

# אקונומטריקה א



$$\{\sqrt{x}\}^2$$



## תוכן העניינים

1. מבוא לרגרסיה פשוטה ..... (ללא ספר)
2. תכונות קו הרגרסיה הפשוטה במדגם ..... (ללא ספר)
3. מובהקות הרגרסיה הפשוטה באוכלוסיה ..... (ללא ספר)
4. מודלים לא ליניאריים ..... 1
5. רגרסיה מרובה ..... 5
6. משתנה דמי ..... 17

# אקונומטריקה א

פרק 1 - מבוא לרגרסיה פשוטה

תוכן העניינים

1. כללי ..... (ללא ספר)

# אקונומטריקה א

פרק 2 - תכונות קו הרגרסיה הפשוטה במדגם

תוכן העניינים

1. כללי ..... (ללא ספר)

# אקונומטריקה א

פרק 3 - מובהקות הרגרסיה הפשוטה באוכלוסיה

תוכן העניינים

1. כללי ..... (ללא ספר)

# אקונומטריקה א

פרק 4 - מודלים לא ליניאריים

תוכן העניינים

1. כללי..... 1

## מודלים לא ליניאריים:

רקע:

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

- המשתנה שיש בו  $LN$  השינוי בו יהיה באחוזים.

תזכורת של חוקי לוגים:

$$LN(e^x) = X$$

$$LN(X^Y) = Y \cdot LN(X)$$

$$LN(X \cdot Y) = LN(X) + LN(Y)$$

$$LN\left(\frac{X}{Y}\right) = LN(X) - LN(Y)$$

## שאלות:

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

$$. MWAGE_t = 139.547 + 118.628 \cdot SCL_t \quad .1$$

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

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

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

א. הסבירו את המשמעות של  $\beta$  בכל אחד מהמודלים.

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

(2) נתונים תוצאות האמידה של המודלים הבאים:

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

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

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

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

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

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

(3) נתונים המודלים הבאים עבור התוצר במשק:

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

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

$$. Q_i = A + K_i^{\beta_1} + e^{u_i} \quad .3$$

$$. Q_i = A + \frac{\beta_1}{L_i} + u_i \quad .4$$

$$. Q_i = A + \beta_1 \sqrt{K_i} + u_i \quad .5$$

$$. Q_i = e^{A + \beta_1 K_i + u_i} \quad .6$$

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

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

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

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.

- מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?
- מי מבין המודלים שלא ניתנים לאמידה בשיטת OLS ניתן להביא למודל ליניארי בפרמטרים ועל כן לאמוד את הפרמטרים שלו?
- עבור כל אחד מהמודלים קבעו מיהו המשתנה המוסבר ומיהו המסביר במשוואת הרגרסיה הליניארית.
- עקומת אנג'ל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנג'ל?

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

- האם ניתן לאמוד את המודל בשיטת OLS?
  - מה המשוואה שצריך לאמוד על מנת לקבל את הפרמטרים למודל זה (כלומר כיצד הופכים את המודל לליניארי בפרמטרים)?
  - נאמד המודל הבא:  $\ln(Q_i) = \alpha_0 + \alpha_1 \ln(K_i) + u_i$ , והתקבלו התוצאות הבאות:  $\hat{\alpha}_0 = 3$ ,  $\hat{\alpha}_1 = 0.8$ .
- מהם האומדנים עבור  $A$ ,  $\beta_1$ ?

- (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.

$$\tilde{\beta} = \frac{\sum_{t=1}^T (\ln X_t - \ln \bar{X}) \ln Y_t}{\sum_{t=1}^T (\ln X_t - \ln \bar{X})^2} : \beta \text{ עבור } \beta$$

- האם האומדן ליניארי?
- האם האומדן חסר הטיה?
- האם האומדן *blue*?
- מהי שונותו?

## תשובות סופיות:

(1) א.1. השינוי השולי. ב.2. אין משמעות כלכלית. ג.3. גמישות. ד.4. שיעור השינוי השולי.

א.1. 0.912 ב.2. 0.77 ג.3. 0.778 ד.4. 0.861

(2) א.1.  $\hat{\ln}(Y) = 4.5 + 0.05 \cdot \ln(X)$  ב.2.  $\hat{\ln}(Y) = 4.5 + 0.05X$

ג.3. אין צורך. ד.4.  $\ln\left(\frac{1-\hat{Y}}{\hat{Y}}\right) = 4.5 + 0.05X$

א.1. 98.45 ב.2. 121.51 ג.3. 4.50833 ד.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:  $\ln(Q_i)$  ב.2. מסביר:  $L_i$ , מוסבר:  $\ln(Q_i)$

ג.3. אינו ליניארי. ד.4. מסביר:  $\frac{1}{L_i}$ , מוסבר:  $Q_i$

א.5. מסביר:  $\sqrt{K_i}$ , מוסבר:  $Q_i$  ב.6. מסביר:  $K_i$ , מוסבר:  $\ln(Q_i)$

ג.7. מסביר:  $K_i = \frac{K_i}{2} + 7$ , מוסבר:  $\ln(Q_i)$  ד.8. מסביר:  $L_i$ , מוסבר:  $Q_i$

א.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.  $\beta_1 = -0.8$ ,  $A = 20$

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

# אקונומטריקה א

פרק 5 - רגרסיה מרובה

תוכן העניינים

1. רגרסיה מרובה ..... 5
2. בחירה בין מודלים לפי קריטריון R בריבוע ..... 14

## רגרסיה מרובה:

רקע:

מבחן T ו-F:

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.  
המודל הקלאסי:  $Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + u_t$ .

- קבוע  $\alpha$  יש אחד.
- מספר ה- $\beta$  טות כמספר המשתנים הב"ת במודל.

מבחן F למובהקות המודל:

$$\begin{aligned} H_0 &= \beta_1 = \beta_2 = \dots = \beta_k = 0 \\ H_1 &: \text{OTHERWISE} \end{aligned}$$

השערות:

סטטיסטי המבחן F וכלל ההכרעה:

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

מבחן t למובהקות ה- $\beta$  טות:

מבחן לבדיקת מובהקות  $\beta$  ספציפית:

$$\begin{aligned} H_0 &= \beta_1 = 0 \\ H_1 &: \beta_1 \neq 0 \end{aligned}$$

השערות:

סטטיסטי המבחן t וכלל ההכרעה:

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

### 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 המוטטה -  $R^2$  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

$$\text{ורק אם: } |t_{\hat{\beta}}| > 1.$$

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:  $t_{\beta} = 0.456$ . 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 W.L.D.

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 התצורות של 500 משקי בית כפונקציה של הכנסה שלהן לפי המשוואה:  $EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$ .
- $EXPENSE_t$  - התצורות 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 בתצורות המוסבר ע"י ההכנסה?
- ג. מהו אומדן לתצורות 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 ₪ לתצורות 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.

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

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

## הגרסיה מרובה:

שאלות:

R בריבוע:

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

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

$$2. \quad y_i = \alpha + u_i$$

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

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

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

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

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

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

$$3. \quad \hat{y}_i = \hat{\gamma}_0 + \hat{\gamma}_2 x_{2i} \quad R^2 = 0.65$$

א. שלושה חוקרים העלו טענה לגבי מקדם  $R^2$  של משוואה מס' (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 :

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

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

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

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

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

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

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

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

$$\text{כי: } w_i = 2x_{3i} + x_{1i} - x_{2i}$$

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:

$$\bar{y} = \bar{x} = 40 \quad \text{למה שווה מקדם המתאם של פירסון בין X ל-Y?}$$

א. 0.09

ב. 0.69

ג. 0.3

ד. 0.72

ה. אף תשובה לא נכונה.

(5) נתון מודל רגרסיה:  $y_i = \alpha + \beta x_i + u_i$ .

$$\text{הוכיחו כי: } SST = SSR + SSE$$

## תשובות סופיות:

- (1)  $4 > 5 > 1 = 3 > 2$
- (2) א. iii. ב. i. לא ניתן לדעת. ii. נכון. iii. נכון.
- (3) 3, 4, 2, 1, 7 ובאופן נפרד: 5, 6.
- (4) ג.
- (5) הוכחה.

# אקונומטריקה א

פרק 6 - משתנה דמי

תוכן העניינים

1. כללי ..... 17

## 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 (הבי"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 WALT 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 WALS):  
 המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכול את  
 המדגם המאוחד:  $W_t = \alpha + \beta X_t + u_t$ .

המודל הלא מוגבל (U) כולל את שני חלקי המדגם:  
 $ESS_U = ESS_f + ESS_m$   
 $DF_U = DF_f + DF_m$

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

סטטיסטי המבחן:  $WALS_{stat}$

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

### סיכום ביניים:

משתנה דמי לכל הפונקציה	משתנה דמי לשיפוע	משתנה דמי לחותך	
$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	המודל
קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).	קיים הבדל בין הקטגוריות בתוספת ל-Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).	ההשערה במילים
מבחן WALS להפרש בין הפונקציות (החותכים והשיפועים): $H_0: \alpha_1 = \beta_1 = 0$ **ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW. אם דוחים את $H_0$ יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALS): $H_0: \alpha_1 = 0$ $H_0: \beta_1 = 0$	מבחן t להפרש השיפועים: $H_0: \beta_1 = 0$	מבחן t להפרש החותכים: $H_0: \alpha_1 = 0$	בדיקת ההשערה

### 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:

$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$

השערות:  $H_1: \text{OTHERWISE}$

המבחן הסטטיסטי – מבחן WALD:

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

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

- שימו לב שהחותך במשוואה המוגבלת איננו  $\alpha_0$  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.

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

המחיר ההתחלתי של הירקות שווה בין עונות השנה ( $\alpha$ ) 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:

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

השערות:

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

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

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

- 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 :

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

בדיקת השערות :

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

המבחן הסטטיסטי - מבחן WALD :

(U)

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

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

אם דוחים את  $H_0$ , יש לבדוק במבחן WALD האם ההבדל הוא בין החותכים

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

אם דוחים את  $H_0$  יש להמשיך לבדוק באמצעות מבחן  $t$  :

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

### 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 :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$$

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

ניתן לבדוק השערות על כל אחד מהמשתנים הבי"ת האיכותיים בנפרד :

$$1. H_0 : \alpha_1 = 0$$

$$2. H_0 : \alpha_2 = 0$$

2. הבדל בחותך עם אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

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 :

$$H_0 : \alpha_3 = 0$$

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.

$$W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה :

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

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

$$WALD = t^2$$

$$PF = Pt$$

**שאלות:****משתנה דמי לחותך:**

- (1) על בסיס מדגם של 50 איש העובדים בחברה מסוימת התקבלו התוצאות הבאות:  

$$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$$
(S.E) (134) (56) (24)  
המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.  
א. מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?  
ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?  
ג. האם הבדל זה מובהק באוכלוסייה?  
ד. בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ-500 ₪ מזה של נשים.  
ה. בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.

**פונקציית רגרסיה המכילה משתנים איכותיים בלבד:**

- (2) על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים.  
תוצאות האמידה:  $W_t = 5200 + 1120 \cdot D$   
נתון:  $S_{\hat{\alpha}_1} = 63$   
בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

**משתנה דמי לשיפוע:**

- (3) על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים לנשים בתוספת לשכר בגין שנות הלימוד.  
תוצאות האמידה נתונות להלן:  

$$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$$
(68) (23) (25)  
בדקו 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:

$$1. \quad NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t} \quad \text{כבישים מהירים בלבד.}$$

$$2. \quad NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t} \quad \text{כבישים לא מהירים בלבד.}$$

$$3. \quad NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t} \quad \text{שני סוגי הכביש (כל המדגם).}$$

$$4. \quad NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t$$

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 WALS?

### משוואה (1) - כבישים מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read 344

Number of Observations Used 344

#### Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4700.81174	4700.81174	89.12	<.0001
Error	342	18039	52.74684		
Corrected Total	343	22740			

Root MSE 7.26270 R-Square 0.2067

Dependent Mean 5.10465 Adj R-Sq 0.2044

Coeff Var 142.27617

#### Parameter Estimates

Variable	DF	Parameter 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 בבלט 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		t Value	Pr >  t
		Estimate	Standard Error		
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 <sup>2</sup>	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 W.L.D.  
 הרגרסיה המוגבלת תחת השערת האפס הינה :
- $$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$$
- 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 :
- $$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$
- ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

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

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

כאשר :

$S$  משתנה דמי : 1 = עבור נשים, 0 = גברים.

$E$  משתנה דמי : 1 = עבור השכלה גבוהה ( $scl > 12$ ), 0 = השכלה נמוכה.

א. רשמו את הפונקציה לחישוב :

- i. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו-10 שנות ניסיון.
- ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.
- iii. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות :

- i. אין השפעה של מגדר והשכלה על השכר.
- ii. השפעת ההשכלה אינה תלויה במגדר.
- iii. אין השפעות השכלה אצל גברים.
- iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

## תשובות סופיות:

- (1) א.  $W_t = 7971$  . ב. 1,043 נח. ג. כן. ד. יש עדות לכך.
- ה. יש עדות לכך. (2)  
יש עדות לכך. (3)  
יש עדות לכך. (3)
- (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.  $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot (AVGD_t + AVGD \cdot TYPE_t) + U_t$ .
- (5) א.i.  $H_0: \alpha_1 = \alpha_2$  . ii. WALD t-1 . 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.  $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$  . ii.  $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$  . iii.  $EXP_t = \frac{-(\alpha_1 + \alpha_3)}{\beta_1 + \beta_3}$   
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.  $H_0: \beta_2 + \beta_3 = 0$