

תרגיל 6

שאלה 1

(4.9) 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, הרגסיות 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 שציינת צריכים להכלל במודל.

שאלה 2

(4.15) הטבלה הבאה מציגה אומדנים וסטטיסטיים שונים (סטיות תקן בסוגריים) המתייחסים

לארבעה מודלים. המשתנה התלוי הוא מחיר המכונית והמשתנים המסבירים הם תכונות

(מאפיינים) 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. כותב השאלה טוען שהסימן של מספר מקדמים "מוטעה" עבור מודל A. עבור כל מקדם גרסיה, 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 ההחלטה. מהי מסקנתך?

שאלה 3

- (4.1) 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

- (4.7) 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 הסירות 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

(4.20) 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 רשיונות 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 באוכלוסיה 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.