

תרגיל 6

שאלה 1

משרד הרשם באחת האוניברסיטאות בארה"ב דגם 427 סטודנטים וקיבל נתונים לגבי ממוצע הציונים של כל סטודנט בקולג' (COLGPA), בבית הספר התיכון (HSGPA), תוצאות המבחן הפסיכומטרי המילולי (VSAT) והמתמטי (MSAT). המודל הבא נאמד (לצורך הפשטות הושמט האינדקס t):

$$\text{COLGPA} = \beta_1 + \beta_2 \text{HSGPA} + \beta_3 \text{VSAT} + \beta_4 \text{MSAT} + u$$

אומדני הפרמטרים וסטיית התקן שלהם נתונים כדלקמן:

	Coefficient	Standard Error	Test Statistic	Significance
$\hat{\beta}_1$	0.423		0.220	
$\hat{\beta}_2$	0.398		0.061	
$\hat{\beta}_3$	0.0007375		0.0002807	
$\hat{\beta}_4$	0.001015		0.0002936	

- הערך של R^2 שהתקבל ללא התאמה לדרגות חופש (unadjusted) הוא 0.22. בגלל שהערך נמוך קיים חשש שהמודל אינו מתאים. בדוק את טיב ההתאמה של המודל (השתמש ברמת מובהקות של 1%). רשום באופן מפורש מהי השערת האפס, ההשערה האלטרנטיבית, והקריטריון שלפיו השערת האפס נדחית או לא נדחית. מהי המסקנה שלך?
- בחן את מובהקות כל אחד מהמקדמים ברמת מובהקות של 1%, כאשר ההשערה האלטרנטיבית היא שהמקדם חיובי. האם ישנם מקדמים שאינם מובהקים? מלא את המקומות החסרים בטבלה וציין האם המקדמים מובהקים. ציין מהי ההתפלגות של המקדמים ומספר דרגות החופש.
- נניח שסטודנט לקח קורס כדי לשפר את השגיו במבחני SAT וכתוצאה מהקורס הוא הגדיל את מספר הנקודות בחלק המתמטי והמילולי של ה-SAT ב-100 נקודות בכל אחד מהחלקים של המבחן. בממוצע, מהו השיפור הצפוי בציון ה-College GPA?
- נניח שברצונך לבדוק את ההשערה שלפיה המקדמים של המשתנים VSAT ו-MSAT שווים, אבל לא בהכרח שווים ל-0. תאר כיצד תבצע את המבחן. פרט מהי השערת האפס וההשערה האלטרנטיבית, הרגרסיות שיש להריץ לצורך ביצוע המבחן, הסטטיסטי שיש לחשב והתפלגותו, והקריטריון לצורך דחייה או אי דחייה של השערת האפס. תאר את הצעדים הללו עבור שלוש השיטות הנדונות בחלק 4.4 בספר הלימוד.
- רשום משתנים נוספים שלדעתך היו צריכים להכלל במודל. הסבר מדוע לדעתך המשתנים שצינת צריכים להכלל במודל.

שאלה 2

הטבלה הבאה מציגה אומדנים וסטטיסטיים שונים (סטיות תקן בסוגריים) המתייחסים ל לארבעה מודלים. המשתנה התלוי הוא מחיר המכונית והמשתנים המסבירים הם תכונות (מ אפיינים) שונים של כל מכונית. נדגמו 82 תצפיות. להלן ניסוח מתמטי של המודל כאשר מכלילים את כל המשתנים המסבירים:

$$\text{PRICE} = \beta_1 + \beta_2 \text{WBASE} + \beta_3 \text{LENGTH} + \beta_4 \text{WIDTH} + \beta_5 \text{HEIGHT} + \beta_6 \text{WEIGHT} + \beta_7 \text{CYL} + \beta_8 \text{LITERS} + \beta_9 \text{GASMPG} + u$$

PRICE - המחיר הרשום של המכונית באלפי דולרים.

WBASE - בסיס (רוחב) הגלגלים באינצ'ים.

LENGTH - אורך המכונית באינצ'ים.

WIDTH - רוחב המכונית באינצ'ים.

HEIGHT - גובה המכונית באינצ'ים.

WEIGHT - משקל המכונית באלפי פאונד.

CYL - מספר הצילינדרים.

LITERS - נפח מנוע בליטרים.

GASMPG - אומדן מספר המיילים לגלון דלק.

להלן טבלת התוצאות:

Variable	Model A	Model B	Model C	Model D
CONSTANT	58.866 (27.33)	54.400 (23.19)	65.476 (20.1)	71.554 (19.93)
WBASE	0.036 0.28			
LENGTH	-0.394 (0.14)	-0.383 (0.117)	-0.391 (0.117)	-0.403 (0.118)
WIDTH	-0.104 (0.24)			
HEIGHT	-0.748 (0.46)	-0.741 (0.43)	-0.703 (0.43)	-0.839 (0.42)
WEIGHT	2.184 (0.47)	2.148 (0.43)	1.926 (0.36)	2.227 (0.31)
CYL	0.959 (1.31)	1.046 (0.691)	1.095 (0.69)	
LITERS	0.264 (1.83)			
GASMPG	0.196 (0.22)	0.194 (0.20)		
ESS	2303.75	2309.978	2337.952	2414.724

R ² (adjusted)	0.559	0.576	0.576	0.568
Variable	Model A	Model B	Model C	Model D
$\hat{\sigma}^2$	31.558	30.394	30.363	30.958
AIC	34.991	32.610	32.210	32.466
FPE	35.022	32.618	32.214	32.468
HQ	38.906	34.999	34.164	34.033
SCHWARZ	45.57	38.889	37.301	36.51
SHIBATA	34.262	32.293	31.989	32.321
GCV	35.449	32.794	32.335	32.546
RICE	35.996	33	32.472	32.631

1. כתב השאלה טוען שהסימן של מספר מקדמים "מוטעה" עבור מודל A. עבור כל מקדם רגרסיה, ובהתעלם מהחותך, ציין לאיזה סימן תצפה ומדוע. לאחר מכן בדוק האם הסימן שנתקבל הוא זה שצפית.
2. בחן את ההשערה המשותפת ברמת מובהקות 5% שלפיה המקדמים של המשתנים: WBASE, WIDTH, CYL, LITTER ו GASMPG כולם שווים לאפס. ציין מהי השערת האפס וההשערה האלטרנטיבית, חשב את הסטטיסטי של המבחן, ציין מהי התפלגות הסטטיסטי המחושב תחת השערת האפס, וזהה את הקריטריון לדחיית ההשערה. מהי מסקנתך?
3. איזה מארבעת המודלים הוא "הטוב ביותר"? הסבר באיזה קריטריון השתמשת כדי לבחור את המודל "הטוב ביותר". פרש את התוצאות של המודל שבחרת.
4. בחן את המובהקות של מודל A ברמת מובהקות של 1%. ציין מהי השערת האפס וההשערה האלטרנטיבית, חשב את הסטטיסטי של המבחן, ציין מהי התפלגותו וציין מהו כלל החלטה. מהי מסקנתך?

שאלה 3

נתון המודל $Y_t = \beta_1 + \beta_2 X_{t2} + \beta_3 X_{t3} + u_t$. תאר כיצד תקבל את האומדן הטוב ביותר ל β_1 - ול β_3 אם ידוע כי מתקיים $\beta_2=1$.

שאלה 4

נתון המודל הפשוט הבא: $CATCH_t = \alpha + \beta BOATS_t + u_t$ כאשר המשתנה CATCH מציין את משקל הדגים שנתפסו באלפי פאונד והמשתנה BOAT מציין את מספר סירות הדייג שהפליגו ביום t. אם מספר הסירות שהפליגו הוא 0 גם ערך הדגים שנתפסו הוא אפס, ולכן $\alpha=0$. מכאן נובע שהמודל הנכון הוא: $CATCH_t = \beta BOATS_t + v_t$. נניח שהשארית את הקבוע במודל ואמדת מודל הכולל חותך ושיפוע. מהי אופי שגיאת הספיציפיקציה שעשית במקרה כזה? האם הוספת משתנה לא רלוונטי או השמטת משתנה רלוונטי? האם האומדנים (1) מוטים (2) עקיבים (3) יעילים? האם בדיקת השערות תקיפה?

שאלה 5

הטבלה הבאה מציגה אומדנים וסטטיסטיים שונים (סטיות תקן בסוגריים) עבור שלושה מודלים. המשתנה התלוי הוא מספר יחידות הדיור להן ניתן רשיון בנייה והמשתנים המסבירים הם תכונות (מאפיינים) שונים של יחידות הדיור. הנתונים מתייחסים ל 40 ערים בארה"ב. להלן ניסוח מתמטי של המודל כאשר מכלילים את כל המשתנים המסבירים:

$$\text{HOUSING} = \beta_1 + \beta_2 \text{VALUE} + \beta_3 \text{INCOME} + \beta_4 \text{LOCATAX} + \beta_5 \text{STATETAX} + \beta_6 \text{POPCHANGE} + u$$

כאשר המשתנה HOUSING הוא מספר רישיונות הבנייה שהונפקו, המשתנה VALUE הוא המחיר החציוני של דירות בבעלות פרטית (במאות דולרים), המשתנה INCOME הוא ההכנסה החציונית של משק הבית (במאות דולרים), LOCALTAX הוא המס המקומי הממוצע לנפש (בדולרים), STATETAX הוא המס הארצי (הפדרלי) לנפש בדולרים ו- POPCHANGE הוא אחוז הגידול באוכלוסייה בין השנים 1980–1982.

להלן טבלת התוצאות:

Variable	Model A	Model B	Model C
$\hat{\beta}_1$ CONSTANT	-420.323 (0.80)	-1071.982 (0.40)	-973.017 (0.44)
$\hat{\beta}_2$ VALUE	-0.724 (0.15)	-0.864 (0.05)	-0.778 (0.07)
$\hat{\beta}_3$ INCOME	111.898 (0.08)	110.193 (0.08)	116.600 (0.06)
$\hat{\beta}_4$ LOCALTAX	0.503 (0.41)	0.491 (0.41)	
$\hat{\beta}_5$ STATETAX	-0.636 (0.15)		
$\hat{\beta}_6$ POPCHANG	28.257 (0.07)	29.662 (0.05)	24.857 (0.08)
ESS	4.886	4.941	5.038
Unadjusted R2	0.332	0.325	0.312
SIGMASQ	1.437	1.412	1.399
AIC	1.649	1.586	1.538
FPE	1.653	1.588	1.539
HQ	1.807	1.712	1.635
SCHWARZ	2.124	1.959	1.821
SHIBATA	1.588	1.544	1.511
GCV	1.691	1.613	1.555
RICE	1.745	1.647	1.574

1. עבור כל מקדם רגרסיה במודל A בחן האם הוא באופן מובהק שונה מאפס, ברמת מובהקות של 10% (התעלם מהחותך). בהתבסס על המבחנים אילו משתנים יש להשמיט מהמודל ומדוע?
השתמש במודל A כמודל ללא מגבלות (unrestricted) ובמודל C כמודל הכולל מגבלות (restricted) ובצע את הצעדים הבאים לצורך ביצוע מבחן סטיסטי:
2. רשום את השערת האפס וההשערה האלטרנטיבית במונחי β .
3. חשב את הסטיסטי המתאים.
4. ציין מהי התפלגות הסטיסטי המחושב תחת השערת האפס, כולל דרגות החופש.
5. רשום את הערך הקריטי של המבחן, וציין האם עליך לדחות או לא לדחות את השערת האפס ברמת מובהקות של 10%.
6. איזה משלושת המודלים הוא "הטוב ביותר"? הסבר את הקריטריונים שלפיהם בחרת את המודל "הטוב ביותר".
7. נתייחס למודל הטוב ביותר שבחרת בסעיף הקודם. ציין האם לחלק מאומדני המקדמים סימן "שגוי". ציין מהו הסימן שתצפה לו ומדוע ובשלב הבא השווה לסימן שהתקבל בפועל.

פתרון תרגיל 6

פתרון שאלה 1

.1

ניסוח ההשערה:

$$H_0 : \rho^2 = 0$$

$$H_1 : \rho^2 > 0$$

בר"מ 1% השערת האפס נדחית, אם

$$F - stat = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} > F_{k, n-k-1, 0.99}$$

$$F - stat = \frac{0.22/3}{(1 - 0.22)/(427 - 3 - 1)} > F_{3, 427-3-1, 0.99} = 3.8$$

בר"מ 1% הרגרסיה מובהקת.

.2

$$H_0 : \beta_2 \leq 0$$

$$H_1 : \beta_2 > 0$$

$$t - stat = \frac{0.398 - 0}{0.061} = 6.52 > t_{427-3-1, 0.99} = 2.33$$

בר"מ 0.01 המקדם β_2 חיובי.

$$H_0 : \beta_3 \leq 0$$

$$H_1 : \beta_3 > 0$$

$$t - stat = \frac{0.0007375 - 0}{0.0002807} = 2.63 > t_{427-3-1, 0.99} = 2.33$$

בר"מ 0.01 המקדם β_3 חיובי.

$$H_0 : \beta_4 \leq 0$$

$$H_1 : \beta_4 > 0$$

$$t - stat = \frac{0.001015 - 0}{0.0002936} = 3.46 > t_{427-3-1, 0.99} = 2.33$$

בר"מ 0.01 המקדם β_4 חיובי.

אם ההפרעה המקרית מתפלגת נורמלית, אזי האומדנים למקדמים מתפלגים נורמלית.

תחת ההשערה $H_0 : \beta_j = 0$, התוחלת של האומדן $\hat{\beta}_j$ היא אפס.

אם כל ההנחות הקלאסיות מתקיימות, אזי האומדן לסטיית התקן של $\hat{\beta}_j$, שסימנו $S_{\hat{\beta}_j}$, הוא א.ח.ה.

$$\frac{\hat{\beta}_j - 0}{S_{\hat{\beta}_j}} \sim t_{n-k-1} \text{ ומתקיים } \hat{\beta}_j \text{ של התקן של } \hat{\beta}_j$$

ה- test-statistic שבטבלה הוא t-value (ובפתרון זה הוא t-stat).

ה- significance הוא p-value. לחישוב ערכו המדויק יש לחשב את השטח מתחת להתפלגות t,

מהערך t-value ועד לאינסוף (או עד למינוס אינסוף – אם ה- t-value שלילי) ולכפול ב-2.

ברגרסיה זו ערכי p-value קרובים מאוד לאפס.

$$\begin{aligned} \Delta COLGPA &= \hat{\beta}_3 \Delta VSAT + \hat{\beta}_4 \Delta MSAT \\ &= 0.0007375 \cdot 100 + 0.001015 \cdot 100 = 0.17 \end{aligned} \quad .3$$

4. ניסוח ההשערה:

$$\begin{aligned} H_0 : \beta_3 - \beta_4 &= 0 & \text{או} & & H_0 : \beta_3 &= \beta_4 \\ H_1 : \beta_3 - \beta_4 &\neq 0 & & & H_1 : \beta_3 &\neq \beta_4 \end{aligned}$$

נציג את הפתרון בעזרת התפלגות t (שתי שיטות) ובעזרת התפלגות F.

נבחר ברמת מובהקות 0.05.

(1) מבחן t, בשיטה המקובלת

$$t - stat = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{S_{\hat{\beta}_3 - \hat{\beta}_4}} \sim t_{n-k-1} \text{ אזי } \hat{\beta}_3 - \hat{\beta}_4$$

אם $|t - stat| > t_{423,0.975}$, השערת האפס נדחית בר"מ 0.05.

כאן, לא ניתן לחשב את t-stat, כיוון שחסר הנתון על השונות המשותפת בין $\hat{\beta}_3$ לבין $\hat{\beta}_4$,

$$t - stat = \frac{(0.0007375 - 0.001015) - 0}{\sqrt{0.000287^2 + 0.0002936^2 - 2 \cdot S_{\hat{\beta}_3, \hat{\beta}_4}}}$$

$$t_{423-3-1, 0.975} = 1.96$$

(2) מבחן t, בשיטה פחות מקובלת

את הביטוי $\beta_3 - \beta_4$ נכנה באות β ונציג את המודל כך ש- β יהיה מקדם של משתנה מסביר.

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 VSAT + \beta_4 MSAT + u$$

$$\beta = \beta_3 - \beta_4 \quad \rightarrow \quad \beta_4 = \beta_3 - \beta$$

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 VSAT + (\beta_3 - \beta) MSAT + u$$

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 VSAT + \beta_3 MSAT - \beta MSAT + u$$

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 (VSAT + MSAT) - \beta MSAT + u$$

נריץ את המודל הזה ונבחן את המובהקות של β .

(3) מבחן F, שיטת WALD

הרגרסיה הלא-מוגבלת היא $COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 VSAT + \beta_4 MSAT + u$

נריץ אותה ונחשב את $\sum e^2$ או את R^2 .

הרגרסיה המוגבלת, שבה $\beta_3 = \beta_4$, היא

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 VSAT + \beta_3 MSAT + u$$

$$COLGPA = \beta_1 + \beta_2 HSGPA + \beta_3 (VSAT + MSAT) + u$$

נריץ אותה ונחשב את $\sum e^2$ או את R^2 .

אם השערת האפס נכונה,

$$F - stat = \frac{\left(\sum e^2 - \sum e^2 \right) / 1}{\sum e^2 / (427 - 3 - 1)} \sim F_{1,427-3-1}$$

וכאן, כיוון שהמשתנה המוסבר בהרגרסיה המוגבלת הוא המשתנה המוסבר בהרגרסיה הלא-מוגבלת, ניתן לחשב את $F - stat$ גם לפי ערכי R^2 ,

$$F - stat = \frac{\left(R^2 - R^2 \right) / 1}{(1 - R^2) / (427 - 3 - 1)} \sim F_{1,427-3-1}$$

אם $F - stat > F_{1,247-3-1,0.95} = 3.84$, השערת האפס נדחית, בר"מ 0.05.

5. הערה כללית: המדגם כולל רק סטודנטים שהתקבלו לקולג' ולומדים בקולג', שהרי המשתנה המוסבר הוא ממוצע הציונים של הסטודנט בקולג'. אם הקולג' מתנה את הקבלה אליו בסף מסוים של ציונים בבית הספר התיכון או בסף מסוים של ציונים במבחנים הפסיכומטריים, הרי שאלה לא יכולים לשמש ככלי מיון טובים להצלחה בקולג', שהרי אין תצפיות על מידת ההצלחה של סטודנטים שהציונים שלהם בבית הספר התיכון ובמבחנים הפסיכומטריים נמוכים.

ובאשר לציונים במבחנים הפסיכומטריים: ידועה הטענה שאלה אינם מצליחים לנבא הצלחה בלימודים הגבוהים, כיוון שהם אינם מודדים את רמת האינטליגנציה. הם מודדים מיומנויות מסוימות, שניתן לשפר אותן בעזרת אימון הנרכש בקורסי הכנה.

ובאשר למשתנים מסבירים נוספים:

מצב סוציו-אקונומי של הורי הסטודנט. ככל שמצבם טוב יותר, סביר להניח שהסטודנט לא יאלץ לעבוד והציונים שלו יהיו גבוהים יותר.

כמו כן, יש להרחיב את המודל ולקבוע כלי מיון ספציפיים יותר עבור כל פקולטה. יכולת מתמטית (המשתנה MSAT) איננה ערובה להצלחה בפסיכולוגיה או במשפטים, כשם שיכולת וורבלית (המשתנה VSAT) איננה ערובה להצלחה במדעים מדויקים.

פתרון שאלה 2

1. אנו מצפים שהמקדמים של כל המשתנים יהיו חיוביים: ככל שהמכונת גדולה יותר, כבדה יותר וחסכונית יותר – מחירה אמור להיות גבוה יותר.

לבחינת ההשערה שהמקדם חיובי, יש לבחון את ההשערה:

$$H_0 : \beta_j \leq 0$$

$$H_1 : \beta_j > 0$$

והמקדם חיובי אם השערת האפס נדחית.

ברור שעבור המקדמים של LENGTH, WIDTH, HEIGHT השערת האפס איננה נדחית, שהרי האומדנים שהתקבלו למקדמים שלהם הם שליליים.

האומדנים שהתקבלו למקדמים של המשתנים האחרים הם חיוביים, אך הבחינה אם הפרמטרים שהם מייצגים הם חיוביים איננה משמעותית, כיוון שהמקדמים אינם מובהקים (המרחק של כל אומדן מהאפס קטן מסטיית תקן אחת).

אחת מהסיבות למקדמים עם סימן שונה מהמצופה, או למקדמים שאינם מובהקים, היא קשר ליניארי חזק מדי בין המשתנים לבין עצמם (בעיית המולטיקוליניאריות). ואכן, ניתן להניח שבנתונים אלה ישנם קשרים חזקים בין המאפיינים של המכונות.

2.

$$H_0 : \beta_{LITTER} = \beta_{CYL} = \beta_{WIDTH} = \beta_{WBASE} = \beta_{GASMPG} = 0$$

$$H_1 : else$$

השערת האפס כוללת יותר מהצהרה אחת ($m = 5$). ניתן לבחון אותה בעזרת מבחן WALT (התפלגות F).

המודל הלא מוגבל הוא המודל המקורי .

המודל המוגבל, שהושמטו בו כל חמשת המשתנים (ורק הם) הוא מודל D.

$$F - stat = \frac{\left(\sum^{RES} e^2 - \sum^{UNRES} e^2 \right) / m}{\frac{\sum^{UNRES} e^2}{(n - k - 1)}} \sim F_{m, n-k-1}$$

$$F - stat = \frac{(2414.724 - 2303.75) / 5}{2303.75 / (82 - 8 - 1)} = 0.7$$

$$F_{5, 82-8-1, 0.95} =$$

ובר"מ 0.05 ההשערה איננה נדחית. סביר להניח שכל חמשת המשתנים אינם תורמים הסבר משמעותי לשונות המחיר של המכונות.

.3

קריטריון מקובל לבחירת המודל הנכון יותר הוא מקדם ההסבר המתוקן, $R^2_{adjusted}$, שסימונו $\overline{R^2}$. בהינתן אותו משתנה מוסבר, ככל שמקדם ההסבר המתוקן גבוה יותר, סביר להניח שהמודל נכון יותר. מקדם ההסבר המתוקן הגבוה ביותר מבין ארבעת המודלים המוצגים כאן הוא 0.576, והוא התקבל עבור שני מודלים: מודל B ומודל C. כיוון שכך, ננסה לאתר את המודל הנכון יותר לפי קריטריונים נוספים. קריטריונים נוספים שנתונים לנו: RICE-ו GCV, SHIBATA, SCHWARZ, HQ, FPE, AIC. ככל שהם נמוכים יותר, סביר להניח שהמודל נכון יותר. בהשוואה בין מודל B לבין מודל C, כל הקריטריונים מצביעים על מודל C כמודל הנכון יותר.

.4

ניסוח ההשערה:

$$H_0 : \rho^2 = 0$$

$$H_1 : \rho^2 > 0$$

בר"מ 1% השערת האפס נדחית, אם

$$F - stat = \frac{R^2/k}{(1 - R^2)/(n - k - 1)} > F_{k, n-k-1, 0.99}$$

את מקדם ההסבר, R^2 , נחשב לפי ערכו של מקדם ההסבר המתוקן, $\overline{R^2}$, ולפי הנוסחה:

$$\overline{R^2} = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

$$0.559 = 1 - (1 - R^2) \frac{82-1}{82-8-1} \Rightarrow R^2 = 0.6026$$

$$F - stat = \frac{0.6026/8}{(1-0.6026)/(82-8-1)} = 13.84 > F_{8,82-8-1, 0.99} = 2.742$$

ובר"מ 1% הרגרסיה מובהקת.

פתרון שאלה 3

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

$$Y_t = \beta_1 + 1 \cdot X_{2t} + \beta_3 X_{3t} + u_t \quad \beta_2 = 1 \text{ תחת המידע}$$

ניתן להשתמש בנוסחאות הקיימות לחישוב האומדנים, רק אם בצד ימין של המשוואה ישנם משתנים מסבירים שהמקדמים שלהם הם פרמטרים לא ידועים. (האמידה ב- OLS מחפשת את האומדנים האופטימליים לפרמטרים, מבלי לקחת בחשבון מגבלה כלשהיא עליהם).

לפיכך, יש להעביר לצד שמאל של המשוואה את $1 \cdot X_{2t}$. נקבל,

$$Y_t - 1 \cdot X_{2t} = \beta_1 + \beta_3 X_{3t} + u_t$$

זהו מודל עם חותך, משתנה מסביר אחד, X_3 , ומשתנה מוסבר $Z = Y - X_2$.

$$\hat{\beta}_3 = \frac{\sum X_3 Z}{\sum X_3^2} \quad \hat{\beta}_1 = \bar{Z} - \hat{\beta}_3 \cdot \bar{X}_3$$

ומשוואת האמידה היא $\hat{Y}_t = \hat{\beta}_1 + X_{2t} + \hat{\beta}_3 X_{3t}$.

פתרון שאלה 4

אם המודל הנכון הוא מודל ללא חותך, ואנו כוללים בו חותך, זוהי שגיאת ספציפיקציה של הכללה של משתנה לא רלוונטי.

האומדנים אינם מוטים. האומדנים עקיבים. האומדנים לא יעילים. כל מבחני ההשערה תקפים.

פתרון שאלה 5

.1

$$H_0 : \beta_2 = 0$$

$$H_1 : \beta_2 \neq 0 \quad t - value = \frac{-0.724 - 0}{0.15} = 4.83 > t_{40-5-1,0.95} = 1.645$$

ובר"מ 0.10 המקדם β_2 מובהק.

$$H_0 : \beta_3 = 0$$

$$H_1 : \beta_3 \neq 0 \quad t - value = \frac{111.898 - 0}{0.08} = 1398.725 > t_{40-5-1,0.95} = 1.645$$

ובר"מ 0.10 המקדם β_3 מובהק.

$$H_0 : \beta_4 = 0$$

$$H_1 : \beta_4 \neq 0 \quad t - value = \frac{0.503 - 0}{0.41} = 1.232 < t_{40-5-1,0.95} = 1.645$$

ובר"מ 0.10 המקדם β_4 לא מובהק.

$$H_0 : \beta_5 = 0$$

$$H_1 : \beta_5 \neq 0 \quad t - value = \frac{-0.636 - 0}{0.15} = 4.24 > t_{40-5-1,0.95} = 1.645$$

ובר"מ 0.10 המקדם β_5 מובהק.

$$H_0 : \beta_6 = 0$$

$$H_1 : \beta_6 \neq 0 \quad t - value = \frac{28.25 - 0}{0.07} = 403.57 > t_{40-5-1,0.95} = 1.645$$

ובר"מ 0.10 המקדם β_6 מובהק.

כיוון שהמקדם β_4 לא מובהק, והוא גם המשתנה היחיד שהמקדם שלו לא מובהק (כך שאין לחשוש שהוא לא מובהק בשל קשר ליניארי בינו לבין המשתנים המסבירים האחרים), ניתן להשמיט מהרגרסיה את המשתנה LOCALTAX.

.2

$$H_0 : \beta_4 = \beta_5 = 0$$

$$H_1 : else$$

.3

$$F - stat = \frac{(0.332 - 0.312) / 2}{(1 - 0.332) / (40 - 5 - 1)} = 0.51$$

4. אם השערת האפס נכונה, $F - stat \sim F_{2,40-5-1}$.

5. הערך הקריטי של F הוא $F_{2,40-5-1, 0.10} = 2.46$.

בר"מ 0.10 השערת האפס איננה נדחית.

6. לבחירת המודל הנכון יותר, נחשב תחילה את מקדמי ההסבר המתוקננים.

$$\overline{R^2} = 1 - (1 - R^2) \frac{n-1}{n-k-1}$$

$$\overline{R^2}_A = 1 - (1 - 0.332) \frac{40-1}{40-5-1} = 0.2338$$

$$\overline{R^2}_B = 1 - (1 - 0.325) \frac{40-1}{40-4-1} = 0.2479$$

$$\overline{R^2}_C = 1 - (1 - 0.312) \frac{40-1}{40-3-1} = 0.2546$$

מקדם ההסבר המתוקנן הגבוה ביותר התקבל עבור מודל C וסביר להניח שזה המודל הנכון יותר. מודל C ייבחר כמודל הנכון יותר גם לפי כל הקריטריונים האחרים הרשומים בפלט, שכן כולם מקבלים ערך נמוך יותר עבור מודל זה.

7. האומדן למקדם של המשתנה VALUE הוא שלילי. המשמעות היא שבעיר שבה מחירי הדירות גבוהים יותר – מספר רישיונות הבנייה שהונפקו נמוך יותר. נראה שסימנו השלילי של המקדם שגוי, שכן, בעיר שבה מחירי הדירות גבוהים יותר, הבנייה כדאית יותר.

האומדן למקדם של המשתנה INCOME הוא חיובי. המשמעות היא שבעיר שבה ההכנסות גבוהות יותר – מספר רישיונות הבנייה שהונפקו גבוה יותר. ואכן, סימנו החיובי של המקדם נראה הגיוני, שכן, בעיר שבה ההכנסות גבוהות יותר, הדירות יוקרתיות יותר והבנייה כדאית יותר.

האומדן למקדם של המשתנה POPCHANGE הוא חיובי. המשמעות היא שבעיר שבה הגידול באוכלוסייה מהיר יותר – מספר רישיונות הבנייה שהונפקו גבוה יותר. סימנו החיובי של המקדם הגיוני בהחלט.