מבוא לאקונומטריקה

פרק 19 - שאלות חזרה למבחן מבוססות על תוכנת STATA

תוכן העניינים

1. כללי 1

שאלות חזרה למבחן מבוססות על תוכנת STATA:

שאלות:

1) כדי לבדוק האם יש קשר בין שכר המורים וההוצאה לתלמיד בבי"ס ציבוריים נאמד המודל הבא: $Pay_i = \beta_1 + \beta_2 Spend + u_i$. כאשר pay הוא שכר המורים ו-Spend מייצג את ההוצאה לתלמיד שנמדדו בדולרים לשנה:

```
. reg pay  spending
```

Source	SS	df	MS
Model	608555015	1	608555015
Residual	264825250	49	5404596.94
Total	873380265	50	17467605.3

Number of obs	51
F(1, 49)	112.60
Prob > F	0.0000
R-squared	0.6968
Adj R-squared	0.6906
Root MSE	2324.8

pay	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
spending	3.307585	.3117043	10.61	0.000	2.681192	3.933978
_cons	12129.37	1197.351	10.13	0.000	9723.204	14535.54

א. פרשו את הרגרסיה (רמז: משמעות האומדים לפרמטרים, האם התוצאות מובהקות?). האם יש הגיון כלכלי לתוצאות שהתקבלו?

ב. מצאו תחזית נקודתית לשכר בממוצע אם ההוצאה לתלמיד היא 5000$.

ג. נבדקו מדדים תיאוריים לגבי המשתנה הב"ת:

```
. su  spending
```

Variable	Obs	Mean	Std. Dev.	Min	Max
spending	51	3696.608	1054.761	2297	8349

ד. הסבירו את משמעות ההנחות הקלאסיות במונחי המשתנים.

ה. האם ההוצאה הממוצעת לתלמיד מסבירה הרבה משונות שכר המורים?

ו. מהו הפער החזוי בשכר המורים בין שני בתי ספר שהפער בהוצאה לתלמיד ביניהם הוא 20$?

ז. האם ניתן לומר על סמך התוצאות כי המודל מובהק? אם כן, באיזה רמת מובהקות?

ח. חוקר ביצע מבחן שמטרתו לבדוק האם המודל יוצא מראשית הצירים:

i. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.

ii. ערך t סטטיסטי לבדיקת המבחן.

iii. מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת סמך של 95%.

ט. חוקר ביצע מבחן שמטרתו לבדוק כי עבור עליה בדולר אחד בהוצאה השנתית הממוצעת לתלמיד שכר המורים הממוצע עולה בלפחות 3 דולר:

i. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.

ii. ערך t סטטיסטי לבדיקת המבחן.

iii. ערך t קריטי לבדיקת המבחן.

iv. מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת ביטחון של 95%.

י. חשבו רווח בר סמך לאמידת השיפוע ברמת סמך של 90%.

יא. חוקר רצה לאמוד את המשוואה בשקלים במקום בדולרים ולכן אמד את המשוואה הבאה: $\hat{pay}1 = \alpha_0 + \alpha_1 spend1$.

הניחו כי שער הדולר עומד על 3.5 ₪.

i. מהי הפקודה ב-stata ליצירת המשתנים החדשים?

ii. מה יהיה האומדן ל- α_1 ? מה יהיה האומדן לסטיית התקן של α_1 ?

על אותם הנתונים נאמדה גם המשוואה הבאה:

$Pay_i = \beta_0 + \beta_1 Spend + \beta_2 Spend2 + u_i$, כאשר: $spend2 = spend^2$.

יב. כתוב את הפקודה ב-STATA ליצירת המשתנה החדש.

יג. מתוצאות האמידה התקבל ש:

3. $\hat{pay}_i = 23128.44 + 1.223 spend_i - 0.0000635 spend2$

(9.55) (10.55) (8.78)

הערכים בסוגריים הם ערכי t סטטיסטי.

השערת האפס לבדיקה האם יש השפעה ליניארית של ההוצאה על התשלום למורים? T סטטיסטי?

מסקנת הבדיקה: דוחים/לא דוחים את H_0 .

יד. מהי הפקודה ב-STATA לבדיקת המתאם בין שני המשתנים הב"ת?

טו. בהינתן תוצאות המודל כיצד נראה פרופיל השכר כפונקציה של ההוצאה על התלמיד (אין צורך בציור מדויק רק במגמה).

טז. מהי ההוצאה לתלמיד שאחריה שכר המורים מתחיל לרדת?

יז. התקבל מתאם של 0.97 בין שני המשתנים הב"ת. לנוכח המתאם הגבוה החוקר טען כי האומדים במשוואה (3) הם מוטים.

הטענה נכונה/הטענה אינה נכונה.

יח. לאור תוצאות האמידה של משוואה (3) ניתן להסיק כי האומדים של משוואה (1) הינם מוטים.

הטענה נכונה/הטענה איננה נכונה.

יט. הניחו כי יש הטרוסקדסטיות במשוואה (3).

מה יהיו ההשלכות לגבי האומדים של המשוואה:

i. האומדים יהיו חסרי הטיה ויעילים.

ii. האומדים יהיו חסרי הטיה אך לא יעילים.

iii. האומדים יהיו מוטים אך יעילים.

iv. האומדים יהיו מוטים ולא יעילים.

2) נתונים 2 המודלים הבאים:

1. $Y_t = \alpha + \beta X_t + u_t$.
2. $Y_t = \beta X_t + u_t$.

א. בהנחה כי R^2 של המודל הראשון גבוה מזה של המודל השני, האם ניתן לומר שהמודל הראשון טוב יותר?
הניחו כי מודל (1) הוא המודל האמיתי.

ב. מהן תכונות האומד לשיפוע של מודל (2)? מוטה/חסר הטיה?

ג. האם ניתן לומר כי האומד הינו עקיב?

ד. חוו דעתכם על הטענה: "על סמך משפט גאוס מרקוב, ניתן להסיק כי אומד הריבועים הפחותים של משוואה (1) הינו יעיל יותר מאומד הריבועים הפחותים של משוואה (2)". הסבירו את תשובתכם.

ה. חשבו את שונות האומד.

ו. לאיזה משני המודלים יהיה אומד לשיפוע גבוה יותר? לראשון/לשני/לא ניתן לדעת.

3) על מנת לאמוד את הקשר שבין השכלה (בשנים) להכנסה (באלפי שקלים) נאמדו שני המודלים הבאים:

```
. reg wage educ
```

Source	SS	df	MS
Model	7888.51144	1	7888.51144
Residual	31092.9858	998	31.1552964
Total	38981.4972	999	39.0205177

Number of obs	1000
F(1, 998)	253.20
Prob > F	0.0000
R-squared	0.2024
Adj R-squared	0.2016
Root MSE	5.5817

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.138517	.0715497	15.91	0.000	.998112	1.278922
_cons	-4.912181	.9667875	-5.08	0.000	-6.80935	-3.015011

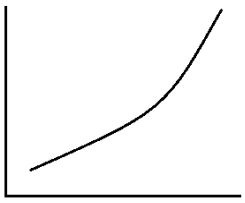
```
. reg lwage educ
```

Source	SS	df	MS
Model	65.5213155	1	65.5213155
Residual	239.767622	998	.240248118
Total	305.288937	999	.305594532

Number of obs	1000
F(1, 998)	272.72
Prob > F	0.0000
R-squared	0.2146
Adj R-squared	0.2138
Root MSE	.49015

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1037608	.0062831	16.51	0.000	.0914313	.1160904
_cons	.7883743	.0848975	9.29	0.000	.6217761	.9549724

א. ידוע כי ניתן לתאר את הקשר בין שכר להשכלה על ידי הגרף הבא:



באיזה מודל כדאי לבחור? תנו 3 נימוקים לבחירתכם.

ב. מהי התשואה להשכלה על סמך המודל הנבחר?

ג. על פי מודל (1), מהי גמישות השכר ביחס להשכלה בנקודת הממוצעים: (7,12)?

ד. על בסיס נתוני המדגם חושב השכר הממוצע עבור התצפיות שלהן 16 שנות השכלה.

```
. su  wage if educ==16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	186	13.30328	7.575015	2.54	60.19

הממוצע הוא 13.303.

שימו לב: סימן ה- "=" משמש להשוואה בין משתנים (למשל: gen price_new= price*1.2), לעומת זאת, סימן ה- "==" משמש להגדרת ערך משתנה (l educ if educ == 8).

i. מהי התחזית לשכר לעובד עם 16 שנות השכלה עבור כל אחד מהמודלים?

ii. איזה מהמודלים נותן תחזית נקודתית מדויקת יותר:
המודל הראשון/המודל השני/ לא ניתן להשוות.

ה. איזה מודל מסביר חלק גדול יותר של השונות של המשתנה התלוי:
המודל הראשון/המודל השני/לא ניתן להשוות.

ו. לפי האומדים בפלט (לא לפי מבחן סטטיסטי) התשואה להשכלה חיובית בהכרח: נכון/לא נכון.

ז. רצו לבחון כיצד השכר מושפע גם מוותק העובד ולכן הכניסו את המשתנה: log(vetek). מה משמעות מקדם השיפוע של המשתנה החדש במודל הראשון ובמודל השני.

ח. בהתייחס למודל השני תוצאות האמידה מראות כי הן התשואה להשכלה והן גמישות הוותק אינם מובהקים. יחד עם זאת הרגרסיה עם שני המשתנים הב"ת יצאה מובהקת. כיצד ניתן להסביר זאת? האם מומלץ להשמיט את שני המשתנים ביחד מהמודל (לאמוד את השכר באמצעות משתנים אחרים)? האם מומלץ לאמוד את המודל ללא משתנה הוותק?

4) על מנת לבחון את פונקציית הייצור של אורז נאמד המודל הבא:

$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 \ln(LABOR) + \beta_4 \ln(FERT) + \varepsilon_i$, כאשר:

PROD - כמויות אורז מדושן (נמדד בטונות).

AREA - גודל החלקות בהם האורז נשתל (נמדד בעשרות דונמים).

LABOR - סה"כ ימי עבודה של עובדים ובני משפחה (של החקלאי).

FERT - כמויות דשן בשימוש (נמדד בק"ג).

להלן תוצאות האמידה:

```
. reg lprod larea llabor lfert
```

Source	SS	df	MS
Model	226.084875	3	75.361625
Residual	40.5653554	348	.116567113
Total	266.65023	351	.759687266

Number of obs	352
F(3, 348)	646.51
Prob > F	0.0000
R-squared	0.8479
Adj R-squared	0.8466
Root MSE	.34142

lprod	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
larea	.3617359	.0639678	5.65	0.000	.2359237	.4875481
llabor	.4328479	.0668825	6.47	0.000	.301303	.5643928
lfert	.2095023	.0382654	5.47	0.000	.1342417	.2847628
_cons	-1.546786	.2556536	-6.05	0.000	-2.049607	-1.043966

א. בחנו את ההשערה כי גמישות הייצור ביחס לגודל החלקות (AREA) שווה לגמישות ביחס לימי העבודה (LABOR). השתמשו ברמת מובהקות של 5%, נסחו את ההשערה בצורה פורמאלית ודווחו את התוצאות תוך שימוש בפלט הבא:

```
. test  larea= llabor

 ( 1)  larea - llabor = 0

       F(  1,    348) =    0.34
            Prob > F =    0.5592
```

ב. כתבו את הפקודות ב-STATA לאמידת הרגרסיה המוגבלת מהסעיף הקודם.

ג. כעת, בחנו את ההשערה המורכבת מההשערה של סעיף א' + ההשערה כי פונקציית הייצור מקיימת תק"ל ($\beta_2 + \beta_3 + \beta_4 = 1$), תוך שימוש בפלט הבא:

```
. gen l_pr_fe=log( prod/ fert)

. gen l_ar_la_fe=log( area* labor/fert^2)

. reg  l_pr_fe  l_ar_la_fe

      Source |       SS       df       MS              Number of obs =     352
-------------+------------------------------           F(  1,   350) =  439.63
       Model |  51.0075377     1  51.0075377            Prob > F      =  0.0000
    Residual |  40.6079092   350  .116022598           R-squared     =  0.5568
-------------+------------------------------           Adj R-squared =  0.5555
       Total |  91.6154469   351  .261012669           Root MSE      =  .34062

------------------------------------------------------------------------------
   l_pr_fe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
l_ar_la_fe |   .3940824    .018795    20.97   0.000     .3571172    .4310477
     _cons |  -1.402958   .0913195   -15.36   0.000    -1.582562   -1.223354
------------------------------------------------------------------------------

.
```

באיזה מבחן אתם משתמשים?

נסחו את ההשערה, את הסטטיסטי והקריטי.

מהי מסקנתכם?

5) חוקרים ביקשו לאמוד את פונקציית החיסכון המצרפי במשק הישראלי.
מפתח שמות משתנים:
GDS87 - חסכון מקומי גולמי.
GDP87 - תוצר מקומי גולמי.
GC87 - הוצאות הממשלה.
כל הנתונים הינם במיליוני ₪ במחירים קבועים של שנת 1987.
נאמדו 2 המודלים הבאים:

```
. reg  gds87  gdp87
```

Source	SS	df	MS
Model	268713647	1	268713647
Residual	60820138.1	24	2534172.42
Total	329533785	25	13181351.4

Number of obs	26
F(1, 24)	106.04
Prob > F	0.0000
R-squared	0.8154
Adj R-squared	0.8077
Root MSE	1591.9

gds87	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp87	.1852456	.0179896	10.30	0.000	.1481169	.2223742
_cons	-2624.027	1018.32	-2.58	0.017	-4725.737	-522.3182

```
. reg  gds87 gdp87 gc87
```

Source	SS	df	MS
Model	307085221	2	153542611
Residual	22448563.7	23	976024.509
Total	329533785	25	13181351.4

Number of obs	26
F(2, 23)	157.31
Prob > F	0.0000
R-squared	0.9319
Adj R-squared	0.9260
Root MSE	987.94

gds87	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp87	.2955591	.0208369	14.18	0.000	.2524547	.3386635
gc87	-.7784411	.1241513	-6.27	0.000	-1.035268	-.5216146
_cons	4217.674	1260.961	3.34	0.003	1609.177	6826.171

א. כתבו את המודל האקונומטרי שנאמד בכל אחת מהאמידות.

ב. איזה מבין שני המודלים הקודמים הייתם מעדיפים? למה?

ג. מהו ההבדל בין המשמעות של האומדן למקדם המשתנה *GDP87* בשני המודלים?

ד. ביחס לאומד במשוואה 2 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
מוטה כלפי מעלה/מוטה כלפי מטה/ חסר הטיה/לא ניתן לדעת

ה. ביחס לאומד במשוואה 3 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
יעיל/לא יעיל/ לא ניתן לדעת

ו. בהינתן תוצאות משוואה 1 הקורלציה בין שני המשתנים הב"ת היא:
חיובית/שלילית/אפס/לא ניתן לדעת

ז. החוקר החליט להוסיף משתנה המודד את הצריכה הממשלתית במיליוני דולרים במקום בשקלים: $GC87 ואמד את המשוואה:
3. $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + \beta_4 \$GC87 + u_t$.
האומדים של משוואה 3 יהיו:
חסרי הטיה ויעילים/מוטים ולא יעילים/חסרי הטיה אך לא יעילים/לא מוגדרים.

ח. החוקר החליט להוסיף מדד נוסף לתוצר המקומי. כתוצאה מהוספת המדד הנוסף משתנה התוצר המקומי הפך להיות לא מובהק.
כיצד ניתן להסביר זאת:
   i. מולטיקוליניאריות מלאה.
   ii. מולטיקוליניאריות חלקית.
   iii. הוספת משתנה לא רלוונטי.
   iv. השמטת משתנה רלוונטי.

ט. החוקר טען כי: "אם נוסיף משתנה נוסף לרגרסיה כלשהי אז האומדן ל- σ^2 לעולם לא יעלה". נכון/לא נכון?

י. החוקר טוען כי "אם נוסיף משתנה נוסף לרגרסיה, אז האומדן ל- $\overline{R}^2$ יעלה בהכרח". נכון/לא נכון.

6) מידע שנאסף לאחרונה על מכירות של 880 בתים בסטוקטון, קליפורניה, נמצא בקובץ *Stockton2*.
המשתנים הם:
PRICE - מחיר בית בדולרים.
SQFT - גודל הבית (square feet).
BEDS - מספר חדרי שינה.
BATHS - מספר חדרי שירותים.
AGE - גיל הבית.
STORIES - הקומה של הבית.
VACANT - משתנה דמי המקבל 1 אם הבית היה פנוי בזמן מכירתו ו-0 אחרת.

להלן תוצאות אמידה המתבססת על המשתנים הנ"ל:

```
. gen price1000= price/1000

. gen lprice1000=ln( price1000)

. gen  sqft100= sqft/100

. reg  lprice1000 sqft100 age beds baths stories vacant
```

Source	SS	df	MS		
Model	92.5168833	6	15.4194805	Number of obs =	880
Residual	32.0322735	873	.03669218	F(6, 873) =	420.24
				Prob > F =	0.0000
Total	124.549157	879	.141694149	R-squared =	0.7428
				Adj R-squared =	0.7410
				Root MSE =	.19155

lprice1000	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft100	.0637441	.0020353	31.32	0.000	.0597495	.0677387
age	-.0024514	.000366	-6.70	0.000	-.0031698	-.0017331
beds	-.0848595	.0133702	-6.35	0.000	-.1111011	-.058618
baths	.0089069	.0181165	0.49	0.623	-.02665	.0444638
stories	-.0182728	.0219074	-0.83	0.404	-.06127	.0247245
vacant	-.0803092	.0132448	-6.06	0.000	-.1063045	-.0543138
_cons	3.994605	.037782	105.73	0.000	3.920451	4.068759

א. מהי משוואת הרגרסיה שנאמדה?
ב. פרשו את תוצאות האמידה. דונו בסימנים ובמשמעויות של כל אחד מהמשתנים.
ג. מה ההבדל במחיר הממוצע בין בית פנוי בזמן מכירתו לבין בית המאוכלס בזמן מכירתו?

ד. חוקר אחר הגדיר את משתנה VACANT= משתנה דמי המקבל 1 אם הבית היה מאוכלס בזמן מכירתו ו-0 אחרת.
האם יש צורך לחשב מחדש את משוואת הרגרסיה? כן/לא.

ה. נניח כי רצו לתרגם את תוצאות המודל מ-sqrft למטרים מרובעים.
יחידה אחת של sqrft שווה ל-0.093 מטר רבוע. מה יהיה הערך של β_1 במודל החדש? מה יהיה ערך t סטטיסטי לדחיית H0 של המשתנה החדש?

ו. מה הפער החזוי בין מחיר של בגודל 130 מטר בת 15 שנה לבין דירה בגודל 115 מטר בת 20 שנה, בהינתן שכל שאר המשתנים נותרים קבועים?

ז. מהי רמת הסמך הנמוכה ביותר עבורה ניתן לדחות את הטענה כי $\beta_5 = 0$?

ח. החוקר החליט לאמוד את המשוואה ללא משתנה VACANT כמשתנה מסביר. להלן תוצאות האמידה:

```
. reg  lprice1000 sqft100 age beds baths stories

      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  5,   874) =  477.40
       Model |  91.1678763     5  18.2335753            Prob > F      =  0.0000
    Residual |  33.3812805   874  .038193685           R-squared     =  0.7320
-------------+------------------------------           Adj R-squared =  0.7304
       Total |  124.549157   879  .141694149           Root MSE      =  .19543

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0654365   .0020569    31.81   0.000     .0613995    .0694734
          age |  -.0021886   .0003708    -5.90   0.000    -.0029163   -.0014608
         beds |  -.0823237   .0136344    -6.04   0.000    -.1090837   -.0555638
        baths |   .0062791   .0184781     0.34   0.734    -.0299876    .0425459
      stories |  -.0260982   .0223123    -1.17   0.242    -.0698901    .0176938
        _cons |   3.925237   .0367376   106.85   0.000     3.853132    3.997341
------------------------------------------------------------------------------
```

כיצד השמטת המשתנה השפיעה על משוואת הרגרסיה ועל מקדמיה?

ט. בנוסף אמדו על בסיס המדגם הנוכחי את שתי הרגרסיות הבאות:

```
. reg  lprice1000 sqft100 age beds baths stories if  vacant==0

      Source |       SS       df       MS              Number of obs =     415
-------------+------------------------------           F(  5,   409) =  233.63
       Model |  50.040644     5  10.0081288            Prob > F      =  0.0000
    Residual |  17.5202976   409  .042836913           R-squared     =  0.7407
-------------+------------------------------           Adj R-squared =  0.7375
       Total |  67.5609416   414   .16319068           Root MSE      =  .20697

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0684748   .0030057    22.78   0.000     .0625662    .0743833
          age |  -.001996    .0005387    -3.70   0.000    -.003055    -.000937
         beds |  -.0977677   .019958     -4.90   0.000    -.1370009   -.0585346
        baths |   .0193179   .0251737     0.77   0.443    -.030168    .0688038
      stories |  -.0654738   .033737     -1.94   0.053    -.1317933   .0008457
        _cons |   3.979725   .0554426    71.78   0.000     3.870737    4.088713
------------------------------------------------------------------------------

. reg  lprice1000 sqft100 age beds baths stories if  vacant==1

      Source |       SS       df       MS              Number of obs =     465
-------------+------------------------------           F(  5,   459) =  242.82
       Model |  37.2518359     5  7.45036717            Prob > F      =  0.0000
    Residual |  14.0830841   459  .030682101           R-squared     =  0.7257
-------------+------------------------------           Adj R-squared =  0.7227
       Total |   51.33492   464  .110635603           Root MSE      =  .17516

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0593134   .0027696    21.42   0.000     .0538708    .0647561
          age |  -.0028851   .0004996    -5.78   0.000    -.0038668   -.0019034
         beds |  -.0678451   .0178018    -3.81   0.000    -.1028283   -.0328619
        baths |  -.0103401    .026461    -0.39   0.696    -.0623399    .0416596
      stories |   .0265265   .0284661     0.93   0.352    -.0294135    .0824665
        _cons |   3.924588   .0472086    83.13   0.000     3.831817     4.01736
------------------------------------------------------------------------------
```

השוו את תוצאות האמידה של שתי הרגרסיות הנ"ל.

י. ערכו מבחן Chow לבדיקת שקילות (יציבות) המקדמים בשתי הרגרסיות מהסעיף הקודם.

7) חוקר מעוניין ללמוד על הקשר בין הכנסה של משפחה לבין מספר שנות הלימוד של הבעל, מספר שנות הלימוד של האישה והימצאות ילדים קטנים בבית. להלן מפתח שמות המשתנים:

FAMINC - הכנסת המשפחה (דולרים בשנה).

HEDU - מספר שנות הלימוד של הבעל.

WEDU - מספר שנות הלימוד של האישה

$$KID6 = \begin{cases} 1 & \text{אם יש ילדים מתחת לגיל 6} \\ 0 & \text{אחרת} \end{cases},$$

הוא אמד שלוש רגרסיות וקיבל את התוצאות הבאות:

```
. */ Table 1

. reg  faminc Heduc Weduc k16

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  3,   424) =   30.43
       Model |  1.4725e+11     3  4.9082e+10           Prob > F      =  0.0000
    Residual |  6.8384e+11   424  1.6128e+09           R-squared     =  0.1772
-------------+------------------------------           Adj R-squared =  0.1714
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40160

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3211.526   796.7026     4.03   0.000     1645.547    4777.504
       Weduc |   4776.907   1061.164     4.50   0.000     2691.111    6862.704
         k16 |  -14310.92   xxxxxxxx    xxxxx   xxxx    xxxxxxxxxx   xxxxxxxxx
       _cons |  -7755.331   11162.93    -0.69   0.488    -29696.91    14186.25
------------------------------------------------------------------------------

. */ Table 2

. reg  faminc Heduc Weduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  2,   425) =   40.87
       Model |  1.3405e+11     2  6.7027e+10           Prob > F      =  0.0000
    Residual |  6.9703e+11   425  1.6401e+09           R-squared     =  0.1613
-------------+------------------------------           Adj R-squared =  0.1574
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40498

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3131.509    802.908     3.90   0.000     1553.344    4709.674
       Weduc |   4522.641   1066.327     4.24   0.000     2426.711    6618.572
       _cons |  -5533.631   11229.53    -0.49   0.622    -27605.97    16538.71
------------------------------------------------------------------------------

. */ Table 3

. reg  faminc Heduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  1,   426) =   61.30
       Model |  1.0455e+11     1  1.0455e+11           Prob > F      =  0.0000
    Residual |  7.2654e+11   426  1.7055e+09           R-squared     =  0.1258
-------------+------------------------------           Adj R-squared =  0.1237
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   41297

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   5155.484   658.4573     7.83   0.000     3861.254    6449.713
       _cons |   26191.27   8541.108     3.07   0.002     9403.308    42979.23
------------------------------------------------------------------------------
```

א. פרשו את משמעויות המקדמים ברגרסיה בטבלה הראשונה, התייחסו למובהקות המקדמים (אין צורך בבדיקה פורמאלית של מובהקות).

ב. התקבל כי ממוצע משתנה KL6 במדגם הוא 0.432. מה משמעות נתון זה?

ג. חשבו את סטית התקן הנאמדת של אומד הריבועים הפחותים ל- β_{KL6} מהמודל הנאמד בטבלה הראשונה. הראו את החישובים שלכם ואת הנוסחאות עליהן אתם מתבססים.
רמז: השתמשו באינפורמציה שניתן לבדוק מובהקות המקדם הנ"ל ביותר מדרך אחת.

ד. הסבירו את הסיבה להבדל המשמעותי בין אומדני β_{HEDU} בשלוש הטבלאות? איך אפשר להסביר את העדר השוני (כמעט העדר שוני) באומדני β_{HEDU} ו- β_{WEDU} בטבלה הראשונה והשנייה?

8) הורצו 2 רגרסיות על מדגם בן 400 תצפיות והתקבלו התוצאות הבאות:

1. $\hat{CM}_i = 263.6416 - 0.0056 \cdot PGNP_i - 2.2316 \cdot FLR_i$
   (s.e) (11.5932) (0.0019) (0.2099) $R^2 = 0.7077$
2. $\hat{CM}_i = 168.3067 - 0.0055 \cdot PGNP_i - 1.768 \cdot FLR_i + 12.8686 \cdot TFR_i$
   (s.e) (11.5932) (0.0019) (0.2099) (?) $R^2 = 0.7474$

כאשר:
$CM = Child\ Mortality$ - מס' מקרי המוות של ילדים מתחת לגיל 5 לכל 1000 לידות חיים.
$PGNP = Per\ Capita\ GNP$ - תוצר לנפש במחירים קבועים בדולרים.
$FLR = Female\ Literacy\ Rate$ - אחוז נשים שיודעות לקרוא ולכתוב.
$TFR = Total\ Fertility\ Rate$ - מס' הלידות הממוצע לאישה במדינה.
הנתונים הם עבור 64 מדינות.

א. כיצד ניתן לפרש את המקדם למשתנה TFR? אפריורית, האם תצפו לקשר חיובי/שלילי בין CM ל-TFR? הסבירו.

ב. האם ערכי המקדמים של המשתנים $PGNP, FLR$ מרגרסיה 1 שונים מאלו ברגרסיה 2? אם כן, מה יכולה להיות הסיבה/ות לשינוי זה?

ג. חוו דעתכם על הטענה הבאה: "כיוון ש: FLR ו-TFR כל כך מתואמים אין לשים אותם באותה הרגרסיה".

ד. באיזה מודל תבחרו מבין השניים? באיזה מבחן סטטיסטי יש להשתמש כדי לענות על שאלה זו? הראו חישובכם. (רמז: הביעו את הסטטיסטי של מבחן F במונחי R^2).

ה. האם תוכלו לחשב את סטיית התקן הנאמדת של המקדם למשתנה TFR? (רמז: היזכרו בקשר בין התפלגות T להתפלגות F).

ו. האם ניתן להשוות את מקדם ההסבר של שתי הרגרסיות? האם ניתן להשוות את מקדם ההסבר המתוקנן? אם כן, השוו ודווחו על התוצאות.

ז. ענו על השאלות התיאורטיות הבאות:

i. איזה מן הגורמים הבאים יכול לגרום לכך שאומדי OLS יהיו מוטים:
   1. הטרוסקדסטיות.
   2. השמטת משתנה מסביר רלוונטי.
   3. מקדם מתאם גבוה מאוד בין שני משתנים מסבירים במודל.

ii. איזה מהגורמים הנ"ל יכול לגרום לכך שסטטיסטי t של OLS לא יהיה תקף?

iii. התייחסו לטענה הבאה: "אם האומדים הינם עקיבים הם יהיו בהכרח גם חסרי הטיה". נכון/לא נכון.

iv. אם נתון ש-u לא מתפלג נורמאלית אז אמידת המשוואה בשיטת OLS תניב אומדים שאינם עקיבים. נכון/לא נכון.

9) חוקרת רצתה לבדוק עונתיות במחירי הירקות. לשם כך הגדירה את משתני הדמי הבאים:

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

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

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

D_4 יקבל את הערך 1 אם מדובר בחורף ו-0 אחרת.

כאשר:

V_t - מדד מחירי הירקות.

p_t - מדד המחירים לצרכן.

לשם כך אמדה את הרגרסיה הבאה על פני 30 שנה:

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 \cdot P_t + u_t .$$

א. מדוע לא הכניסה החוקרת למשוואת הרגרסיה את משתנה D_4?

תוצאות האמידה שהתקבלו הן:

$$V_t = 1379.11 + 99.18 D_{1t} + 2209.47 D_{2t} - 476.56 D_{3t} + 489.92 \cdot P_t \quad , \quad R^2 = 0.0844 .$$

נאמדה בנוסף גם המשוואה הבאה: $V_t = \beta_0 + \beta_1 \cdot P_t + \varepsilon_t$.

תוצאות האמידה שהתקבלו: $V_t = 11114.14 + 536.36 \cdot P_t$, $R^2 = 0.0670$.

ב. על סמך התוצאות שהתקבלו, דרגו את העונות לפי רמת המחיר הבסיסית שלהן. הציגו לכל עונה את המיקום שלה ואת רמת המחיר הבסיסית כפי שבא לידי ביטוי במודל.

ג. בדקו את ההשערה כי עונתיות לא משפיע על מחיר הירקות.

ד. כתבו את ההשערות הבאות:

i. מדד מחירי הירקות זהה בחורף ובאביב.

ii. מדד מחירי הירקות גבוה בקיץ מאשר בחורף ביותר מ-600.

ה. האם יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן (בהנחה שהמחיר ההתחלתי של הירקות זהה בין עונות השנה)?

10) חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.
החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את המשוואות הבאות:

1. $NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t}$.

2. $NUM_t = \alpha + \beta_1 \times TYPE_t + \beta_2 \times AVGD_t + \beta_3 \times (AVGD \times TYPE)_t + U_t$.

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 מוצגות להלן:

1. $NUM_t = 0.739 + 0.0233 \cdot AVGD_t$.

2. $NUM_t = 0.14978 + 1.40311 \cdot TYPE_t + 0.002877 \cdot AVGD_t - 0.008 \cdot (AVGD \cdot TYPE)_t$.

1. $ESS = 20963$, $Pt_{\hat{\alpha}} = 0.0019; Pt_{\hat{\beta}} = 0.0001$.

2. $ESS = 20759$, $Pt_{\hat{\alpha}} = 0.6534; Pt_{\hat{\beta}_1} = 0.0067; Pt_{\hat{\beta}_2} = 0.0001; Pt_{\hat{\beta}_3} = 0.1283$.

א. בדקו את טענת החוקר.

ב. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?

ג. מהי השערת האפס לבדיקת הטענה?

ד. הרגרסיה המוגבלת "תחת H_0" למבחן F (WALD) הינה:

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

כאשר:

$Z_0 =$ _____

$Z_1 =$ _____

$Z_2 =$ _____

$Z_4 =$ _____

11) ברשותכם נתונים על מחירי ארוחת ביג מק במסעדות מקדונלדס ברחבי הארץ ב-1/1/2008 וב-1/1/2009. חלק מהחנויות מנוהלות ע"י הרשת והשאר מנוהלות ע"י זכיינים. אתם מעוניינים לבדוק את השפעתה של פסיקת בית משפט מ-7/2008 אשר מאפשרת לרשת לקבוע מחיר מקסימום עבור מוצרים הנמכרים בחנויות המנוהלות ע"י זכיינים.

לפניך הנתונים הבאים:

$Price$ - מחיר ארוחת ביג-מק.

$$D_{2009} = \begin{cases} 0 & year = 2008 \\ 1 & year = 2009 \end{cases}$$

$$D_{Franchise} = \begin{cases} 0 & \text{חנות רשת} \\ 1 & \text{חנות זכיין} \end{cases}$$

נתון המודל הבא המתאר את המחיר:

$$Price_i = \beta_1 + \beta_2 D_{2009} + \beta_3 D_{Franchise} + \beta_4 D_{2009} * D_{Franchise} + \varepsilon_i .$$

א. מה המקדם שמגלם את הפרש השכר הצפוי בין חנות רשת לחנות זכיין כשכל יתר המשתנים מוחזקים קבוע?

ב. נסחו את H_0 ו-H_1 לבדיקת ההשערות הבאות:

i. מחיר ארוחת ביג מק ב-2008 היה זהה בחנויות המנוהלות ע"י הרשת ובחנויות המנוהלות ע"י זכיינים.

ii. פסיקת בית המשפט שינתה את פער המחירים בין חנויות המנוהלות ע"י הרשת לבין חנויות המנוהלות ע"י זכיינים בין השנים 2008 ו-2009.

ג. הגדירו את משתנה האינטראקציה על ידי הגדרת 4 הקבוצות הבאות:

D_1 - רשת לפני 2009.

D_2 - רשת אחרי 2009.

D_3 - זכיין לפני 2009.

D_4 - זכיין אחרי 2009.

i. מהי הפקודה הרלוונטיות ב-STATA ליצירת D_1?

ii. כמה ממשתני הדמי יש להכניס לתוך הרגרסיה?

iii. כתבו את משוואת הרגרסיה.

iv. נסחו שוב את ההשערות של סעיף א'.

v. חוקר טוען כי התוצאות שהתקבלו בבדיקת ההשערות על ידי הגדרת משתני הדמי בשני האופנים יהיו זהות. נכון/לא נכון.

12) היצע העבודה של נשים נשואות היה נושא מרכזי במחקר הכלכלי.
לצרוך אמידת היצע זה נבחר המודל הבא:

$$HOURS_i = \beta_1 + \beta_2 WAGE_i + \beta_3 EDUC_i + \beta_4 AGE_i + \beta_5 KIDSL6 + \beta_6 KIDS618 + \beta_7 NWIFEINC_i + u_i$$

כאשר:

HOURS - היצע העבודה בשעות.

WAGE - שכר לשעה.

EDUC - מספר שנות הלימוד.

AGE - גיל.

KIDSL6 - מספר הילדים בבית מתחת לגיל 6.

KIDS618 - מספר הילדים בגילאים 6-18.

NWIFEINC - הכנסת משק הבית ממקורות שאינם מעבודתה של האישה.

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

ב. הסבירו מדוע לא ניתן לאמוד את משוואת ההיצע הנ"ל בשיטת הריבועים הפחותים. איזה מתכונות אר"פ עלולות להיפגע?

ג. אם כתוצאה מאמידת המשוואה בשיטת OLS התקבל כי $\hat{\beta}_2 > 0$, האם ניתן להסיק כי $\text{cov}(wage_i, hours_i) > 0$ במדגם?

ד. החוקר טען שכדאי להשמיט את משתנה ה-wage בשל היותו מתואם עם הטעויות במודל. לטענתו, השמטתו תאפשר לאמוד את המשוואה ב-OLS ולקבל אומדים חסרי הטיה. חוו דעתכם על טענתו.

ה. 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 וכיצד ניתן לבדוק זאת.

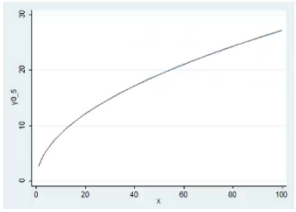
ו. האם ניתן לאמוד את משוואת ההיצע על ידי שימוש במשתני העזר הנ"ל? הסבירו.

ז. תארו את השלבים (לא בפקודות מחשב) שתבצעו כדי לקבל את האומדים בשיטת TSLS.

ח. כתבו את פקודת המחשב לביצוע TSLS ב-STATA.

תשובות סופיות:

1) א. משמעות: ראו סרטון, תוצאות: מובהקות, היגיון כלכלי: יש.
ב. $\hat{pay} = 28,666.87$. ג. $\hat{pay} = 24,355.9$. ד. ראו סרטון.
ה. כן. ו. 66.15. ז. כן, $pv = 0.00$.
ח.i. $H_0 : \beta_1 = 0$ / $H_1 : \beta_1 \neq 0$. ii. $t = 10.13$. iii. דוחים.
ט.i. $H_0 : \beta_2 = 3$ / $H_1 : \beta_2 > 3$. ii. $t = 0.98$. iii. $t_{(0.05,48)} = 2$.
iv. לא דוחים.
י. $(2.90 < \beta_2 < 3.71)$. יא.i. $gen\ pay_1 = pay * 3.5$ / $gen\ spend_1 = spend * 3.5$. ii. אין שינוי.
יב. $gen\ spend_2 = spend * spend$. יג. $H_0 : \beta_2 = 0$, $t = -8.78$, דוחים.
יד. $Corr\ spend\ spend_2$. טו. $X^* = 9,629.92$. טז.



יז. אינה נכונה. יח. נכונה. יט. ii.

2) א. לא ניתן להשוות. ב. מוטה. ג. לא. ד. ראו סרטון.
ה. $V(\hat{\beta}) = \frac{\sigma_u^2}{\sum X_t^2}$. ו. לשני.

3) א. במודל השני. ב. ראו סרטון. ג. $\eta_{y,x} = 1.95$.
ד.i. $WAGE = 13.304$, $WAGE_F = 11.571$. ii. המודל הראשון.
ה. לא ניתן להשוות. ו. לא נכון. ז. ראו סרטון.
ח. לא מומלץ.

4) א. מבחן WALD: $prob(F - ST) = 0.5592 > 0.05$, לא דוחים.
ב. $Gen\ Z_1 = \ln(AREA) + \ln(LABOR)$ / $\text{Reg}\ \ln(PROD)\, Z_1 \ln(FERT)$.
ג. מבחן F.
השערה: $H_0 : \beta_2 + \beta_3 + \beta_4 = 1, \beta_2 = \beta_3 \Leftrightarrow \beta_4 = 1 - 2\beta_2, \beta_2 = \beta_3$ / $H_1 : else$.
סטטיסטי: $F - st = 0.1825$.

5) א. המודל הראשון: $GDS87_t = \alpha_1 + \alpha_2 GDP87_t + \varepsilon_t$.
המודל השני: $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + u_t$.
ב. המודל השני. ג. ראו סרטון. ד. מוטה כלפי מטה.

ה. לא יעיל. ו. שלילית. ז. לא מוגדרים.
ח. ii. ט. לא נכון. י. לא ניתן לדעת.

6) א.

$$\ln(PRICE1000) = \beta_0 + \beta_1 sqft100 + \beta_2 age + \beta_3 beds + \beta_4 baths + \beta_5 stories + \beta_6 vacant + \varepsilon_i$$

ב. ראו סרטון. ג. $e^{-0.080309} = 0.9228$. ד. לא.
ה. $sqrft1 = \frac{sqrft}{0.093}$, t לא ישתנה. ו. 0.957. ז. 0.404.
ח. ראו סרטון. ט. ראו סרטון. י. ראו סרטון.

7) א. ראו סרטון. ב. פרופורציה. ג. $se(\hat{\beta}_{KL6}) = 5003.818$.
ד. ראו סרטון.

8) א. ראו סרטון, חיובי. ב. שונים, מולטיקולינאריות. ג. ראו סרטון.
ד. השני, מבחן F. ה. כן. ו. מקדם ההסבר: אין טעם להשוות,
מקדם ההסבר המתוקנן: כן. ז. i. 2. ii. 1 ו-2.
iii. לא נכון. iv. ראו סרטון.

9) א. הייתה נוצרת מולטיקולינאריות מלאה.
ב. סתיו (902.58), חורף (1379.11), אביב (1478.29) וקיץ (3588.58).
ג. אין עדות. ד. i. $\begin{matrix} H_0: \alpha_1 = 0 \\ H_0: \alpha_1 \neq 0 \end{matrix}$. ii. $\begin{matrix} H_0: \alpha_2 = 600 \\ H_0: \alpha_2 > 600 \end{matrix}$.
ה. לא ניתן לבדוק.

10) א. ראו סרטון. ב. $NUM_t = 1.532398$. ג. $\begin{matrix} H_0: \beta_2 + \beta_3 = 2 \cdot \beta_2 \\ H_0: \beta_3 = \beta_2 \end{matrix}$.

ד. $\begin{matrix} Z_0 = NUM_t \\ Z_1 = TYPE_t \\ Z_2 = (AVGD_t + AVGD \cdot TYPE)_t \\ Z_4 = 0 \end{matrix}$.

11) א. $\beta_3 : D_{franchise}$. ב. i. $\begin{matrix} H_0: \beta_3 = 0 \\ H_1: else \end{matrix}$. ii. $\begin{matrix} H_0: \beta_4 = 0 \\ H_1: \beta_4 \neq 0 \end{matrix}$.
ג. i. $\begin{matrix} genD_1 = 1 \; if \; D_{franchise} == 0 \; and \; D_{2009} == 0 \\ replace \; D_1 == 0 \; if \; D_1 == \end{matrix}$.
ii. 3. iii. $Price = \alpha_1 + \alpha_2 D_1 + \alpha_3 D_2 + \alpha_4 D_3 + u_i$.
iv. i. $\begin{matrix} H_0: \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_2 - \alpha_4 \neq 0 \end{matrix}$. ii. $\begin{matrix} H_0: \alpha_3 - \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_3 - \alpha_2 - \alpha_4 \neq 0 \end{matrix}$. v. נכון.

12) א. *wage* חיובי, *educ* לא ניתן לדעת, *age* חיובי או שלילי תלוי במדגם, *kidsl6* שלילי, *kids*618 חיובי, *nwifeinc* שלילי. ב. ראו סרטון.
ג. לא. ד. שגויה. ה. ראו סרטון.
ו. כן. ז. ראו סרטון.
ח.

Ivreg HOURS EDUC AGE KIDSL6 KIDSS618 NWIFEINC (WAGE = EXPER, EXSPER²).