**תרגול מס' 12**

**רגרסיה ליניארית מרובה**

**רגרסיה ליניארית מרובה** הינה הרחבה של מודל רגרסיה ליניארית פשוטה (רל"פ).   
במקום משתנה מסביר אחד בלבד, ברגרסיה מרובה יש מספר רב של משתנים מסבירים: . למודל רגרסיה מרובה יישומים רבים כאשר מעוניינים להעריך את ההשפעה הסימולטנית של מספר גורמים על משתנה תגובה כלשהו (המשתנה המוסבר).

נקבל אוסף של תצפיות: ,   
כאשר הוא המשתנה התלוי.



**לדוגמה:** רוצים לבחון את הפרמטרים המשפיעים על מחיר דירה.

המשתנה התלוי הוא מחיר הדירה.

המשתנים המסבירים הם מספר החדרים בדירה, שטחה, והקומה שהיא נמצאת בה.

מחיר הדירה מושפע מהקומבינציה של המשתנים המסבירים (שילוב של מספר חדרים, שטח וקומה).

**מודל הרגרסיה הליניארית המרובה**

הנחות מודל הרגרסיה הליניארית המרובה מהוות הכללה של ההנחות שראינו עבור רל"פ. בפרט:

 - התצפית ה- של המשתנה המוסבר

 רעש אקראי. שימו לב שהשונויות שוות לכל .

 הם מקדמי הרגרסיה, כאשר:

* הוא החותך - התוחלת של כאשר כל המשתנים המסבירים שווים ל-0.
* הוא השיפוע של מישור הרגרסיה בכיוון של , כלומר התוספת לתוחלת של כתוצאה מהגדלת ביחידה אחת, כאשר שאר המשתנים מוחזקים קבועים

כמו ברל"פ, גם כאן המקדמים אינם ידועים, ולכן מייצרים אומדי ריבועים פחותים (האומדים מסומנים כ-). הפיתוח הוא די מייגע, ולכן לא נבצע אותו ידנית, אלא רק ניעזר בפלטים של כלים סטטיסטיים ממוחשבים.

**בדיקת מובהקות מודל הרגרסיה השלם – מבחן F**

היינו רוצים לבדוק אם המשתנים המסבירים  משפיעים על המשתנה התלוי ().

מספיק קשר עם משתנה מסביר אחד (עם אחד מה – x – ים) כדי להגיד שקיימת השפעה כזו!   
פורמלית, מערכת ההשערות שלנו היא:



סטטיסטי המבחן, כמו ברל"פ, הוא .

מוגדרים באופן דומה לרל"פ, אבל מספר דרגות החופש שלהם שונה:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Mean Square** |  | **Sum of Squares** | **Source of Variation** |
|  |  |  |  | Regression |
|  |  |  |  | Error |
|  |  |  |  | Total |

לכן **כלל ההכרעה בר"מ** : דחה אם

כמו ברל"פ, מהווה אומד חסר-הטייה ל- (שונות הרעש).

**בעיה לדוגמא – איי גלפגוס**

מעוניינים למצוא את הגורמים המשפיעים על מספר זני בעלי החיים בכל אחד מהאיים בקבוצת איי גלפגוס. הגורמים שהוצעו הינם:

Area – שטח האי (קמ"ר)

Elevation – גובה הנקודה הגבוהה ביותר באי (מטרים)

Nearest – מרחק האי הקרוב ביותר (ק"מ)

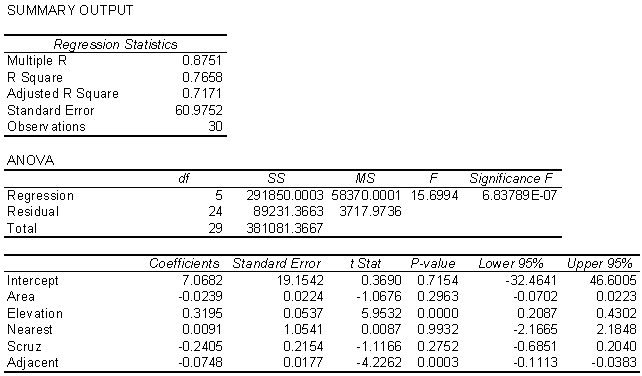
Scruz – המרחק מהאי Santa Cruz (ק"מ)

Adjacent – שטח האי הקרוב ביותר (קמ"ר)

Species – מספר הזנים שניתן למצוא על האי.

נאספו נתונים על 30 איים, מרוכזים בטבלה בעמוד הבא.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **island** | **Species** | **Area** | **Elevation** | **Nearest** | **Scruz** | **Adjacent** |
| **i** | **Yi** | **X1** | **X2** | **X3** | **X4** | **X5** |
| Baltra | 58 | 25.09 | 346 | 0.6 | 0.6 | 1.84 |
| Bartolome | 31 | 1.24 | 109 | 0.6 | 26.3 | 572.33 |
| Caldwell | 3 | 0.21 | 114 | 2.8 | 58.7 | 0.78 |
| Champion | 25 | 0.1 | 46 | 1.9 | 47.4 | 0.18 |
| Coamano | 2 | 0.05 | 77 | 1.9 | 1.9 | 903.82 |
| Daphne.Major | 18 | 0.34 | 119 | 8 | 8 | 1.84 |
| Daphne.Minor | 24 | 0.08 | 93 | 6 | 12 | 0.34 |
| Darwin | 10 | 2.33 | 168 | 34.1 | 290.2 | 2.85 |
| Eden | 8 | 0.03 | 71 | 0.4 | 0.4 | 17.95 |
| Enderby | 2 | 0.18 | 112 | 2.6 | 50.2 | 0.1 |
| Espanola | 97 | 58.27 | 198 | 1.1 | 88.3 | 0.57 |
| Fernandina | 93 | 634.49 | 1494 | 4.3 | 95.3 | 4669.32 |
| Gardner1 | 58 | 0.57 | 49 | 1.1 | 93.1 | 58.27 |
| Gardner2 | 5 | 0.78 | 227 | 4.6 | 62.2 | 0.21 |
| Genovesa | 40 | 17.35 | 76 | 47.4 | 92.2 | 129.49 |
| Isabela | 347 | 4669.32 | 1707 | 0.7 | 28.1 | 634.49 |
| Marchena | 51 | 129.49 | 343 | 29.1 | 85.9 | 59.56 |
| Onslow | 2 | 0.01 | 25 | 3.3 | 45.9 | 0.1 |
| Pinta | 104 | 59.56 | 777 | 29.1 | 119.6 | 129.49 |
| Pinzon | 108 | 17.95 | 458 | 10.7 | 10.7 | 0.03 |
| Las.Plazas | 12 | 0.23 | 94 | 0.5 | 0.6 | 25.09 |
| Rabida | 70 | 4.89 | 367 | 4.4 | 24.4 | 572.33 |
| SanCristobal | 280 | 551.62 | 716 | 45.2 | 66.6 | 0.57 |
| SanSalvador | 237 | 572.33 | 906 | 0.2 | 19.8 | 4.89 |
| SantaCruz | 444 | 903.82 | 864 | 0.6 | 0 | 0.52 |
| SantaFe | 62 | 24.08 | 259 | 16.5 | 16.5 | 0.52 |
| SantaMaria | 285 | 170.92 | 640 | 2.6 | 49.2 | 0.1 |
| Seymour | 44 | 1.84 | 147 | 0.6 | 9.6 | 25.09 |
| Tortuga | 16 | 1.24 | 186 | 6.8 | 50.9 | 17.95 |
| Wolf | 21 | 2.85 | 253 | 34.1 | 254.7 | 2.33 |

**פלט הרגרסיה:**

**מבחן F לכל המודל**

**מבחן t**

**לכל מקדם בנפרד**

 **>** 

נבדוק את כלל הדחייה: ולכן נדחה את .   
כלומר: מספר הזנים השונים על איי הגלפגוס **כן** מושפע מלפחות אחד מהמשתנים שבדקנו.

שיקול נוסף שאפשר להפעיל: PV של סטטיסטי המבחן נמוך מאוד (סדר גודל של ) ולכן לכל רמת מובהקות סבירה (למשל, 5%) נדחה את השערת האפס.

**הסקה על מקדמי הרגרסיה**

לאחר מבחן כולל למובהקות המודל הרב משתני, נרצה לבדוק השערות לגבי פרמטרים בודדים או קבוצות של פרמטרים. **קיימות שתי גישות לבדיקת השערות של המקדמים: מבחן t, ומבחן F חלקי.**

1. **מבחן :** זהה למבחן המבוצע ברגרסיה פשוטה.מתבסס על ההתפלגות הנורמלית של אמדי הריבועים הפחותים ועל אמדי סטיות התקן המתקבלים בתהליך האמידה. מאפשר בחינה של כל פרמטר בנפרד.

אם הפרמטר ה- שווה ל-0, אין למשתנה ה- השפעה ליניארית על המשתנה המוסבר.



סטטיסטי המבחן הוא , כאשר היא סטיית התקן של האומד למקדם של המשתנה ה- (ערך זה מופיע בפלט של הרגרסיה).

כלל ההכרעה: דחה אם .

**הערה (1) :** רב"ס דו"צ לפרמטר ברמת סמך : [

**הערה (2):** ניתן לבצע מבחני השערות על כל שיפוע בנפרד (ח"צ / דו"צ), באופן דומה לתרגול 11 עמ' 5, למעט העובדה שמספר דרגות החופש ברגרסיה ליניארית מרובה הוא .

1. **מבחן F** **חלקי** – בוחן את התרומה של קבוצת משתנים מסבירים להסבר שהרגרסיה מספקת (ניתן לבדוק פרמטר יחיד, או קבוצת פרמטרים בו"ז). משווים את בלי קבוצת המשתנים, לעומת הערך של רגרסיה שכוללת אותם.

לצורך הדגמת הסימון שנשתמש בו, נניח שאנו בוחנים שלושה משתנים מסבירים - .

- השונות המוסברת ע"י מודל רגרסיה הכולל את שלושת המשתנים.

– השונות המוסברת ע"י מודל רגרסיה הכולל רק את המשתנה .

אז ניתן לומר ש"תוספת ההסבר" שהמשתנים מספקים היא:

נרצה לבדוק אם הערך הזה אכן מובהק בעזרת מבחן השערות.

**באופן כללי, נגדיר:**

המודל "המלא" – כולל קבוצת המשתנים שאנחנו בוחנים

המודל המצומצם – ללא קבוצת המשתנים שאנחנו בוחנים

בכל מודל, מחשבים בנפרד את . מספר דרגות החופש () של בכל מודל הוא מספר המשתנים המסבירים הנכללים בו.

סטטיסטי המבחן:

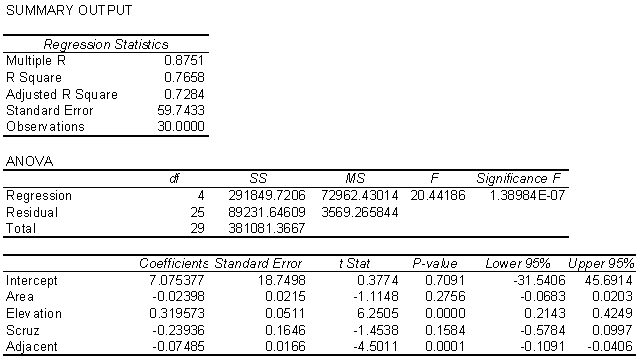
שימו לב שניתן לחלק גם את המונה וגם את המכנה ב-, *ולקבל את הביטוי החלופי הבא ל-:*

כלל ההכרעה בר"מ : דחה את השערת האפס אם .

**למשל,** בדוגמת גלפאגוס:

נבדוק אם למשתנה Nearest יש תרומה מובהקת למודל.

פלט הרגרסיה ללא Nearest:



|  |  |  |
| --- | --- | --- |
| במודל המלא (עמ' 3): |  |  |
| במודל החלקי (עמ' 5): |  |  |

ולכן לא דוחים את , כלומר **לפי מבחן F חלקי** למשתנה אין השפעה ליניארית מובהקת (בר"מ 0.05) על מספר הזנים באי.

נבדוק אם מגיעים לאותה מסקנה לפי מבחן במודל הרגרסיה המלא (עמ' 3):

כלל הדחייה של מבחן : דחה אם , ולכן לא נדחה את - גם לפי *מבחן אין למשתנה המסביר השפעה מובהקת על מספר הזנים באי.*

ניתן להגיע למסקנה דומה גם בעזרת של המשתנה שהוא גבוה מאוד!

**מקדם ההסבר ברגרסיה ליניארית מרובה**

באופן כללי, ככל שמספר המשתנים המסבירים גדל, כך הרגרסיה יכולה להתאים משוואה טובה יותר לנתונים. לכן, מקדם ההסבר לעולם לא יירד כאשר נוסיף משתנים מסבירים למודל. לכאורה, סך השונות המוסברת גדל (וזה טוב), אולם הגידול הזה הינו טכני ומלאכותי, ולא נובע בהכרח מ"הסבר" טוב יותר של המשתנים באופן מהותי.

לכן, שימוש ב- כמדד ליכולת ההסבר של המודל הינו בעייתי במודל רגרסיה ליניארית מרובה.

מדד חלופי הינו , אשר כולל "מעניש" כנגד הכללת משתנים מסבירים רבים במודל:

– מספר התצפיות

– מספר המשתנים המסבירים במודל

ניתן להראות שבכל מודל רגרסיה מתקיים , כך שיש פחות תמריץ להוסיף הרבה משתנים למודל.

בדוגמת גלפגוס: ניתן לראות כי המודל ללא Nearest הינו בעל גבוה יותר מאשר המודל המלא, לכן נעדיף את המודל הזה. בשלב זה ניתן לנסות להוסיף/להוריד משתנים נוספים במטרה למצוא מודל טוב יותר.

דוגמה לחישוב במודל ללא Nearest (על סמך הפלט שמופיע בעמ' 5):

**מולטיקוליניאריות**

**מולטיקוליניאריות היא תופעה שבה קיים מתאם ליניארי חזק בין שני משתנים מסבירים (או יותר).**

במקרה כזה, המשתנים המסבירים מספקים מידע יתיר לגבי המשתנה המוסבר. מסתבר שלמולטיקוליניאריות יש השלכות בעייתיות על האמינות והיציבות של האומדנים שהרגרסיה מספקת.

**השלכות אפשריות של מולטיקוליניאריות** - אלו בד"כ סימנים מעידים לנוכחות של מולטיקוליניאריות בנתונים:

1. **האומדים לשונות של מקדמי הרגרסיה () "מתנפחים".** דברזה גורם לכך שהרב"סים למקדמים האמיתיים () הרבה יותר רחבים מכפי שהם אמורים להיות, ואז הרגרסיה פחות אמינה.
2. **תוצאות "מוזרות": סימני המקדמים הפוכים מהצפוי.**
3. **תוצאות "מוזרות": מבחן F מעיד על מודל מובהק, בזמן שאף מבחן t אינו מובהק.**

למה זה קורה? מבחן F בודק אם המשתנה המוסבר מוסבר בצורה מספקת ע"י כלל המשתנים המסבירים. מבחן t לכל משתנה בנפרד בודק אם למשתנה זה יש תוספת הסבר מובהקת למשתנה המוסבר כאשר כלל המשתנים האחרים כלולים כבר במודל. כאשר המשתנים המסבירים מסבירים אחד את השני, ייתכן שהתרומה של כל אחד מהם בנפרד לא תהיה מובהקת, בעוד שבפועל הם אכן מסבירים בצורה טובה את המשתנה המוסבר.

**מבחנים למולטיקוליניאריות**

**דרך ראשונה - בדיקת המתאם בין כל זוג משתנים מסבירים בנתונים**: ראינו (בתרגול 12) שאומד למתאם בין שני כל משתנים לפי מדגמים שלהם הוא .

באופן דומה, ניתן לחשב אומד למתאם בין כל שני משתנים מסבירים:

אם המתאם גבוה (קרוב ל-1 בערך מוחלט), כדאי לוותר על אחד מהמשתנים הללו.

**דרך שנייה – מדד VIF (Variance Inflation Factor):** לכל משתנה מסביר מחשבים את הערך , כאשר הוא מקדם ההסבר עבור מודל רגרסיה שבו המשתנה המוסבר הוא והמשתנים המסבירים הם שאר המשתנים המסבירים.

ערך מדד מעיד על בעיה של מולטיקוליניאריות (המשתנה המסביר מוסבר ע"י שאר המשתנים מסבירים, כך שיש תיאום).

הערך של מדד אומר פי כמה גדלה השונות של המקדם בגלל התלות של המשתנה המסביר במשתנים המסבירים האחרים.

**דוגמה**

מעוניינים לבנות מודל רגרסיה ליניארית שבו מסבירים את גובהו של אדם () באמצעות שני משתנים מסבירים – גודל כף הרגל הימנית (), וגודל כף הרגל השמאלית (). קובץ הנתונים שישרת אותנו בבעיה (מופיע באתר הקורס) הוא בעל המבנה הבא:

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Right Foot (Inch)** | **Left Foot (Inch)** | **Height (Inch)** |
| (1) | 14.43 | 14.49 | 77.31 |
| (2) | 11.21 | 11.23 | 67.58 |
| … |  |  |  |
| (105) | 12.02 | 12.10 | 69.57 |

נתחיל מניתוח תוצאות מודל רגרסיה מרובה שכולל את שני המשתנים:



הרגרסיה מסבירה היטב את הנתונים (לפי ו-) והיא גם מובהקת לפי מבחן F. עם זאת, קיבלנו שתי תוצאות לא צפויות: המקדם של המשתנה שלילי (בניגוד לאינטואיציה שלנו), ומבחן t שלו אינו מובהק. זה צריך לעורר אצלנו נורה אדומה שישנה מולטיקוליניאריות בנתונים, לכן נבצע את הבדיקות המתאימות....

מבחן ראשון - מטריצת המתאם: בתא () מופיע , האומדן למתאם בין ו-:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **גובה** | **גודל כף רגל ימין** | **גודל כף רגל שמאל** |
| **גובה** | 1 |  |  |
| **גודל כף רגל ימין** | 0.903 | 1 |  |
| **גודל כף רגל שמאל** | 0.900 | 0.999 | 1 |

שני המשתנים המסבירים מתואמים באופן (כמעט) מושלם.

מבחן שני – נחשב את ה- של : נבצע רגרסיה שבה המשתנה המוסבר הוא והמשתנה המסביר הוא . ברגרסיה הזו ואז . זו עדות נוספת לקיומה של מולטיקוליניאריות בנתונים.

**אז מה עושים?**

הפתרון הפשוט ביותר הוא להשמיט את המשתנה המסביר שתלוי במשתנים האחרים. במקרה הזה שני המשתנים תלויים אחד בשני, ואכן בכל פעם שמשמיטים אחד מהם מקבלים רגרסיה מובהקת.

הגובה כפונקציה של גודל כף רגל ימין:



הגובה כפונקציה של גודל כף רגל שמאל:



בשתי הרגרסיות מקדם ההסבר גבוה, המודל כולו מובהק וגם מבחני t עבור המקדמים מובהקים.

ישנם גם פתרונות אפשריים נוספים למולטיקוליניאריות, אבל יריעתנו קצרה מלהכיל....

**משתני דמי = משתנים קטגוריאליים**

מודלי הרגרסיה שעסקנו בהם עד כה היו מבוססים על **משתנים כמותיים**, כלומר משתנים הנמדדים על ציר מספרי, לדוגמא: טמפרטורה, מרחק, גיל, עלות וכו'. לעיתים יש צורך בשילוב של משתנים איכותניים במודל הרגרסיה, לדוגמה שיוך לשכונה מסוימת.

כל ערך של משתנה כזה נקרא "רמה", והשיטה המקובלת להערכת ההשפעה של רמות שונות של משתנה איכותני על משתנה תלוי היא **שימוש באינדיקטורים**. באופן כללי, מייצגים משתנה איכותני בעל רמות על ידי אינדיקטורים המקבלים את הערכים 0 או 1.

**לדוגמה:** מעוניינים לבדוק את רמת ההשכלה כמשתנה מסביר עבור השכר של בוגרי תואר בהנדסת תעשייה. זהו משתנה מסביר קטגוריאלי עם שלוש רמות: בוגר תואר ראשון, בוגר תואר שני, בוגר תואר שלישי.

על מנת למדל משתנה קטגוריאלי עם 3 רמות, נשתמש בשני משתני דמי ו-, אותם נקודד באופן הבא:

|  |  |  |  |
| --- | --- | --- | --- |
| בוגר תואר שלישי | בוגר תואר שני | בוגר תואר ראשון |  |
| 0 | 1 | 0 |  |
| 1 | 0 | 0 |  |

המודל לבחינת השפעת ההשכלה על השכר :

תוחלת השכר של בוגרי תואר ראשון:

תוחלת השכר של בוגרי תואר שני:

תוחלת השכר של בוגרי תואר שלישי:

המשתנה משמעו שצריך להוסיף **"קפיצה"** בגובה בערך של כאשר מדובר על שכר של בוגרי תואר שני ביחס לבוגר תואר ראשון.

המשתנה משמעו שצריך להוסיף **"קפיצה"** בגובה בערך של כאשר מדובר על שכר של בוגרי תואר שלישי ביחס לבוגר תואר ראשון.

**תרגיל**

מהנדס מכונות מעוניין לחקור את הקשר שבין מהירות החיתוך (RPM) של מחרטה לבין טיב פני השטח (טפ"ש) של החלקים המיוצרים בה. הנתונים שנאספו מוצגים בטבלה בעמוד הבא.

1. הציגו את הנתונים על פני גרף. האם מודל רגרסיה יכול להתאים לתיאור הנתונים?
2. מצאו (בעזרת ) את משוואת הרגרסיה של המודל שהצעתם.
3. האם המודל מובהק? האם השפעת כל אחד מהמשתנים שהצעתם מובהקת? הניחו רמת מובהקות נדרשת של 0.01.
4. מהי התחזית לטיב פני השטח של חלק שנחרט במהירות 270 במחרטה 416 ?

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| סוג המחרטה | RPM | טיב פני השטח | מספר תצפית |  | סוג המחרטה | RPM | טיב פני השטח | מספר תצפית |
| 302 | 225 | 45.44 | 11 |  | 416 | 224 | 33.50 | 1 |
| 302 | 200 | 42.03 | 12 |  | 302 | 245 | 48.75 | 2 |
| 416 | 212 | 31.23 | 13 |  | 416 | 248 | 37.52 | 3 |
| 416 | 238 | 33.92 | 14 |  | 416 | 260 | 37.13 | 4 |
| 302 | 235 | 47.92 | 15 |  | 302 | 250 | 50.10 | 5 |
| 416 | 243 | 34.70 | 16 |  | 302 | 218 | 44.78 | 6 |
| 302 | 265 | 52.26 | 17 |  | 416 | 224 | 32.13 | 7 |
| 302 | 259 | 50.52 | 18 |  | 302 | 237 | 47.79 | 8 |
| 302 | 221 | 45.58 | 19 |  | 416 | 232 | 33.49 | 9 |
| 416 | 251 | 35.47 | 20 |  | 416 | 216 | 32.29 | 10 |

**פתרון**

1. מהתבוננות בגרף, ניתן להבחין בשני דברים:
2. הנתונים מחולקים לשתי קבוצות – הקבוצה העליונה כוללת את הנתונים על הטפ"ש במחרטה 302, והקבוצה התחתונה כוללת את הנתונים שקשורים לטפ"ש במחרטה 416.
3. בכל קבוצה כזו יש קשר שנראה (בעין) ליניארי, עם שיפוע זהה.

מודל רגרסיה אפשרי לתיאור הנתונים יכול להיות מודל הרגרסיה הליניארית המרובה הבא, אשר מנבא את טיב פני השטח כפונקציה של מהירות החותך ושל סוג המחרטה:

כאשר:

- טיב פני השטח

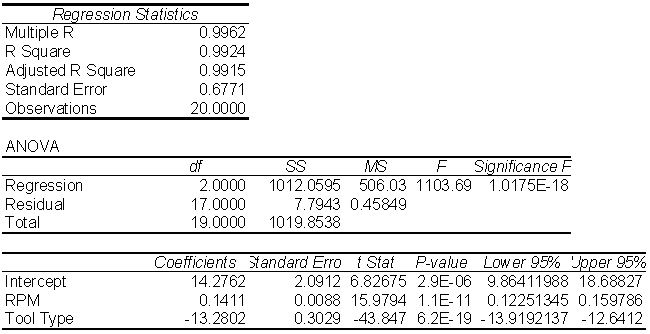
- מהירות החיתוך (משתנה רציף)

-  סוג המחרטה (משתנה קטגוריאלי)

1. כאמור, קשה למצוא מקדמים של מודל רגרסיה ליניארית מרובה באופן ידני, ולכן ניעזר באקסל. נייצר טבלה בסגנון הבא:

|  |  |  |  |
| --- | --- | --- | --- |
| X2 | X1 | Y | i |
| סוג המחרטה | RPM | טיב פני השטח | מספר תצפית |
| 0 | 225 | 45.44 | 1 |
| … | … | … | … |
| 1 | 224 | 33.50 | 11 |
| 1 | 212 | 31.23 | 12 |
| … | … | … | … |

פלט הרגרסיה:



מתוך פלט הרגרסיה נקבל את משוואת הרגרסיה:

בפרט, עבור מחרטה 302 (כאשר מתקיים הקשר הליניארי הבא:

ועבור מחרטה 416 (כאשר ), מתקיים הקשר הליניארי הבא:

שימו לב! כאשר מדובר במחרטה 416 ישנה קפיצה כלפי מטה בגודל 13.28, אבל בשתי המחרטות, מהירות החיתוך משפיעה באותו **קצב** (עם אותו שיפוע) על טיב פני השטח המתקבל.

1. את מובהקות תוצאות הרגרסיה (כולו) ניתן לבדוק באמצעות הפלט של מבחן . נמוך (מאוד), באופן שמצביע על מודל מובהק סטטיסטית (לכל רמת מובהקות סבירה). באופן דומה,   
   את מובהקות המקדמים ניתן לבדוק באמצעות מבחן מפלט הרגרסיה, וגם שם נמוך מאוד, באופן שמצביע על מקדמים מובהקים.
2. התחזית לטיב פני השטח של חלק שנחרט במהירות 270 במחרטה 416 :

**אינטראקציות**

**אינטראקציה** היא השפעה משולבת של מספר משתנים מסבירים. בכל המודלים שראינו עד עכשיו, הנחנו שגודל ההשפעה (=השיפוע) של כל משתנה מסביר אינו תלוי בערכים של המשתנים המסבירים האחרים. אבל לעתים מעניין לבדוק אם לשילוב של ערכים שונים של המשתנים המסבירים ישנה השפעה שונה, זוהי בדיקה של האינטראקציה בין המשתנים המסבירים.

נדגים באמצעות השאלה הקודמת (המקדחות):

1. **הציעו מודל שלוקח בחשבון השפעה משולבת של שני הגורמים, נוסף על ההשפעה של כל אחד מהגורמים בנפרד.**

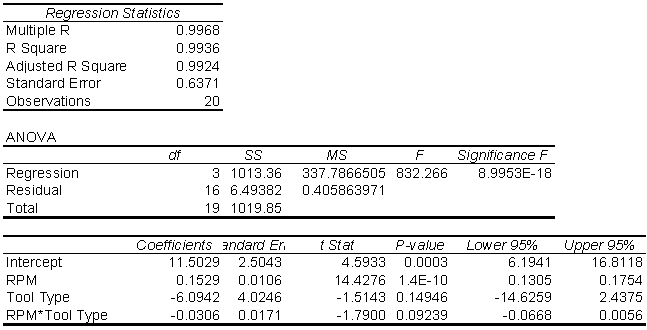
במודל שהוצג בסעיף ב', בדקנו האם לשתי המחרטות השפעה שונה על טיב פני השטח (נקודות חיתוך שונות עם הצירים), תוך הנחה כי למהירות החיתוך השפעה זהה בשתי המחרטות, כלומר לשני הישרים יש את אותו השיפוע.

כעת אנחנו מעוניינים לבדוק האם לשילוב של סוג המחרטה ומהירות החיתוך יש השפעה נוספת – **אינטראקציה, כלומר לשני הישרים אין בהכרח את אותו השיפוע.**

משוואת המודל תהיה:

שימו לב! למרות שישנם משתנים מסבירים שמוכפלים אחד בשני, זהו מודל רגרסיה **ליניארית,** משום שהרגרסיה צריכה להיות ליניארית במקדמים!

למעשה, המשמעות של מודל כזה היא שמגדירים משתנה מסביר חדש: , ואז בוחנים את המודל.



**מה המשמעות של ערכי המקדמים במקרה כזה?**

|  |  |  |
| --- | --- | --- |
| משוואת המודל: |  | |
| עבור חלקים שיוצרו במחרטה 302 (נציב ): | |  |
| עבור חלקים שיוצרו במחרטה 416 (נציב ): | |  |

בשונה מסעיף ב', במודל הזה תחת כל מחרטה, מהירות החיתוך משפיעה באופן שונה על טיב פני השטח (השיפוע של שונה!).