה א' בכלכלה צפוי לעלות בכ-5.11 נקודות.
ב. אם ממוצע הציונים בשנה א' יורד בנקודה אחת אז ציון הפסיכומטרי צפוי לרדת בכ-10.22 נקודות.
ג. ציון הפסיכומטרי החזוי של סטודנטית עם ממוצע ציונים 90.02 בשנה א' של כלכלה הוא 736.
ד. ממוצע הציונים בשנה א' של סטודנט לכלכלה עם צעון פסיכומטרי של 720 צפוי להיות 73.6.

**5)** איזה מבסיסי הנתונים הבאים הינו בסיס נתוני אורך (סדרה עיתית)?
א. נתונים אודות הריבית הנומינלית בארה"ב בין השנים 1950-1960.
ב. נתונים אודות שיעורי האבטלה לפי רשות מקומית בישראל בשנת 2015.
ג. נתונים אודות יבוא חיטה במדינות אירופה ברבעון הראשון של 1982.
ד. נתונים אודות מספר הנשים בעלות הקביעות בנחלקות השונות באוניברסיטת בר אילן בשנת 2003.

**6)** נתונה הרגרסיה הבאה: $y_i=\beta_0+\beta_1 x_i+u_i$ . איזו מהטענות הבאות אינה נכונה?
א. אומד הריבועים הפחותים $\hat{\beta}_1$ יהיה אומד חסר הטיה ל-$\beta_1$ אם $E\left(u_i\right)\neq 0$.
ב. אומד הריבועים הפחותים $\hat{\beta}_1$ יהיה אומד חסר הטיה ל-$\beta_1$ גם אם $u_i$ לא מתפלג נורמלית.
ג. אומד הריבועים הפחותים $\hat{\beta}_1$ לא יהיה אומד חסר הטיה ל-$\beta_1$ אם התוחלת של גורמים שאינם כלולים $\left(E\left(u_i\right)\right)$ תלויה בערכים של $x_i$, זאת אומרת $Cov\left(X,u_i\right)\neq 0$.
ד. אומד הריבועים הפחותים $\hat{\beta}_1$ יהיה אומד יעיל רק כאשר $Var\left(u_i\right)$ היא קבועה.

**7)** חוקר אמד את הרגרסיה הבאה: $\hat{X}_i = 50 - 0.2P_i$, כאשר $X$ היא הכמות שנרכשה ממוצר מסוים ו- $P$ המחיר ב-₪. החוקר החליט שהוא עשה שגיאה כשאמד את המחיר ב-₪. לכן הוא הגדיר משתנה חדש, $D_i = \frac{p_i}{\varepsilon_i}$, כאשר $D$ הוא המחיר בדולרים ו- $\varepsilon_i$ שער החליפין שקל-דולר.
שער החליפין בתקופה שעליה היו תצפיות נע בין 3.8 ל-4.2 ₪ לדולר, עם ממוצע 4 ₪ לדולר.
טענה א': ברגרסיה: $\hat{X}_i = \alpha + \beta D_i + u_i$, האומד ל- $\beta$ שווה 0.8-.
טענה ב': ברגרסיה: $\hat{X}_i = \alpha + \beta D_i + u_i$, האומד ל- $\beta$ שווה 0.05-.
טענה ג': ברגרסיה: $\hat{X}_i = \alpha + \beta D_i + u_i$, לא ניתן לדעת את האומד ל- $\beta$ ללא נתונים על שער החליפין במשך כל התקופה.

א. טענה ג' נכונה.
ב. טענה א' נכונה.
ג. טענה ב' נכונה.

**8)** נניח כי אמדו את הרגרסיה (בסוגריים סטיות תקן):

$$\hat{Y}_i = 3.2 + 1.4 * X_i + 3.9 * Z_i$$
$$(1.2)\quad (2.3)\qquad (2.1)$$

איזו מבין הטענות הבאות היא טענת אמת:

א. מתאם פירסון בין $X$ ל- $Z$ $(rX,Z)$ קטן, בערך מוחלט, מ-1.
ב. אם נוסיף עוד משתנה לרגרסיה, הערך של $\bar{R}^2$ לא יפחת.
ג. נניח כי מוסיפים לרגרסיה עוד משתנה נוסף, $P$. ידוע כי מקדמי המתאם של פירסון בין $P$ לבין $X$, ובין $P$ לבין $Z$, קטנים שניהם, בערך מוחלט, מ-1 (כלומר: $rX,P<1, rZ,P<1$). לכן ניתן להוסיף את המשתנה $P$ לרגרסיה ואין חשש למולטיקולינאריות מושלמת.
ד. אם נוסיף עוד משתנה לרגרסיה, הערך של $\bar{R}^2$ יגדל.

**9)** נאמדה רגרסיה והתקבלו התוצאות הבאות: $\hat{Y}_i = 5 + 2X_i$. כעת חוקר חושב שהוא אמד את הרגרסיה תוך שימוש בנתונים לא נכונים. לכן הוא מגדיר משתנה חדש, $Z_i = 3Y_i$. הרגרסיה החדשה שקיבל:

א. $\hat{Z}_i = 15 + 6X_i$.
ב. $\hat{Z}_i = 5 + 6X_i$.
ג. $\hat{Z}_i = 15 + 2X_i$.
ד. $\hat{Z}_i = 3X_i$.

**10)** חוקר מעוניין לאמוד את ההשפעה של המשתנה $X$ על המשתנה $Y$. הוא אמד את הרגרסיה: $Y_i = \alpha + \beta X_i + u_i$ (1). הוא שוקל להוסיף לרגרסיה משתנה מסביר נוסף, $Z$. איזו מהטענות הבאות נכונה?

א. אם $Cov(X,Z)$ שונה מאפס ו- $Cov(Z,Y)$ שונה מאפס, האומד ל- $\beta$ ברגרסיה (1) מוטה.

ב. אם $Cov(X,Z)$ שווה אפס ו- $Cov(Z,Y)$ שונה מאפס, האומד ל- $\beta$ ברגרסיה (1) מוטה.

ג. אם $Cov(X,Z)$ שונה מאפס ו- $Cov(Z,Y)$ שווה אפס, האומד ל- $\beta$ ברגרסיה (1) מוטה.

ד. האומד ל- $\beta$ ברגרסיה (1) איננו מוטה בשום מקרה (זהו אומדן OLS).

**11)** שלושה סטודנטים דנו בתכונות של אומדי הריבועים הפחותים (OLS).
להלן טענותיהם:

ג׳ורדי: המדד לטיב ההתאמה של OLS, ה- $R^2$, שווה תמיד למקדם המתאם פירסון בריבוע.

סטטיק: ב-OLS אי אפשר לדעת את השונות של $u_i$ ולכן אנחנו נאלצים לאמוד אותה.

בן-אל: אומדי OLS הם אומדים יעילים כיוון שהשונות שלהם אינה גדלה כאשר מוסיפים משתנים מסבירים שאינם רלוונטיים.

א. רק סטטיק צודק.

ב. רק ג׳ורדי צודק.

ג. רק בן-אל צודק.

ד. גם סטטיק וגם בן-אל צודקים.

**12)** חוקרים בודקים את הקשר בין אי-שוויון (GINI) לתוצר (GDP) בעזרת מדגם של כל מדינות העולם. על פי התיאוריה של קוזנץ (Kuznets), ברמות תוצר נמוכות, עלייה בתוצר מובילה לעלייה באי-שוויון עד לנקודה מסוימת ואז עלייה בתוצר מובילה לירידה באי-שוויון.
איזו מהרגרסיות הבאות מייצגת את הקשר כפי שמתואר על ידי קוזנץ?

א. $gini = 0.23 + 0.04 * gdp - 0.0065 * gdp^2$.

ב. $gini = 0.23 + 0.04 * gdp - 0.0065 * \ln(gdp)$.

ג. $gini = 0.23 - 0.04 * gdp + 0.0065 * gdp^2$.

ד. $\ln(gini) = 0.23 + 0.04 * \ln(gdp)$.

**13)** חוקר אמד את הרגרסיה: $Y_i = \alpha + \beta_1 X_i + u_i$.
אילו מבין הטענות הבאות היא טענת אמת:
א. דחייה של השערת האפס: $H_0: \beta_1 = 0 \, , H_1: \beta_1 \neq 0$ ברמת מובהקות של 5%, משמעה שנדחה גם את השערת האפס על השיפוע ברגרסיה שבה $X$ הוא המשתנה המוסבר ו- $Y$ המשתנה המסביר (באותה רמת מובהקות).
ב. דחייה של השערת האפס: $H_0: \beta_1 = 0 \, , H_1: \beta_1 \neq 0$ ברמת מובהקות של 5%, משמעה שיש קשר באוכלוסייה בין $X$ ו- $Y$.
ג. קבלה של השערת האפס: $H_0: \beta_1 = 0 \, , H_1: \beta_1 \neq 0$ ברמת מובהקות של 5%, משמעה שאין קשר באוכלוסייה בין $X$ ו- $Y$.
ד. דחייה של השערת האפס: $H_0: \beta_1 = 0 \, , H_1: \beta_1 \neq 0$ ברמת מובהקות של 5%, משמעה שאין קשר באוכלוסייה בין $X$ ו- $Y$.

**14)** המודל האמיתי באוכלוסייה הוא: $y_i = \lambda_0 + \lambda_1 x_{1i} + \lambda_2 x_{2i} + u_i$,
והסימונים: $\hat{\lambda}$ , $\hat{y}$ מסמנים את אומדי ה-OLS במדגם.
סמנו את הנוסחה השגויה:
א. $y_i = \hat{\lambda}_0 + \hat{\lambda}_1 x_{1i} + \hat{\lambda}_2 x_{2i} + u_i$.
ב. $y_i = \hat{\lambda}_0 + \hat{\lambda}_1 x_{1i} + \hat{\lambda}_2 x_{2i} + u_i$.
ג. $y_i = \hat{\lambda}_0 + \hat{\lambda}_1 x_{1i} + \hat{\lambda}_2 x_{2i}$.
ד. $E\left(y_i \middle| x_{1i}, x_{2i}\right) = \hat{\lambda}_0 + \hat{\lambda}_1 x_{1i} + \hat{\lambda}_2 x_{2i}$.

**15)** היתרון של $R^2_{adjusted}$ על פני $R^2$ הוא:
א. $R^2_{adjusted}$ מגלם את המחיר של הוספת משנים מסבירים.
ב. $R^2_{adjusted}$ קל יותר לחישוב מאשר $R^2$.
ג. $R^2_{adjusted}$ מצמצם את סכום הסטיות מהממוצע.
ד. אין ל- $R^2_{adjusted}$ יתרון על $R^2$.

**16)** נתונות התוצאות של רגרסיה (בסוגריים סטיות תקן):
$$\underset{(1.91)}{\hat{Y}_i = 19.74} - \underset{(0.052)}{0.27 X_i}.$$
כמו כן ידוע כי מספר התצפיות הוא 59 וכי: $\sum\left(Y - \bar{Y}\right)^2 = 1008.52$.
הערך של סכום השגיאות הריבועיות $\left(\sum \hat{u}^2\right)$ הוא:
א. 683.8.
ב. 324.7.
ג. 1234.25.
ד. 12.3425.

## תשובות סופיות:

| | | | | |
|---|---|---|---|---|
| 1) א' | 2) א' | 3) א' | 4) א' | 5) א' |
| 6) א' | 7) א' | 8) א' | 9) א' | 10) א' |
| 11) א' | 12) א' | 13) א' | 14) א' | 15) א' |
| 16) א' | | | | |

# מבוא לאקונומטריקה

## פרק 29 - פתרון מודרך של מבחן - מועד א - 03.02.2019

תוכן העניינים

# פתרון מודרך של מבחן – מועד א – 03.02.2019:

## שאלות:

**1)** נתונים שני מודלים:

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$

$$y=\alpha_0+\alpha_1x_1+v$$

טענה א': אם: $Cov(x_1,x_2)=0$, אזי: $\hat{\alpha}_1=\hat{\beta}_1$.

טענה ב': אם: $\hat{\alpha}_0=\hat{\beta}_0$ ו- $\hat{\alpha}_1=\hat{\beta}_1$, אזי יש מולטיקולינאריות מושלמת בין $x_1$ ו- $x_2$.

טענה ג': אם: $x_2=x_1^2$, אזי: $\hat{\alpha}_1>\hat{\beta}_1$.

א. רק טענה א' נכונה.
ב. רק טענות א' ו-ג' נכונות.
ג. כל הטענות נכונות.
ד. רק טענה ג' נכונה.

**2)** נאמדה הרגרסיה: $Y_i=\hat{\gamma}_0+\hat{\gamma}_1G_i+\hat{u}_i$. בהינתן שכל שאר הגורמים אינם משתנים, אילו מהגורמים הבאים אינו משפיע על הרווח בר סמך של $\hat{\gamma}_1$?

א. הערך הממוצע של המשתנה המוסבר, $\bar{Y}$.
ב. גודל המדגם, $N$.
ג. השונות במדגם של המשתנה המסביר, $\text{var}(G_i)$.
ד. השונות של הסטייה, $\text{var}(\hat{u}_i)$.

**3)** נתון המודל: $\ln(y)=\beta_0+\beta_1x+u$.
כאשר $y$ מייצג את רמת ההשקעה ו- $x$ את מדד המחירים.
מהו השינוי הצפוי באחוזים בהשקעה בעקבות עלייה של 1% במדד המחירים?
(אם מניחים ש- $\beta_1$ קטן).

א. $\beta_{1x}$.
ב. $\beta_1$.
ג. $\beta_1\dfrac{1}{y}$.
ד. $\beta_1\dfrac{x}{y}$.

**4)** כאשר אומדים את מודל הרגרסיה הפשוטה: $y=\beta_0+\beta_1 x+u$ , ב-OLS איזו טענה נכונה.

טענה א': אם אומדני OLS מוטים, אזי: $Cov(x,\hat{u})\neq 0$.

טענה ב': אם אומדני OLS מוטים, אזי: $\sum_{i=1}^{n}(x_i-\bar{x})\neq 0$.

טענה ג': אם אומדני OLS מוטים, אזי: $Var(u|x)$ משתנה עם $x$.

א. אין טענות נכונות.
ב. רק טענות א' ו-ג' נכונות.
ג. רק טענה ב' נכונה.
ד. רק טענה ג' נכונה.

**5)** נתונה פונקציית ביקוש: $y=\beta_0+\beta_1 x+u$.
כאשר $y$ מייצג את מספר היחידות הנמכרות ו-$x$ את המחיר ב-₪.
טענה א': אם המחיר נמדד בדולרים (3.5 ₪=1 דולר), אזי האומדן החדש של $\beta_1$ גדול יותר מהאומד המקורי.
טענה ב': אם המחיר נמדד ב-$\ln$ והכמות הנמכרת נמדדת גם כן ב-$\ln$, אזי אומדן האפקט של מחיר על הכמות לא משתנה עם יחידות המדידה.
טענה ג': אם הכמות נמדדת באלפים, אזי האומדן החדש של $\beta_1$ גדול יותר מהאומד המקורי.

א. רק טענות א' ו-ב' נכונות.
ב. רק טענות א' ו-ג' נכונות.
ג. כל הטענות נכונות.
ד. רק טענה ב' נכונה.

**6)** חוקר טוען שאומד אפקט סיבתי של משתנה מסביר $x$ על משתנה מוסבר $y$ בעזרת שיטת OLS ברגרסיה לינארית מרובה.
טענה א': החוקר מניח שאין מתאם בין ההפרעה לבין המשתנה המסביר.
טענה ב': החוקר מניח שלמשתנה המסביר יש אפקט קבוע על פני ערכים של $x$.
טענה ג': אם $R$ בריבוע ברגרסיה קרוב ל-0 האפקט אינו סיבתי.

א. רק טענה א' נכונה.
ב. רק טענות א' ו-ב' נכונות.
ג. כל הטענות נכונות.
ד. רק טענה ב' נכונה.

**7)** נתון המודל הבא: $y=\beta_0+\beta_1 x+u$.

חוקר אמד את המודל ב-OLS וקיבל $\beta_0$ ו- $\beta_1$.

טענה א': אם: $\beta_1 \neq \beta_1$, אזי האומדן של $\beta_1$ מוטה.

טענה ב': אם: $Var(\beta_1)=Var(\beta_1)$, אזי האומדן של $\beta_1$ יעיל.

טענה ג': אם: $\sum_{i=1}^{n} u_1=0$, אזי האומדן של $\beta_1$ לא מוטה.

א. אין טענות נכונות.
ב. רק טענות א' ו-ב' נכונות.
ג. כל הטענות נכונות.
ד. רק טענה א' נכונה.

**8)** ברגרסיה: $y=\beta_0+\beta_1 x_1+\beta_2(x_2+x_3)+u$.

טענה א': יש מולטיקולינאריות מושלמת ולכן אי אפשר לאמוד את $\beta_2$.

טענה ב': אי אפשר לאמוד את האפקט של $x_2$ על $y$.

טענה ג': האומד של סטיית התקן של $\beta_2$ יהיה גדול כך ש- $\beta_2$ לא יהיה מובהק.

א. רק טענה ב' נכונה.
ב. רק טענות א' ו-ב' נכונות.
ג. רק טענה א' נכונה.
ד. כל הטענות נכונות.

**9)** בהתייחס למחקר 1 החוקרים התבקשו לחזות את משקל היילוד עבור שתי נשים הזהות בכל הנתונים למעט בכמות הסיגריות שהן מעשנות, כאשר אישה א' מעשנת סיגריה אחת לשלושה ימים (שליש סיגריה ליום) ואילו אישה ב' מעשנת 20 סיגריות ביום.

א. מידת הדיוק של התחזית של אישה א' תהיה גבוהה יותר ממידת הדיוק של התחזית עבור אישה ב'.
ב. מידת הדיוק של התחזית של אישה ב' תהיה גבוהה יותר ממידת הדיוק של התחזית עבור אישה א'.
ג. התחזיות עבור שתי הנשים הן בעלות מידת דיוק זהה.
ד. כיוון שה- $R^2_{adjusted}$ של הרגרסיה בה נכללה כמות הסיגריות הוא נמוך, לא ניתן לחזות את משקל היילוד.

**10)** מרצה בחר מספר סטודנטים מהקורס שלו ובאופן אקראי סידר להם שיעורים פרטיים. אחר כך הוא אמד ב-OLS את המודל: $y = \beta_0 + \beta_1 x + u$.
כאשר $y$ מייצג את הציון הסופי שכל סטודנט קיבל בקורס ו-$x$ אם הסטודנט קיבל שיעורים פרטיים.

טענה א': $\hat{\beta}_1$ הוא אומד סיבתי של האפקט של $x$ על $y$.
טענה ב': רק אם מאפייני הסטודנטים כמו מין, גיל, מוטיבציה וכישרון נכנסים לרגרסיה כמשתנים מסבירים נוספים, אזי $\hat{\beta}_1$ הוא אומד סיבתי של האפקט של $x$ על $y$.
טענה ג': המדגם אינו מייצג ולכן $\hat{\beta}_1$ הוא לא אומד סיבתי של האפקט של $x$ על $y$.

א. רק טענה א' נכונה.
ב. רק טענות ב' ו-ג' נכונות.
ג. רק טענה ב' נכונה.
ד. רק טענה ג' נכונה.

**11)** חוקר אמד רגרסיה וקיבל: $\ln(wage) = 0.13 + 0.09educ + 0.04\exp er - 0.0007\exp er^2$.
כאשר $wage$ שכר לשעה בדולרים, $educ$ שנות השכלה ו-$\exp er$ שנות ניסיון עבודה.
טענה א': השכר המנובא לאדם בעל 12 שנות השכלה ו-17 שנות ניסיון הוא 1.6877.
טענה ב': הגמישות של השכר ביחס לשנות ניסיון לאדם בעל 16 שנות השכלה ו-5 שנות ניסיון הוא 0.165.
טענה ג': לניסיון יש השפעה שולית שלילית לאדם בעל 30 שנות ניסיון.

א. רק טענות ב' ו-ג' נכונות.
ב. כל הטענות נכונות.
ג. רק טענה ב' נכונה.
ד. רק טענות א' ו-ב' נכונות.

**12)** חוקר אמד רגרסיה בעזרת 25 תצפיות וקיבל:
$$\underset{(0.62)}{\hat{y} = 7.54} - \underset{(0.06)}{0.25x}$$
, כאשר המספרים בסוגריים מראים את טעויות התקן וידוע כי: $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 125$.
טענה א': המקדם של $x$ מובהק ברמת מובהקות של 1%.
טענה ב': אנו דוחים כי המקדם של $x$ גדול מ-0 ברמת מובהקות של 5%.
טענה ג': אין מספיק נתונים כדי לחשב: $SSR = \sum_{i=1}^{n} u_i^2$.

א. רק טענה א' נכונה.
ב. רק טענה ב' נכונה.
ג. רק טענות ב' ו-ג' נכונות.
ד. רק טענות א' ו-ג' נכונות.

**13)** חוקרת רוצה לאמוד את המודל: $y=\beta_0+\beta_1 x+u$.
אבל בגלל סיבות חיצוניות היא מדדה במקום $x$ את משתנה: $z=100+x, z$,
כך שהיא קיבלה: $\hat{y}=10+5z$.
אם היא הייתה אומדת את המודל המקורי, היא הייתה מקבלת כי:
טענה א': $\beta_0=10$.
טענה ב': $\beta_1=5$.
טענה ג': $\beta_1=500+x$.
א. רק טענה ב' נכונה.
ב. רק טענות א' ו-ג' נכונות.
ג. רק טענה א' נכונה.
ד. רק טענה ג' נכונה.

**14)** חוקר אמד את המודל: $y=\beta_0+\beta_1 x_1+\beta_2 x_2+u$ והוא קיבל:
$$\underset{(0.2309)}{\hat{y}=-1.06}+\underset{(0.0877)}{1.001x_1}+\underset{(0.016)}{0.5x_2}$$
, כאשר המספרים בסוגריים מראים את טעויות
התקן, מספר התצפיות 524 ו-$SSR=116.7$. ברמת מובהקות של 5%.
טענה א': אנו דוחים ש-$\beta_1=-1$.
טענה ב': אם: $Cov\left(\hat{\beta}_1,\hat{\beta}_2\right)=0$ אנו דוחים ש-$\beta_1+\beta_2=1$.
טענה ג': אם: $Cov\left(\hat{\beta}_1,\hat{\beta}_2\right)=0$ אנו דוחים ש-$\beta_1-\beta_2=0$.
א. רק טענות א' ו-ב' נכונות.
ב. רק טענות ב' ו-ג' נכונות.
ג. רק טענה א' נכונה.
ד. כל הטענות נכונות.

**15)** אנו רוצים לאמוד את שני המודלים הבאים:
$$Savings=\alpha_0+\alpha_1 Income+u$$
$$Consumption=\beta_0+\beta_1 Income+v$$
כאשר ידוע כי: $Consumption+Savings=Income$.
טענה א': $\alpha_0+\beta_0=0$.
טענה ב': $\alpha_1+\beta_1=1$.
טענה ג': $\alpha_1=\beta_1$.
א. רק טענות א' ו-ב' נכונות.
ב. רק טענה א' נכונה.
ג. רק טענות ב' ו-ג' נכונות.

**16)** חוקרת אמדה שלוש רגרסיות:

$$y = \beta_0 + \beta_1 x + \beta_2 z + u$$

$y = \alpha_0 + \alpha_1 z + v$ וחישבה את השאריות: $\hat{w}$ , $\hat{v}$ , $\hat{u}$.

$$x = \delta_0 + \delta_1 z + w$$

טענה א': באמידת הרגרסיה הליניארית: $\hat{u} = \lambda_0 + \lambda_1 x + \mu$ , $\lambda_1 = 0$.

טענה ב': באמידת הרגרסיה הליניארית: $\hat{v} = \lambda_0 + \lambda_1 \hat{w} + \mu$ , $\lambda_1 = \hat{\beta}_1$.

טענה ג': באמידת הרגרסיה הליניארית: $\hat{w} = \lambda_0 + \lambda_1 \hat{u} + \mu$ , $\lambda_1 = 0$.

א. רק טענות א' ו-ב' נכונות.
ב. רק טענה ב' נכונה.
ג. רק טענה א' נכונה.
ד. כל הטענות נכונות.

**17)** נתון המודל הליניארי: $y = \beta_x + u$ כאשר כל ההנחות הקלאסיות מתקיימות.

נתונים שני אומדנים: $b_2 = \frac{\bar{y}}{\bar{x}}$ , $b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$. עם שגיאות האמידה $\hat{u}_1$ ו- $\hat{u}_2$ בהתאם.

טענה א': שני האומדים הינם חסרי הטיה.
טענה ב': השונות של $b_1$ גדולה מזו של $b_2$.
טענה ג': $\left|\sum_{i=1}^{n} \hat{u}_{1,i}\right| \neq \left|\sum_{i=1}^{n} \hat{u}_{2,i}\right|$.

א. רק טענות א' ו-ג' נכונות.
ב. רק טענה א' נכונה.
ג. כל הטענות נכונות.
ד. רק טענות ב' ו-ג' נכונות.

**18)** חוקר אמד את המודל הבא: $\ln(y) = \beta x + u$ כאשר $y$ מייצג צריכה ב-₪ ו- $x$ הכנסה ב-₪ עבור 30 פרטים. נתון כי: $\sum_{i=1}^{n} x_i^2 = 310969.94$ , $\sum_{i=1}^{n} \ln(y_i) = 127.256$ ,

$\sum_{i=1}^{n} \ln(y_i)^2 = 540.6$ , $\sum_{i=1}^{n} x_i \ln(y_i) = 12882.057$ , $\sum_{i=1}^{n} x_i = 3019.1885$.

מה יהיה אומד OLS לנטייה השולית לצרוך במודל (כלומר: $\frac{\partial Y}{\partial X}$) אם הצריכה שווה ל-100 ₪?

א. 4.14.
ב. 0.029.
ג. 2382.
ד. 1.05.

## תשובות סופיות:

| | | | | |
|---|---|---|---|---|
| 1) א' | 2) א' | 3) א' | 4) א' | 5) א' |
| 6) א' | 7) א' | 8) א' | 9) א' | 10) א' |
| 11) א' | 12) א' | 13) א' | 14) א' | 15) א' |
| 16) ד' | 17) א' | 18) א' | | |

# מבוא לאקונומטריקה

## פרק 30 - פתרון מודרך של מבחן מה - 11.02.2020

תוכן העניינים

# פתרון מודרך של מבחן מה – 11.02.2020:

## שאלות:

**1)** חוקר אמד את המודל: $Y = B_0 + B_1X_1 + u$ על סמך מדגם של 150 תצפיות

וקיבל: $\underset{(0.5)}{\hat{Y} = 40.5} + \underset{(8.1)}{0.8X_1}$ (בסוגריים סטיית תקן).

אחרי האמידה התברר לחוקר כי עקב טעות הקלדה כל ערכי ה- $X$ ים במדגם היו גדולים ב-100 מהערך האמיתי שלהם וכי היה עליו להריץ את המודל: $Y = \gamma_0 + \gamma_1X_2 + u$ כאשר: $X_2 = X_1 - 100$.

טענה A: $\hat{\gamma}_1 = 0.8$.

טענה B: $\hat{\gamma}_0 = 120.5$.

א. שתי הטענות נכונות.
ב. רק טענה A נכונה.
ג. רק טענה B נכונה.
ד. שתי הטענות אינן נכונות.

**2)** המודל: $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + u$ נאמד לפי OLS.

טענה A: $\sum \hat{u}_i = 0$.

טענה B: רק אם באוכלוסייה: $E(u_i) = 0$ אזי: $\sum \hat{u}_i = 0$.

טענה C: אם בנתוני המדגם: $X_1 = X_2 + 1$ אזי: $\sum \hat{u}_i = 0$.

א. רק טענה A נכונה.
ב. כל הטענות לא נכונות.
ג. רק טענה A ו-C נכונות.
ד. רק טענה C נכונה.

**3)** חוקר הניח שפונקציית הייצור היא מסוג קוב-דאגלאס:

$\ln Q = \alpha + \beta_1 \ln L + \beta_2 \ln K + u$.

$Q$ - תפוקה.

$L$ - עבודה.

$K$ - הון.

להלן תוצאות האמידה לפי מדגם מקרי של 87 משקים (בסוגריים - סטיות תקן):

(1) $\ln Q = \hat{\alpha} + \underset{(0.09)}{0.4 \ln L} + \underset{(0.1)}{0.54 \ln K}$

(2) $\ln K = const. + \underset{(0.00585)}{0.072 \ln L}$

טענה A: מקדם ההסבר $R^2$ בין $\ln L$ ו- $\ln K$ הוא 0.64.
טענה B: אם החוקר היה משמיט מרגרסיה (1) את $\ln L$ אזי גמישות הייצור ביחס להון הייתה גדלה.
טענה C: אם החוקר היה משמיט מרגרסיה (1) את $\ln L$ אזי גמישות הייצור ביחס להון הייתה קטנה.

א. טענות A ו-B נכונות.
ב. רק טענה B נכונה.

4) חוקר אמד את המשוואה הבאה: $Y_i = \beta X_i + u_i$.
טענה א': רק אם: $\bar{X} = 0$ קו הרגרסיה יעבור דרך נקודת הממוצעים.
טענה ב': מודל זה מקיים: $SST = SSR + SSE$.
טענה ג': רק אם: $\bar{Y} = 0$ קו הרגרסיה יעבור דרך נקודת הממוצעים.

א. כל הטענות אינן נכונות.
ב. רק טענה א' נכונה.
ג. רק טענה ב' נכונה.
ד. כל התשובות האחרות אינן נכונות.

5) החוקר אמד את מודל הרגרסיה הבא על סמך שלושה מדגמים בגודל קבוע של 100 תצפיות (הריץ 3 פעמים את הרגרסיה): $Y_i = \alpha + u_i$.

א. $R^2 = 0$ בכל המדגמים.
ב. $R^2 > 0$ רק במדגמים בהם: $\bar{Y} > 0$.
ג. $R^2 = 1$ בכל המדגמים.
ד. אף תשובה לא נכונה.

6) הנח כי המודל הנכון הוא: $Y = \theta + \delta X + v$.
החוקר אמד את המודל הבא בשיטת OLS: $Y = \alpha + \beta X + \gamma Z + u$.
והנח שכל ההנחות הקלאסיות מתקיימות, לרבות: $E(v|X,Z) = 0$.

טענה א': $E(\hat{\gamma}) = 0$.

טענה ב': $V(\hat{\alpha}) = V(\hat{\theta})$.

טענה ג': אם ידוע כי $R^2$ ברגרסיה בין $X$ ל- $Z$ (עם חותך) שווה ל-0, אזי: $V(\hat{\beta}) = V(\hat{\delta})$.

א. רק טענות א' וג' נכונות.
ב. כל הטענות נכונות.
ג. רק טענה א' נכונה.
ד. כל התשובות האחרות אינן נכונות.

**7)** בהתבסס על פלט 1:

ידוע כי $X$ מייצג את רמת הפרסום (באלפי ₪) ואילו $Y$ מייצג את הרווחים (במאות אלפי ₪). בהינתן תוצאות האמידה:

א. אם ההשקעה בפרסום עומדת כיום על 3,000 ₪, כדאי להגדיל אותה, אבל רק עד לרמה של 5,014 ₪.

ב. השפעת הפרסום על הרווחים היא תמיד חיובית.

ג. אם ההשקעה בפרסום עומדת כיום על 6,000 ₪ כדאי להגדיל אותה, אבל רק עד לרמה של 10,028 ₪.

ד. אף תשובה לא נכונה.

**8)** בהתבסס על פלט 1:

ידוע כי $X$ מייצג את רמת הפרסום (באלפי ₪) ואילו $Y$ מייצג את הרווחים (במאות אלפי ₪). פרסומאי מעלה השערה שההשפעה השולית של הפרסום על הרווחים קבועה.

טענה א': כדי לבדוק את השערת הפרסומאי חייבים להריץ מודל מוגבל.

טענה ב': ההשערה נדחית ברמת מובהקות של 5%.

טענה ג: $F$ סטטיסטי לבדיקת השערת הפרסומאי הוא 24,567.

א. רק טענות ב' ו-ג' נכונות.

ב. רק טענה א' נכונה.

ג. רק טענה ב' נכונה.

ד. כל התשובות האחרות שגויות.

**9)** עבור מודל רגרסיה בפלט 2:

ידוע כי $X$ מייצג את רמת הפרסום (באלפי ₪) ואילו $Y$ מייצג את רמת המכירות (במאות אלפי ₪). בהינתן תוצאות האמידה.

טענה א': ל- $X$ יש השפעה שולית חיובית על $Y$.

טענה ב': בהינתן תוצאות האמידה, רמת המכירות המרבית (מקסימלית) הינה נמוכה מ-1,811,000 ₪.

טענה ג': ל- $X$ יש השפעה שולית שלילית על $Y$.

א. רק טענות א' ו-ב' נכונות.

ב. רק טענה א' נכונה.

ג. רק טענה ג' נכונה.

ד. כל התשובות האחרות שגויות.

**10)** ברגרסיה המרובה הבאה: $Y=\beta_0+\beta_1X_1+\beta_2X_2+u$, איזו טענה נכונה?

א. $\hat{\beta}_0=\bar{Y}-\hat{\beta}_1\bar{X}_1-\hat{\beta}_2\bar{X}_2$.

ב. $\hat{\beta}_0=-\hat{\beta}_1\bar{X}_1-\hat{\beta}_2\bar{X}_2$.

ג. $\hat{\beta}_0=\dfrac{Cov(Y,X_1)+Cov(Y,X_2)}{\sqrt{Var(X_1)+Var(X_2)}}$.

ד. אף טענה לא נכונה.

**11)** יועץ הציע לחברה מסוימת להעלות את המחיר של המוצר שהיא מייצרת $p$ באחוז אחד כי המכירות $q$ ירדו ב-0.3 אחוזים בלבד ללא תלות במחיר.
איזה מודל הכי עקבי עם המודל וההצעה של היועץ?

א. $\ln q=\beta_0+\beta_1\ln p+u$, $\hat{\beta}_1=-0.3$.

ב. $q=\beta_0+\beta_1p+u$, $\hat{\beta}_1=-3$.

ג. $q=\beta_0+\beta_1\ln p+u$, $\hat{\beta}_1=-0.3$.

ד. $\ln q=\beta_0+\beta_1\ln p+u$, $\hat{\beta}_1=-3$.

**12)** אנו יודעים שלוג-השכר $\ln w$ הוא פונקציה ליניארית של שנות ההשכלה $e$ ושל רמת המוטיבציה $m$, כך שהמודל האמיתי הוא: $\ln w=\beta_0+\beta_1e+\beta_2m+u$ כשהמודל מתקיים את ההנחות של מודל הרגרסיה הקלאסי. לחוקרת לא היו נתונים על מוטיבציה ולכן היא אמדה את המודל הבא: $\ln w=\gamma_0+\gamma_1e+v$.
אם אנחנו יודעים שמוטיבציה מתואמת באופן חיובי עם שנות ההשכלה ועם השכר, איזו טענה נכונה:

א. $E\left[\hat{\beta}_1\right]<E\left[\hat{\gamma}_1\right]$.

ב. $E\left[\hat{\beta}_1\right]>E\left[\hat{\gamma}_1\right]$.

ג. $E\left[\hat{\beta}_1\right]=E\left[\hat{\gamma}_1\right]$.

ד. $\hat{\gamma}_1$ יהי מובהק.

**13)** המודל שמגדיר את אחוז המהגרים $Y$ במדינה מסוימת הוא:
$Y=\beta_0+\beta_1\ln X+u$ כש- $X$ מסמן את השכר הממוצע במדינה וכשהמודל מקיים את ההנחות הקלאסיות. חוקר יצירתי טוען שגם מחיר הגלידות $P$ משפיע על $Y$, אבל טענתו אינה נכונה. החוקר אמד את המודל הבא: $Y=\gamma_0+\gamma_1\ln X+\gamma_2 P+u$. החוקר מצא כי $\gamma_1$ אינו מובהק ברמת מובהקות של 5 אחוזים. איזו טענה תמיד נכונה:

א. $E[\hat{\gamma}_1]=\beta_1$.

ב. לא ייתכן כי $\hat{\beta}_1$ תהיה מובהק ברמת מובהקות של אחוז אחד.

ג. $\hat{\gamma}_1=\hat{\beta}_1$.

ד. $\hat{\gamma}_2=0$.

**15)** במדינה מסוימת התושבים מוציאים את כל הכנסתם $Y$ על רכישת אוכל $F$ וחינוך $E$ בלבד. חוקר אסף מדגם מייצג של התושבים והריץ כמה רגרסיות. איזו טענה נכונה?

א. באמידת הרגרסיה: $Y=\beta_0+\beta_1F+u$, $\hat{\beta}_1$ אומד מוטה של $\beta_1$.

ב. באמידת הרגרסיה: $Y=\beta_0+\beta_1F+\beta_2E+u$, $E\left[\hat{\beta}_1\right]=E\left[\hat{\beta}_2\right]$.

ג. באמידת הרגרסיה: $Y=\beta_0+\beta_1F+u$, $E\left[\hat{\beta}_1\right]=1$.

ד. באמידת הרגרסיה: $Y=\beta_0+\beta_1F+\beta_2E+u$, $R^2<1$.

**16)** חוקרת אמדה מודל ב-OLS במדגם מייצג גדול וקיבלה: $Y=300+50\ln X+u$ כאשר $Y$ מייצג את הצריכה ו- $X$ מייצג את ההכנסה. מה תהיה הנטייה השולית לצרוך, כלומר: $\frac{\partial Y}{\partial X}$, לפרט שצריכתו היא 100 והכנסתו היא 200?

א. 0.25.

ב. 50.

ג. 25.

ד. 0.5.

**17)** חוקרת אמדה את הרגרסיה הבאה: $Y = \beta_0 + \beta_1 X + \beta_2 Z + u$ על סמך מדגם של 1000 תצפיות כשכל ההנחות הקלאסיות מתקיימות וקיבלה:

$$\hat{Y} = 0.5 - 9X + 0.2Z$$
$$(5.6) \quad (1.2) \quad (0.01)$$

המספרים בסוגריים מייצגים את סטיות התקן.

איזו תשובה נכונה בוודאות לגבי המודלים הנ"ל אם: $R^2 = 0.8$?

א. ה- $F$ סטטיסטי של מובהקות הרגרסיה גדול מ-3.

ב. ניתן לדחות את ההשערה כי: $\beta_1 = \beta_2$.

ג. ניתן לדחות את ההשערה כי: $\beta_1 \le -7$ ברמת מובהקות של 0.01.

ד. כל התשובות האחרות אינן נכונות.

## תשובות סופיות:

| | | | | |
|---|---|---|---|---|
| 1) א' | 2) א' | 3) א' | 4) א' | 5) א' |
| 6) א' | 7) א' | 8) א' | 9) א' | 10) א' |
| 11) א' | 12) א' | 13) א' | 15) א' | 16) א' |
| 17) א' | | | | |