

תרגיל 12

שאלה 1

(7.1) נניח שמתקיים הקשר הבא בין כמות הכסף שהמדינה מוציאה על תכניות רווחה (Y) והכנסת המדינה (X):

$$Y = \alpha_1 + \alpha_2 D1 + \alpha_3 X + \alpha_4 (D1 * X) + u$$

כאשר D1 הוא משתנה דמי המקבל את הערך 1 אם בית המחקר של המדינה נשלט ע"י דמוקרטים ו 0 אחרת. משמעות הסימן * הוא כפל. עוזר המחקר של החליט להגדיר משתנה דמי אחר, D2, המקבל את הערך 1 אם בית המחקר שלט ע"י לא-דמוקרטים ו 0 אחרת. עוזר המחקר אמד את המודל הבא:

$$Y = \beta_1 + \beta_2 D2 + \beta_3 X + \beta_4 (D2 * X) + u$$

תאר צעד אחרי צעד כיצד ניתן לקבל אומדנים של α מהאומדנים של β מבלי להריץ רגרסיה נוספת. במילים אחרות: עליך לגזור ביטויים עבור α במונחי β .

שאלה 2

(7.3) חוקר הריץ רגרסיה בהתבסס על מדגם מקרי של 1,629 מעסיקים. נתקבלו התוצאות הבאות:

$$\widehat{LNWAGE} = 0.448 + 0.0795SCHOOL + 0.059EXP - 0.00076EXP^2 \\ + 0.18725URBAN - 0.04393SOUTH \\ - 0.01172HEALTH + 0.03173UNION$$

כאשר LNWAGE הוא הלוגריתם של תעריף השכר לשעה, SCHOOL הוא מספר שנות הלימוד, EXP הוא מספר שנות הנסיון, EXP^2 הוא הריבוע של EXP, URBAN=1 אם המעסיק גר בעיר ו 0 אחרת, SOUTH=1 אם המעסיק גר במדינה דרומית בארה"ב, HEALTH=1 אם למעסיק יש בעיות בריאותיות, ו UNION=1 אם המעסיק הוא חבר איגוד.

1. "סביר שיש קורלציה גבוהה בין המשתנים EXP ו EXP^2 לכן, האומדנים מוטים וכל מבחן השערות המבוסס עליהם אינו תקף". האם הטענה הזאת נכונה? נמק את תשובתך.
2. רשום את היחס הנאמד עבור מעסיק חבר ארגון, ללא בעיות בריאותיות, המתגורר במדינה דרומית (ברשותך כל המידע הדרוש כדי לענות על השאלה).
3. גזור ביטוי לגמישות החלקית של השכר ביחס לנסיון, וחשב את הערך המספרי שלו כאשר הנסיון הוא חמש שנים.

שאלה 3

- (7.5) נתונה פונקציית תצרוכת של משקי בית (נסמן אותה כמודל A): $C = \alpha + \beta Y + u$, כאשר המשתנה C מוגדר כהוצאות משק הבית לתצרוכת ו Y היא ההכנסה הפנויה (הכנסה בניכוי המס, באלפי דולרים). נתייחס עתה להשערות הבאות לגבי המקדמים של פונקציית התצרוכת: (1) הערך של המקדם α תלוי בגודל המשפחה N ובבעלות על הדירה שבה מתגוררת המשפחה. (2) הערך של המקדם β תלוי לינארית ברמת הכנסה, בגודל המשפחה N ובבעלות על הדירה שבה מתגוררת המשפחה. ברשותך נתונים לגבי 40 משקי בית.
- נסח את המודל הכללי ביותר שיאפשר לך לבדוק את שתי ההשערות (נסמנו כמודל B). כדי לנסח את המודל וודא שהגדרת את כל המשתנים החדשים הדרושים לבדיקת ההשערות. שים לב לנסח את המודל בזיהירות. טעות בסעיף זה יגרור טעויות בכל הסעיפים הבאים של השאלה.
 - רשום את המשוואה המתייחסת למשפחות שהן שוכרי דירות. חזור על הסעיף עבור משפחות שהן בעלות דירה.
 - בהתייחס למודל B, ברצונך לבדוק את ההשערה שלפיה לגודל המשפחה אין השפעה מובהקת על התצרוכת. ציין מהי השערת האפס וההשערה האלטרנטיבית, ורשום מהו המודל עם מגבלות (restricted).
 - לצורך ביצוע את מבחן Wald, תאר אילו רגרסיות תריץ ואיזה סטטיסטי תחשב (הצב ערכים מספריים בנוסחה, היכן שערכים אלה ידועים).
 - ציין מהי ההתפלגות של הסטטיסטי המחושב, והערכים המספריים של דרגות החופש.
 - תאר כיצד תשתמש בגישת ערך ה-p (p-value approach) כדי להחליט האם לדחות את השערת האפס ברמת מובהקות של 5%.
 - במודל B גזור את הנטייה השולית לצרוך: (1) ביחס להכנסה (2) ביחס לגודל המשפחה. ערוך את החישוב בנפרד עבור: (1) שוכרי דירות. (2) בעלי דירות.
 - ציין מהי השערת האפס וההשערה האלטרנטיבית, אם נתייחס למודל A כמודל עם מגבלות (restricted) ולמודל B כמודל ללא מגבלות (unrestricted).
 - בהנחה שמודל A הוא מודל הבסיס, ברצונך לבצע מבחן LM לצורך הוספה של משתנים חדשים במודל B. תאר אילו רגרסיות עליך להריץ ואילו סטטיסטיים עליך לחשב. הסבר בבהירות את הצעדים לביצוע המבחן והצב ערכים מספריים בנוסחה במידת האפשר.
 - מהי התפלגות הסטטיסטי המחושב? מהם הערכים המספריים של דרגות החופש? מהו הערך הקריטי ברמת מובהקות של 5%?
 - תאר את כלל ההחלטה לדחיית השערת האפס.
 - ניח שדחיית את השערת האפס. אילו צעדים נוספים תצטרך לנקוט כדי לקבל מודל יותר כללי ממודל A מבלי להשתמש בגישת Wald?
 - ניח ש Y^* הוא ההכנסה במונחי מאות דולרים והמשתנה Y^* מחליף את המשתנה Y במודל B. רשום מהו היחס בין כל אחד מהמקדמים לפני ואחרי השינוי.

שאלה 4

(7.8) נתון המודל הבא: $P_t = \alpha + \beta S_t + u_t$, כאשר S ו P הם הרווחים והמכירות של חברה בתקופה t . מנהלת המכירות של החברה חושבת שהיחס המתואר במודל משתנה בהתאם לארבע עונות השנה (קיץ, חורף, סתו, אביב). היא סיפקה לך נתונים לכל עונה במשך 20 שנה ($n=80$) והיא מעוניינת שתבחן את השערת האפס שלפיה אין הבדל בפרמטרים כאשר העונה משתנה.

1. נסח מודל שיאפשר לך לבדוק את השערת האפס.
2. רשום מהי השערת האפס וההשערה האלטרנטיבית.
3. תאר כיצד תבדוק את ההשערה תוך שימוש בגישת Wald. בפרט, ציין אילו רגרסיות תריץ, כיצד תחשב את הסטטיסטי, מהי התפלגות הסטטיסטי ומספר דרגות החופש, מהו הקריטריון לדחייה או אי דחייה של השערת האפס.
4. חזור על השאלה תוך שימוש בגישת LM.

שאלה 5

(7.15) המודל המפורט להלן הוצע ע"י Patel, Sikula, and McKenna (1983). המודל מתאר את המוביליות של פרופסורים למנהל עסקים:

$$AVGYR = \alpha + \beta PAPERS + \gamma AGE + u$$

כאשר AVGYR הוא מדד המוביליות המוגדר כמספר השנים הממוצע שהפרופסור מועסק במכון מחקר או באוניברסיטה, PAPERS הוא מספר המאמרים שהתפרסם, AGE הוא גיל הפרופסור בשנים. קיים גם מידע זמין לגבי: (1) דרגת כל אחד מהפרופסורים במדגם (עוזר הוראה, פרופסור חבר, פרופסור מן המניין), (2) האם קיימת הזדמנות טובה לקידום (3) שביעות הרצון של הפרופסור מהאדמיניסטרציה.

1. נסח מודל שיאפשר לך לבחון האם α זהה עבור כל הפרופסורים באותה דרגה ובעלי אותו תכונות איכותיות אחרות כמתואר בשאלה. תאר כיצד תבצע את המבחן.
2. חזור על השאלה עבור שלושת הפרמטרים במשותף (α, β, γ) .

שאלה 6

(7.16) נגדיר את E_t כתצרוכת אנרגיה בזמן t ואת T_t כטמפרטורה בזמן t . נניח קיומו של היחס הבא בין המשתנים $E_t = a + bT_t + u_t$ וזאת כמתואר בסעיף 7.5 בספר הלימוד. סביר להניח, שהשפעת הטמפרטורה על תצרוכת האנרגיה תלויה בזמן במשך היום (בוקר, צהריים, ערב, לילה). תאר כיצד תשתמש במבחן LM כדי לבחון את ההשערה המתוארת לגבי כל הפרמטרים.

שאלה 7

(7.21) בשאלה זאת עליך להעזר בקבצי הנתונים הנמצאים ע"ג הדיסק המצורף לספר הלימוד.

השאלה דורשת הרצת פרוצדורות סטטיסטיות באמצעות מחשב. קובץ DATA7-10 מכיל נתונים לגבי איכות האוויר ב 30 איזורים מטרופוליטניים בארה"ב (SMSA), וכן נתונים לגבי משתנים המסבירים את איכות האוויר. נתוני החתך מתייחסים לתקופה שבין השנים 1970-1972 עבור המשתנים שונים. SMSA הם יחידות חברתיות וכלכליות משולבות הכוללות לפחות 5,000 תושבים. רוב האיזורים הם מחוזות או אפילו איזורים גדולים יותר ממחוזות.

1. הסבר בפירוט מדוע כל אחד מהמשתנים המסבירים בקובץ עשויים להשפיע על איכות האוויר. ציין מהו הסימן שתצפה בהתייחס למקדם של כל משתנה. אם לדעתך קיימים משתנים שאינם רלוונטיים, ציין זאת.
2. נסח מודל בסיסי עם איכות האוויר כמשתנה תלוי ורשימה של משתנים מסבירים (אל תכלול משתנים בינריים בסעיף זה של השאלה). הצדק את בחירתך. ע"י שימוש בתכנת רגרסיה, אמוד את המודל.
3. האם בנתונים שלך יש בעיית מולטיקולינאריות? אם תשובתך חיובית, האם התוצאות שקיבלת מושפעות ע"י בעיית המולטיקולינאריות?
4. תאר כיצד משתני הדמי יכולים להשפיע על היחס בין המשתנים, כפי שניסחת בסעיף 2. נסח מודל חדש ואמוד אותו. בצע מבחן סטטיסטי מתאים כדי להשוות את שני המודלים. מהי המסקנה שלך?