

# אקונומטריקה ב



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



## תוכן העניינים

1. חזרה על רגרסיה פשוטה ומרובה תוך שימוש בפלטי SSPS ..... (ללא ספר)
2. מבחן ML ..... 1
3. בעיות ספציפיקציה ..... 5
4. תיאוריה מולטיקוליניאריות ..... 6
5. סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות ..... 9
6. משתנה דמי ..... 14
7. תיאוריה הפרת ההנחות קלאסיות ..... 31
8. הטרוסקדסטיות ..... 32
9. מתאם סדרתי ..... 40
10. סיכום מתאם סדרתי והטרוקדסטיות ..... 51
11. רגרסיה לוגיסטית ..... 52

# אקונומטריקה ב

פרק 1 - חזרה על רגרסיה פשוטה ומרובה תוך שימוש בפלטי SPSS

תוכן העניינים

1. תוספות לרגרסיה פשוטה ..... (ללא ספר)

# אקונומטריקה ב

פרק 2 - מבחן LM

תוכן העניינים

1. כללי ..... 1

## מבחן LM:

## רקע:

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

## לדוגמא:

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

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

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

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

השערות:  $H_1 : \text{OTHERWISE}$

המודל המוגבל (RESTRICTED):  $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t}$

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

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

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

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

חישוב הסטטיסטי: ( $R^2$  של רגרסיית העזר \* מספר התצפיות)  $LM_{stat} = R^2 \cdot T$ .

כלל הכרעה: אם  $LM_{stat} > \chi_m^2$  נדחה את  $H_0$  (מס' ההגבלות ב- $H_0$  =  $m$ ).

## • שימו לב כי:

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

## שאלות:

(1) נניח מודל הכולל 4 משתנים מסבירים (UNRESTRICTED):

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

UNRESTRICTED

Dependent variable: Y

## Analysis of Variance

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

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

## Parameter Estimates

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

**RESTRICTED**

Dependent variable: Y

## Analysis of Variance

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

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

## Parameter Estimates

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

רגרסיית עזר

Dependent variable :RES

## Analysis of Variance

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

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

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

א. בדוק את הטענה כי לפחות אחד מן המשתנים הנוספים רלוונטי למודל בשתי דרכים.

ב. איזה מן המשתנים הנוספים רלוונטי למודל?

ג. הסבירו את הקשרים בין שלוש המשוואות:  $U$ ,  $R$ , ועזר ואת הקשר בין מבחן WALS ומבחן LM.

$$U: \hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{\beta}_4 x_{4t} + \hat{v}_t$$

$$R: \hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{u}_t$$

$$\text{עזר: } \hat{u}_t = \hat{\delta}_0 + \hat{\delta}_1 x_{1t} + \hat{\delta}_2 x_{2t} + \hat{\delta}_3 x_{3t} + \hat{\delta}_4 x_{4t} + \hat{w}_t$$

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

ה. שחזרו בעזרת המשוואה האחרונה (רגרסיית העזר) את  $WALS_{stat}$ .

### תשובות סופיות:

1) א. מבחן LM ומבחן WALS, יש עדות לכך.

ב.  $pt_{\hat{\beta}_3} = pt_{\hat{\beta}_4} = 0.00$

ג. i. עזר  $U=R+$

ii.  $ESS_U = ESS_Y$

iii.  $ESS_R = TSS_Y$

iv.  $R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{ESS_U}{ESS_R} = \frac{ESS_R - ESS_U}{ESS_R}$

ד.  $LM_{stat} = 43.489$

ה.  $WALS_{stat} = 26.962$

# אקונומטריקה ב

פרק 3 - בעיות ספציפיקציה

תוכן העניינים

1. תיאוריה ..... 5

## בעיות ספציפיקציה:

### רקע:

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

1. הוספת משתנה לא רלוונטי:

$$\text{למשל, המודל האמיתי: } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

$$\text{המודל הנאמד (הטעותי): } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon$$

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:

$$\text{למשל, המודל האמיתי: } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

$$\text{המודל הנאמד (הטעותי): } Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$$

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

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

# אקונומטריקה ב

פרק 4 - תיאוריה מולטיקוליניאריות

תוכן העניינים

1. כללי ..... 6

## מולטיקוליניאריות:

### רקע:

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

נבחין בין מולטיקוליניאריות מלאה לחלקית.

### מולטיקוליניאריות מלאה:

מתאם מלא בין המשתנים המסבירים במודל.

הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של

המשתנה המסביר השני:  $x_1 = a + bx_2$  (הוא קומבינציה ליניארית מלאה של  $x_2$ )

מכאן ש:  $r_{12} = 1$ .

- שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל:  $x_1 = x_2^2$ ), אז בהכרח:  $r_{12} \neq 1$ .

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני. מדוע זה בעייתי?

כיוון שלא ניתן לאמוד את המודל שכן אר"פ אינם מוגדרים. פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

### מולטיקוליניאריות חלקית:

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

מכיוון שיש מתאם גבוה בין המשתנים הב"ת לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי. כל אחד מהמשתנים הב"ת "יגזול" מן ההשפעה הייחודית שיש למשתנה הב"ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב"ת יהיה מובהק, התרומה הייחודית של כל משתנה ב"ת לניבוי התלוי לא תהיה מובהקת.

זיהוי מולטיקוליניאריות חלקית :

1. כאשר קיימת סתירה בין התוצאה במבחן  $F$  למובהקות המודל (המודל מובהק) לבין מבחני  $t$  למובהקות השיפועים (אף אחד מן השיפועים איננו מובהק).

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

$$\text{למובהקות השיפועים: } S_{\hat{\beta}_1}^2 = \frac{MSE}{SSX_1(1-r_{12})}, \quad t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$$

2. רגישות לספציפיקציה – הורדת משתנה ב"ת שאיננו מובהק תהפוך משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקוליניאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים 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 מולטיקוליניאריות חלקית :

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

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

פתרונות למולטיקוליניאריות חלקית :

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

2. ניתן לעיתים לאחד את שני המשתנים למשתנה אחד.

שלבי בדירת ההשערות :

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

# אקונומטריקה ב

פרק 5 - סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות

תוכן העניינים

1. כללי ..... 9

## סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות:

רקע:

הבעיה	הגדרה	זיהוי	השלכות	פתרון												
הוספת משתנה לא רלוונטי	המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$ המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$	קבלת $H_0$ במבחן $t$ למובהקות $\beta_2$	ניתן לבצע בדיקת השערות אר"פ $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ חסרי הטיה אומדי השונות $(S_{\hat{\alpha}}^2, S_{\hat{\beta}_1}^2, S_{\hat{\beta}_2}^2)$ חסרי הטיה	הורדת המשתנה*												
השמטת משתנה רלוונטי	המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$	דחיית $H_0$ במבחן $t$ למובהקות $\beta_2$	<table border="1"> <thead> <tr> <th>בהיעדר : <math>x_2</math></th> <th>אומד ל-<math>\beta_1</math></th> <th>אומד ל-<math>\alpha</math></th> <th>אומד לשונות הפרמטרים</th> </tr> </thead> <tbody> <tr> <td><math>S_{12} = 0</math></td> <td>חסר הטיה</td> <td>מוטה אלא אם: <math>\bar{x}_2 = 0</math></td> <td>מוטה חיובית</td> </tr> <tr> <td><math>S_{12} \neq 0</math></td> <td>מוטה חיובית: <math>S_{12}</math> ו-<math>\beta_2</math> שויי סימן מוטה שלילית: <math>S_{12}</math> ו-<math>\beta_2</math> מנוגדי סימן</td> <td>מוטה</td> <td>מוטה חיובית</td> </tr> </tbody> </table> <p>לא ניתן לבצע בדיקת השערות</p>	בהיעדר : $x_2$	אומד ל- $\beta_1$	אומד ל- $\alpha$	אומד לשונות הפרמטרים	$S_{12} = 0$	חסר הטיה	מוטה אלא אם: $\bar{x}_2 = 0$	מוטה חיובית	$S_{12} \neq 0$	מוטה חיובית: $S_{12}$ ו- $\beta_2$ שויי סימן מוטה שלילית: $S_{12}$ ו- $\beta_2$ מנוגדי סימן	מוטה	מוטה חיובית	הוספת המשתנה
בהיעדר : $x_2$	אומד ל- $\beta_1$	אומד ל- $\alpha$	אומד לשונות הפרמטרים													
$S_{12} = 0$	חסר הטיה	מוטה אלא אם: $\bar{x}_2 = 0$	מוטה חיובית													
$S_{12} \neq 0$	מוטה חיובית: $S_{12}$ ו- $\beta_2$ שויי סימן מוטה שלילית: $S_{12}$ ו- $\beta_2$ מנוגדי סימן	מוטה	מוטה חיובית													
מולטיקוליניאריות מלאה	מתאם מלא בין המשתנים המסבירים במודל $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ כאשר: $r_{12} = \pm 1$	אם: $x_1 = a + bx_2$ אז: $r_{12} = 1$	לא ניתן לבצע בדיקת השערות אר"פ $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ בלתי מוגדרים.	הורדת אחד המשתנים												
מולטיקוליניאריות חלקית	מתאם חזק בין המשתנים המסבירים במודל $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ כאשר: $0.7 <  r_{12}  < 1$	א. סתירה בין מבחן F ל- $t$ ב. רגישות לספציפיקציה ג. סימנים הפוכים	ניתן לבצע בדיקת השערות אין פגיעה בתכונות אר"פ ושונותם	הורדת אחד המשתנים או איחודם												

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

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

## שאלות:

(1) להלן מודל של שכר  $W_t$ , כפונקציה של שנות לימוד  $S_t$ :

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

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

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

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:

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

חוה דעתך על המשוואה השלישית.

(2) נתונות ארבע משוואות הרגרסיה הבאות (כאשר הסטיות במודל האמיתי

מקיימות את הנחות הרגרסיה הקלאסיות):

$$.1 \quad X_{2t} = \lambda + \delta \cdot X_{1t} + V_t \quad \text{כאשר התקבל: } \sum \hat{V}_t^2 = \sum (X_{2t} - \bar{X}_{2t})^2$$

$$.2 \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t \quad (19.8) \quad (10.3)$$

$$.3 \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + W_t \quad (0.37) \quad (17.3) \quad (9.9)$$

$$.4 \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \sum_t \quad (6.3)$$

(המספרים בסוגריים הם ערכי  $t$  של אומדני המקדמים).

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

א. האומד של  $\beta_1$  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 השונות 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).

$$(3) \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$$

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

א. בהנחה כי מתקיים:  $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$   $R^2 = 0.92$   
(0.5) (0.3)

- הערכים בסוגריים הם ערכי t.  
למובהקות הבטות יש טעות במודל  
כי המודל מובהק והמקדמים לא :  
נכון/לא נכון/לא ניתן לדעת
- 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:  $t_{\hat{\beta}_1} = 1.31$ ,  $t_{\hat{\beta}_2} = 1.45$ , מה יהיה הפתרון הטוב ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף ii)?
1. להוריד את  $x_1$ .
  2. להוריד את  $x_2$ .
  3. להוריד את שני המשתנים.
  4. להותיר את שני המשתנים.

**תשובות סופיות:**

- (1) א. ii. ב. לא נכון. ג. קיימת בעיית מולטיקוליניאריות מלאה.  
(2) א. לא נכון. ב. לא נכון. ג. נכון. ד. נכון. ה. נכון.  
ו. לא נכון. ז. נכון. ח. לא נכון. ט. נכון.  
(3) א. לא נכון. ב. נכון. ג. לא נכון. ד. הוכחה. ה. i. לא נכון.  
ii. מולטיקוליניאריות חלקית. iii. 4.

# אקונומטריקה ב

פרק 6 - משתנה דמי

תוכן העניינים

1. כללי ..... 14

## 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 :  
 $ESS_U = ESS_f + ESS_m$   
 $DF_U = DF_f + DF_m$

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

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 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 (בחותר ובשיפוע).	קיים הבדל בין הקטגוריות בתוספת ל-Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותר).	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 במבחני 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 מבחני  $t$  :

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

### 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 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)  
 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:

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

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

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

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

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 :  

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

$$\ln(Y)_i = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_i$$
 האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

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

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

כאשר :

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

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

א. רשמו את הפונקציה לחישוב :

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

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות :

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

## תשובות סופיות:

- (1) א.  $W_t = 7971$  . ב. 1,043 נח. ג. כן. ד. יש עדות לכך.
- ה. יש עדות לכך. (2)  
יש עדות לכך. (3)  
יש עדות לכך. (3)
- א. יש עדות לכך, מבחן CHOW : 1, 2 ו-3, משתנה דמי : 3 ו-4. (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.  $H_0 : \alpha_3 = \alpha_1 - \alpha_2$  או  $H_0 : \alpha_2 = \alpha_1 - \alpha_3$  .  
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.  $H_0 : \beta_2 + \beta_3 = 0$  .

# אקונומטריקה ב

פרק 7 - תיאוריה הפרת ההנחות קלאסיות

תוכן העניינים

1. כללי ..... 31

## הפרת ההנחות הקלאסיות:

### רקע:

ארבעת הנושאים הבאים עוסקים במצב של הפרת אחת ההנחות הקלאסיות הדרושות לאמידת הפרמטרים בשיטת OLS :

- הטרוסקדסטיות (הפרת הנחה מס' 5) – שונות קבועה ויחידה לאורך קו הרגרסיה :  $V(u_t) = \sigma^2$ .
- מתאם סידרתי (הפרת הנחה מס' 6) – אי תלות בין הטעויות :  $\text{cov}(u_t, u_s) = 0$ .
- מודלים דינמיים ו-משוואות סימולטניות (הפרת הנחה מס' 4) – אי תלות בין המשתנים הב"ת לטעויות :  $\text{cov}(x, u) = 0$ .

בכל אחד מן הנושאים נלמד :

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

# אקונומטריקה ב

פרק 8 - הטרוסקדסטיות

תוכן העניינים

1. כללי ..... 32

## הטרוסקדסטיות:

### רקע:

הטרוסקדסטיות הוא מצב שבו מופרת הנחת ההומוסקדסטיות, הגורסת כי שונות הטעויות היא אותה שונות עבור כל תצפית ותצפית:  $V(u_t) = \sigma^2$  לכל  $t$ , כלומר התצפיות מפוזרות באופן אחיד סביב קו הרגרסיה. במצב של הטרוסקדסטיות שונות הטעויות של כל תצפית היא שונה:  $V(u_t) = \sigma_t^2$ .

### ההשלכות של הטרוסקדסטיות על אומדי OLS:

בהינתן הטרוסקדסטיות מופרת תכונת היעילות של אומדי הריבועים הפחותים שכן בכדי לחשב שונות יעילה של האומדים השתמשנו בהנחה של שונות קבועה.

### מבחנים לזיהוי הטרוסקדסטיות:

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות – באיזה אופן השונות של הטעויות משתנה בין תצפית לתצפית. שיטות לזיהוי הטרוסקדסטיות: מבחן GQ (Goldfeld-Quandt) ומבחן White. מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונות שונות בלבד. ואילו מבחן White מניח כי לכל תצפית ותצפית שונות שונה של טעויות.

#### 1. מבחן GQ:

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

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

- סטטיסטי זה מתפלג:  $F_{(\alpha; T_1-K-1, T_2-K-1)}$
- כלל ההכרעה: אם  $F_{stat} > F_C$  אז דוחים את  $H_0$ .

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

## 2. מבחן White:

ההנחה העומדת בבסיס מבחן זה כי לכל תצפית ותצפית שונות שונה של טעויות. הביטוי המתמטי של הנחה זו היא היותה של השונות פונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:

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

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

האומד ל- $\sigma_t^2$  הוא  $\hat{u}_t^2$ .

המבחן הוא מבחן LM:

- אומדים את המודל המקורי ומקבלים את הסטיות מקו הרגרסיה  $\hat{u}_t$  (המכונה בתוכנה הסטטיסטית RES-SAS).
- אומדים את  $\hat{u}_t^2$  כפונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:  $\hat{u}_t^2 / x_j, x_j^2, x_j x_j$  זוהי רגרסיית העזר.
- נחשב את סטטיסטי LM:  $LM_{stat} = T_y \cdot R_y^2$ .
- אם  $LM_{stat} > \chi_m^2$  כאשר  $m =$  מס' המשתנים ברגרסיית העזר.

- השערות:  $H_0: \alpha_j = \beta_j = \gamma_{jj} = 0$   
 $H_1: OTHERWISE$

## פיתרון בעיית ההטרוסקדסטיות – ריבועים פחותים משוקללים (WLS):

נניח שאנו רוצים לאמוד את המודל:  $Y_t = \alpha + \beta X_t + u_t$  וידוע כי לכל קריזה שונות אחרת. ההנחה שעומדת בבסיס שיטת ה-WLS היא כי השונות המשתנה כוללת בתוכה מרכיב קבוע ומרכיב משתנה:  $\sigma_t^2 = Z_t \cdot \sigma^2$  את המרכיב המשתנה בשונות ( $Z_t$ ) יש לנטרל. לשם כך ניצור משתנה חדש  $W_t$  שיהווה השורש ההופכי

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

נכפיל כל תצפית במשתנה החדש  $W_t$  וניצור משוואה שהיא קומבינציה ליניארית של

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

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

בצורתה המפורשת המשוואה החדשה נראית כך:  $\frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$

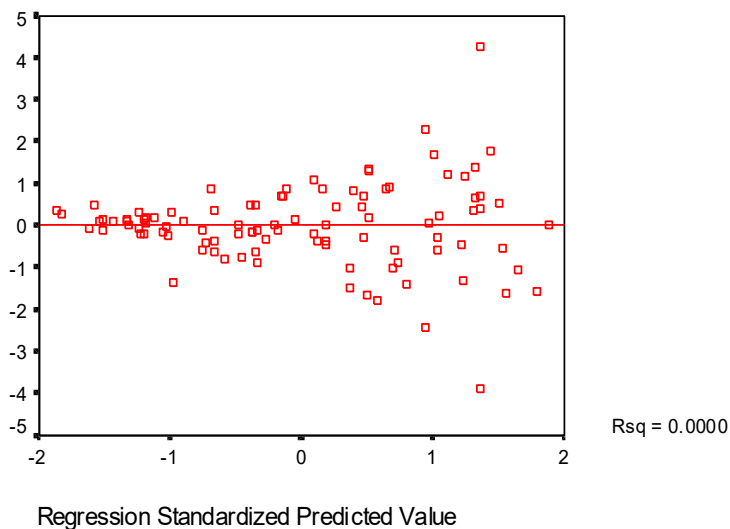
**שאלות:**

**מבחן GQ:**

(1) נאמד הקשר שבין הכנסה לתצרוכת:  $Y_t = \alpha + \beta X_t + u_t$ .  
גרף השאריות של הרגרסיה הנ"ל נתון להלן:

Scatterplot

Dependent Variable: Y



מגרף זה אנו למדים כי השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה – זהו מצב של הטרוסקדסטיות.  
בכדי לבצע מבחן GQ:

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

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

### משוואה (1)

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read 16  
Number of Observations Used 16

#### Analysis of Variance

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

### משוואה (2)

The REG Procedure

Model: MODEL1

Dependent Variable: y

Number of Observations Read 16  
Number of Observations Used 16

#### Analysis of Variance

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

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

(2) על מנת לבחון את פונקציית הייצור בענף מסוים נאספו נתונים על 150 פירמות. נסמן:

$Q$  - תפוקה שנתית באלפי שקלים.

$L$  - מספר עובדים.

המודל הנאמד:  $\ln(Q) = \alpha + \beta \cdot \ln(L)$ .

החוקר חשש שההפרעה המקרית איננה הומוסקדסטית. לשם כך הוא מניין את

התצפיות בסדר עולה של מספר העובדים, השמיט  $\frac{1}{3}$  מהתצפיות האמצעיות

והריץ שתי רגרסיות נפרדות עם מספר שווה של תצפיות:

ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא

קיבל:  $R^2 = 0.403$ ,  $ESS = 279.3$

ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא

קיבל:  $R^2 = 0.238$ ,  $ESS = 493.8$

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

### מבחן WHITE:

3 על אותו הקשר שבין הכנסה לתצרוכת:  $Y_t = \alpha + \beta X_t + u_t$   
שאנו חושדים על פי גרף השאריות כי קיים בו מצב של הטרוסקדסטיות.  
בכדי לבצע את מבחן WHITE:

- נחשב את השאריות של הרגרסיה:  $\hat{u}_t = Y_t - \hat{Y}_t$
- נעלה את השאריות בריבוע:  $\hat{u}_t^2$
- נאמוד את המשוואה:  $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + v_t$

תוצאות האמידה מוצגות להלן:

```

The REG Procedure
Model: MODEL1
Dependent Variable: RES2
Number of Observations Read      48
Number of Observations Used      48

Analysis of Variance
Source      DF      Sum of Squares      Mean Square      F Value      Pr > F
Model
Error
Corrected Total

Root MSE      R-Square      0.390763
Dependent Mean      Adj R-Sq

```

בצעו מבחן White (השערות, סטטיסטי המבחן, כלל הכרעה ומסקנה).

- 4) חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות והאפשרות של מכירת עיתונים.  
נסמן:  
 $Y\_SALES$  - מכירות חודשיות (ש).  
 $X1\_SQUARES$  - שטח החנות (מ"ר).  
 $X2\_RENT$  - דמי שכירות (\$).  
 $PAPERS$  - משתנה איכותי המקבל 1 אם החנות מוכרת גם עיתונים ו-0 אם לא.  
 החוקר חשד כי קיימת בעיה של הטרוסקדסטיות בנתונים.  
 החוקר ביצע מבחן לזיהוי הטרוסקדסטיות שתוצאותיו נתונות להלן:

Dependent Variable:					
		Number of Observations Read	20		
		Number of Observations Used	20		
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model					
Error					
Corrected Total					
		Root MSE	R-Square	0.086942	
		Dependent Mean	Adj R-Sq		
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept					
X1_SQUARE					
X1_SQUARE^2					
X1_SQUARE*X2_RENT					
X2_RENT					
X2_RENT^2					
X2_RENT*PAPERS					
PAPERS					

- הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: \_\_\_\_\_  
 על ידי מבחן: \_\_\_\_\_  
 המשתנה התלוי הינו: \_\_\_\_\_  
 המשתנים הב"ת: \_\_\_\_\_  
 ההשערות הינן: \_\_\_\_\_  
 גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): \_\_\_\_\_  
 המסקנה המתקבלת היא: \_\_\_\_\_

## שיטת WLS:

(5) נתון המודל:

$$.1 \quad Y_t = \alpha + \beta \cdot X_t + U_t$$

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

א. מהי הבעיה שנוצרת באמידת משוואה (1)?

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

כדי לפתור את הבעיה שנוצרה, נאמדה המשוואה הבאה:

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

ג. מהו  $W_t$  שבעזרתו ניתן לאמוד את  $\alpha$  ו- $\beta$  בצורה יעילה?ד. מהו האומד היעיל של  $\sigma^2$ ?ה. האם ניתן להשוות בין המודלים על בסיס  $R^2$ ? אם לא, האם ניתן

להחליט בכל זאת איזה מודל טוב יותר?

ו. חוו דעתכם על הטענות הבאות, ונמקו:

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

ללא שינוי.

ii. המשוואה הנורמאלית:  $\sum \hat{\varepsilon}_t = 0$  (כאשר:  $\varepsilon_t = U_t \cdot W_t$ ) היא

אחת המשוואות הנורמאליות לאמידת משוואה (2).

(6) ענו על השאלה הקודמת, כאשר נתון כי:  $VAR(U_t) = \sigma^2 \cdot X_t^2$ .(7) נתון המודל:  $Y_t = \alpha + \beta X_t + u_t$ . וקיים מדגם של 100 תצפיות כאשר נתון

כי:  $V(u_t) = \sigma_t^2 = \begin{cases} X_t \sigma^2 & \Leftrightarrow t \geq 50 \\ X_t^2 \sigma^2 & \Leftrightarrow t \leq 50 \end{cases}$ . (שאר ההנחות הקלאסיות מתקיימות).

א. במשוואה מס' 1 יש בעיה של: \_\_\_\_\_.

ב. אמידת משוואה (1) תניב אומדים

בלתי מוטים ועקיבים: נכון/ לא נכון/ לא ניתן לדעת

ג. פתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה

הבאה:  $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t$ כאשר:  $W_t = \underline{\hspace{2cm}}$ .

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

האם ישתנו תשובותיכם לסעיפים א' ו-ב': כן/לא/לא ניתן לדעת

## תשובות סופיות:

(1) יש עדות לכך, השערות:  $H_0: \sigma_1^2 = \sigma_2^2$ , חישוב סטטיסטי: 17.62, כלל הכרעה: 2.48.  
 $H_1: \sigma_1^2 > \sigma_2^2$

(2) יש עדות לכך, השערות:  $H_0: \sigma_1^2 = \sigma_2^2$ , חישוב סטטיסטי: 1.77, כלל הכרעה: 1.69.  
 $H_1: \sigma_1^2 > \sigma_2^2$

(3) יש עדות לכך, השערות:  $H_0: \alpha_1 = \beta_1 = 0$ , חישוב סטטיסטי: 18.75,  
 $H_1: OTHERWISE$

כלל הכרעה: 5.991

(4) קיום הטרוסקדסטיות בנתונים.

$(LM)_{WHITE}$ .

$.RES^2$

$X1\_SQUARE$ ,  $X1\_SQUARE^2$ ,  $X1\_SQUARE * X2\_RENT$ ,

$.X2\_RENT$ ,  $X2\_RENT^2$ ,  $X2\_RENT * PAPERS$ ,  $PAPERS$

$.LM_{stat} = 1.73$

אין עדות לכך.

(5) א. הטרוסקדסטיות. ב. ראו סרטון. ג.  $W_t = \frac{1}{\sqrt{Z_t^2}} = Z_t$ .

ד.  $\sigma^2 = \frac{ESS}{T-K}$ . ה. לא, המודל השני. ו. i לא נכון. ii לא נכון.

(6) א. הטרוסקדסטיות. ב. ראו סרטון. ג.  $W_t = \frac{1}{X_t}$ .

ד.  $S^2 = \frac{ESS}{T-k-1}$ . ה. ראו סרטון. ו. ראו סרטון.

(7) א. הטרוסקדסטיות. ב. נכון.

ג.  $W_t = \frac{1}{\sqrt{X_t^2}}$   $t \leq 50$ ,  $W_t = \frac{1}{\sqrt{X_t}}$   $t \geq 50$ . ד. לא.

# אקונומטריקה ב

פרק 9 - מתאם סדרתי

תוכן העניינים

1. כללי ..... 40

## מתאם סדרתי:

### רקע:

מתאם סדרתי עוסק במצב שבו מופרת ההנחה (מס' 6) של אי תלות בין הטעויות:  
 $cov(u_t, u_s) = 0$  ונוצרת תלות סטטיסטית בין הטעויות במודל:  $cov(u_t, u_s) \neq 0$ .  
 תלות כזו בין הטעויות קיימת בדרך כלל כאשר הנתונים הנאספים הם נתוני סדרות  
 עיתיות ולא נתוני חתך בהם עסקנו עד כה. בנתוני סדרות עיתיות, מאחר ומדובר  
 באותו הפרט הנמדד בזמנים שונים סביר שהטעויות בניבוי שלו תהיינה תלויות אחת  
 בשנייה.

### השלכות על אומדי הריבועים הפחותים (OLS):

מבין התכונות של אר"פ (ליניאריות, חוסר הטיה, עקיבות ויעילות) היחידה שמופרת  
 כאשר קיים מתאם סדרתי היא: תכונת היעילות.  
 משום שתכונת היעילות היא היחידה מבין תכונות אר"פ התלויה להוכחתה בקיומה  
 של הנחת אי התלות בין הטעויות. משום הפגיעה בתכונת היעילות, בדיקת ההשערות  
 לא תהיה תקפה.

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

### מבנה המתאם הסדרתי:

מתאם סדרתי מסדר ראשון:

ההנחה היא כי יש מתאם בין הטעויות במרחק אחד, כלומר  $u_t$  תלוי ישירות רק

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

את המתאם בין הטעויות מסדר ראשון ניתן לנסח באופן הבא:  $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

כך ש:

- $\rho \neq 0$  (כי אם  $\rho = 0$  אין מתאם סדרתי).
- $-1 < \rho < 1$  (כי אם חורג מ-1 הטעות הולכת וגדלה עם הזמן).
- $\rho$  חיובי פירושו מתאם סדרתי חיובי ואילו  $\rho$  שלילי פירושו מתאם סדרתי  
 שלילי (לא נפוץ).

4.  $\varepsilon_t$  מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma_\varepsilon^2 \quad \text{בניגוד ל- } u_t \text{ כך ש:}$$

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

המודל יכיל שתי משוואות-המשוואה העיקרית והגדרת המתאם הסדרתי (מסדר

$$\text{ראשון): } \begin{aligned} Y_t &= \alpha + \beta X_t + u_t \\ u_t &= \rho \cdot u_{t-1} + \varepsilon_t \end{aligned}$$

מלבד  $\alpha$  ו- $\beta$  נרצה לאמוד גם את  $\rho$ .

מתאם סדרתי מסדר שני:

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

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

מתאם סדרתי מסדר P:

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

$u_t$  מושפע מתקופות שונות בעבר.

**תכונות המתאם הסדרתי:**

מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן:  $r_{(u_t, u_{t-s})} = \rho^s$

המתאם של  $u_t$  הולך ופוחת עם הזמן:  $\rho_{u_t, u_{t-1}} > \rho_{u_t, u_{t-2}}^2 > \rho_{u_t, u_{t-3}}^3 > \dots > \rho_{u_t, u_{t-s}}^s$   
בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t) = 0$$

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

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

## מבחנים לזיהוי מתאם סדרתי:

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

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

$$0 \text{---} \rho > 0 \text{---} d_L \text{---} d_U \text{---} \rho = 0 \text{---} 4 - d_U \text{---} 4 - d_L \text{---} \rho < 0 \text{---} 4$$

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

$$\begin{aligned} H_0: \rho = 0 & \text{ השערות:} \\ H_1: \rho > 0, \rho < 0 & \end{aligned}$$

$$\text{חישוב הסטטיסטי: } DW_{stat} \cong 2 \cdot (1 - \hat{\rho})$$

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

למבחן DW יש שתי בעיות עיקריות:

1. מתאים רק למתאם סדרתי מסדר ראשון.
2. יש אזורים "מתים" בטווח בהם לא ניתן לדעת האם יש מתאם סדרתי.

בנוסף לכך על מספר תנאים להתקיים כדי שאפשר יהיה להשתמש במבחן DW:

1. הרגרסיה כוללת חותך.
2. ה-Xים קבועים ולא משתנים.
3. אין משתנים מסבירים שהם פיגור של המשתנה המוסבר.
4. אין תצפיות חסרות באמצע.
5. אם קיים מתאם סדרתי מסדר ראשון אז הוא מהצורה:  $u_t = \rho \cdot u_{t-1} + \varepsilon_t$ .

מבחן LM :

לעומת מבחן DW מבחן LM מתאים גם לבחינת קיומו של מתאם סדרתי מסדרים

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

גבוהים יותר מסדר ראשון :

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

המודל הלא מוגבל :

$$Y_t = \alpha + \beta \cdot X_t + \rho \cdot u_{t-1} + \varepsilon_t \quad (U)$$

ניתן להתייחס לבחינת קיומו של מתאם סדרתי כהוספת משתנה מסביר :  $\hat{u}_{t-1}$ .  
השלים לביצוע המבחן :

- נאמוד את המודל המקורי ונחשב  $\hat{u}_t$  ו-  $\hat{u}_{t-1}$ .
- נאמוד את רגרסיית העזר :  $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}$ .
- נחשב סטטיסטי LM :  $LM_{stat} = T \cdot R^2$ .
- נדחה את  $H_0$  כאשר :  $LM_{stat} > \chi_m^2$  כאשר  $m =$  סדר המתאם הסדרתי.  
אם נדחה את  $H_0$  נדע את סימנו של המתאם הסדרתי לפי המקדם של  $\hat{u}_{t-1}$   
ברגרסיית העזר ששווה ל-  $\hat{\rho}$ .  
שימו לב כי אם נרצה לבדוק מתאם סדרתי מסדרים גבוהים יותר :

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0$$

ההשערות :

$$H_1 : OTHERWISE$$

$$\text{גרסיית העזר : } \hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}, \hat{u}_{t-2}, \dots, \hat{u}_{t-s}$$

### פתרון בעיית המתאם הסדרתי – רגרסיית הפרשים (שיטת קוקרן-אורקט):

ניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית שבה לא יהיה מתאם סדרתי ולכן ניתן יהיה לאמוד אותה בשיטת הריבועים הפחותים, האומדים יהיו יעילים וניתן יהיה לבצע בדיקת השערות.

$$\text{משוואה (1): המודל בזמן } t : Y_t = \alpha + \beta X_t + u_t$$

$$\text{משוואה (2): המודל בזמן } t-1 \text{ מוכפל ב- } \rho : \rho \cdot Y_{t-1} = \rho \cdot \alpha + \rho \cdot X_{t-1} + \rho \cdot u_t$$

החסרת משוואה (2) ממשוואה (1) :

$$Y_t - \rho Y_{t-1} = \alpha(1 - \rho) + \beta(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$$

כדי לאמוד את הפרמטרים של רגרסיית הפרשים נגדיר :

$$Y_t^* = Y_t - \rho Y_{t-1}$$

$$\alpha^* = \alpha(1 - \rho)$$

$$\beta^* = \beta$$

$$X_t^* = X_t - \rho X_{t-1}$$

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

$$\text{כך "נטרלנו" את המתאם הסדרתי : } \varepsilon_t = u_t - \rho \cdot u_{t-1}, u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

$\varepsilon_t$  מקיים את כל ההנחות הקלאסיות ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.  
 המשוואה "המתוקנת" אותה נאמוד:  $Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$ .  
 לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים:  $\hat{\alpha}$ ,  $\hat{\beta}$ .  
 מאחר ש- $\rho$  איננו ידוע יש צורך לאמוד אותו.

אמידת  $\rho$  בשיטת קוקרן אורקוט:  
 שיטת קוקרן אורקוט לאמידת  $\rho$  היא שיטה איטראטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות.  
 התהליך הממוחשב נקרא אוטורגרסיה (AUTOREGRESION) מסדר ראשון, שני, שלישי וכו' (תלוי בסדר המתאם הסדרתי). התיקון למתאם הסדרתי יתבצע על ידי הרצת רגרסיה עם משתנה AR(1) (אוטו רגרסיה מסדר ראשון), AR(1) ו-AR(2) (אם מניחים קיום אוטורגרסיה מסדר שני) וכו'. אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

## שאלות:

## תכונות המתאם הסדרתי:

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

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

מצאו את:

א. המתאם בין  $u_t$  ל- $u_{t-1}$ .ב. המתאם בין  $u_t$  ל- $u_{t-4}$ . הסבר את ההבדל בין המתאמים (סעיף א' ו-ב').ג. השונות  $\sigma_u^2$ .ד. חזרו על סעיפים א' עד ג' עבור  $\rho = 0.4$ . הסבירו את ההבדל בין התוצאות.

## מבחן DW:

(2) חוקר רצה לאמוד את מחיר סגירה של מניה כפונקציה של הזמן שעובר:

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

כאשר:

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

$$TIME_t = \text{משתנה זמן שמקבל את הערכים: } 1, 2, 3, \dots$$

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

Analysis of Variance					
F Value	Prob>F	Mean Square	Sum of Squares	DF	Source
	55.78		-----	1	Model
			-----	151	Error
			-----	152	C Total
		0.181	R-square	----	Root MSE
		-----	Adj R-sq	----	Dep Mean
				----	C.V.
Parameter Estimates					
T for H0:	Standard Error	Parameter Estimate	DF	Variable	
Parameter=0	0.0148	1.3474	1	INTERCEP	
91.047	0.0000	-0.00075	1	TIME	
0.0000	-7.468				
Durbin-Watson D	0.150				

האם קיים מתאם סדרתי?

TABLE 12 Cutoff Points for the Distribution of the Durbin-Watson Test Statistic

Let  $d_\alpha$  be the number such that  $P(d < d_\alpha) = \alpha$ , where the random variable  $d$  has the distribution of the Durbin-Watson statistic under the null hypothesis of no autocorrelation in the regression errors. For probabilities  $\alpha = .05$  and  $\alpha = .01$ , the tables show, for numbers of independent variables,  $K$ , values  $d_L$  and  $d_U$  such that  $d_L \leq d_\alpha \leq d_U$ , for numbers  $n$  of observations.

$\alpha = .05$										
$n$	$K$									
	1		2		3		4		5	
	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$	$d_L$	$d_U$
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	0.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

$$Y_t = \alpha + \beta X_t + \varepsilon_t \quad (3)$$

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

$u_t$  - הפרעה מקרית קלאסית.

$T = 100$  וידוע כי :

$$u_t = 0.9u_{t-1}$$

$$d_L = 1.57$$

$$d_U = 1.65$$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

**מבחן LM:**

(4) עבור הדוגמא הקודמת – ניבוי מחיר סגירה של מניה כפונקציה של הזמן :

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

נבחן את קיומו של מתאם סדרתי מסדר ראשון באמצעות מבחן LM.

$$u_t = \gamma_0 + \gamma_1 \cdot TIME_t + \rho \cdot u_{t-1} + \varepsilon_t$$

תוצאות האמידה שהתקבלו :

Dependent variable: RES

		Analysis of Variance			
F Value	Prob>F	Mean Square	Sum of Squares	DF	Source
			-----	2	Model
			-----	150	Error
			-----	152	C Total
		0.855	R-square	-----	Root MSE
		-----	Adj R-sq	-----	Dep Mean
				-----	C.V.

Parameter Estimates

Parameter	Estimate	DF	Variable
	-.000096	1	INTERCEP
	1.16331E-05	1	TIME
	0.927172	1	RES1

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסייה?

## תיקון המתאם הסדרתי:

(5) סטודנט הניח כי במודל:  $Y_t = \alpha + \beta X_t + u_t$  קיים מתאם סדרתי מסדר ראשון בשאריות כך שמתקיים:  $u_t = 0.7u_{t-1} + \varepsilon_t$  ולכן במקום לאמוד את המודל המקורי אמד את המודל:  $Y_t - 0.7Y_{t-1} = \alpha(1-0.7) + \beta(X_t - 0.7X_{t-1}) + u_t$ . הסטודנט טען כי במודל החדש לא קיים מתאם סדרתי. טענת הסטודנט: נכונה / לא נכונה / לא ניתן לדעת

(6) נמשיך עם הדוגמא של ניבוי מחיר סגירה של מניה כפונקציה של הזמן:  $CLOSE_t = \alpha + \beta \cdot TIME + u_t$ . נניח כי קיים מתאם סדרתי מסדר ראשון בנתונים. תוצאות האמידה בשיטת אוטורגרסיה מסדר ראשון מוצגות להלן:

Parameter Estimates					
Prob> T	T for H0: Parameter=0	Standard Error	Parameter Estimate	DF	Variable
			1.333	1	INTERCEP
			-0.0006	1	TIME
0.000			0.927	1	(1)AR
Durbin-Watson D		2.235			

א. בדקו האם נפתרה בעיית המתאם הסדרתי.  
ב. מהי המשוואה לאמידת מחיר הסגירה הצפוי ביום המסחר הבא?

## תרגול מסכם:

(7) נאמד הקשר שבין הכנסה לתצרוכת לתקופה ינואר 1994 עד דצמבר 1997 (T=48). המודל הינו:  $C_t = \alpha + \beta Y_t + u_t$ . בניסיון לבדוק האם מתקיים קשר מהסוג הבא:  $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$ . נאמדה המשוואה הבאה:  $\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \gamma_2 \hat{u}_{t-2} + \gamma_3 \hat{u}_{t-3} + \gamma_4 Y_t + \omega_t$ . תוצאות האמידה מוצגות להלן:

Depended Variable: RES

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob> T
INTERCEP	1	-54.709	710.85	-0.076	0.939
RESID1	1	0.705	0.152	4.631	0.000
RESID2	1	-0.0066	0.188	-0.035	0.972
RESID3	1	-0.337	0.167	-2.012	0.051
Y	1	0.0027	0.032	0.085	0.932
Durbin-Watson D		1.954			

נתון בנוסף כי:  $R^2 = 0.479$ .

- א. הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: \_\_\_\_\_  
 על ידי מבחן: \_\_\_\_\_  
 ההשערות הינן: \_\_\_\_\_  
 גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): \_\_\_\_\_  
 המסקנה המתקבלת היא: \_\_\_\_\_

בהנחה כי קיים מתאם סדרתי מסדר שלישי בנתונים נאמד מחדש הקשר שבין ההכנסה לתצרוכת בהתאם לשיטתם של קוקרן ואורקוט. תוצאות האמידה מוצגות להלן:

Depended Variable: C

Parameter Estimates						
		:T for H0	Standard	Parameter	DF	Variable
	Error	Parameter=0	Parameter=0	Prob> T Estimate		
	1128.38	-1.864	0.069	-2103.53	1	<u>INTERCEP</u>
	0.0510	14.086	0.000	0.71944	1	Y
AR(1)	1	0.7070	0.1525	4.634	0.000	
AR(2)	1	-0.0064	0.1889	-0.034	0.972	
AR(3)	1	-0.3282	0.1669	-1.966	0.0562	
Durbin-Watson D	1.954					

ב. רשמו את המשוואה המתוקנת המשמשת לעריכת תחזיות.

- 8) חוקר רצה לאמוד את עקומת הביקוש לטיסות לאירופה. לרשותו נתונים שבועיים לאורך 3 שנים (52 שבועות). נסמן:
- $Y_t$  - מספר כרטיסי הטיסה לאירופה שנמכרו בשבוע  $t$ .
  - $p_t$  - מחיר ממוצע ב-\$ של הכרטיסים שנמכרו בשבוע  $t$ .
- החוקר אמד את המודל:  $Y_t = e^\alpha \cdot P_t^{\beta_1} \cdot P_{t-1}^{\beta_2} \cdot e^{u_t}$ .
- וקיבל לאחר הטרנספורמציה הלוגריתמית:  $R^2 = 0.81$ . לבדיקת ההשערה כי קיים מתאם סדרתי בנתונים מסדר ראשון הוא חישב את ערכי  $\hat{u}_t$  ולאחר מכן חישב את הרגרסיה:  $u_t = \gamma_0 + \gamma_1 \ln P_t + \gamma_2 \ln P_{t-1} + \gamma_3 u_{t-1} + v_t$ . מקדם ההסבר המרובה ברגרסיה זו הוא 0.282.
- א. נסח את ההשערה ובחן אותה בר"מ של 0.05. החוקר מניח שיש מתאם סדרתי מסדר ראשון.
- לאחר תיקון Cochrane-Orcutt התקבל:  $\ln \hat{Y}_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$ ,  $\hat{\rho} = 0.2$ . הניחו שהשבוע ובשבוע שעבר מחיר ממוצע של כרטיס היה \$500. השבוע נמכרו 6,185 כרטיסים. בשבוע הבא צפוי מחיר של \$400.
- ב. כמה כרטיסים יימכרו?
- החוקר גם מנסה לקבוע האם בנתונים אלה קיים מתאם מסדר שני.
- ג. רשמו את המשוואה הנוספת שעליו לאמוד. במשוואה הנוספת התקבל מתאם מרובה השווה ל-0.12.
- ד. מהי המסקנה בר"מ של 0.05?

## תשובות סופיות:

- (1) א.  $r_{(u_t, u_{t-1})} = 0.9$  . ב.  $r_{(u_t, u_{t-4})} = 0.6561$  . ג.  $\sigma_u^2 = 5.263$  .
- ד.  $\sigma_u^2 = 1.19$  ,  $r_{(u_t, u_{t-4})} = 0.0256$  ,  $r_{(u_t, u_{t-1})} = 0.4$  .
- (2) יש עדות לכך.
- (3) יש עדות לכך.
- (4) א. יש עדות לכך. ב.  $\hat{\rho} = 0.927$  . ג. חיובי.
- (5) לא נכונה.
- (6) ראו סרטון.
- (7) א. קיומו של מתאם סדרתי מסדר שלישי בנתונים.  
LM  
 $H_0 : \rho_1 = \rho_2 = \rho_3 = 0$   
 $H_1 : OTHERWISE$   
 $LM_{stat} = 22.99$   
יש עדות לכך.
- ב.  $\hat{C}_t = -2103.53 + 0.719 \cdot Y_t + 0.707 \cdot \hat{u}_{t-1} - 0.0064 \cdot \hat{u}_{t-2} - 0.328 \cdot \hat{u}_{t-3}$  .
- (8) א.  $H_0 : \rho = 0$  , יש עדות לכך. ,  $H_1 : \rho \neq 0$  . ב.  $Y_{t+1} = 5,568$  .
- ג.  $u_t = \gamma_0 + \gamma_1 \ln P_t + \gamma_2 \ln P_{t-1} + \gamma_3 u_{t-1} + \gamma_4 u_{t-2} + \omega_t$  . ד. אין עדות לכך.

# אקונומטריקה ב

פרק 10 - סיכום מתאם סדרתי והטרוקדסטיות

תוכן העניינים

1. כללי ..... 51

## סיכום מתאם סדרתי והטרוקדסטיות:

רקע:

הטרוסקדסטיות	מתאם סדרתי	
	למשל: $Y_t = \alpha + \beta X_t + u_t$	המשוואה העיקרית של המודל
$V(u_t) = \sigma^2$	$\text{cov}(u_t, u_{t-s}) = 0$	ההנחה הקלאסית המופרת
$V(u_t) = \sigma_t^2$	$\text{cov}(u_t, u_{t-s}) \neq 0$	המצב לאחר ההפרה
$V(u_t) = W_t \sigma^2$	$u_t = \rho u_{t-1} + \varepsilon_t$	המשוואה המאפיינת את ההפרה
מתקבלים אומדים חסרי הטיה ועקיבים, אך תכונת היעילות נפגעת.		מה קורה אם אומדים OLS-ב
מבחן GQ מבחן White	מבחן DW מבחן LM	זיהוי הבעיה
שיטת WLS	שיטת קוקרן – אורקוט (רגרסיית ההפרשים) הכנסת משתנה מוסבר בפיגור (מודל דינמי)	פתרון הבעיה

# אקונומטריקה ב

פרק 11 - רגרסיה לוגיסטית

תוכן העניינים

1. רגרסיה לוגיסטית.....52

## הגרסה 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 ה"ב"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<sup>a</sup>

Observed				Predicted		Percentage Correct
				whether mom believes course will help		
				no	yes	
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

$$ODDS = \frac{p}{1-p} \text{ : יקרה לא אירוע}$$

ODDS=1 – הסיכוי שהאירוע יתרחש שווה לסיכוי שהוא לא יתרחש  $(\frac{0.5}{0.5})$ .

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

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

: ODDS RATIO (OR)

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

כיצד משתנה ההסתברות במעבר מקבוצה A לקבוצה B.

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

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

OR<1 – הסיכוי להתרחשות האירוע בקבוצה A נמוך מאשר בקבוצה B – קשר שלילי.

## טבלת המקדמים – תרומות ייחודיות של כל מנבא:

(מקביל לטבלת Coefficients בגרסיות לינאריות)

		Variables in the Equation					
		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	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:

$$\log\left(\frac{p}{1-p}\right) = \hat{\alpha} + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\beta}_4 x_{4i}$$

$$p = \frac{1}{1+e^{-\log odds}} : (p) \text{ חישוב הניבוי במונחי הסתברות}$$

$$ODDS = e^{\log odds} : (ODDS) \text{ חישוב הניבוי במונחי סיכויים}$$

4. Exp(B) - יחס הסיכויים (Odds Ratio):  
מבטא את העלייה (אם גדול מ-1) או את הירידה (אם קטן מ-1) בסיכויים להיות בעלי ערך '1' ב-Y כאשר הערך במשתנה המנבא גדל ביחידה אחת.

$$\log \text{Exp}(B) = B \quad ; \quad e^B = \text{Exp}(B) : \text{Exp}(B) \text{ ל-} B \text{ היחס בין } B$$

## 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:

		women		Total
		.00	1.00	
working	.00	13	130	143
	1.00	338	345	683
Total		351	475	826

- על סמך הטבלה חשבו:
- מה ההסתברות של אישה לעבוד?
  - מה הסיכוי של אישה לעבוד?
  - מה ההסתברות של גבר לעבוד?
  - מה הסיכוי של גבר לעבוד?
  - מה יחס הסיכויים (OR) של נשים לעבוד לעומת גברים?
  - מה הלוגריתם של יחס הסיכויים?
  - מה יהיה ערך מקדם השיפוע B בגרסיה הלוגיסטית לניבוי תעסוקה על פי מגדר ומה משמעותו?
  - מה יהיה ערך  $Exp(B)$  בגרסיה 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:

Observed	Predicted	apartm		Percentage Correct
		.00	1.00	
		Step 1 apartm .00	22	11
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 <sup>a</sup>	.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 <sup>a</sup> 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 הבי"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 הבי"ת?
- ז. לאיזה מהמשתנים הבי"ת יש תרומה ייחודית מובהקת לניבוי בעלות על דירה? מהי משמעות מקדם B ו- $\text{Exp}(B)$  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:

	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 <sup>a</sup>	.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 <sup>a</sup> 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, הקורס, הביצוע במהלך הקורס.

- האם למודל הכולל את שלושת המנבאים יכולת הסבר משמעותית?
- כמה אחוזים מתוך השונות של Y מצליח המודל להסביר?
- מהי משוואת הניבוי?
- לאיזה מן המשתנים הב"ת תרומה מובהקת לניבוי?
- הסבירו את משמעות המקדמים (b) שהתקבלו עבור המשתנים הב"ת: מגדר, השכלה קודמת והביצוע במהלך הקורס.
- בטאו את המקדמים במונחי הסיכויים להצלחה בקורס (odds) והסבירו אותם.
- 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.
- 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.

$$\text{נאמד המודל: } p = \frac{1}{1+e^{-z}} \text{ כאשר } p = P(Y=1).$$

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 שח, 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 שח כדי שהסיכוי שלו לצאת למסעדה יהיה גבוה פי 3 מהסיכוי שלו לא לצאת למסעדה.
  - ז. עבור odds ratio של משתנה "שמרטפיות" התקבל רווח בר סמך הבא: [0.123 ; 1.01] ברמת ביטחון של 95%.
- לפיכך ניתן לומר כי למשתנה "שמרטפיות" 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.

$$\text{המודל הנאמד הינו: } p = \frac{1}{1+e^{-z}} \text{ כאשר } p = P(Y=1).$$

$$z = \alpha + \beta_1 age + \beta_2 age^2 + \beta_3 S$$

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
<u>בגרות ריאלית</u>	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 להצלחה 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 להצלח 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. 0.24    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.  $\ln(odds) = -2.734 - 0.498status(1) - 0.52status(2) - 0.18status(3) + 0.93 \cdot incom$   
 ז. משתנה "הכנסה"  
 יא. קטנה.    יב. 0.42    יג. 29,400    יד. 0.093    ט. 3    י. 0.3
- (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".  
 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.